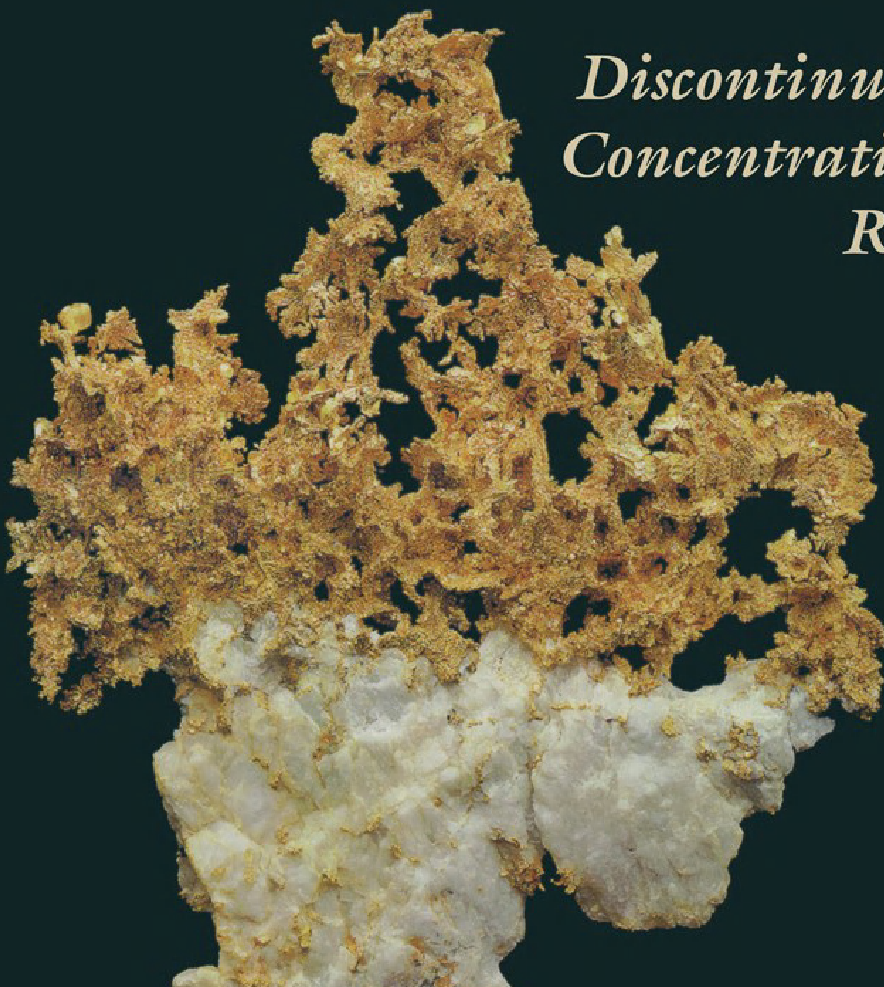
Benoit B. Mandelbrot

FRACTALS *and* SCALING *in* FINANCE

Discontinuity, Concentration, Risk



FRACTALS *and* SCALING *in* FINANCE

SELECTED WORKS OF BENOIT B. MANDELBROT

REPRINTED, TRANSLATED OR NEW

WITH ANNOTATIONS AND GUEST CONTRIBUTIONS

COMPANION TO *THE FRACTAL GEOMETRY OF NATURE*

Benoit B. Mandelbrot

FRACTALS *and* SCALING *in* FINANCE

Discontinuity, Concentration, Risk

SELECTA VOLUME E

With Foreword by R.E. Gomory

and Contributions by P.H. Cootner, E.F. Fama, W.S. Morris, H.M. Taylor, and others

Springer

Benoit B. Mandelbrot
Mathematics Department
Yale University
New Haven, CT 06520-8283 USA
http://www.math.yale.edu/mandelbrot
IBM T.J. Watson Research Center
Yorktown Heights, NY 10598-0218 USA

Library of Congress Cataloging-in-Publication Data
Mandelbrot, Benoit B.
Fractals and scaling in finance : discontinuity, concentration, risk / Benoit Mandelbrot.
p. cm.
Includes bibliographical references.
ISBN 978-1-4419-3119-1 ISBN 978-1-4757-2763-0 (eBook)
DOI 10.1007/978-1-4757-2763-0
1. Finance—Statistical methods. 2. Fractals. 3. Scaling (Social sciences)
I. Title.
HG176.5.M36 1997
332'.01'514742—dc21 97-37930

ISBN 978-1-4419-3119-1 Printed on acid-free paper

9 8 7 6 5 4

springeronline.com

To my sons, Laurent and Didier,
I dedicate this intellectual fruit of mine,
their demanding sibling

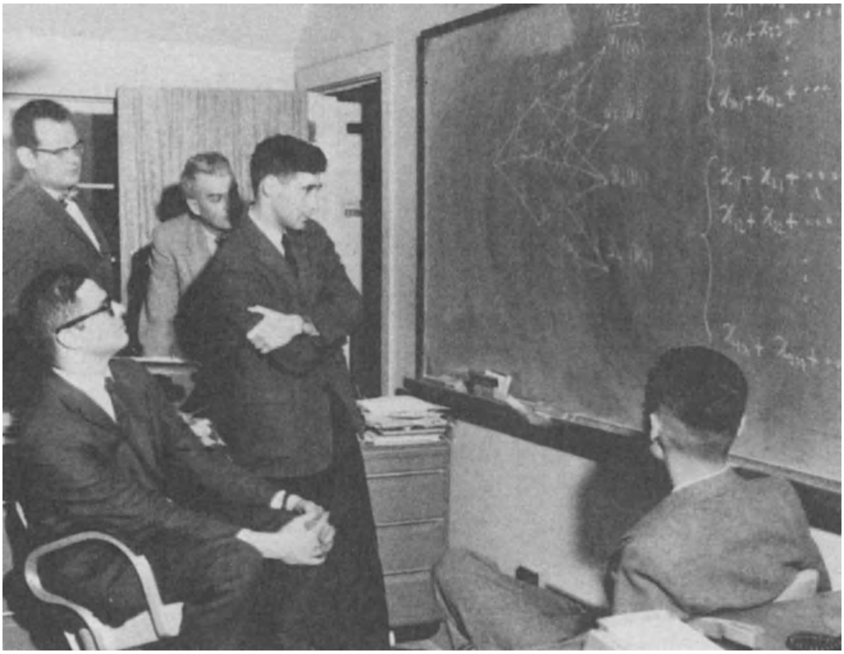
List of Chapters

In this List of Chapters, *the sources given after the titles include (in parentheses) the letter M followed by the year of publication and by the lower case letter that the* Bibliography *uses to distinguish different texts published in the same year. In the* Bibliography, *the items reproduced in this book and in Volumes N and H are marked by a star followed by a chapter number, which in some cases is incomplete or only tentative.*

Foreword

IN 1959-61, while the huge Saarinen-designed research laboratory at Yorktown Heights was being built, much of IBM's Research was housed nearby. My group occupied one of the many little houses on the Lamb Estate complex which had been a sanatorium housing wealthy alcoholics.

The picture below was taken about 1960. It shows from right to left, T.C. Hu, now at the University of California, Santa Barbara. I am next, staring at a network I have just written on the blackboard. Then comes Paul Gilmore, late of the University of British Columbia, then (seated) Richard Levitan, now retired, and at the left is Benoit Mandelbrot.



Even in a Lamb Estate populated exclusively with bright research-oriented people, Benoit always stood out. His thinking was always fresh, and I enjoyed talking with him about any subject, whether technical, political, or historical. He introduced me to the idea that distributions having infinite second moments could be more than a mathematical curiosity and a source of counter-examples. This was a foretaste of the line of thought that eventually led to fractals and to the notion that major pieces of the physical world could be, and in fact could only be, modeled by distributions and sets that had fractional dimensions. Usually these distributions and sets were known to mathematicians, as they were known to me, as curiosities and counter-intuitive examples used to show graduate students the need for rigor in their proofs.

I can remember hearing Benoit assert that day-to-day changes of stock prices have an infinite second moment. The consequence was that most of the total price change over a long period was concentrated in a few hectic days of trading and it was there that fortunes were made and lost. He also asserted that the historical data on stock prices supported this view, that as you took longer and longer historical data, the actual second moments did not converge to any finite number.

His thinking about floods was similar.

Benoit's ideas impressed me enormously, but it was hard to get this work recognized. Benoit was an outsider to the substantive fields that his models applied to, for example economics and hydrology, and he received little support from mathematicians who saw only that he was using known techniques. Benoit's contribution was to show that these obscure concepts lie at the roots of a huge range of real world phenomena.

Lack of recognition, however, never daunted Benoit. He stuck to his ideas and worked steadily to develop them and to broaden their range of applicability, showing that one phenomenon after another could be explained by his work. I was very pleased when I was able to get him named an IBM Fellow, and later was successful in nominating him for the Barnard Medal. After that the floodgates of recognition started to open and Benoit today is one of the most visible of scientific figures.

Surely he has earned that visibility, both for his world-changing work, and for his courage and absolute steadfastness.

Ralph E. Gomory
President, Alfred P. Sloan Foundation

Preface

PUBLIC OPINION AND THE PROFESSION know better than ever before that prices often change with hair-raising swiftness. In 1996, IBM experienced two large near-*discontinuities:* it fell by more than 10% and later rose by 13.2%. Furthermore, even in the absence of actual jumps, price change does not occur more or less evenly over time, but tends to *concentrate* in short "turbulent" periods. Concentration is a familiar concept in the study of industrial organization; it will be shown to generalize to the variation of prices. Wide awareness of discontinuity and closely related forms of concentration is reinforced by failures of portfolios that claimed to be free of *risk*.

This book is largely devoted to the above topics. Stock Market "chartists," as contrasted with the "fundamentalists," believe that charts embody everything needed to predict the future and devise winning strategies. Irrespectively of one's position on this dispute, I believe that one must understand the structure of charts thoroughly, including features that fail to bring positive returns to the investor. The understanding gained by a thorough exploration is bound to bring significant knowledge about the mechanisms of the financial markets and about the laws of economics. Most importantly, this knowledge is essential for evaluating the unavoidable risks of trading.

This book is symbolized on its cover by a spongiform object that can be admired in Paris, at the Muséum National d'Histoire Naturelle (gift of Fondation Elf). This object is a *natural fractal*; as such, it exhibits holes of many different sizes (ranging from large down to a "lower cutoff"), just as financial data exhibit "cycles" of many different durations. This object is a polysynthetic aggregate, the end-product of ill-known natural causes that only *randomness* can describe. Finally, and appropriately, its substance is almost pure *gold,* near-completely extracted from the white quartz within which it hid underground before its recent discovery in California.

The search for winning trading strategies is a goal of every study in finance, but will not be addressed here. Ultimate scientific explanation is attractive and important, but financial engineering cannot wait for full explanation. In any event, this book strives towards a second best: it seeks a "descriptive phenomenology" that is organized tightly enough to bring a degree of order and understanding. Indeed, a multitude of consequences that can be tested empirically will be drawn from somewhat abstract statistical algorithms called "scaling". They look simple, even simple-minded, but prove to be "creative," insofar as they generate unexpectedly complicated and structured behavior. This will establish that a wealth of features beloved by chartists *need not* be inserted "by hand", but *may well* follow inevitably from suitable forms of totally random variability.

The word "model" shall *not* denote the mathematical expression of an economic relationship, rather a statistical algorithm meant to fulfill an apparently extravagant ambition that was stated eloquently by Einstein. "The grand aim of all science is to cover the greatest number of empirical facts by logical deduction from the smallest number of hypotheses or axioms." Thus, an ideal model of price variation is one that produces sample data streams that are hard to distinguish from actual records, either by eye or by algorithm, and achieves a good part of this goal without ad-hoc "patch" or "fix."

Let us return to the word "scaling:" this book shows that many empirical facts are consequences of the following property assumed as "axiom." Starting from the rules that govern the variability of price at a certain scale of time, higher-frequency and lower-frequency variation is governed by the same rules, but acting faster or more slowly. This axiom has several increasingly broad implementations, which is fortunate, because financial data are of such awesome variety, that no single model can fit every market without being too complicated in some cases.

The preceding thoughts, combined with acceptance of the market maxim that it is better to be approximately right than certifiably wrong, led through successive innovations to several models of increasing power and generality; determining the model that applies helps sort out the markets into several distinct "states of variability."

At this point in time, my most general model is described by the words "fractional Brownian motion of multifractal time;" it successfully accounts for the variation of foreign exchange rates.

The first and most specialized model I put forward in 1963 is "Lévy stable motion;" it suffices to account for the "tail-dominated" variation of cotton spot prices.

A model I put forward in 1965 is "fractional Brownian motion;" it accounts for prices whose variation is "dependence-dominated."

In 1967, with H. M. Taylor, I injected the process of "subordination", and in 1972 announced a more general approach involving multifractals.

This book explains all those technical terms and motivates the broad family of processes that they denote.

All my models' ambition is to provide more effective ways to handle relatively rare events that have very strong effects. Man tends to react by either overestimation or neglect, and the discussion will necessarily involve several specific issues of wider relevance to the understanding and use of statistics.

Having been very active in finance and economics in the late fifties and early sixties, I longed to bring out a selection of my papers on those topics. The project kept being postponed, and eventually evolved into a hybrid: a newly-written book that is followed by long "appendices" reproducing old papers that give historical depth and add technical detail.

Alternatively, this book may be said, like the clothing worn by a traditional English bride, to incorporate something old, something new, something borrowed and even something blue.

The *old* refers to papers I devoted to economics in the 1960s. They are occasionally described as centered on "cotton prices and the Nile River," even though their scope was far wider. Key words that continue to be in use include "L-stability," "fractional Brownian motion," and "subordination." Key words that were superseded include "Lévy stability," "Lévy distributions," "Pareto-Lévy law," and "Paretian phenomena." My most important, best-known and most influential old paper is "The Variation of Certain Speculative Prices," M 1963b{E14}. In the reference style explained in the first lines of the *List of Chapters* and of the *Cumulative Bibliography*, M stands for "Mandelbrot" and {E14} means "Chapter 14 of this book." M 1963b{E14} and its sequel clearly show that I never believed that prices follow exactly a random walk.

The *borrowed* refers to *Comments* and a *Guest Contribution*.

As to *blue*, in three shades, it bears continuing witness of the official colors of several institutions: I worked at IBM for half a life, continued at Yale, and, long ago, started at the Paris laboratory of Philips.

Last, but not least, the *new* refers to the fact that many significant steps beyond M 1963b{E14} were taken after the 1960s, some of them recently; hence roughly half of the chapters were specially written for this book. The contents of three papers by myself, Fisher & Calvet (in different permutations) will be summarized, but this book could not wait for those papers to be completed.

In no way does this book claim to be a balanced exposition of the present status of my approach to finance. Books and other publications that either challenge it, or adopt and expand its conclusions, are rapidly increasing in numbers, but I never managed to read more than a few of them. (Some are mentioned in M, Fisher & Calvet 1997.) Extensive recent discussions with many references are Mirowski 1990, 1996, McCulloch 1996, Ballie 1996, Rachev 1996, and forthcoming books by Lévy-Véhel & Walter and Rachev & Mittnik.

A few words about data and method are in order. Data was scarce in the 1960s, but I took great pains to analyze them in detail. Today, data are extremely abundant, but I am ill-equipped for empirical work

This book's level of mathematical difficulty varies from high to low. Large parts are readable without much mathematical training, but many sections involve mathematics that, without being difficult, is delicate and familiar to few, and in many cases is published here for the first time. But mathematics is not pursued for its own sake.

To those who may find some parts to be overly mathematical, or otherwise baffling, my advice is "do not give up, merely skip ahead; the next topic may well be familiar or otherwise reassuring, and the difficulties you encounter may be gone on second reading."

As to the development of new statistical tools, I do not dwell on it, though statistical fitting is performed with care, and the need for new tools is evident throughout.

All told, my methods of investigation are those of a practicing theoretical and computational physicist. As a matter of fact, this has been the case in every substantive field in which I worked.

But there are significant wrinkles. I do not propose to pursue the adaptation to economics of an existing theory of equilibrium and of "mild" fluctuations. To the contrary, my tools were not to reach physics proper until later, as shall be told in this Preface. Therefore, my forty-five years in science can be viewed as being unified in giving a broader scope to the spirit of physics.

A proper quantitative study of financial markets began in the early 1960s, as witnessed by an influential anthology, Cootner 1964. That book includes a reprint of M 1963b{E14}, and interesting comments were contributed by the editor, Paul H. Cootner (1930-1978). (See also Chapter E16.) One reads that "Mandelbrot ... has forced us to face up in a substantive way to those uncomfortable empirical observations that there is little doubt most of us have had to sweep under the carpet up to now. With determination and passion he has marshalled, as an integral part of his argument, evidence of a more complicated and much more disturbing view of the economic world than economists have hitherto endorsed. Furthermore, this new view has a strong attraction for many of us."

Elsewhere in Cootner 1964, one reads that "There can be no doubt that Mandelbrot's hypotheses are the most revolutionary development in the theory of speculative prices since Bachelier 1900." This last reference can be viewed as the point of departure of a rational approach to finance, since it was the first to describe the Brownian motion model, which will be discussed momentarily.

Needless to say, my view of the economic world's complexity was not adopted in 1964, its implications were not faced, and the study of finance continued to rely on the "1900 model" of Bachelier, which pointedly denies discontinuity and concentration. Those obvious defects are becoming unacceptable, bringing me back to the study of finance.

To tackle discontinuity and concentration, I conceived in the late fifties a tool that was already mentioned, but deserves elaboration. I concluded that much in economics is *self-affine;* a simpler word is *scaling*. This notion is most important, and also most visual (hence closest to being self-explanatory), in the context of the financial charts. Folklore asserts that "all charts look the same." For example, to inspect a chart from close by, then far away, take the whole and diverse pieces of it, and resize each to the same horizontal format known to photographers as "landscape". Two renormalized charts are never identical, of course, but the folklore asserts that they do not differ in kind. The scholarly term for "resize" is to "renormalize" by performing an "affinity," which motivated me in 1977 to coin the term "self-affinity." (Scaling can also be self-similar, as we shall see in Section 2 of Chapter E6.) The scholarly term for "to look alike" is "to remain statistically invariant by dilation or reduction."

I took this folklore seriously, and one can say that a good portion of this book studies financial charts as geometric objects.

Brownian motion is scaling also, but it is best to restrict this term to non-Brownian models. Non-Brownian forms of scaling are sometimes called "anomalous;" a better term is "non-Fickian."

Viewing the renormalized pieces of a chart as statistical samples from the same underlying process, I identified, one after another, several non-Brownian implementations of scaling, and tested them - successfully - in one or another financial context.

From the viewpoints of economics and finance, the most striking possible consequence of "non-Fickian" scaling is discontinuity/concentration.

In this spirit, the "M 1963 model," centered on M 1963b{E14}, concerned speculative prices for which long-term dependence is overwhelmed by discontinuities or periods of very fast change.

As soon as advances in computer graphics made it possible, fractal "forgeries" of price records were drawn for the M 1963 model. These forgeries prove to be realistic. This significant discovery became an early exhibit of a surprising and fundamental theme common to all aspects of an area to be discussed momentarily, fractal geometry.

In a later "M 1965 model" (M 1965h{H}, M & Van Ness 1968{H}, and many other chapters in M 1997H, the key factor is long-term dependence.

The "M 1967 model," described in M & Taylor 1967{E21}, soon introduced the notion of trading time and pointed out the relevance of a once esoteric mathematical notion called "subordination."

Finally, M 1972j{N14} introduced the notion of *multifractal*, and concluded (page 345) by immediately pointing out this notion's possible implications to economics. This is why the term "M 1972 model" will be applied to the approach that I developed in recent years on the basis of multifractals.

The scaling principle of economics incorporates these and all other forms of variability and promises further generalizations. I also used scaling in other contexts of economics, such as the distribution of income and of firm sizes.

Moving on from the visual to the analytic aspect of the study of scaling, it involves diverse "scaling exponents." One is Pareto's income exponent, which is familiar to many economists.

M 1963b{E14} attracted attention, witnessed by the above quotes from Cootner 1964, praise from W.H. Morris, and a special session of the Econometric Society (see Chapter E16), witnessed also by doctoral dissertations that I supervised, Fama 1965 and Zajdenweber 1976, and the dissertations of Fama's students, especially Blume 1968 and Roll 1970. In addition, I was elected to Fellowship in the Econometric Society (and listed in a *Who's Who in Economics.)* But until the strong revival that we are now witnessing, interest in the M 1963 model waned (and the second edition of that *Who's Who* dropped me.)

It was clear that I addressed concerns that were not being faced, that is, answered questions that were not being widely asked, and that my reliance on ccmputers was prohibitive. More importantly, the conceptual tools of my work, and its perceived consequences, were resisted. In a way, one may say that I was "victimized" by a case of unexpected historical primacy of financial economics over physics.

An earlier case of such primacy was already mentioned twice. Odd but true, and implied in the quote from Cootner 1964, a maverick named Louis Bachelier (1870-1946) discovered Brownian motion while studying finance, five years before Albert Einstein and others independently rediscovered and developed it in physics. Eventually, but not until the 1960s, many hands brought Brownian motion back into economics.

Quite similarly, scaling and renormalization were central to my work in finance several years before they were independently discovered and developed in the study of critical collective phenomena of physics, through the work of Fisher, Kadanoff, Widom, Wilson, and others. In fact, "scaling" and "renormalization" are those physicist's terms, and replace my weaker terminology. This "rerun" of the story of Brownian motion confirms that, while economics finds it easy to borrow from established physics, the cases of historical primacy of economics over physics start by being a handicap, if they involve overly unfamiliar and untested tools.

Interrupting the study of finance, I went on to investigate scaling and renormalization in more hospitable contexts scattered over mathematics, physics, astronomy, the earth sciences, and elsewhere. This explains yet another word, already mentioned in this preface and featured on this book's title page. *Fractal,* a word I did not coin until M 1975o, denotes a new geometry of nature that, in a word, makes it possible to quantify many forms of *roughness*. It brought together diverse bits of scattered knowledge coming from opposite and distant sides, and made them interact and bear fruit. Thanks to fractal geometry, those bits of knowledge became understood, acquired a clear identity, and ceased to be "homeless" by becoming part of a new field.

One side brought diverse and scattered empirical observations (some of them stated in the form of "folklore.") All raised questions that cry out to be answered, but were finding no answer with the existing tools. In a real sense, many drifted into the dustbin of the history of science.

The other side brought several notions that mathematicians had contrived with no motivating application in sight. Quite the contrary, those notions were specifically introduced, and continued to be presented, as examples of "exceptional" or "pathological" behavior. Their sole role was to demonstrate that mathematics has a creative freedom that requires no concrete need or use to be unleashed. I turned those notions around, so to speak, and matched those "answers without questions" to the "questions without answers" mentioned in the preceding paragraph.

Needless to say, this match between the "dustbin" and the "psychiatric hospital" brought immediate need for new tools and new data, hence advances in mathematics as well as in observation. Furthermore, the move of fractal geometry from finance to physics and mathematics was deeply helped by an unexpected novelty: the computer and computer graphics began to play an increasingly essential role. The reader may know M 1982F{*FGN*} for what we call "computer forgeries" of mountains that share the "extravagant ambition" expressed early in this preface,... and fulfill it partially, but better than anyone expected.

Acquaintance with fractals is *not* a prerequisite to reading this book, but a sketch is found in Section 2 of Chapter E6. In any event, this book uses the term *fractal* sparingly, but even the reprinted chapters are thoroughly impregnated with the idea. In fractal technical terms, my successive models in finance signify that, once again, the charts generated by my models are self-affine. Being self-affine involves several specific embodiments of increasing generality and power.

Yet another reason stalled the writing of this book in the mid-sixties and made me move out of economics. When P. H. Cootner described M 1963b{E14} as a "revolutionary development," did he think mostly of destruction or reconstruction? The answer is found in Cootner 1964: "Mandelbrot, like Prime Minister Churchill before him, promised us not utopia, but blood, sweat, toil and tears. If he is right, almost all our statistical tools are obsolete ..., past econometric work is meaningless ... It would seem desirable not only to have more precise (and unambiguous) empirical evidence in favor of Mandelbrot's hypothesis as it stands, but also to have some tests with greater power against alternatives that are less destructive of what we know."

The wish to see improved statistical tests can be applauded without reservation. But it is prudent to fear that "what we know" is not necessarily the last word. Allow me just one example. When Fast Fourier algorithms became available around 1964, spectral analysis created great interest among economists. Tests showing some price series to be nearly "white" were interpreted as implying the absence of serial dependence. The result made no sense and was forgotten. But (as reported in Chapter E6) a promising non-Gaussian model of price variation is white, despite the presence of strong serial dependence! It often seems that novel uses of known statistics may at the same time test a hypothesis and test a test!

Be that as it may, the study of financial fluctuations moved on since 1964, and became increasingly refined mathematically while continuing to rely on Brownian motion and its close kin. Therefore, Cootner's list of endangered statistical techniques would now include the Markowitz mean-variance portfolios, the Black-Sholes theory and Ito calculus, and the like.

In Cootner's already quoted words, my "view of the economic world is more complicated and much more disturbing than economists have hitherto endorsed." This implies that the erratic phenomena to which this book is devoted deserve a special term to denote them, and explain why I chose to call them *wildly random.* By contrast, Brownian motion and most models used in the sciences deserve to be characterized as *mildly random,* and lognormality and some other treacherous forms of randomness that are intermediate between the mild and the wild will be described as *slow.*

The point is this: the specificity of slow or wild randomness in economics could be disregarded for many years, but no longer. In particular, new statistical tools are urgently needed.

Attention can now turn to a last word on the title page. This is one of several volumes under preparation collectively called *Selecta,* and meant to accompany my book-length *Essays,* M 1975o{*LOF*}, M 1977F and M 1982F{*FGN*}. Those Essays elucidate a strong and long-term unity of purpose, but their technical references were hard to locate, read and relate to each other. The technical *Selecta* originated in a set of reprints sorted out in several categories, typically centered around a key paper. (It was a relief to observe that all my papers could be reprinted without being embarrassing, even though in fact only a few will be reprinted.)

The topics of those *Selecta* books are deeply interdependent, but to a large extent they can be read independently of each other, in no logical order. Therefore, each is denoted by the first letter of a key word. Here, this key word is economics, which is why all chapter numbers begin by E. At least three of these books, the present one, M 1997H, and M 1997N, concern the broad concept of *self-affine fractality*. Those other books focus on turbulence, noises, hydrology, and other physical phenomena; however, their discussions of multifractality and self-affinity are essential to understand, study, and develop my models in finance.

The diverse obstacles that made my work "premature" in the 1960s have vanished. Computers are everywhere. Physicists and fractalists developed new modeling tools that can be applied to finance. Abundant financial data are readily available. Events insure that concern with discontinuity is near-universally shared, and my work is pointedly addressed to the many "anomalies" that bedevil prevailing financial models.

Overall, my ideas now fit in the framework of fractals (hence also of chaos), and of experimentation with new financial products. An effect of these changes is to bring me back to the study of finance.

Yet, this book barely scratches its topic. In particular, the M 1963 and M 1965 models are *linear*. This is unrealistic, but convenient, and some techniques of econometrics extend to linear models with nonGaussian "error terms." However, non-linearity entered with the model in M 1966b{E19} and the M 1967 and M 1972 models; it is at the center of current work that would not fit in this book.

New Haven, CT, & Yorktown Heights, NY, May 1997

Acknowledgements

Other volumes of these *Selecta* record a multitude of difficulties and (near) disasters with journal editors and referees. The editors of journals on finance and economics (and their referees) were a glowing exception, especially (listed alphabetically) L. R. Klein, E. Malinvaud, and M. Miller. In nearly every case, critics spoke and wrote only *after* a work was published, and did not interfere with publication.

In creating this book out of a pile of articles, unpublished research reports and drafts, invaluable help was provided by my secretaries, L. Vasta, followed by P. Kumar, then K. Tetrault. The last stages were also helped by M. McGuire. In the 1960s, however, I worked largely alone. More precisely, I had help for the computation of tables (M & Zarnfaller 1961) and many drafts were discussed in varying detail with my Ph.D. student E. F. Fama. I also learned much from my later economics students, D. Zajdenweber and C. Walter. Additional individual acknowledgements are found after each paper.

For over thirty-five years, the Thomas J. Watson Research Center of the International Business Machines Corporation provided a unique haven for a variety of investigations that science needed, but Academia neither welcomed nor rewarded. As a beneficiary through most of that period, I am deeply indebted to IBM for its continuing support of mavericks like myself, Richard Voss, with whom I worked closely, and Rolf Landauer. Many other names rush to mind, and among long-term friends and colleagues, the Yorktown of its scientific heyday will remain most closely associated with Martin Gutzwiller and Philip Seiden.

As to management, at a time when my work in finance was starting and was perceived as a wild gamble, it received whole-hearted support from Ralph Gomory, to whom I reported for twenty-five years in his successive capacities as first-line manager of a small group, later Department Director, and finally IBM Senior Vice-President and Director of Research.

Last but not least, this book would not have been written without the constant and extremely active participation of my wife, Aliette, and her unfailingly enthusiastic support. My earlier books were dedicated to my parents and to her, and she also shares automatically in this book's dedication to our two sons.

Organization of the book

Granted time, assistance, and freedom from other tasks, this book would have become less idiosyncratic and bulky. But the present format has redeeming features: the reader can choose the sequence of chapters that is most suitable, and can start either with the history represented by the reprints, or the present represented by the new chapters.

The old and new halves of the book are organized differently.

The reprinted chapters. The second half of the book is straightforward. The reprints of major old papers follow a logical order that happens to be close to the order of original publication. To avoid clutter, several short publications follow the chapters to which they are closest in subject, as *Pre-* or *Post-publication Abstracts* or *Appendices*. *Chapter Forewords* and *Annotations* were added, and many chapter titles were updated from the originals found in the *Cumulative Bibliography*.

As the originals are readily available in standard periodicals, and were criticized for poor English style, it was felt best to copy-edit them carefully, with added subtitles, but of course no change whatsoever in the original thinking. A few afterthoughts and/or corrections were inserted as *Postscripts* placed within braces { }. More importantly, the notation and terminology were unified – more or less systematically: original terms that did not take root were replaced by terms chosen to be easy on the tongue, like *scaling, fractal, L-stable, mild, slow* and *wild*. Once again, they were not part of my vocabulary when those old papers were written.

The remaining overlap between the original articles will (once again) assist those who will not read the reprints in sequential order.

The newly written chapters. Unavoidably, the first half of this book is organized in somewhat complicated fashion. It is best viewed as a house that boasts not one but several welcoming entrances, between which the self-sufficient reader is free to choose. To make this task easier, the new chapters, like the old, involve repeats; those are easily skipped, but decrease the need for cross references. Thanks to this policy, it became possible to order the chapters for the sake of one category of readers: those who prefer to soften the harder mathematics by beginning with extensive verbal explanations and light mathematical arguments.

PART I: NONMATHEMATICAL PRESENTATIONS

Chapter E3, an early expository work, appeared the same month as "The Variation of Some Speculative Prices," M 1963b{E14}, and remains of more than historical interest, but of course is not up-to-date. The detailed introduction prepared for this book is divided into three chapters written at a level of detail intermediate between the breezy Preface and the reprints. Additional nonmathematical preliminaries are found in the first sections of Chapters E5 and E6.

&&&&&&&&&&&&&&&&&&&&&&&&&&&&

E1

Introduction

✦ **Abstract.** A pragmatic attitude towards the slippery notion of randomness is described in Section 1. A premise of my work is that graphics must not be spurned; Section 2 argues that it remains indispensable even beyond the first stage of investigation.

A more pointed premise of my work is that the rules of price variation are *not* the same on all markets, hence a single statistical model may not describe every market without unacceptable complication. I agree that "it is better to be approximately right than certifiably wrong," and worked, in succession or in parallel, with several distinct fractal models of finance of increasing generality. Adapting the style of reference used throughout this book, those models will be denoted by the letter M followed by the year of original publication or announcement.

The core of this chapter is made of Sections 5 to 9. A first glance at scaling is taken in Section 5. Section 6 introduces the M 1963 model, which deals with tail-driven variability and the "Noah" effect and is based on L-stable processes. Section 7 introduces the M 1965 model, which deals with dependence-driven variability and the "Joseph" effect and is based on fractional Brownian motion. Old-timers recall these models as centered on "cotton prices and the River Nile." Section 8 introduces the M 1972 combined Noah-Joseph model, which this book expands beyond the original fleeting reference in M 1972j{N14}. That model is based on fractional Brownian motion of multifractal trading time.

The M 1965 and M 1972 models have a seemingly peculiar but essential feature: they account for the "bunching" of large price changes indirectly, by invoking unfamiliar forms of serial dependence with an *infinite* memory. All the alternative models (as sketched and critized in Section 4 of Chapter E2) follow what seems to be common sense, and seek to reach the same goal by familiar short memory fixes. Infinite memory and infinite variance generate many paradoxes that are discussed throughout this book, beginning with Section 8.3.

Here are the remaining topics of this chapter: Brownian motion (the 1900 model!) and martingales are discussed in Section 3. The inadequacies of Brownian motion are listed in Section 4. Section 9 gives fleeting indications on possible future directions of research. Finally, the notions of "creative model" and of "understanding without explanation" are the topics of Section 10. ✦

IN ADDITION TO THIS CHAPTER, the book also has other welcoming entrances, each geared to a different constituency. Chapters E1 to E4 use mathematics sparingly and proceed leisurely. Their point of arrival will not be described in mathematical terms until Chapter E6, but was mentioned at the beginning of the Preface.

1. AN ENTIRELY PRAGMATIC VIEW OF THE SLIPPERY NOTION OF RANDOMNESS

Randomness is an intrinsically difficult idea that seems to clash with powerful facts or intuitions. In physics, it clashes with determinism, and in finance it clashes with instances of clear causality, economic rationality and perhaps even free-will. It is easy to acknowledge that randomness can create its peculiar regularities. But it is difficult to acknowledge that such regularities either could be interesting or could arise in physics or finance. As a result, the fact that any statistical model could be effective seems a priori inconceivable and is difficult to acknowledge.

The problem is mitigated in physics because the atoms in a gas are not known individually, and much exaggerated in finance for the opposite reason. Furthermore, it is difficult in finance to disentangle the roles of the observer and the active participant. Most persons seek financial knowledge for the purpose of benefitting from it, and thereby modifying what they benefit from. But merely describing the markets does not perturb them, and the ambition of this book is simply to observe, describe some degree of order, and thereby gain some degree of understanding. This leads to a pragmatic view described in Chapter 21 of M 1982F{*FGN*}, titled "Chance as a Tool in Model Making." Several paragraphs of that chapter will now be paraphrased. In short, the notion of randomness is both far more effective and far less assertive than is often assumed or feared.

It is necessary to first comment on the term, *random*. In everyday language, a fair coin is called random, but not a coin that shows *head* more often than *tail*. A coin that keeps a memory of its own record of *heads* and *tails* is viewed as even less random. This mental picture is present in the term *random walk*, especially as used in finance (Section 3).

Of course, statisticians hold a broader view of randomness, which includes coins that are not fair or have a memory. However, more or less explicitly, statisticians ordinarily deal with observations that fluctuate around a "normal state" that represents equilibrium. A good picture of that classical scenario is provided by the edge of a razor blade. When greatly enlarged, it presents many irregularities, but from the user's viewpoint, a high quality blade is practically straight overall, therefore its description splits naturally into a fluctuation and a highly representative "trend." In Chapter E5, such fluctuations will deserve to be called *mild*.

For contrast, consider a coastline like Brittany's or Western Britain's. Taking into account an increasingly long portion will average out the small irregularities, but at the same time inject larger ones. A straight trend is *never* reached and interesting structures exist at every stage.

All too often, however, "to be random" is understood as meaning "to lack any structure or property that would single out one object among other objects of its kind." The question arises, is this a fair characterization of randomness or, to the contrary, is the mathematical notion of chance powerful enough to bring about the strong degree of irregularity and variability encountered in coastlines as well as in financial charts?

The answer to that question came as a surprise: not only is chance powerful enough, but in many cases it goes *beyond* the desired goal such as, for example, the case for this generalized processes I introduced in M1967b{N10} and proposed to call "sporadic". In other words, I perceive a tendency to grossly underestimate the ability of chance to generate extremely striking structures that had not been deliberately inserted in advance.

However, fulfilling this goal demands forms of randomness that are far broader than is acceptable in the bulk of statistics. Once again, the physicists' concept of randomness is shaped by mild chance that is essential at the microscopic level, while at the macroscopic level it is insignificant. In the scaling random fractals that concern us, to the contrary, the importance of chance remains constant over a wide range of levels, including the macroscopic one. This nonaveraging change is described in Chapter E5 as *wild*; I continually explore it, and claim that it, and *not* mild chance, is the proper tool in finance and economics.

Be that as it may, the relationship between unpredictability and determinism raises fascinating questions, but this work has little to say about them. It makes the expression "at random" revert to the intuitive connotation it had at the time when it entered medieval English. The original French phrase "un cheval à randon" is reputed to have been unconcerned with causes and the horse's psychology, and merely served to denote an irregular motion the horseman could not fully *predict* and *control.*

Thus, while chance evokes all kinds of quasi-metaphysical anxieties, I am little concerned with whether or not Einstein's words, "the Lord does not play with dice, " are relevant to finance. I am also little concerned with mathematical axiomatics. The reason I make use of the theory of probability is because *there is no alternative*: it is the only mathematical tool available to help map the unknown and the uncontrollable. It is fortunate that this tool, while tricky, is extraordinarily convenient and proves powerful enough to go well beyond mild randomness to a wild and "creative" state of that concept.

Let us draw some other consequences from the preceding combination of a *credo* and a promise. To be entitled to use probability theory, there is *no need* to postulate that *every* financial and economic event is generated by chance, rather than by cause. *After the fact* (ex-post), one may find uncontroversial or reasonable causes for at least some features. But *before the fact* (ex-ante), the situation is very different. In order to move towards a quantitative approach to economics (a "rational economics" to echo the grand old term "rational mechanics"), one must *unavoidably* resort to probability theory. Further subtle but significant differences between ex-post and ex-ante are found throughout this book.

To believe that the concept of randomness has far more power than it is credited with is essential, but is not enough. One must construct actual random processes that use few inputs, are tightly organized and fit the data so well as to yield a degree of understanding. To show this is feasible is an ambition of this book.

2. GRAPHICS, THE COMPUTER, STATISTICS AND BEYOND: "INDEX NUMBERS" AND OTHER SUMMARIES OF THE DATA

"To see is to believe," and to prove that the last paragraph of the preceding section is valid, the easiest and most convincing path consists in allowing the eye to compare the actual data to the outputs of random processes, without waiting for the technical details that will not be given until Chapter E6. The reader is, therefore, encouraged to compare Figures 1 and 2 of this chapter. The latter is a simulation of a surprisingly simple process (which took a surprisingly long time to be identified!) and it can be "tuned" to achieve a wide range of different behaviors.

My claim is that ability to imitate is a form of understanding. To preempt a challenge, imagine that a statistical test proclaims that the data in Figure 1 are actually very different from the simulation in Figure 2. If so, what should the proper response be: to dismiss the evidence of the eye or to look carefully for hidden assumptions that may have biased the statistical test? This second possibility is the one that this section proposes to examine in detail. Many readers may want to proceed to Section 3.

Through most of history, computation and pictorial representation were at best prohibitively expensive and mostly impractical or even inconceivable. Deep philosophical reasons were put forward to justify the fact that the eye became suspect and was almost completely banished from "hard science." In my opinion, those philosophical reasons were unconvincing, and the real reasons for foresaking the eye was simply practical. Therefore, the advent of cheap computing and graphics will have a profound and increasing impact throughout the sciences. Of special concern to us is its impact on probability theory and statistictive fields are related but separate and best examined in turn.

In the past, random processes could only be investigated through formulas and theorems. The resulting knowledge is invaluable, but three episodes made me conclude, long ago, that it is incomplete.

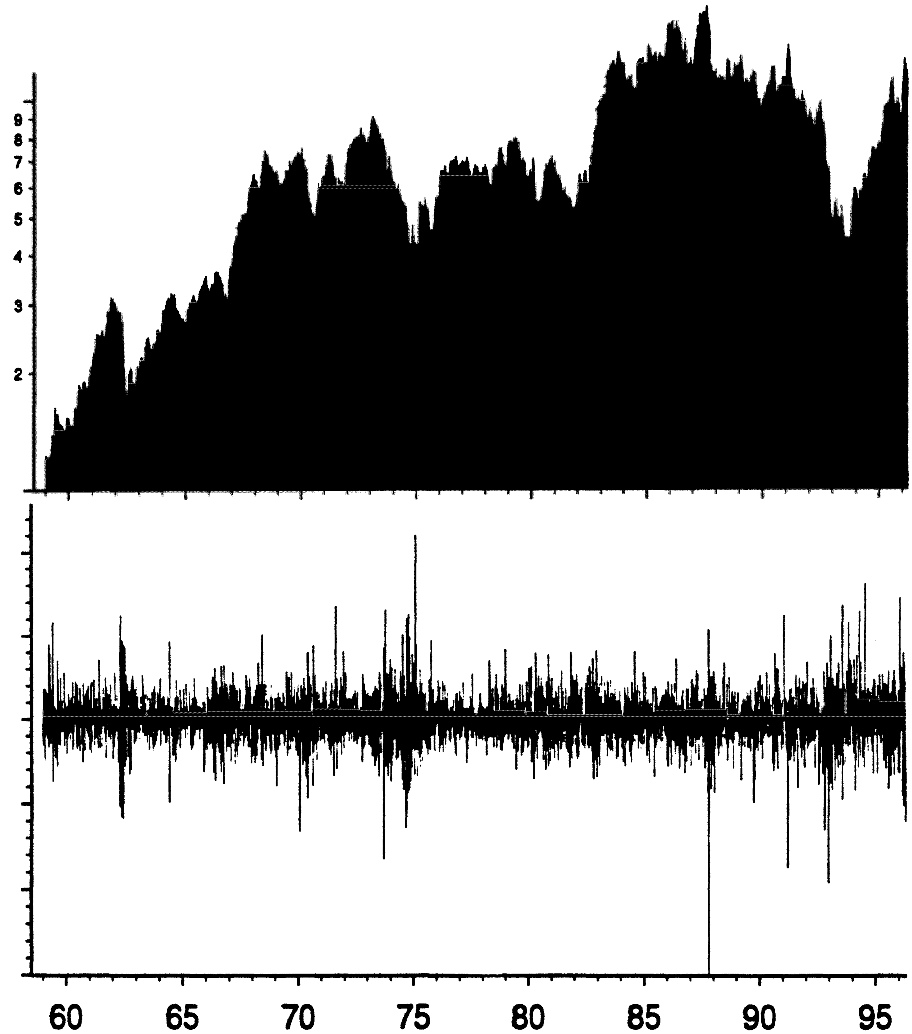


FIGURE E1-1. Top: IBM stock from 1959 to 1996, in units of $10, plotted on logarithmic scale. Bottom: the corresponding relative daily price changes, in units of 1%.

The shallowness of my earlier understanding of random walk first became obvious after I spent hours dreaming while examining, again and again, the sole illustration in Feller 1950 (Volume 1). This illustration, a

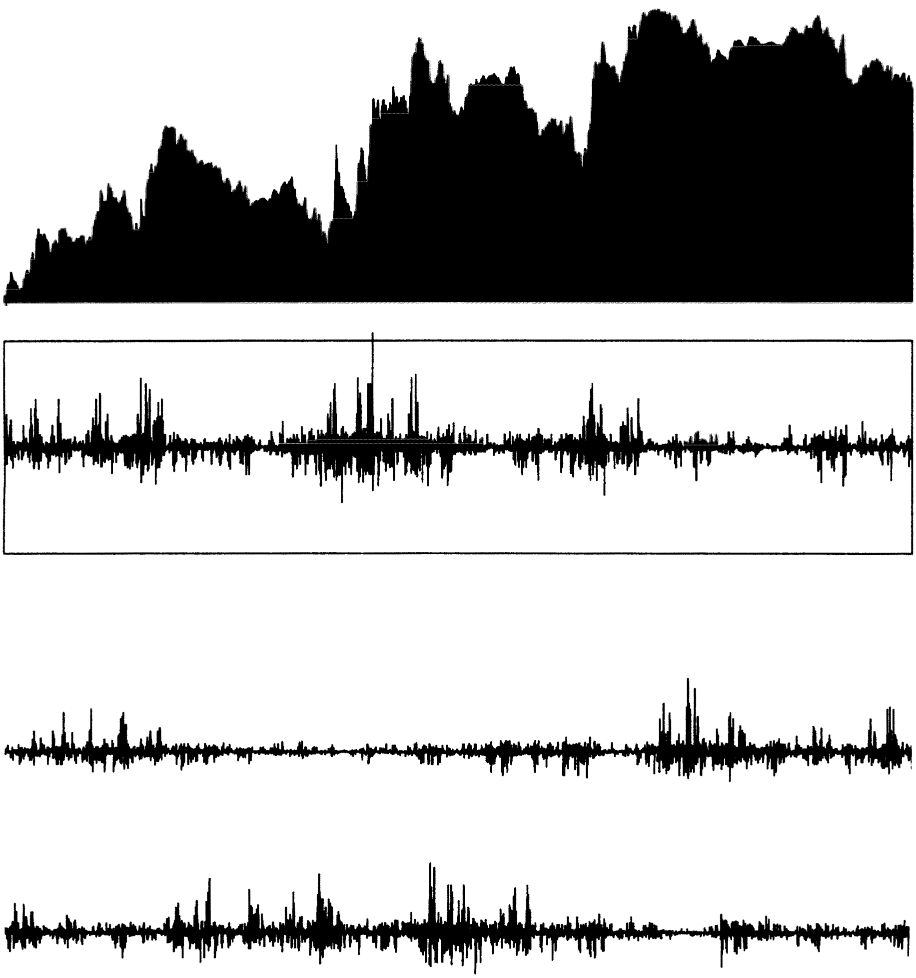


FIGURE E1-2. From top to bottom: plot of a computer simulation of the current fractal model of the variation of prices followed by a plot of its increments, and additional samples of the increments for different seeds. This figure is included here to make a point in anticipation of Chapter E6, where Section 4 explains the underlying construction and Figure 5 exhibits additional samples.

record of painstaking actual throws of a coin, is reproduced, with comments, in Figure 4 of M 1963e{E3}, and also in Plate 241 of M 1982F{*FGN*}. William Feller later confided to me that this was one of several possible illustrations prepared for him, and was chosen because the alternatives were even further removed from the readers' prejudices. It was disappointing that a figure I found inspiring should exemplify the fact known to everyone, that Man often use pictures to *disguise* reality, instead of illustrating it. But Man also uses words for the same purpose! Be that as it may, I think that the publication of *all* of Feller's pictures would have provided clearer "intuitive" or "in the fingers" understanding.

Second episode: Following immediately upon M 1963b{E14}, Berger & M 1963{N} faced the challenge of making it obvious that a certain physical phenomenon involved a degree of randomness well beyond the mild. To force conviction, they had to resort to a hand-cut wire model.

Third episode: As soon as crude "Calcomp" tracing tables could be attached to a computer, M & Wallis 1969a,b,c{H} hastened to put them to use in illustrating a process that will be discussed momentarily. Comparing the data with the sample functions of the M 1965 model and other models, we saw instantly that certain models could not possibly be correct, while other models seemed adequate. The "objective" statistical tests available at that time provided less clear-cut distinctions, confirming that they had been devised to deal with a context substantially different from the context of fractals. Only half in jest, we thought that the calculations involving existing statistical techniques were not only a way to test a model, but also to test the tests.

Given the minimal cost of sample graphs, I never felt we erred by producing too many, but often fear (for example, when preparing this book) that we err by producing too few.

The same issue can be seen under another light. Under the old technological constraints, the long lists of data provided by observation and measurement could not be graphed or manipulated usefully, and it was imperative to begin by compressing them drastically. In economics, of course, this compression yields diverse "indicators" or "index numbers." The most classical and simplest index numbers are (weighted) averages, but it is useful to use the same term more widely, in particular, for moments of order higher than 1.

Needless to say, the proper selection of index numbers is an endless source of controversy, and skeptics claim that suitable weights can yield any result one wishes. Nevertheless, it is widely believed that moments are an intrinsic concept. But are they really? In mechanics, it is indeed true that the first and second moments, weighted by mass, yield a body's center of gravity and radius of gyration. In finance, the value of a portfolio is a non-controversial first sample moment; more controversial is "cost-of-living." But, ex-ante, second and higher moments are far from unquestionable. It is true that they play a central role in Taylor series (to be touched upon at the end of Chapter E5), but their main role is to provide quantitative "index numbers," whose usefulness must *not* be viewed as obvious but must instead be established separately in each case.

I believe that, ex-post, the compression implicit in the moments may, but need not, be useful, depending on specific circumstances to be distinguished in Chapter E5. In the cases of *mild* randomness, as exemplified above all by independent Gaussian variables, very simple compression is "sufficient" and preserves the important information. However, much of this book will argue that economics and finance (as well as many fields of natural science) are characterized by forms of randomness that are *not mild at all*. A prime characteristic of *wild* randomness is that familiar index numbers altogether cease to be representative in their case. What should be done until new and more appropriate statistical tools become available? I think that one will have to live with graphical tests, and learn to perform them with care and without haste, taking full advantage of computer graphics.

"When exactitude is elusive, it is better to be approximately right than certifiably wrong." To be unquestionably correct is a nice idea but is not an option, and the Stock Market wisdom quoted in this paragraph's title is an excellent characterization of one aspect of my approach. In particular, the Brownian model of Section 3 claims to be valid without restrictions, which is certainly quite wrong, while the scaling models of Sections 6 to 8 only claim to be approximately exact over a limited range of applicability. That is, when the data are compared with my theoretical distributions, one should expect systematic errors of specification that are *larger* than errors due to statistical fluctuations.

More detailed reasons for tolerating the imperfection described in this subsection's title will be discussed in Section 5 and in Chapter E2. A discussion of this "necessary tolerance" should be part of statistics, but is not.

Even the best objective statistical tests are not *of universal validity. Who is testing the testers?* My attitude is deeply colored by publications that reexamined my M 1963 model quantitatively and concluded that "objective" tests contradict the "subjective" claims I based on graphical evidence. Actually, that graphical evidence was confirmed again and again, while the objective tests are forgotten, as they deserved to be. Indeed, there is no theorem without assumption, and even the best statistical test can only be used under certain conditions of validity.

Take spectral analysis: this is unquestionably a powerful tool, but a very tricky one: Chapter E6 (Section 3.5) will describe a significant "blind spot" of spectral whiteness that concerns nonGaussianity. It is not a matter of mathematical nitpicking, but it directly concerns finance. More generally, if my models are in fact close to being correct, reality lies beyond the domain of applicability of many universally accepted statistical tests, and we should expect to find that these tests will conclude against my models' validity. All too often, strange as it may sound, the conclusions yielded by such statistical criteria evaluate both the model and the test in some inextricable combination, from which little of use can be inferred.

Once again, while statistics works out the above challenges, we have no choice but to rely on graphics. It is worth noting that fully fleshed-out and detailed pictures – not skeletal diagrams – put *no* premium on concision, therefore on compression. But they put a heavy premium on the ability of the eye to recognize patterns that existing analytic techniques were not designed to identify or assess.

We discussed only a small aspect of a wider shift that computer graphics brought on the role of pictures in the hard sciences (and in my own life). The question of whether "in the beginning" was the word or the picture is a fascinating topic best discussed elsewhere.

3. "RANDOM WALK DOWN THE STREET," MILD RANDOMNESS, BROWNIAN MOTION (THE "1900" MODEL), AND MARTINGALES

Scaling models are meant to replace the simplest model of price variation, which Malkiel 1973 breezily called "Random walk down the street." Every version assumes that prices change randomly and each price change is *statistically independent* of all past ones. The probabilists' original random walk proceeds in equal steps, up or down, equally spaced in time. In Figure 3, the steps are so small as to be indistinct.

Another basic version assumes that price changes follow the Gaussian ("bell curve") distribution, which allows for a "mild" level of scatter. Typical generalizations assume or imply that individual price changes need not be Gaussian, but are only mildly scattered. The technical meaning of the term "mild" is sketched in Chapter E2 and described in Chapters E5 and E6. Quite appropriately, those examples interpret the word *walk* to denote a motion that proceeds in *steps*, while the alternative M 1963 model proceeds in *jumps*.

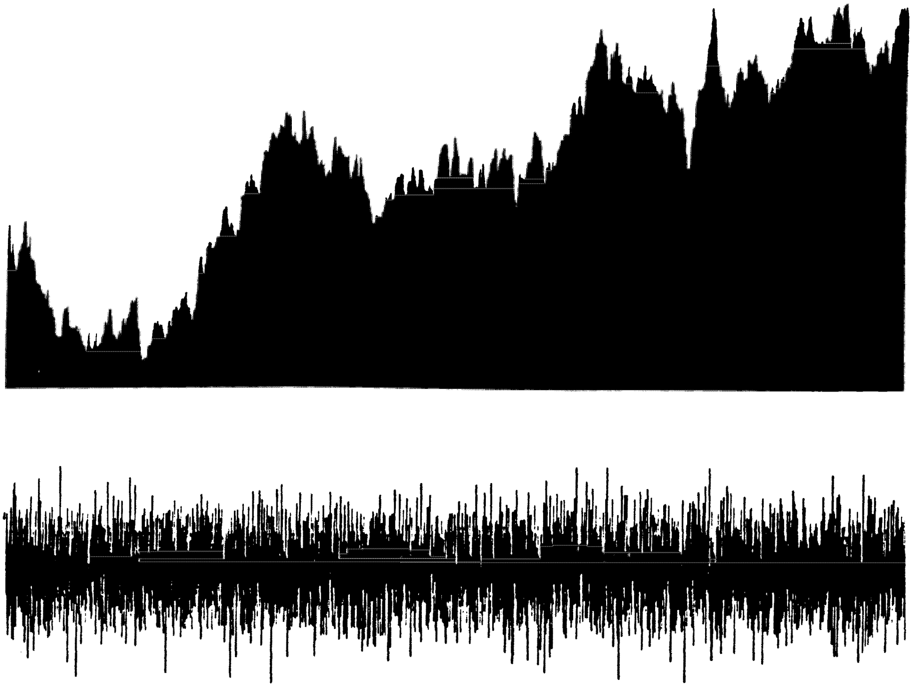


FIGURE E1-3. Graph of a sample of Brownian motion (top), and its white noise increments in units of 1 standard deviation (bottom).

3.1 The "ordinary" Wiener Brownian motion

The continuous-time counterpart of random walk was advanced in Bachelier 1900, is now called *Brownian motion,* and will be denoted as $B(t)$. Bachelier's discovery of Brownian motion in financial speculation occurred years before physicists discovered it in the motion of small particles, and decades before a mathematical theory of $B(t)$ was provided by Norbert Wiener. The tale is recounted in M 1982F{*FGN*}, p. 392. Hence, the composite term *Wiener Brownian motion* (WBM) will be used in case of ambiguity, in particular, when one must provide contrast with *fractional* Brownian motion (Section 7); a term used on occasion is "B 1900 model".

The main properties of Wiener Brownian motion are best listed in two categories, as follows.

3.1 Invariance properties of Wiener Brownian motion

Invariances are familiar in the hard sciences. Thus, classical geometry begins by investigating what can be done when the shapes one deals with reduce to lines, planes, or spaces. And the simplest physics arises when some quantity such as density, temperature, pressure, or velocity is distributed in homogeneous manner. The line, plane or space, and the homogeneous distribution on them, are *invariant* under both *displacement* and *change of scale*; in technical terms, they are both *stationary* and *scaling*.

Both properties extend to Wiener Brownian motion.

• *Statistical stationarity of price increments.* Equal parts of a straight line can be precisely superimposed on each other, but this is not possible for the parts of a random process. However, samples of Wiener Brownian motion taken over equal time increments can be superimposed in a statistical sense.

• *Scaling of price.* Moreover, parts of a sample of Wiener Brownian motion corresponding to non-overlapping time increments of *different* durations can be suitably rescaled so they too can be superimposed in a statistical sense. This key property implements the principle of scaling: except for amplitude and rate of change, the rules of higher- and lower-frequency variation are the same as the rules of mid-speed frequency variation.

3.2 More specialized properties of Wiener Brownian motion

Stationarity and scaling do not suffice to determine Brownian motion. It also has the following properties.

• *Independence of price increments.* Knowing the past brings no knowledge about the future.

• *Continuity of price variation.* A sample of Brownian motion is a continuous curve, even though it has no derivative anywhere. (A technicality deserves to be mentioned once: the above properties only hold almost surely and almost everywhere.)

• *Rough evenness of price changes.* The eye and the ear are more sensitive to records of changes than of actual values. A record of Wiener Brownian price changes, over equal time increments Δt, is a sequence of independent Gaussian variables. A telling term for this process is "white noise." The ear hears it like the hum on a low-fidelity radio not tuned to any station. The eye sees it as a kind of evenly spread "grass"that sticks out nowhere. A telling sample is shown in the bottom of Figure 3.

• *Absence of clustering in the time locations of the large changes.*

• *Absence of cyclic behavior.*

3.3 The martingale assumption, pro or con

Another basic property of $B(t)$ must be discussed separately. Bachelier 1900 introduced $B(t)$ as the easiest example he knew of a far broader class of processes now called *martingales,* which embody the notion of "efficient market" and successful arbitraging.

Prices are said to follow a martingale if they somehow acquire the following desirable property: whether the past is known in full, in part, or not at all, price changes over all future time spans have zero as expectation.

This definition allows properties other than the expectation to depend upon the past (as it does in M 1966b{E19}.) The notion of margingale was a bold hypothesis and a major breakthrough in 1900 and eventually received wide notice, witnessed by Cootner 1964. It does remain attractive and enlightening, but raises serious difficulties.

A first difficulty is this. A positive martingale always converges; that is, it eventually settles down and ceases to vary randomly (Samuelson 1965.) Conversely, a martingale that continues to vary randomly must eventually become negative. For example, a random walk eventually becomes negative. True, a "proportional effect" argument makes it customary to postulate that it is the *logarithm of price* that is Brownian, or at least a martingale. When this is the case, price itself cannot become negative, but ceases to be a martingale. Therefore, the "efficient market" justification for martingales disappears.

A second problem is more serious. While a martingale implements the ideal of an efficient market, is it possible to implement it by arbitraging, even under ideal conditions? The answer is *yes* under the conditions postulated in M 1966b{E19}. But the answer is *no* in M 1971e{E20}, which postulates that non-arbitraged price follows fractional Brownian motion, a generalization of $B(t)$ to be discussed in Section 7. This function is not a martingale, and *cannot* be arbitraged to become one.

4. BROWNIAN MOTION'S INADEQUACIES AS MODEL OF PRICE VARIATION, AND SKETCH OF PROPOSED REPLACEMENTS

Brownian motion is far and away more manageable than any alternative. An immense mathematical literature grew around it, and recently developed "financial mathematics" draws extremely long mathematical inferences from the assumptions that prices follow a martingale, and/or that $B(t)$ applies very exactly to prices. Unfortunately, $B(t)$ is an extremely poor approximation to financial reality. Soon after 1900, Bachelier himself saw that the data are nonGaussian and statistically dependent. When the model in Bachelier 1900 was "rediscovered" and confronted with reality, those discrepancies were independently observed by many authors. Thus, Osborne 1963 describes trading as tending to come in "bursts" and Alexander 1964 concluded that price variation is non-stationary. In the editor's comments of Cootner 1964, p. 193, one finds the suggestion that the bursts may be linked to the model of Berger & M 1963; not surprisingly, I had the same idea, found it difficult to implement, but implemented it in due time – as will be seen in Section 8.

4.1 A list of discrepancies between Brownian motion and the facts

• *Apparent non-stationarity of the underlying rules.* The top diagram in Figure 1 is an actual record of prices. Different pieces look dissimilar to such an extent that one is tempted *not* to credit them to a generating process that remains constant in time. While a record of Brownian motion changes looks like a kind of "grass," a record of actual price changes (bottom diagram of Figure 1) looks like an irregular alternation of quiet periods and bursts of volatility that stand out from the grass.

• Repeated instances of *discontinuous change.* On records of price changes, discontinuities appear as sharp "peaks" rising from the "grass."

• Clear-cut *concentration*. A significant proportion of overall change occurs within clear-cut periods of high price variability. That is, the "peaks" rising from the "grass" are not isolated, but bunched together.

• Conspicuously *cyclic* (but not periodic) *behavior*. For example, the real price series shown in Figure 1 shows conspicuous "cycles."

It will be seen that the preceding discrepancies can be traced to two characteristics of a more theoretical nature.

• *The long-tailed ("leptokurtic") character of the distribution of price changes*. An especially sharp numerical test of the instability of the sample variance is provided by the analysis of cotton data in Figure 1 of M 1967j{E15}: over 50 sub-samples of 30 days, the sample variance ranged a hundred-fold.

• *The existence of long-term dependence.*

4.2 The Noah and Joseph effects, taken singly or in combination

Those and other flaws of Brownian motion are widely acknowledged, but the usual response is to disregard them or to "fix" WBM piecemeal, here and there, as needed. The resulting "patchwork" is discussed in Section 4 of Chapter E2. My approach is very different. Instead of seeking a grand "model of everything," I moved in successive piecemeal steps, adding generality and versatility as suitable tools became available.

This strategy began by tackling nonGaussian tails and long dependence separately. Reflecting two stories in the Bible, those of the Flood and of the Seven Fat and Seven Lean Cows, the underlying phenomena were called, respectively, *Noah* and *Joseph Effects* (M & Wallis 1968{H}). The M 1963 model (Section 6) concerns cases where serial dependence is unquestionable, but a stronger driving feature resides in non-Gaussianity. The M 1965 model (Section 7) concerns cases where non-Gaussianity is unquestionable, but the strongest driving feature resides in serial dependence.

Because of their simplicity, the M 1963 and M 1965 models remain instructive and essential, but they are obviously oversimplified. The M 1967 model simply rephrases the symmetric case of the M 1963 model, and the M 1972 model goes further and tackles non-Gaussianity and serial dependence simultaneously. The models in Sections 6 and 7 generalize Brownian motion in two directions one may call orthogonal to each other, and Section 8 brings those two generalizations together again, as special cases of the M 1972 model. Ways to separate Noah and Joseph features in a record of real data are tackled at the end of Section 7.

4.3 It is fruitful to emphasize "the exceptional", even at the cost of temporary and comparative neglect of "the typical"?

The distinction between the typical and the exceptional is ancient, and my stress on discontinuity and concentration has been criticized. Clearly, when faced with rare events, Man finds it difficult to avoid oscillating between overestimation and neglect.

Most common is a stress on the typical. It motivated Quételet 1835 to his concept of "average man," and we read the following in the Preface of a famous treatise, Marshall 1890. "Those manifestations of nature which occur most frequently, and are so orderly that they can be closely watched and narrowly studied, are the basis of economic as of most other scientific work; while those which are spasmodic, infrequent, and difficult of observation, are commonly reserved for special examination at a later stage: and the motto *Natura non facit saltum* is specially appropriate to a volume on Economic Foundations ... [T]he normal cost of production ... can be estimated with reference to 'a representative firm' ..."

At first, these words seem to contradict an opinion expressed by Jacques Hadamard, that "it is the exceptional phenomena which are likely to explain the usual ones." But the case of a very concentrated industry suggests that the two viewpoints need not be in contradiction. Many believe, as I do, that emphasis on the largest firms agrees with Hadamard's opinion when "exceptional" is interpreted as meaning "concerning few entries in a list of all firms." But it agrees with Marshall's when "representative" is interpreted as meaning "concerning a large proportion of persons on the list of all the employees of those firms."

5. INVARIANCE PRINCIPLES: STATIONARITY AND SCALING

The fractal approach to finance and economics rests on two features.

One is a profound faith in the importance of invariances and in the possibility of identifying stationarity and scaling as invariance principles in economics. This will be elaborated upon in Chapter E2.

The second feature is the recognition that probability theory is more versatile than generally believed, and the willingness to face several distinct "states of randomness." When suitably chosen, a scaling random process can allow variation that will be described as "wild," a term to be explained in Chapter E5. The sample functions of wild processes contain significant features that were *not* deliberately incorporated into the input, yet, without a special "fix," achieve one or several of the following properties.

• *Repeated instances of sharp discontinuity* can combine with *continuity*. The fact that non-Brownian scaling random processes can be discontinuous and concentrated is extraordinarily fortunate; it is not a mathematical pathology that would be source of concern.

• *Concentration* can automatically and unavoidably replace *evenness*.

• *Non-periodic cycles* can automatically and unavoidably follow from long-range statistical dependence.

5.1 Principles of invariance

In mathematics and physics, such principles are a staple and the key to a wonderful wealth of consequences drawn from one simple idea. But they are not an established part of economics. I recall the eloquence of Jacob Marshak (1898-1977), when proclaiming that the single economic invariance he could imagine concerned the equality between the numbers of left and right shoes, ... and even that could not be trusted. Marshak was doubtless thinking of the basic invariance of mathematics and theoretical physics, which are stated as absolute. However, as already mentioned in Section 2, other parts of physics find it extraordinarily useful to work with invariances that are approximate and have a limited range of applicability.

Thus, I propose to abandon Wiener Brownian motion as a model, but endeavor to preserve stationarity and scaling as basic invariance principles.

5.2 Scaling under an especially critical form of conditioning

The probabilists' notation used throughout this book. This notation represents random elements by capital letters, and their actual values by corresponding lower case letters. Pr{"event"} will denote the probability of the "event" described between the braces. EX will denote the expected value of the random element X. Furthermore, this book uses words like *scaling, Gaussian, lognormal* as substantives, to mean scaling, Gaussian or lognormal *distributions.*

The scaling distribution. As applied to a positive random variable, the term *scaling,* is short for *scaling under conditioning.* To condition a random variable U specified by the tail distribution $P(u) = \Pr\{U > u\}$, suppose that it becomes known that U is at least equal to w. This knowledge changes the original unconditioned U to a conditioned random variable W. Using a vertical slash to denote conditioning, the tail distribution of W is

$$P_W(u) = \Pr\{W > u\} = \Pr\{U > u \mid U > w\} = \frac{P(u)}{P(w)}.$$

Now take tail distribution $P(u) = Cu^{-\alpha} = (u/\tilde{u})^{-\alpha}$. When $w > \tilde{u}$, conditioning yields $P_W(u) = (u/w)^{-\alpha}$. This expression is functionally identical to $P(u)$. The sole response to conditioning is that the scale changes from $\tilde{u}$ to w. Hence, the tail distribution $P(u) = Cu^{-\alpha}$ is denoted by the term *scaling*. In this wording, as seen in many places in this book, "Pareto's law" is an established empirical finding that states that the "frequency distribution of personal income is scaling." Conversely, $P(u) = Cu^{-\alpha}$ is the only distribution that is scaling under this particular conditioning.

By the logarithmic transformation $V = \log_e U$, this invariance property reduces to a well-known invariance property of the exponential distribution $\Pr\{V > v\} = \exp[-\alpha(v - \tilde{v})]$. Conditioned by $V > w > \tilde{v}$, the tail distribution becomes $P_w(v) = \Pr\{V > v \mid V > w\} = \exp[-\alpha(v - w)]$, which is identical to $\Pr\{V > v\}$, except for a change of *location* rather than *scale*.

Starting from an exponentially distributed V, a scaling U is obtained as $U = \exp V$. The logarithmic transformation is simple, the exponential is well-known, and the passage from V to U is obvious. Therefore, one may presume that no conclusion that is at the same time new *and* interesting can be obtained concerning the scaling U. The interesting surprise is that this presumption is totally unwarranted. The transformation from U to V raises new and difficult questions that go beyond technical detail to deep and concretely relevant issues. Therefore, the scaling property that follows from $P(u) = (u/\tilde{u})^{-\alpha}$ has far-reaching consequences. This book is largely devoted to studying them, yet is far from exhausting the topic.

Experimental measurement of the scaling exponent, and practical lack of meaning of high values of α. The exponent α is typically measured on the straight portion of a graph of $\log \Pr(U > u)$ versus $\log u$. The discussion that accompanies Figure 1 in M 1963p{E5} underlines that experimental work should give little or no credence to high values of α. Such values are entirely determined by observations for which the range of values of $\log u$ is small, making it harder to ascertain the straightness of doubly logarithmic graphs, and errors in α are far larger when α is large than when α is small.

5.3 Wild randomness and its surprising creativity

Stationarity and scaling suffice to derive many facts from few assumptions. Ex-ante, as already observed, they seem simple-minded. Ex-post, they turn out to be surprisingly "predictive" or "creative," in a sense developed in Section 10 of this chapter and elsewhere in this book. Implemented properly and helped by the evidence of computer graphics, they suffice to account for an extraordinary wealth of complicated behaviors, all bound together in what will be described as a "tightly organized phenomenology." The surprising possibility of such organization is essential from the viewpoint of "understanding," and Section 10 will argue that it is a second best to full explanation. This "creativity" is the second feature underlying the fractal approach to finance, and expresses a mathematical possibility that is central to every aspect of fractal geometry.

More specifically, the mathematical concept of *stationary randomness* is far less restrictive than generally believed. Properly tuned, it generates structures whose richness is well beyond the power of Brownian and near-Brownian randomness, which will be described as "mild." That is, stationary and scaling processes also extend to the very different forms of randomness that will deserve to be described by the provocative term, "wild." What we shall see is that many of the observed facts that motivate other writers to propose diverse "fixes" to Brownian motion (see Section 4 of Chapter E2) can also be accounted for by suitable wild randomness.

6. FRACTALS IN FINANCE, STAGE I: THE "M 1963"MODEL FOR TAIL-DRIVEN VARIABILITY AND THE "NOAH EFFECT"

The M 1963 model assumes that successive price changes are independent and highly non-Gaussian but stationary and scaling. In practice, it adequately addresses price records in which the long-tailedness of the changes is dominant, and their serial dependence can be studied as a later and closer approximation. This situation turned out to be a good approximation for commodity prices and other examples examined in M 1963b{E14} and M 1967j{E15}.

Given a price series $Z(t)$, write $L(t, T) = \log_e Z(t + T) - \log_e Z(t)$. The M 1963 model assumes that $L(t, T)$ follows a probability distribution called *L-stable*. When successive $L(t, T)$ are independent, $\log_e Z(t)$ is said to follow a random process called *L-stable motion* ("LSM"). The significant parameter is an exponent α ; its range could be $[0, 2]$, but in the case of price changes, it narrows down to [1, 2]. Wiener Brownian motion is the very atypical limit case of L-stability for $\alpha = 2$.

The limitation to $\alpha < 2$ is a significant irritant. It makes L-stability inappropriate for certain prices, and perhaps also for certain forms of income investigated in M 1963i{Appendix IV to E10}. Section 8 will show how a generalized model extends the range of α beyond 2.

6.1 The original evidence for the M 1963 model: the case of cotton

Figure 4, which provided the earliest evidence, first appeared in M 1962c. It was promptly reproduced with detailed explanations in M 1963b{E14}, then in many references including p. 340 of M 1982F{*FGN*}. This original empirical test of L-stability used Pareto-style log-log plots. The following description is translated in slight paraphrase from M 1962c.

"Denote by $Z(t)$ the spot price of cotton, namely, the price for immediate delivery on day t.

"Curves 1a and 2a represent, for the period 1900-1904, the empirical frequencies $\text{Fr}\{L(t, T = \text{one day}) > u\}$ and $\text{Fr}\{L(t, T = \text{one day}) < -u\}$.

"Curves 1b and 2b represent, for the period 1944-1958, the empirical frequencies $\text{Fr}\{L(t, T = \text{one day}) > u\}$ and $\text{Fr}\{L(t, T = \text{one day}) < -u\}$.

"Curves 1c and 2c represent, for the period 1880-1940, the empirical frequencies $\text{Fr}\{L(t, T = \text{one month}) > u\}$ and $\text{Fr}\{L(t, T = \text{one month}) < -u\}$.

"Both coordinates are logarithmic for all nx curves. To my knowledge, the evidence concerning price variation was never presented in this way.

"Those various curves quickly become straight lines having the same slope of approximately $\alpha = 1.7$. Therefore, we can write

$$\log [\text{Fr}\{L(t, T) > u\}] \sim -\alpha \log u + \log C'(T),$$
$$\log [\text{Fr}\{L(t, T) < -u\}] \sim -\alpha \log u + \log C''(T).$$

"Thus Figure 4 suggests that the tails are asymptotically ruled by the scaling distribution (see section 5.2), with the same α exponent throughout. We also observe that $C' \neq C''$, which reveals a slight asymmetry. The average value of $L(t, T)$ is practically zero.

"We see that 1a is parallel to 1b, and 2a is parallel to 2b. This shows that between 1904 to 1958 the distribution of $L(t, 1)$ did not change, except for scale. There is also evidence (not shown here) that the distribution of $L(t, 1)$ changed little from 1816 to 1940. The fact that curves 1a and 1c and

2a and 2c are parallel shows that the distributions of $L(t, T = \text{one month})$ and of $L(t, T = \text{one day})$ are identical, except for a change of scale.

"In a first approximation, the six curves displayed in Figure 4 can be superposed on each other by horizontal translation, showing that the distribution of $L(t, T)$ is *L-stable under change of T*. This feature will be interpreted as a strong quantitative symptom of scaling."

Deviations from exact superposition are full of meaning, as shown in Chapter 14, both in the text which reproduces M 1963b, and in Appendix III which reproduces M 1972b.

6.2 A deviation from invariance can be significant statistically, without being significant scientifically

As mentioned in Section 2, careful testing will doubtless show that prices exhibit statistically significant deviations from stationarity and scaling.

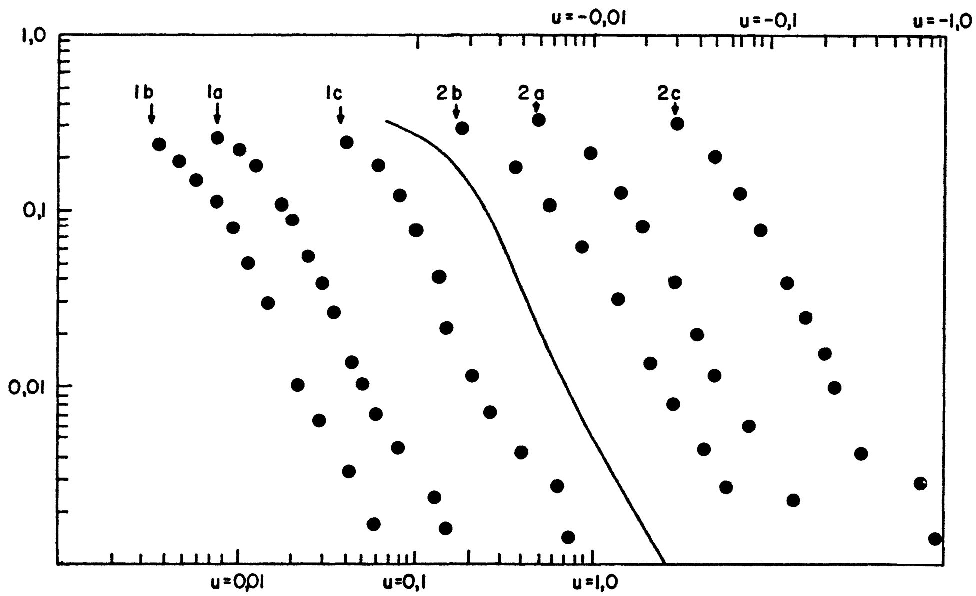


FIGURE E1-4. These figures on the spot price of cotton are explained in the text. They provided the first empirical evidence of scaling in finance. Reproduced from Figure 5 of M 1963b{E14}.

To elaborate, each scaling model of price variation claims to describe properties that apply (up to size factors) to effects at short, middle or longish time scales. Moreover, the M 1963 model involves one basic parameter, α. Specific properties such as a probability of a given kind of ruin can be evaluated, as seen in Section 3.2 of Chapter E6, and the results can be affected dramatically by the value of α. Thus, in the range around $\alpha = 1.7$, the probability of ruin may be *approximately* 10^{-1}, while in the (Brownian) limit case $\alpha = 2$, it may be *exactly* 10^{-20}. Now, what about the actual α? By visual inspection, the cotton prices yielded $\alpha = 1.7$, but one must expect that short, medium and longish time spans will yield quantitative estimates of α that slightly differ from each other. Hence, the above-mentioned probability of ruin may, in fact, differ from 10^{-1} by a factor ranging between 1/2 and 2, or perhaps even between 1/3 and 3.

As seen in Chapter E16, my prudent vagueness about the value of α was criticized by P. H. Cootner. I reported those reservations to William S. Morris (who will be quoted in Chapter E16), and he had no difficulty convincing me that the resulting uncertainty about the probability of ruin pales into insignificance. There is a far greater difference between the uncertain value of "approximately 10^{-1}" relative to $\alpha \sim 1.7$, and the certainly incorrect value of "exactly 10^{-20}" relative to $\alpha = 2$. It is fair to criticize the M 1963 model for being insufficiently precise, but only after praising it for providing a correct order of magnitude.

6.3 Beyond the M 1963 model

Be that as it may, the M 1963 model was pointedly only meant to to account for *certain* prices. Given the messiness of the data, it would be reckless not to fear that strict invariances are never encountered. But my systematic policy is to first seek improved models that preserve and generalize stationarity and scaling.

Sections 7 and 8 will, I hope, convince the reader that this has been a fruitful strategy. But Chapter E2 will argue that a sensible person should expect some features of the markets to contradict scaling. That is, ad-hoc "fixes" or "touch-ups" may eventually become necessary. But, as Section 4 of Chapter E2 will argue, those fixes must not be applied to Brownian motion, but instead to a model that is already part way to the truth.

7. FRACTALS IN FINANCE; STAGE II: THE"M 1965"MODEL FOR DEPENDENCE-DRIVEN VARIABILITY AND THE "JOSEPH EFFECT"

M 1963b{E14}, where the M 1963 model was first described, specifically acknowledges the existence of serial dependence in price changes, but the model itself approximated by postulating independence. This attitude was criticized for many reasons. In particular, every form of so-called "static" description is viewed as less desirable than a "dynamical" one that promises to be a possible basis for both portfolio management and conceptual understanding. I do not share this scorn for statics, but moved beyond the M 1963 model in several steps. In a first step, M 1965h{H} addressed records in which change is dominated by global (long-run) dependence and the deviation of the margins from the Gaussian can be studied separately and later.

The M 1965 model has a generic and specific aspect. Generic aspect: it introduces infinite memory into statistical modeling. Specific aspect: it introduces fractional Brownian motion ("FBM"), a process that has one significant parameter: the Hurst or Hölder exponent H satisfying $0 < H < 1$. The Wiener Brownian motion WBM is the atypical special case corresponding to the value $H = 1/2$. Early references on FBM are M & van Ness 1968{H} and the papers by M & Wallis{H}; the use of FBM in economics was pioneered in M 1970e, M 1971n, M 1971q, M 1972c and M 1973j. Bras & Rodriguez-Iturbe 1993, Baran 1994 and Samorodnitsky & Taqqu 1944 (Section 7.2) are among many textbooks that discuss FBM.

The original empirical test of long-run dependence, once again, used Pareto-style log-log plots, but did not apply them to the tail distribution but instead, to either the correlation or the spectrum. The latter take a very characteristic scaling form, described as "$1/f$," which is mentioned in Chapter E6 and discussed in detail in M 1997H and M 1997N.

7.1. Concrete justification for the idea of infinite memory, through the behavior of high-dimensional systems, and a metaphor from physics

While FBM implies an infinite memory, it may be reassuring that a model based on FBM *need not* imply belief in action at a distance.

Indeed, consider a very high-dimensional system (physical or economic) that is Markovian when viewed in its full glory. Such a system's one-dimensional or few-dimensional coordinates *need not* be Markovian at all. To think of them as following FBM involves no paradox whatsoever.

A useful metaphor is suggested at this point by the statistical physics of magnets. Infinite range dependence controlled by power-law expressions is the rule in systems such that actual interactions only occur between immediate neighbors. Those systems must be of high enough dimensionality and observed under conditions that physicists describe as "critical".

Be that as it may, infinite memory in finance calls for explanation. It also proved to require a new frame of thinking, but, to my delight, was accepted more readily than infinite variance. To my knowledge, no writer went as far as P.H. Cootner did, when (see the Preface) he described the M 1963 model as promising "blood, sweat and tears."

7.2. Historical digression: the hydrology connection

The intellectual path that led from LSM to FBM brings light on the similarities and differences between the M 1963 and M 1965 models, therefore remains interesting. When I was a Visiting Professor of Economics at Harvard and it became known that I was able to deal with the "pathology" of price variation, I was flooded with examples of other pathologies. Most proved beyond my skills, but two exceptional "hits" led me far away from finance for a while.

In 1962, a pattern of very anomalous noise led to Berger & M 1963{N5}, which implicitly introduces the notion of *fractal time* to which we shall turn in Section 8.

In 1963, hearing of the "Hurst puzzle" of hydrology (Hurst 1951, 1955, Hurst et al.1965), I identified it immediately as a new example of scaling, and briefly believed that it required a straight replay of the M 1963 model. But this beautiful theory was soon demolished by a mere fact: while LSM was the proper tool to deal with price variability driven by long tails, yearly river discharges are *not* far from Gaussian. This led me to conclude that the Hurst puzzle was driven by the accumulation of variables that may even be Gaussian, yet exhibit serial correlation of infinite time span. The form of infinity raises many questions, but they are best discussed in Section 4 of Chapter E2 while tackling *ARMA* representations.

7.3. The widespread confusion between the M 1963 and M 1965 models

Such confusion occurs despite the sharp differences between the Noah and Joseph Effects and the LSM and FBM processes. In the M 1963 (LSM) model, depending on which feature is singled out, fractal dimension is either $D_G = 2 - 1/\alpha$ or $D_T = \alpha$. In the M 1965 (FBM) model, depending on which feature is singled out, fractal dimension is either $D_G = 2 - H$ or $D_T = 1/H$. Even a competent mathematician sees between those two processes a number of parallelisms that some describe as "mysterious."

Help is on the way. Section 4 of Chapter E6 generalizes the standard self-affine models further and presents the resulting family of possibilities in very graphic fashion. As a result, order and simplicity are restored, and confusion decreases.

The Noah and Joseph effects often coexist; this fact raises two issues. The following subsection sketches an effective way to disentangle the two effects' contribution to a given record, and Section 8 sketches a versatile and effective way to build random processes that combine long-tails and long-dependence in "tunable" proportions.

7.4. A way to disentangle the Noah and Joseph contributions to a record: *R/S* analysis of global dependence, and its application in finance

"*R/S* analysis" is concerned with the kind of global dependence that the eye perceives as clear-cut cycles having no determined periodicity. This statistical technique, very different from spectral analysis, originated in the work on river discharges that is described in several papers by M & Wallis; see also M 1975h. This method started attracting wide attention, as exemplified by Feder 1988, and is one of the main topics of M 1997H. The details lie well beyond the scope of this book, but it deserves a comment that paraphrases M 1970e.

"It is obviously important to know whether dependence in price change records vanishes, is positive or is negative. Unfortunately, the empirical investigations disagree, and all are unconvincing, because they invariably use statistical tools that imply that the underlying process is nearly Gaussian. Before any statistical test of dependence is used, its robustness with respect to infinite variance must be investigated. For this purpose, M & Wallis proposed R/S analysis, and I used it on financial data. It is not foolproof, in fact has not yet been extensively explored. But it should be added to the classical tools and promises to provide more applicable results. It measures the intensity of global R/S dependence by a single parameter J.

"From the viewpoint of R/S, all independent random processes and a variety of martingales behave identically in yielding $J = 1/2$. They can be called "*R/S* independent." A striking fact is that the notion of R/S independence is robust with respect to infinite variance. In the case of price changes, it can serve as a useful surrogate for market efficiency, which is far from easy to handle statistically.

"Using R/S, I analyzed interest rate series from Macaulay 1932, 1936, prices of commodities from various sources and series of daily and monthly returns on securities. I found that different kinds of "price" series fall into different categories.

"Certain prices, and also the rate of call money, exhibit global persistence with, for example, an exponent of $J = 0.7$. This result was expected: since call money was itself a tool of arbitraging, its price cannot itself be arbitraged to take advantage of inefficiency. Therefore, its behavior should follow closely that of the various exogenous quantities that affect the economy. There is strong evidence that economic time series other than price changes (Adelman 1965, Granger 1966) and various physical (e.g., climatic) triggers of the economy are globally persistent, and the J observed for call money rates is typical of exogenous economic quantities.

"At the other extreme, British Consuls, cash wheat and some securities have R/S independent increments. The reason for this behavior is unclear. The data may be dominated by what may be called "market noise." However, spot commodity prices are not subject to thorough arbitraging. As a result, the absence of persistence in wheat is a puzzle. An explanation may be sought in institutional features; the arbitraging that is present in future prices may have an indirect effect on spot prices.

"Intermediate cases that exhibit a small degree of global dependence include prices of spot cotton and many securities. Closer investigations showed in many instances that the observed R/S dependence is wholly due to small price changes, which are both more difficult and less worthwhile to arbitrage. Large changes are practically R/S independent, even though they occur at highly non-independent (clustered) instants of time.

This and some of my other results leave many issues open. In particular, it is questionable whether or not the actually observed dependence is precisely compatible with efficiency. It is also unknown why there are so many differences between different series, and so many series in which the dependence is negligible."

A warning: No less than spectral analysis, R/S is a delicate statistical technique. There are rumors that the "Hurst's exponent" has become well-known in finance. However, recent developments reveal that R/S is an even more delicate technique than I believed in the 1960s. See M 1997H.

8. FRACTALS IN FINANCE, STAGE III: THE "M 1967" AND "M 1972" MODELS; TRADING TIME AND THE "NOAH-JOSEPH" EFFECT

We are now ready to perform the crucial task of combining the non-Gaussian distribution of the M 1963 model with the dependence rule of the M 1965 model. The task took time and was not easy. Therefore, while this section is merely a preview of Chapter E6, it is unavoidably more technical than the rest of this chapter.

"Fractional Lévy flight" is a tempting obvious combination of scaling margins and long dependence. It is mathematically interesting, but fails to fit the actual records. Thirty years ago, its inadequacy set me to search for other broad methods in many different directions.

8.1 Uniform or variable Hurst-Hölder exponents, the distinction between physical (clock) and trading time, and the notion of compound process

The Noah-Joseph combination that best fits in this book is sufficiently general to include many important special cases: the B 1900 model, the M 1963 model without asymetry, and the M 1965 model. The key step is to introduce an auxiliary quantity called *trading time*. The term is self-explanatory and embodies two observations. While price changes over fixed clock time intervals are long-tailed, price changes between successive transactions stay near-Gaussian over sometimes long time periods between discontinuities. Following variations in the trading volume, the time intervals between successive transactions vary greatly. This suggests that trading time is related to volume, but testing this empirical relation should be separated from an exploration of the model itself. Perhaps one could save Brownian motion by allowing price change to be due to extraneous impulses that are bunched in clock time.

To provide an alternative motivation of trading time, let us summarize very informally some properties of existing models concerning the "order of magnitude" of the price change Δx over a time increment Δt. Since Δt is assumed to be small, we deal with *local* behavior.

- For the B 1900 model, $\Delta x \sim \sqrt{\Delta t} = \Delta t^H$. The exponent is time invariant and $H = 1/2$.

- For the M 1963 and M 1965 models, $\Delta x \sim \Delta t^{1/\alpha}$ or $\Delta x \sim \Delta t^H$, respectively. The exponent $1/\alpha$ or H is again time invariant but $\neq 1/2$.

Unifractality versus multifractality. Because their scaling exponent is unique, the preceding models can be called *uniscaling* or *unifractal*. The generalization to which we now proceed can, by contrast, be called *multi-*

scaling or *multifractal*, because it consists in allowing the exponent to depend on t, and to be chosen among an infinity of possible distinct values.

Since we deal with local behavior of small Δt, large or small values of $H(t)$ express, respectively, that $x(t)$ varies slowly or rapidly near the instant t. Trading time is an alterative way of thinking about this variability of the exponent H. One imagines that $x(t)$ varies more or less unformily in its own intrinsic time, but the latter varies non-uniformly in clock time.

The preceding two comments should suffice for motivation. As a strictly mathematical idea, every non-decreasing function $\theta(t)$ of physical time provides a formal representation of $Z(t)$ as a *compound* process $Z(t) = \tilde{Z}[\theta(t)]$. But the result will be a useless increase in complication, unless special circumstances prevail. In the spirit of my work and of this book, I took both $\tilde{Z}(\theta)$ and $\theta(t)$ to be scaling, namely self-affine. To be practical, the only case I examined thus far is where $\tilde{Z}(\theta)$ and $\theta(t)$ are statistically independent (see Section 9.) Specifically, I allowed $\tilde{Z}(\theta)$ to be a Wiener or fractional Brownian function, and $\theta(t)$ to be a fractal or multifractal time. As was hoped, the resulting generalization of Wiener Brownian motion provides a sensible approximation to interesting data that combine long tails and dependence. Let us take up the topic in historical sequence, which also corresponds to increasing difficulty.

8.2 Subordination and the "M 1967" model: the "symmetric M 1963" model is representable as a Wiener Brownian motion in fractal time

The simplest form of compounding was pointed out formally by H. M. Taylor (Section 1 of M & Taylor 1967{E21} and I went on to interpret it concretely (Section 2 of M & Taylor 1967{E21}). The shortened form "M 1967 model" will be used, but it is not meant in any way to distract from the merits of Howard M. Taylor.

The M 1967 concerns the special case where $\theta(t)$ is a random function with *independent increments*; for historical reasons, it is called *subordinator*. The mathematical aspects of the notion of "subordination" (due to S. Bochner) are discussed in several places in Feller 1950 (Volume II.) It came to play an important role in many aspects of fractal geometry, therefore the concrete aspects are discussed in detail in Chapter 32 of M 1982F{*FGN*}, where it is illustrated and interpreted in a variety of contexts.

Specifically, M & Taylor 1967 takes price to be a Wiener Brownian motion of trading time. In order for physical time to be a non-decreasing self-affine function of trading time, it is necessary for the graph of $t(\theta)$ to be the so-called Lévy devil staircase. This object is defined in Chapter 3 of M 1982F{*FGN*}, as the simplest randomized form of a Cantor devil staircase. The points where the staircase moves up form a "Lévy dust" characterized by an exponent that is the dust's fractal dimension, a concept to be sketched in Chapter E6. Each discontinuity of the inverted staircase corresponds to a step of the staircase and collapses a finite interval of trading time into an instant of physical time. Conversely, trading time followed as function of physical time reduces to a series of mutually independent jumps of widely varying size. Price followed as a function of physical time also undergoes jumps.

Surprisingly, the above procedure simply reproduces the symmetric M 1963 model. The exponent α of the L-stable motion is "fed in" by choosing a Lévy staircase of dimension $\alpha/2$. (Some messy details will be discussed momentarily.)

Devil staircases are standard examples of *self-affine fractals*, a concept described in Chapter E6. Therefore, a trading time ruled by a devil staircase is called a *fractal time*. Section 4 of Chapter E2 mentions that Clark 1973 preserved subordination, but with a trading time that is not fractal. M 1973c {Section 3 of E21} argued against Clark's non-fractal substitute, but never implied that M & Taylor 1967{E21} said the last word. Let us now proceed beyond.

A goal for generalizations of the M 1963 model: it is necessary to correct its unrealistic prediction, that large price changes are statistically independent ex-ante, therefore isolated ex-post. In the M 1967 model, jumps with independent positions and amplitudes are inherent to the definition of subordination. Unfortunately, such jumps are unacceptable in the study of finance. Clear-cut bunching of large price changes is noted in M 1963b{E14}, but could not be seriously taken into account until a natural solution presented itself in the altogether different context of the study of turbulence.

A stepping stone towards generalization of the M 1963 model. Since it brings no new prediction or property, M & Taylor 1967 is best described as providing a "representation." All too often, such formal representations are mathematically important, but of limited practical interest. A glowing exception, subordination opened the gates to generalizations to which we now proceed. They are new, even from the mathematical viewpoint.

An easily described generalization of the M 1967 model, replacing the Wiener Brownian motion by fractional Brownian motion. The result combines long tails and long-range dependence. It is defined by two main parameters:

an α exponent that is twice the α exponent of the Lévy staircase, and the H exponent of the subordinated B_H. Little is known about it.

8.3 Compounding and the M 1972 model: Wiener Brownian motion of multifractal time; the turbulence connection and the paradoxical character of the perception of infinite memory by the ear and the eye

The most direct replacement for subordination replaces fractal trading time by a construct that is more general and more richly structured (also, more complicated), called *multifractal time*.

The step from fractality to multifractality was first taken in my very first full publication on turbulence, M 1972j{N14}, to which we shall return momentarily. Today, every field takes this step near-automatically at some point in time, following a general pattern advocated in Chapter ix of M 1975o and in an entry on "relative intermittence " on p. 375 of M 1982F{*FGN*}. Broadly speaking, patterns that seem fractal in a first approximation tend on a second look to be multifractal.

Returning to M 1972j{N14}, it ends (p. 345 of the original) with the following words:

"The interplay ... between multiplicative perturbations and the lognormal and [scaling] distributions has incidental applications in other fields of science where very skew probability distributions are encountered, notably in economics. Having mentioned the fact, I shall leave its elaboration to a more appropriate occasion."

The concept first introduced in M 1972j{N14} is a family of many-parameter *multifractal functions* to be denoted by $M(t)$. They are non-decreasing and continuous but non-differentiable. Their increments are called *multifractal measures*, and Figure 5 reproduces parts of the original example in M 1972j{N14}, for two values of a basic parameter μ. As is typical of the most interesting multifractal measures, the corresponding integral $M(t)$ is represented by a graph that is monotone increasing but lacks the flat steps characteristic of a devil staircase. It follows that the inverse of $M(t)$ has no jumps; like $M(t)$, it is continuous but non-differentiable.

Originally, the increments of $M(t)$ were meant to model the gustiness of the wind and other aspects of the intermittence of turbulence. An earlier *fractal* model of gustiness assumed that the wind comes in sharp isolated peaks. M 1972j{N14}, M 1974f{N15} and M 1974c{N16} put forward a more realistic *multifractal* picture of the wind's gustiness. After a delay of fifteen years, experiments (largely performed or supervised by K. Sreenivasan) confirmed the validity of that picture.

In addition, however, Figure 5 reminded me instantly of something entirely different, namely Figure 1 of M 1967j{E15}, which represents the variance of cotton price increments over successive time spans. After a long delay, this initial hunch proves to be an astonishingly good approximation. It was not elaborated until recently and is published for the first time in this book. The elaboration will, nevertheless, be called the "M 1972" model. Thus, my theoretical views of turbulence in the wind and the stock market were immediately and completely parallel.

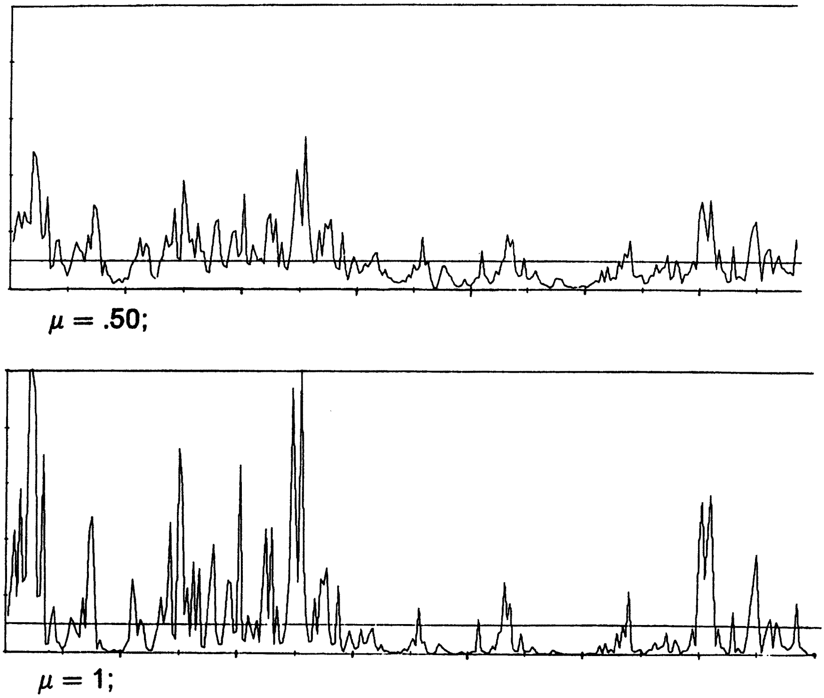


FIGURE E1-5. Both graphs are reproduced from Figure 1 of M 1972j{N14}; they illustrate the original multifractal measure for two parameter values.

As intended and achieved, this graph represents the variability of energy dissipation in a turbulent fluid. But it instantly brought to mind the graphs representing the variability of the variance of price increments, thus revealing a deep link between the uses of the fractal approach in the study of turbulence and of finance.

Graphs analogous to Figure 1 of M 1967j{E15} are, of course, very familiar in finance, and their ubiquity motivates the "patchworks of quick fixes" called ARCH models, which are outlined and criticized in Section 4 of Chapter E2. In the fractal context, to the contrary, the same resemblance immediately suggested a very different thought: a change in trading time from fractal to multifractal may generalize the M 1963 model.

Explanation, using acoustics, of the "unreasonable effectiveness" of infinite-memory models in accounting for bursts. ARCH-type models closely follow common sense. Even a casual look at diagram I of Figure 1 shows that large price changes are clustered. It is natural to attribute this fact to the presence of "high-frequency" serial dependence between price changes over neighboring time spans. Low-frequency serial dependence at a distance does not even come to mind.

To the contrary, my models start by accounting for very low frequencies, but they also succeed in accounting for the perceived high-frequency effects. Before we go on, we must establish that this paradoxical claim is not absurd. This is best done by injecting yet another physical science connexion, namely, some phenomena called $1/f$ noises. Among them, the first to come to mind is the derivative of FBM, simply because it is Gaussian, but the most appropriate illustration a definitely nonGaussian phenomenon called "flicker noise." As indicated by this name (or alternative names like "frying" and "popping"), the human ear perceives such a noise as a sequence of bursts separated by quieter periods. Hence, the ear seems to inform the brain that it deals with an "intermittent high-frequency phenomenon." Unfortunately, attempts to model this vague description failed. More importantly, spectral analysis yields a more objective diagnosis: a smooth spectrum concentrated in very low frequencies.

In a nutshell, the M 1972 model claims that the variance of $L(t, T)$ is a kind of flicker noise. As can be seen in M 1997N, I developed from 1964 to 1974 the technical know-how required to handle such noises, and the goal of the developments to be described from now on is to apply this know-how to finance.

8.4 From scaling to multiscaling: from marginal distributions that "collapse" to long-tailed distributions that shorten under averaging

Let us return to a sober examination of the marginal distribution of price change. "Data collapse" is said to occur when $L(t, T)$ has the same distribution for all values of T, except for scale. This is predicted by the M 1963 model and observed in Figure 4. But other price series proved to behave in a different and more complicated fashion (Officer 1972). For them, the distribution of $L(t, T)$ is reasonably close to being L-stable for small T, but the tails become markedly shorter than predicted.

The psychological impact of those findings was surprisingly strong among students of speculation. The consensus became that, as T increases, $L(t, T)$ might eventually converge to the Gaussian, so that the M 1963 model does not matter. As to the "transient" behavior before the Gaussian is reached, it was to be handled by the ARCH-type models to be discussed in Section 4 of Chapter E2, and/or other "quick fixes."

No one could dispute that the observed drift is incompatible with the combination of scaling with statistical independence of price increments. Therefore, the presence of a drift means that it is necessary to face serial dependence. The point is that this can be done using a multifractal scenario developed from the above-quoted remarks in M 1972j {N14}.

This replacement led to the M 1972 model whose most striking prediction is as follows. As $T \to 0$, the distribution of $L(t, T)$ becomes *increasingly* sharp-peaked and long-tailed. Qualitatively, this prediction matches the empirical evidence.

A technical illustration of drift away from collapse. A key feature of multifractality concerns the scale factors $\sigma(q) = \{E[L^q(t, T)]\}^{1/q}$. In the fractal case, the scale factors for $q < \alpha$ are powers of T, with an exponent independent of q, which is why this case is called *uniscaling*. In the multifractal case, to the contrary, the exponents of the scale factors depend on q, which is why this case is called *multiscaling*.

A useful mental picture is suggested by the limit lognormal multifractals introduced in M 1972j{N14}). The picture consists in a sequence of lognormal variables Λ_σ such that $E\Lambda_\sigma = 1$, while $\log \Lambda_\sigma$ is characterized by an increasing variance σ^2. The probability densities $p_\sigma(u)$ of those Λ_σ *cannot* be collapsed by using linear transformation. The M 1972 model predicts that the same is true of the distributions of $L(t, T)$ for different T. The resulting drift is slight if the exponent of the scale factor $\sigma(q) = \{E[L^q(t, T)]\}^{1/q}$ increases little and slowly as with $q \to \infty$. But the drift's intensity can be "tuned" at will.

The careful reader may observe that, while the lognormal probability densities $p_\sigma(u)$ cannot be collapsed linearly, the corresponding expressions $\log p_\sigma(u)$ are easy to collapse. This is pretty much all there is to the multifractal function $f(\alpha)$ to be mentioned in Section 3.9 of Chapter E6.

An alternative scenario behind the drift away from data collapse. The long-term dependence characteristic of multifractals implements the following scenario, which I often heard mentioned, but do not recall seeing in print. In the case of statistical independence, a price increment of small probability p cannot be observed, unless the sample is at least equal to a few times $1/p$. But suppose that larger changes of $L(t,T)$ are strongly clustered for all values of T. If so, a price increment of probability p can only be observed on a sample containing many more than $1/p$ roughly independent values. A shorter sample should be expected to include far too few values in the tails.

When a sample for $t = 0$ to $t = t_{max}$ is examined for increasing values of T, the sample size t_{max}/T decreases. Therefore, as T grows, the probability of hitting upon large deviations decreases. So does the histogram's tail.

8.5 Simulations of the multifractal M 1972 model, and empirical tests

As mentioned in the Preface, those tests were neither prompt, nor complete. Before they are described, it is good to restate a basic point already made in Section 2. The eye tells us that the behavior exemplified by Figure 1 is mimicked reasonably in Figure 2. We can now add an explanation: each line of Figure 2 is a separate implementation of a very simple "surrogate" to the M 1972 model – as explained fully in Section 4 of Chapter E6. "Fine tuning" the surrogate algorithm also tunes the output.

As to quantitative comparisons, I am overcommitted, and at present lack any competitive advantage in handling financial data. However, an exploratory study of foreign exchange rate changes is extremely promising. The results will be sketched in Chapter E6, after some technical tools are described. Full details will be published in free-standing form in M, Fisher & Calvet 1997, Calvet, Fisher & M 1997 and Fisher, Calvet & M 1997.

9. BEYOND ALL THE MODELS DESCRIBED IN THIS BOOK

Flaws of the preceding models should be immediately acknowledged.

Independently of whether fractional Brownian motion is followed in clock or multifractal time, its increments have a symmetrical distribution. Therefore, the M 1965, M 1967, and M 1972 models do not apply to cases where the distribution is clearly asymetric, for example to cotton price changes.

The compound process in Section 8 assumes independence between trading time as function of clock time and price as function of trading time. It will be nice to go beyond independence.

For the above reasons and others, the M 1972 model does not claim to be the last word. In fact, my investigation of the combination of long tails and long dependence also tried out two very different approaches that remain little explored and may reward a fresh look.

M 1966b{E19} describes an interesting and surprisingly realistic form of rational market "bubbles." The key is a form of arbitraging whose input has short tails and long dependence, while its output is a martingale with long tails and remaining long dependence.

The "fractal sums of pulses" {FSP} processes are described in M 1995n and several papers I co-authored with R. Cioczek-Georges; they are listed in the Bibliography. The FSP are used in Lovejoy & M 1985{H} to simulate the shapes of clouds, and also show promise in finance, but this book chose to explore a very different approach.

10. OLD-FASHIONED CHEMISTRY AS EXAMPLE TO EMULATE; VALUE OF "UNDERSTANDING" SHORT OF EXPLANATION

Section 4 of Chapter E2 will describe alternatives to the models announced in the preceding sections. Those alternatives are based on "fixes" of Brownian motion, to which I see many defects. Some are practical, while others may be called esthetic. In my view, even if an accumulation of quick "fixes" were to yield an adequately fitting "patchwork", it would bring no understanding. That concept will be taken up now, in the form of a methodological digression meant to contrast the fractal approach in finance and the piecemeal approaches using fixes.

At one extreme on the scale of perceived achievement in the sciences is the ideal of determinism and "reductionism," whose model is Newton's law of gravitation. To call a law "explanatory" is to give up any attempt to seek an even deeper truth behind it. Reductionism seeks general principles of independent value that account for ("explain") the individual observations, and also make correct predictions. The history of science tells us that this ideal was often achieved, mostly in physics, hence it is useful to keep it in mind. The goal of Quételet 1835 was to rework it into a "social physics." But in the study of finance, reduction proves elusive.

At the opposite extreme, on the scientific scale from the die-hard reductionists, one finds those resigned to accept that the variation of prices follows rules that continually change in time and are not written down. Such alternatives are so open-ended that they cannot be proven wrong, but they are non-scientific, and to accept them would be to admit that a theoretical and quantitative approach to finance is not possible.

The extremes of renunciation and reductionism bracket every phenomenological approach to finance, whether unorganized, like the step-by-step fixes described in Section 4 of Chapter E2, or tightly organized by insistent adherence to stationarity and scaling, like the models described in Sections 6 to 8.

The said "fixes" are close to the non-reductionist extreme, and unavoidably bring to mind the representation of the planets' trajectories provided by the *Ptolemaic system.* There, each feature was specifically inserted to account for some aspect of observed reality that earlier models neglected. The construction started with "cycles," then corrected for the cycles' inadequacies by adding "epicycles." When epicycles also proved inadequate, yet another fix moved the center of the cycle away from the center of the system! The final outcome gave a good fit to the data available in its time (Gingerich 1993). One could achieve perfect fit by adding epi-epi-cycles ad infinitum, and some historians of mathematics argue that the Ptolemaic system was an early step towards the representation of arbitrary almost-periodic functions by Fourier series.

The exclusive reliance on an accumulation of separate fixes, which characterizes the Ptolemaic system and the majority approach to finance, makes them examples of "unorganized descriptive phenomenology."

The fractal approach is closer to the reductionist extreme, and it would be tempting to take a last step and view scaling as that demanding no further explanation. But this is impossible, if only because scaling takes on many different forms, even in finance. In addition, only a few instances of scaling in the social sciences led to convincing explanations.

In physics, the situation is far more satisfying, as many instances of scaling are well-understood, at least in principle. Several are explained by full "renormalization" arguments, as alluded in Section 6 of Chapter E4. Additional fully implemented models exist, and special attention should be drawn to a widely valid reductionist explanation that I proposed in Chapter 11 of M 1982F{*FGN*}. The claim is that the solutions of the partial differential equations of physics invariably end by having fractal singularities.

Outside of physics, the situation is less rosy. A few examples of explanation, each a few lines long, will be given in Chapter E8. Other explanations of scaling and also a famous but (I think) grossly overrated "explanation" of lognormality will be discussed in Chapter E9. I feel, with regret, that when those arguments are fully understood, they cease to be convincing. Among other difficulties, some arguments justify the Gaussian probability distribution on the basis of a limit theorem of probability. Other explanatory arguments invoke steady states attained by in random processes, that is, the fixed points of corresponding transformations. All such approaches appeal to some kind of long-run. Macroscopic physics is concerned with assemblies of a colossal number of items and the long-run is meaningful and effective. But in finance or economics, such assemblies are not given time to develop, and (once again) references to them fail to convince.

Altogether, the fractal approach ro finance unavoidably brings to mind two distinct metaphors from different physical sciences.

The first brings in the organized phenomenology of Johannes Kepler. Ptolemaic fixes are not creative and bring no understanding whatsoever. To the contrary, *Kepler's laws* are "creative," insofar as they predict many consequences that no person had inserted by design. While short of explanation, they bring real "understanding." This last notion is important, or so I think, and is encountered in many observed instances of fractal behavior. It is not easy to define, even in mathematics, and calls for elaboration.

Let us be generous and allow that separate but satisfactory "fixes" may eventually be discovered for every defect of the Brownian motion in finance. If so, each fix will cry out for a separate explanation, and then the separate explanations will cry out to be organized. By contrast, fractality "bundles" everything from the start.

A second metaphor from physical science is even more compelling. From the reductionists' viewpoint, *chemistry* was not respectable until quantum mechanics transformed it, at least in principle, into a chapter of core physics. The question is, how should a scientist behave during a possibly interminable wait for reduction. Viewed as a temporary phenomenology, modest old-fashioned chemistry was remarkably well-organized, robust, and creative. In contrast to more ambitious efforts that ended in failure, chemistry developed into a body of knowledge and understanding that deserves to be viewed as a shining example, never to copy, of course, but to emulate.

E2

Discontinuity and scaling: their scope and likely limitations

✦ **Abstract.** Chapter E1 stated emphatically my view that Gaussianity, random walks and martingales are attractive hypotheses, but disagree with the evidence concerning price variation. This chapter presents, in largely non-mathematical style, the processes I propose as replacements for Brownian motion. Their foundations include an evidence-based theme and a conceptual tool.

The theme is discussed in Section 1: it is *discontinuity* and the related notions of *concentration* and *cyclicity*. The tool, *scaling*, is discussed in Section 2. The possible limitations of scaling expressed by *cutoffs* and *crossovers* are discussed in Section 3. Section 4 comments on alternative approaches that contradict scaling, and instead replace Brownian motion by a "patchwork" of step-by-step "fixes." Section 5 describes some paradoxes of scaling.

Stationarity and scaling express invariances with respect to translation in time and change in the unit of time. Diverse principles of invariance are essential to my work, in economics as well as in physics. ✦

IN A WAY, THIS CHAPTER IS AN "OVERFLOW" that helps prevent Chapter E1 from becoming unbearably long. The intention is to meditate in greater detail on several issues that passed by too briefly, and examine in passing some alternatives I perceive as unfortunate.

While the concepts of discontinuity, concentration and cyclicity are intimately connected, Section 1 tackles them one by one, contrasting them with the corresponding features of standard models: the Random Walk on the Street and Brownian motion.

Section 2 describes the principle of scaling. Scaling is present in finance when short pieces of a chart are like down-sized versions of longer pieces. The original evidence of scaling in cotton prices is an aspect of Figure 3 of Chapter E1. That chapter also points out that scaling is of great versatility and "creativity," that is, allows many sharply quite distinct possibilities. My strategy began by sharply restricting those possibilities in various increasingly general ways and studying the restricted processes in detail, one after another.

Section 3 follows M 1982F{*FGN*} in acknowledging the limitations that fractal geometry near-invariably sets to its own applicability.

Often, there is an *inner cutoff.* For example, trading never proceeds in mathematically continuous time, therefore a very short piece of the record includes few transactions and could not possibly look like a longer piece of the record.

An *outer cutoff* is also expected, right or wrong, if only because it is hard to believe that the rules of price variation are not modified as one moves from time scales dominated by "mere speculation," to longer scales dominated by more fundamental economic effects.

Thirdly, one scaling model may not suffice to account for all the data. That is, distinct ranges of the time increment may call for distinct scaling models, separated by *crossovers.*

Those limitations were acknowledged when the discovery of scaling in finance was reported in M 1963b{E14}. The real surprise resided in the existence surprisingly broad and clear-cut intermediate scaling zones governed by surprisingly simple rules.

Section 4 surveys a few of the many publications by other authors, which concern the same facts but attempt to account for them, not by scaling, instead, by a "patchwork" of "fixes."

Section 5 illustrates the "creativity" inherent to random scaling phenomena. It begins seriously, by describing the paradoxes of expectation that bedevil the scaling distribution, then moves on in jocular tone, through several fanciful but enlightening stories.

1. CONTRASTED BASIC TENETS OF THE RANDOM WALK ON THE STREET AND OF THE FRACTAL MODELS OF FINANCE

1.1 The Brownian dogma of continuity versus the often discontinuous character of actual market prices

Once again, the Bachelier 1900 model predicts that price varies continuously, and continuity is also a basic assumption in the overwhelming bulk of financial and economic literature. It is often left unmentioned because it is felt to be obviously appropriate. Technically, continuity is incomparably attractive because, were it a description of reality, the study of finance could borrow heavily from existing techniques of mathematics and physics.

As background to continuity, Alfred Marshall (1842-1924) gave to every edition of his *Principles of Economics* (Marshall 1890) the motto, *Natura non facit saltum,* "Nature does not undergo discontinuities." I may attribute too much weight to a famous but very old book, or misinterpret its intent through ignorance. But there is no question that Marshall's motto fits in a long tradition of philosophical thought that extends well beyond economics. It is the tradition of the *Great Chain of Being* favored by Aristotle and Leibniz (Lovejoy 1936; see also pages 405-8 and 412-13 of M 1982F{*FGN*}.) Marshall comes to my mind each time that an "unanticipated" price *saltus* defeats a beautiful trading scheme founded on continuity.

For contrast, consider the M 1963 fractal model, as proposed in M 1963b{E14} and sketched in Section 6 of Chapter E1. When interpolated to continuous time, this model predicts that price variation *must* be discontinuous within every interval of time. Many other fractal models mix continuous and discontinuous variation.

Of course, price quotes only include a finite (but increasing!) number of digits. Therefore, continuity can only mean that successive prices differ by amounts of the order of magnitude of an irreducible minimum. Smaller discontinuities embodied in a model are not part of reality, only a mathematical fiction. The concrete contents of the M 1963 model reflects an unquestionable fact: Instantaneous price changes far larger than the minimum are clearly visible on almost every financial chart.

Do the discontinuities actually hide a deeper "latent" reality that is continuous and can be "revealed" by proper data analysis? My view, to the contrary, is that the possibility of absolutely sharp discontinuity is an essential ingredient that sets finance apart from classical physics. If prices on financial markets are to fulfill their role, they *must* be able to change instantly when important new information becomes available. For example, when Company A is bidding to buy Company B, the price of either party must be able to change *on the spot* by an unboundedly large percentage. And that is indeed how prices behave; a small company's price that instantly doubles or halves is by no means an exception. And the sheer size of IBM did not prevent its stock from dropping by 10% near-instantly in early 1996, and later from rising by 13.2%.

If one is prepared to assume that it is legitimate to think of "latent" prices, one might well imagine that they are *even more* discontinuous than those recorded and shown by financial charts. After all, many markets employ "specialists," whom the United States Securities Exchange Commission enjoins to "insure the continuity of the market." While they cannot always fulfill this task, they successfully smooth out small discontinuities by buying or selling on their own account. In effect, they are enjoined to keep two sets of books; one corresponding to an undisturbed "latent" interplay of market forces, and a second reporting a smoother behavior.

Insofar as the fractal approach helps explain the need for the institution of specialists, it is "predictive," because durable and time-seasoned institutions are unlikely to arise without causes that are rooted in durable characteristics of the market. I see such a characteristic in discontinuity.

Here is a relevant story. The M 1963b model includes a "tunable" parameter α, such that continuity is achieved by choosing $\alpha = 2$, and any desired degree of discontinuity, by letting α decrease to 1. Therefore, the computer made it possible to perform the following multiple-choice test. Show to a sophisticated subject a series of sample plots of the same model, but with different values of α, and pretend that one plot is a genuine financial chart while the others are samples from a theory. Challenged to identify the real chart, the sophisticated subject never chooses the continuous sample function. Needless to say, as acknowledged in Section 2 of Chapter E1, arguments that involve the eye had come to be held in low regard in science. In particular, "eye-ball curve-fitting" came to attract deep scorn from statisticians. It tends to be viewed as subjective and far inferior to objective numerical fitting methods. Now that the eye has been greatly empowered by computer graphics, the safest is to use both methods together as needed.

1.2 Concentration of price change, versus the Brownian dogma of evenness

Consider a price that changed markedly over some long period of duration T days (or hours, or weeks...). Should one expect this total change to be more or less evenly distributed over the T contributing days? Or should one expect in many cases to find that the total change is concentrated in, or dominated by, a few contributing periods of particularly brisk change?

The original Random Walk on the Street proceeds in equal up or down steps, hence predicts that the absolute value of the daily step is precisely even. For larger T, both Random Walk and Brownian motion predict a less sharp form of evenness: a daily change that is the fraction $1/T$ of the total change, plus or minus a Gaussian "error term" that is independent of the total price change over T days. Thus, even if the total change is large, the Gaussian model predicts rough equality between the orders of magnitude of the contributions, each individual contribution becoming negligible as T increases.

The M 1963 model makes an altogether different prediction. When the total monthly change is large, not only the contribution from any single day will fail to be asymptotically negligible, but the contribution from *one or a few* days will most likely be of the same order of magnitude as the cumulative contribution of *all the other* days. Similarly, a large daily change will be dominated by contributions from one or a few contributing hours or minutes.

When this form of concentration is refined to apply to increasingly small time increments, it unavoidably creates discontinuities.

A fundamental distinction: difference between "ex-ante" identity in distribution and "ex-post" order-of-magnitude identity of samples. A subtle point is involved here. Both the Random Walk and the M 1963 model assume that T contributing price changes are identically distributed *ex-ante*. Ex-ante identity results in *ex-post* near-identity in the Brownian model, but fails to do so in the fractal models. Thus, the distinction between the concepts of ex-ante and ex-post is bound to become important beyond economics.

Musings concerning the "theory of errors:" the attractiveness of its foundations and its failure in the presence of concentration. Instead of disaggregating a price change into parts contributed by successive time intervals, one may disaggregate it into the effects of separate "causes." Centuries ago, observational astronomers developed a "theory of errors" that holds those effects to be numerous, additive and individually negligible, both ex-ante and ex-post. This is why errors are expected to follow the Gaussian distribution.

Similarly, one can (at least in a first approximation) view the separate possible causes of price change as numerous and additive. This being the case, consider the ex-post contributions of various causes to a large price change. In the classical (Gaussian) theory of errors, a large change would typically result from the rare chance simultaneity of many large contributing causes, each of them individually negligible. In economics, this inference is indefensible. Typically, the occurrence of a large effect means that one contributing cause, or at most a few turn out ex-post to be large. It may be worth reporting that I once heard an economist condemn *all* of statistics for being founded on the classical theory of errors, therefore on a misinterpretation of reality. I interrupted his tirade to point out that he was describing the phenomenon of concentration, which is indeed beyond the scope of elementary statistics books, but is a key feature of the M 1963 model of price variation.

In the same spirit but a different context, a sharp large change in the value of an index or portfolio is often overwhelmingly due to a very large change in one or a few components.

1.3 Concentration without discontinuity for dependent increments

Having correctly predicted concentration, the M 1963 model goes on to predict that concentration is mostly due to *isolated* discontinuities. In the real world, to the contrary, isolated discontinuities are not the rule. A more typical scenario involves so-called periods of "market turbulence" that include discontinuities following each other in close sequence, intertwined with small price changes.

Indeed, the strongest short argument to account for discontinuity (like IBM moving down, and later up, by 10%, then 13.2%) involves the fact that exogenous stimuli can be very large. But a large exogenous stimulus need not come out as one single piece of news. Thus, I like to speak of concentration as involving the Noah effect; but the Bibilical flood was not a discontinuity, since rain fell for forty nights and forty days. Furthermore, markets need not be able to respond instantly. Trading limits and liquidity problems may play a role. A realistic scenario is incompatible with independent price increments.

Based on this type of thinking, the last section of M 1963b{E14} acknowledges instances where a large change is not concentrated in one day, but spread over successive days. A toy model is even presented, but only tongue-in-cheek. The basic fact is that the M 1963 model cannot be expected to hold exactly. It shows that one needs a model with serial dependence.

It may be recalled that a very versatile way of introducing dependence while preserving scaling is sketched in Section 8 of Chapter E1 and discussed in Chapters E6 and E21. The idea is that price variation is best not followed in physical "clock time," rather in an auxiliary "trading time." To implement this idea in a scaling world, one must identify price variations as a scaling function of trading time, and trading time as a scaling function of clock time. The burden of accounting for discontinuity and concentration can be made to fall on the choice of trading time. M &

Taylor 1967{Section 1 and 2 of E21} noted that the M 1963 model can be represented as Brownian motion in "fractal time."

This last fact is only mildly interesting, but it suggests two very important generalizations. One replaces Brownian motion by Fractional Brownian motion. A further generalization is the M 1972 model, which allows Brownian or fractional Brownian motion to proceed in "multifractal" trading time that allows for both rapid but continuous changes and arbitrarily sharp discontinuities.

1.4 Practical consequences of discontinuity and/or concentration, concerning hedging and the size and nature of risk

Price discontinuity and concentration are major ingredients in a realistic evaluation of risks. To stress their importance, take a portfolio and compare the risks using the Random Walk on the Street and the M 1963 model. The former estimates the risks as small, say, one thousandth, one millionth or less, while the latter may estimate the same risk to be a hundredth, a tenth or more.

In particular, price continuity is an essential (but seldom mentioned) ingredient for all trading schemes that prescribe at what point one should buy on a rising price and sell on a sinking price. Being discontinuous, actual market prices will often jump over any prescribed level, therefore, such schemes cannot be implemented. An early example was very specific and clear-cut: it was Sidney S. Alexander's "filter method" (Alexander 1961). I viewed filters as "perpetual financial machines" that challenged the theorist to identify the subtle hidden flaw that seemed to make them work. That is, they reminded me that thermodynamics first showed its bite by pinpointing the subtle flaws in diverse mechanical "perpetual motions." M 1963b{E14} argued, indeed, that filters rely on the fallacy of continuity, and – on average – the nearest implementable trading method may yield no benefit. This prediction was confirmed, as seen in Appendix II of Chapter E14. The filters' use and repeated failure might have sufficed to discredit them, but one cannot be sure. In any event, it is good to know that their failure was precisely and correctly predicted before they had had time to be used, on the fundamental theoretical ground of continuity versus discontinuity.

1.5 One can account for the observed non-periodic cycles by stationarity combined with globality of dependence and intermittency

Statisticians and historians find it convenient to describe price records as involving random fluctuations that add to trends and a diversity of "cycles" of short, medium and long duration. Most economists view those cycles as significant, but Keynes asserted in jest that their main utility is to help long treatises of economic history be broken into manageable smaller volumes.

The following observation in M & Wallis 1969a suggests that Keynes had a strong point. The key finding concerns the fractional Gaussian noise, which is defined as the sequence of increments of fractional Brownian motion, and can range between independent events (white Gaussian noise) and Brownian motion. From the very beginning of the study of fractional Gaussian noise, a central ingredient consisted in graphics examined with great care, combining respect with skepticism. Using primitive tracing tables, M & Wallis 1969a synthesized and visualized long samples of fractional noise, and found it exhibits an already familiar feature: there is a fundamental difference between ex-ante absence of deliberately built-in cycles and ex-post obvious visual cyclicity. That is, the ex-ante generating mechanism involves no periodic component, nor any privileged time scale, only a built-in form of "global" or "long-run" dependence.

Surprisingly, ex-post examination of samples from this process revealed that *every* sample seems to exhibit three cycles. This striking "rule of three" is true for all sample durations, because it is an aspect of self-affinity, an implementation of the notion of scaling that will be discussed in Section 2. Now let us go back to data. Could it be a simple coincidence that the "long cycles" Kondratiev observed in a sample of a hundred-odd years consisted in three oscillations? Similar cycles are claimed to exist in weather and hydrological records, and it is in their context that I pioneered an indirect approach to long-run dependence that throws deep doubt on the "reality" of long cycles. See M 1997H.

If cycles' relation to self-affinity becomes confirmed, the basic problem will be displaced from determining cycles' lengths and making concrete use of them, to determining the rules of self-affinity and identifying their significance from the viewpoint of prediction and control.

1.6 Cause versus chance as accounting for the wealth of sharply defined features one observes in financial and economic data

Once again, every contrast between ex-ante simplicity and ex-post complexity exemplifies a surprising and fundamental theme of fractal geometry, mentioned repeatedly in this book: misleadingly simple-looking algorithms (random or even non-random) typically generate unexpectedly complicated but highly structured behavior. Even when there is *no* dependence between the increments, and *no* prediction is possible, a sample from a suitable fractal random process can exhibit features on which a technical analyst would base buy or sell recommendations. Other fractal random processes exhibit global dependence but involve no preferred time scale, yet generate swings reminiscent of economic cycles.

To provide an intellectual background to those findings, it is good to mention a parallel discovery I made concerning galaxies; M 1982F{*FGN*} describes this problem in Chapters 9, 32, 34 and 35. Random processes that involve no preferred spatial scale, hence no trace of built-in hierarchy, nevertheless generate patterns that the eye invariably and spontaneously organizes into a hierarchy of clusters and super-clusters.

2. THE PRINCIPLE OF SCALING IN FINANCE AND ECONOMICS, AND APPARENT LIMITATIONS THAT DISAPPEAR WHEN THE MODEL IS SUITABLY GENERALIZED

Scaling was on my mind when I was working on the distribution of income (Part III of this book), but it did not come into its own until 1963, when it imposed itself in two altogether different ways.

On the one hand, M 1963e{E3} describes a peculiar theoretical argument that concludes as follows: When seeking a statistical description of economic reality, the alternative one faces is not "scaling against a multitude of other possibilities," but "scaling against a form of lawlessness."

On the other hand, there is quantitative empirical evidence for scaling, under different forms that depend upon the context.

2.1 From practitioners' folklore to specific and verifiable statistical predictions

A central role is played in the preceding sections by examples of "folklore" and "anecdote". These strong words usually denote a mess of unverified qualitative observations that languish beyond the boundaries of organized scientific discourse. The bulk of it is nonsense, often the harmful fruit of fabrication or careless observing, reading or listening. But much of my scientific work is grounded on anecdotes and pieces of folklore that are incontrovertible and would not have remained beyond science, if a way had been known to fit them in. One such observation is *empirical scaling*, and was made separately in many different areas.

A first example far removed from finance and economics may help the reader view this statement with greater objectivity. Geologists are firmly instructed to take a multitude of photographs, to view them seriously, and *always* include an object of known size: a pencil, a hammer, or a colleague. They are warned by their teachers that, in the absence of such an object, the scale of the photograph may be impossible to guess. The counterpart in the field of finance is that, if one does not exert great care to label the axis of time and the axis of price, "all financial charts tend to look the same," and are readily mistaken one for another.

2.2 Quantitative evidence, old and new, of scaling in prices

The eye would not suffice to establish scaling, of course; in any event, M 1963b{E14} was written before computer graphics could help.

Original quantitative evidence for scaling in finance. Let us review the original evidence concerning the spot price of cotton, as presented in Figure 5 of M 1963b{E14}, which is reproduced as Figure 3 of Chapter E1. To judge whether or not "the financial charts are really the same," I chose the method of "data collapse," to use the physicists' term for a method many other scientists also use. Monthly and daily price changes were plotted suitably, and it was found that six separate curves can be superposed by performing translations that amount to changes of scale. This is close to the the standard way to reduce all Gaussian distributions to be of mean 0 and variance 1. It is the principle of scaling that suggested the suitable plotting procedure.

An example of recent evidence for scaling in finance: zero crossings. A standard problem of probability is raised by the distribution of the time interval between zero crossings. In a random walk or Brownian motion, this time interval is scaling with the exponent $\alpha = 1/2$. Chapter E8 mentions this topic in Section 1.1.3 and Appendix I. More generally, when a random process is scaling, and its exponent H – as defined in Chapter E6 – satisfies $H < 1$, the time intervals between zero crossings are often scaling with the exponent $1 - H$. The test is easy to perform and Zajdenweber 1994 reports a value of H that is reasonably well-determined and clearly distinct from the Brownian $1/2$.

A second and more abundant recent argument for scaling is best stated at the end of Section 2.4.

2.3 A tale that should serve as warning.

Since no one predicted the degree of stationarity that I found in records of certain price change, no one expressed surprise when Figure 4 of Chapter E1 (identical to Figure 5 of Chapter 14) suggested that the price of cotton was more volatile near 1900 than near 1950. In time, however, this perceived difference in volatility proved to be based on an error on my part. As described in M 1972b{Appendix I of E14}, those figures analyzed handwritten sheets in which slots corresponding to Sundays seemed to report prices based on actual trading. But in fact those slots were reporting weekly averages! When the same test was redone with actual data, it showed little change between 1900 and 1950!

More generally, everyone expected to find that the change in overall economy between 1816 and 1950 provoked drastic changes in the rules of variation of cotton prices. In fact, as shown in Figure 6 of M 1963b{E14}, those changes were not overwhelming at all.

2.4 An apparent limitation of an invariance in finance and economics may disappear when this invariance is suitably modified: from stationary to conditionally stationary and from (uni) scaling to multiscaling

In the study of finance, stationarity is all-too-often replaced by a strawman that is easy to argue against. First, it is interpreted in old-fashioned and narrow fashion, as implying that the ex-post appearance of samples must not differ "too much" from white Gaussian noise. The patent discrepancies between this narrow interpretation and the facts are routinely used to dismiss the strawman, but also every conceivable form of stationarity. My point is that *one must not* rush to take those discrepancies as irremediable. Indeed, even the M 1963 model suffices to demonstrate that the apparent failure of one implementation of stationarity *does not* exclude a more general implementation, one that may eventually fail, but only at a later stage of the process of data accumulation and modeling.

For invariance with respect to contractions, namely scaling, the all-too-ready acceptance of limitations is necessarily a more recent phenomenon. The question, "is scaling a property of price series?" was on no one's mind. But assume for the sake of argument that it was asked at the time when modeling used only one scaling process, the Wiener Brownian motion. The answer would have been that, since actual price series are not Brownian, they cannot be scaling. But, once again, the M 1963 model showed that they are scaling, albeit in "anomalous," nonBrownian fashion.

Eventually, a variety of discrepancies between M 1963 model and the evidence were identified. Firstly, α is bounded by $\alpha 2$ in the M 1963 model, but perhaps not in the data. Secondly, large discontinuities do not occur in isolation, but in "bunches."

Some authors used those imperfections to disqualify fractal modelling in general. Since those critics gave no convincing counter-model, the net effect was to encourage a return to a "default" model, Random Walk on the Street, together with "fixes" of the kind discussed in Section 4. Other authors directed "fixes" to scaling. For example, they acted to "save" the M 1963 model by introducing sharp cutoffs, crossovers or a variable value for the leading exponent α.

The models listed in Sections 6 to 8 of Chapter E1 show that I read the same evidence very differently. The M 1963 and M 1965 models moved beyond Brownian scaling, but that concept's scope was not exhausted and it is constantly expanded by fresh alternative implementations. Many have not yet been tried, as shown in the "Panorama" presented in Chapter E6. Each generalization organizes an increased wealth of evidence. An important generalization that moved from Sections 6 and 7 to Section 8 of Chapter E1 deserves to be emphasized by an example.

The limitations of one form of scaling may not apply to a more refined form; the example of foreign exchange rates. Olsen 1996 is a collective term the Bibliography uses for an abundant and varied collection of papers and preprints, some of which test a specific model, while others are not model specific but concern scaling itself.

The diagrams that test scaling are invariably of impressive straightness. Naively, one might have expected that this straightness indicates that foreign exchange rates follow the M 1963 model, but in fact they clearly fail to do so.

From (uni) scaling to multiscaling. The appearance of contradition in the preceeding example suffices to show that the M 1963 model is *less general* than scaling.

A way to resolve this discrepancy is to introduce a basic distinction that was mentioned in Section 8 of Chapter E1 and will be central to Section 3.8 of Chapter E6. Scaling can take two forms, *uniscaling* and *multiscaling*, and the proper model of foreign exchange rates appears to be a multifractal one.

Conditional stationarity. The preceding distinction is linked to two further notions, which cannot be discussed here: *conditional stationarity* and *sporadic process*. Both are discussed in M 1997N.

3. IRREDUCIBLE LIMITATIONS OF STATIONARITY AND SCALING

Once again, the basic shapes in geometry and physics, line, plane and space, are invariant under displacement and change of scale. Those simplest invariances are invaluable, but geometry and physics soon move beyond their domain of validity, and at that point everything becomes more complicated.

Inevitably, the same is true in finance, and the two themes of this section reflect the complication of economic reality. Section 2.4 argues that it is unwise to rush and conclude that a phenomenon is not invariant. An apparent deviation from invariance may disappear when one introduces a suitably generalized model. Nevertheless, deviations from stationarity and/or scaling are facts of life, in every concrete use of fractals. That is, a single intellectual path led to both the scaling nature of financial time series, and a variety of clear-cut deviations from scaling. This subtlest and most significant one is discussed Section 5.4 of Chapter E6.

3.1 A lesson from physics: a scientific principle can be useful even when there are limitations to its applicability

In times past, physics boasted of principles that envisioned no exception. Sometime around 1800, Pierre-Simon Laplace used especially eloquent words to proclaim that everything in the world will eventually fall under the sway of Newton's "inverse square of distance" law of gravitational attraction and will become fully predictable. This confidence was rewarded as mechanics became filled with very long deductive arguments that correctly predict the outcome of careful experiments. Diverse "principles of least (or greatest) something," some in physics and others in economics, also claimed they could bear the burden of long mathematical arguments.

However, physics abounds in examples of grand principles that contradict one another, hence cannot all be strictly true. Joseph Plateau (1801-1883) discovered an extraordinary example of such "frustrations." One expects the symmetries of a folded wire to be preserved in the soap bubbles formed on that wire. But this symmetry is not implemented (it is said to be "broken") when it contradicts another law of nature, which proclaims that no more than 4 bubbles can meet at one point. Physics also abounds in theories that are successful but based on assumptions that are theoretically incompatible, but close enough to compatibility under limited conditions. In effect, much of statistics places an equal degree of confidence in the Gaussian distribution. For example, many refined theorems concerning risk and the probability of ruin only hold in a universe in which the Gaussian holds with *absolute exactitude.*

Even in "hard" sciences, the claims of the principle of scaling are infinitely more modest than those of Newton's law. Thus, the study of turbulence in fluids attributes a central role to a combination of scaling with cut-offs and crossovers. Cut-offs and crossovers also play a central role in disputes concerning the validity of my fractal model of the distribution of galaxies.

3.2 Cutoffs of scaling due to impenetrable bounds

Most quantities encountered in nature are bounded, while most theoretical random variables are without bound. But it is generally understood that this contradiction need not be serious at all.

The example of the Gaussian distribution. The Gaussian is unbounded, while humans' heights are by definition positive and do not range up to twice the average height. But the Gaussian is useful, because its built-in self-contradiction is extraordinarily unlikely to be expressed.

Here is a quick and dirty estimate of the probability of its being a nuisance. Take average height to be 6 ft. and the standard deviation 10 times smaller, namely, 7.2 inches. Height would fall below 0 or above 12 feet when a Gaussian exceeds 10 σ, an event of probability roughly equal to the inverse of the Avogadro number 10^{23}, not a reason to worry.

Personal incomes and interest rates. More to the point, M 1960i{E10} proposed an L-stable model for the distribution of personal income. This model allocates a small probability to negative income: this unintended implication's quality of fit cannot be tested, but it is harmless – indeed, rather charming!

M 1967b{E15} investigated the distribution of several interest rates, and found a clear indication of approximate L-stability. But two bounds come into the picture. Once again, L-stability would allow interest rates to become negative. This possibility is theoretically absurd, but so unlikely that in practice it is harmless. Secondly, can rates increase without bound? My first test in M 1967j{E15} concerned the rate of "call money," which could become very large – and occasionally did. With arithmetic interest the rate that would require daily payments equal to the principal would be 36500%; the M 1963 model would take forever to reach that level but in any event it is clearly inapplicable, because changes in fast-moving interest rates are surely not independent.

Of course, the interest rates practiced today have lower upper bounds that do not allow the L-stable model to be extrapolated. Some markets fight volatility by setting an upper bound to the daily variation of a price. Such bounds inevitably create serial dependence and introduce diverse complications on which I have little to say.

Sharp cutoffs and reasons for postulating them only when unavoidable are discussed in Section 4.1. The simplicity inherent in a sharp cut-off is misleading.

3.3 Crossovers between different scaling regimes

Many phenomena that are not scaling are usefully represented by introducing "crossovers" between several distinct "scaling regimes" corresponding to different ranges of size. We shall begin by describing this notion in a non-financial context, then comment on it.

Different dimensions implicit in a ball of thread. This homely topic is discussed in M 1982F {*FGN*}, p. 17. In slight paraphrase, one reads the following:

"A ball of 10 cm diameter made of a thick thread of 1 mm diameter possess (in latent fashion) several distinct effective dimensions.

"To an observer placed far away, the ball appears as a zero-dimensional a point. From a distance of 10 cm resolution, it is three-dimensional. At 10 mm, it is a mess of one-dimensional threads.

"At 0.1 mm, each thread becomes a column and the whole becomes three-dimensional again. At 0.01 mm, each column dissolves into fibers, and the ball again becomes one-dimensional, and so on, with the dimension crossing over repeatedly from one value to another. When the ball is represented by a finite number of atom-like pinpoints, it becomes zero-dimensional again.

"An analogous sequence of distinct dimensions separated by crossovers is encountered in a sheet of paper."

The underlying idea is that it is often best to subdivide the total range of sizes corresponding to a concrete quantity and tackle each sub-range separately. This is routine in physics, where, to take an example, phenomena at the molecules' level are first studied separately from those at the level of atoms or fluids.

The preceding comments also apply to the two principles of invariance in finance. A single scaling rule of scaling may not apply to $L(t, T)$ for all values of the time lag T. If so, it is a good idea to proceed as with the ball of thread and identify different ranges of T, each characterized by separate rule of scaling. The surprise is not that crossovers as needed, but that they are needed far less often that expected.

Crossovers in high-frequency foreign-exchange data. The existence of crossovers is suggested in Olsen 1996, and M, Fisher & Calvet 1997 confirm it. The case of the dollar/deutschmark exchange rate will be mentioned in Section 3.15 of chapter E6.

The anticipated breakdown of scaling due to the crossover between finance and economics. It is irresistibly tempting to believe that short-run fluctuations are dominated by contingent constraints of speculation and the organization of financial markets, while long-run fluctuations concern fundamental rules of economic change. A priori, speculation and economics could follow entirely different rules. Therefore, the folklore that all financial charts look the same should be expected to fail for charts corresponding to long time periods.

I shared this expectation and knew how to test it. The original test concerned cotton, and the result is part of Figure 3 of Chapter E1. It came as a major surprise and an unexpected triumph for the idea of scaling. Roughly speaking, the distances between curves 1a and 1c (or 2a and 2c) seemed to indicate that the model used of daily changes, if rescaled, also represents monthly changes. But I was careful to note that the superposition of the daily and monthly curves in Figures 1 and 2 is imperfect, except if one argues that a month contains an "effective number" of days smaller than the actual number. This pointed out to a negative "correlation" between changes within a month. Eventually, the M 1972 model extended scaling to statistical interdependence between successive price increments and scaling again became a contender.

However, this apparent success is not established in a definite fashion. A more careful investigation may reveal breaks between short-run speculation and long-run economics. This possibility could not be tested in the 1960s, though the data on cotton were exceptionally extensive.

Conclusion. While crossovers are likely to be observed sooner or later, Section 4 will argue that one must *not* rush into "fixes" that involve sharp cutoffs or exponential terms.

4. STEP-BY-STEP CORRECTIONS FOR SPECIFIC INADEQUACIES OF BROWNIAN MOTION, AND "PATCHWORKS" OF "FIXES"

Once again, the notions of stationarity and scaling were inspired by properties of the homogeneous mass distribution on the line. Science soon moves beyond the line, the plane and other flat spaces, invariances remain valuable, and must not be given up without trying. The problems that involve invariances are not only the most attractive for the physicist, but also, as an unexpected "gift," many are surprisingly realistic. Unfortunately, every invariance eventually ceases to be verified. At this point, two strategies are available.

My strategy is to follow the example of physics, and view invariance as so valuable, that it is best to preserve the idea as long as feasible. When one invariance proves to have broken down, I seek better invariances, farther and farther beyond Brownian motion. At some point, this search will fail, but this point must be postponed as far as possible.

In sharp contrast, scaling and even stationarity are of no concern to the authors of the multitude of alternatives to Brownian motion that are continually being put forward.

Most (but not all) of those alternatives begin with the Random Walk on the Street, then apply "patches" or "quick fixes" as needed, without overall plan. (Even successful physics often needs a patch, denoted by the word *Ansatz*.) Inevitably, the result is more complicated than the original, and serious problems arise, especially if graphics are allowed as evidence.

Firstly, each discrepancy between the data and $B(t)$ is handled by a separate and specific amendment, and no feature is present in the resulting pyramid, unless it was purposefully inserted. Section 10 of Chapter E1 argued that the fixes are not "creative," and provide no "understanding."

Secondly, graphic ecidence *must* be allowed, as argured in Section 2 of Chapter E1. I take the liberty of challenging the proponents of the Brownian motion and of ARCH-like processes to examine how the graphic outputs of their algorithms compare with the actual data. I have no doubt about the outcome.

The discussion of the fixes should not be viewed as entirely negative, but as providing an opportunity to characterize my own approach more precisely. Sections 4.1 and 4.2 comment on fixes concerning marginal probability distributions. Section 4.3 concerns the rules of serial dependence. Section 4.4 expresses my attitude towards fixes, from the viewpoint of the distinction to be made in Chapter E5 and E6 between several distinct "states of randomness."

4.1 Fixes that neglect serial dependence and invoke a transient marginal distribution: truncation and multiplicative exponential decay

Two obvious but unfortunate fixes must be considered first: rather than with Brownian motion, they begin with the scaling distribution and the M 1963 model.

Assertions that the scaling variable is in fact "truncated" that is, does not exceed a finite maximum u_{max}. It is known, as already stated repeatedly, that on certain markets the distribution tails are shorter than implied by the scaling distribution or the M 1963 model. A strong temptation is to account for those discrepancies by assuming the existence of a sharp finite cut-off that has no basis in either regulation or economics and is merely meant to account for the tails' shortness.

Sometimes, u_{max} is assumed to be so large that its value does not directly affect any observable quantity. If so, the assumption $u_{max} < \infty$ does insure finite population moments, but otherwise is a form of empty posturing that also distracts from consideration of genuine problems.

Transients, and the pitfalls that make them misguided and misleading. In most cases when a finite u_{max} is invoked, it is meant to be small enough to affect observable quantities of interest. For example, define U as a L-stable distribution that was truncated to $u_{max} < \infty$ and consider the sum of N such variables. For small values of N, the distribution of the properly normalized sum is unaffected by u_{max}. For large N, a different normalization predicts convergence to the Gaussian. Therefore, the tails become increasingly short as N increases.

The scenario based on truncation implies the belief that interesting values of N fall between those two zones of simplicity. Unfortunately, the intermediate zone called "transient" is complicated and hard to control. Instances where this scenario is confirmed fail to contribute to the goal of "understanding" described in Section 10 of Chapter E1. Therefore, while this scenario that may be unavoidable if everything else fails, it is extremely unattractive and I do my best to avoid it. My attitude towards transients is elaborated upon when the case is made against the lognormal in Chapter E7.

A suggestion that goes back to V. Pareto: a scaling density multiplied by an exponential decay that yields a progressive cut-off. The standard method for achieving a crossover from scaling is one hundred years old! Indeed, just after Pareto's law of income distribution, $\Pr\{\text{income} > u\} \sim u^{-\alpha}$, Pareto 1896 also included $u^{-\alpha}\exp(-\beta u)$ as a *possible* improvement. One may, instead, replace density $\sim \alpha u^{-\alpha-1}$ by one $\sim u^{-\alpha-1}\exp(-\beta u)$. When β is small, $\exp(-\beta u)$ is ~ 1 for small u, but becomes predominant for large u.

There may be cases when a term $\exp(-\beta u)$ is actually imposed by the data. But I think that in most cases it is simply an easy "fix." From a traditional statistical viewpoint, a density $\sim u^{-\alpha-1}\exp(-\beta u)$ with $\beta > 0$ shares the main virtue of truncation with $u_{max} < \infty$. Assuming $\beta > 0$ all moments are finite, while assuming $\beta > 0$ the q-th moment is infinite when $q > \alpha$. This whole book argues, however, that such infinite moments are *not* "improper" in any way. They are *not* abominations to be hidden or papered over at all cost, but important symptoms of an underlying reality. They must be faced squarely.

This exponential decay with $\beta > 0$ is eminently natural, since many decades after Pareto, it independently entered physics. Even later, several physicists unacquainted with Pareto proposed to "import" $\beta > 0$ back into finance. Every new parameter injected in a formula is bound to improve the fit, but the justification offered for $\beta > 0$ in physics does not carry over to economics. I view Pareto's $\exp(-\beta u)$ as nothing but ad-hoc and unconvincing *deus ex machina*.

Comment on price changes. The preceding two ancient and crude procedures relative to the distribution of income returned to life as would-be improvements addressed to cases when the M 1963 model's tails are perceived to be too long. As the reader knows by now, the same discrepancies led me to change over from scaling to multiscaling.

4.2 Student's and hyperbolic distributions, and other fixes that neglect serial dependence and solely concern the marginal distribution

The Student's distribution. Blattberg & Gonedes 1974 questioned the M 1963 model, and proposed to replace the L-stable distribution, Λ, by Student's distribution, Σ. It is true that

a) statisticians view Σ as familiar, and Λ as exotic;

b) a simple analytic expression exists for Σ, namely a negative power of $(1 + x^2)1$ but none exists for Λ;

c) the α parameter is common to both distributors; its value is unrestricted for Σ; for Λ, to the contrary, it is bounded by a maximum value $\alpha = 2$;

d) lastly, some statistical tests, when applied to daily price changes concluded that Σ gives a better fit than Λ.

I have doubts about item d), because of a "qualitative" feature: when plotted as in Figure 3 of Chapter E1, Σ yields a cap-convex curve, while the curves in Figure 3 are not cap-convex. But I prefer to limit my response to Blattberg & Gonedes to more basic considerations of two kinds. The first was originally phrased in the context of a different standard statistical distribution, namely, one that Clark 1973 described as preferable to the L-stable in this context examined at the end of Section 4.3 and in Section 3 of Chapter E21.

Statistical fit need not be optimal for each individual feature of the problem, but must be acceptable for every quantity under consideration. An observational scientist, after he sets aside the study of daily "seasonals," *must not* focus exclusively on daily price changes, but must study in parallel the price changes over periods different from 1 day.

From this viewpoint, Λ has the virture of being invariant under addition of independent quantities, which is why the M 1963 model and other fractal constructs can be interpolated and extrapolated, and the result confronted with the facts. Moreover, one knows a variety of rules of serial dependence that preserve Λ. By contrast, I do not know how to interpolate Student's Σ from daily price changes to T less than one day. The extrapolation of Student's to T = several days is feasible, but analytically unmanageable. Furthermore, I do not know any rule of serial dependence that preserves Σ.

To conclude, the unquestionable assets of Student's distribution, its analytic simplicity and familiarity, apply only to $T = 1$, which is not enough. While the predictions of the M 1963 model are likely to be only approximate, they form a coherent and fully specified whole. Careful statistics is worthless if it involves distributions that fail to belong to a coherent and fully specified whole.

The parameter α *and role of possible values* $\alpha \geq 2$. In Student's case, there is nothing atypical or critical about $\alpha = 2$. To the contrary, the L-stable distribution behind the M 1963 model sets $\alpha = 2$ as an atypical critical value that cannot be exceeded. On this criterion, Student's seems preferable, so why label it a "fix," while L-stable is called a model? One reason is that $\alpha = 2$ is a genuinely critical value, and the difference between $\alpha < 2$ and $\alpha > 2$ is genuinely critical and not made to vanish if it is "papered over." For example, in cases where Brownian motion in multifractal trading time applies, $\alpha < 2$ is compatible with either serial independence or dependence, while $\alpha > 2$ calls for strong serial dependence.

The "hyperbolic distribution." This term has been proposed for the probability density $p(u)$ such that the graph of $\log p(u)$ is one-half of a hyperbola (Barndorff-Nielsen & Blaesdild 1983.) Its theoretical motivation is inconsequential, in my opinion, and its sole unquestioned merit is that it is analytically convenient. This distribution combines a near-Gaussian bell and two exponential tails into a hybrid or "chimera" for which the graph of log $p(u)$ is cap-convex. Therefore, it could not fit those data for which the graph of log $p(u)$ is emphatically not cap-convex.

Be that as it may, the invariances of the hyperbolic are not additive; hence, in cases where daily data are satisfactorily fitted by the hyperbolic, serial dependence cannot be neglected, even in a first approximation. The hyperbolic distribution's sole virtue, namely, analytical convenience, is not expected to be preserved under extrapolation to changes over $T > 1$ days.

4.3 Fixes that start with Wiener Brownian motion and inject serial dependence: high-order Markov, ARMA and ARCH representations

To make a point, this subtitle uses the term "representation," instead of "model." Mathematical analysis tells us that a wide variety of functions can be approximated by ordinary or trigonometric polynomials or represented by trigonometric series. Similarly, there is no question that a wide class of random processes of all kind can be accurately approximated by high-order Markov, a wider class falls under the ARMA or ARCH algorithms, and even wider classes fall under further generalizations.

However, what do we learn from those approximations? This question arises again and again, in one field after another. The earliest examples are ancient; two that I know date back to Gauss himself and are discussed in M 1982F{*FGN*}, pages 417-8. Early authors proposed representations that were in the spirit of high-order Markov, and this proposal provoked sharp rebukes from two great physicists. In effect, though obviously not in actual wording, J. C. Maxwell and L. Boltzmann argued that those examples manifest long-time serial dependence, and that high-order Markov processes were the wrong way to tackle them. Let us now look closer.

An examplary tale: high-order Markov processes in the context of hydrology. As mentioned in Section 7 of Chapter E1, I first encountered this issue in the context of the "Hurst puzzle" that is taken up in detail in M 1997H and deals with the persistence manifested by river discharges. Hydrology is far from finance but is characterized by equally erratic fluctuations. The "persistence" in the successive yearly discharge of a river like the Nile was promptly recognized as reflecting serial dependence and, naturally enough, the first model to be tried was ordinary Markov process of memory $M = 1$. This idea was found inadequate, but a renowned probabilist ventured, in Feller 1951, that one ought to try a Markov process of longer memory $M > 1$.

As I recall, the search for a detailed geophysical motivation of the Markov property was never more than casual. Indeed, like the Gaussian distribution, Markov processes are so profoundly imbedded in science as to be taken for granted, as the probabilistic expression of direct causality.

By contrast, M 1965h interpreted the Hurst's puzzle as a symptom implying that the dependence is not Markovian, but scaling. This implies that the memory is not finite, but infinite, with a single main parameter, H, that is readily measured from R/S graphs (see Section 7 of Chapter E1.) To be scaling, a Gaussian process must be a fractional Brownian motion, which is also mentioned in Section 7 of Chapter E1 and in greater detail in Chapter E6.

William Feller adopted my argument as soon as he heard it, but those hydrologists who were not informed followed up on the ill-considered suggestion found in Feller 1951. They went on to estimate the memory M from the data, and soon made a striking discovery: as the sample length T increases, a good statistical fit requires a memory M that increases without limit, staying in the range between $T/3$ and $T/2$.

This shocking conclusion made it clear that the Markov property is an artifact, and M is a non-intrinsic and worthless characterization that provides no scientific insight and has no pragmatic predictive value. As a result, the Markov model was forgotten, and action moved to ARMA, which raises the same issues, but in a more opaque way. The references are scattered throughout M 1997H.

The eye versus high-order Markov, ARMA and ARCH. Let us step aside for comments that bring in the eye. They may seem overly "soft" and qualitative, but in factare basic and concern a matter of principle.

Once again, as argued in Section 1 of Chapter E1, the validity of Markov models *must* be subjected to test by eye. Within a very long sample, take two sub-samples of length M separated by an interval of the same duration M. If the process is Markov of memory M, those samples are near-independent. It follows that the large sample includes no "feature" of duration longer than M. In other words, seen from a distance much larger than M, a Markov process of memory M looks near indistinguishable from the white noise in Figure 2 of Chapter E1. To the contrary, once again, financial data like those of Figure 1 of Chapter E1 unquestionably include a wealth of long duration structures. Back to the case of the Nile River, the absence of structure predicted by Markov clearly contradicts well-founded legend and historical facts. The Bible describes "seven fat and seven lean years " and history records periods of seventy wet and fat years of famous Pharaohs followed by seventy dry years ruled by forgotten dynasties.

To represent such data by Markov, ARMA, ARCH, or any other short-term dependent process would be highly unadvisable, unless very strong independent reasons exist to believe that long duration phenomena are distinct and different from the short range phenomena to be modeled by ARMA or ARCH. To the contrary, the M 1965 model attempted to account for all data, and this "extravagant ambition" (to quote the first pages of the Preface) was at least partly successful.

ARMA representations. The comments that follow are addressed to readers already acquainted with the "auto-regressive moving average" processes. The notion caused no stir when advanced by Herman Wold in the 1930s, but in the 1960s the Box-Jenkins computer programs made nearly everyone rush to try ARMA, then a variant called ARIMA.

As expected, every sample of hydrologic and any other data can be fitted satisfactorily in this manner. Generality is always billed as being a great asset, but in many cases it is actually a major liability. Once again, like a Fourier series, ARMA is not a model, only a versatile representation. A satisfactory statistical fit is of no use in science unless the fitting parameters are consistent in time and have an intrinsic meaning. Unfortunately, the ARMA or ARIMA parameters obtained from successive samples are near-invariably mutually contradictory, and have no intrinsic meaning. Furthermore, because of the lack of long term structures, they have no predictive value whatsoever.

In that respect, the M 1965 model is very different. It does not have a multitude of parameters, but a single parameter H. It is not a versatile representation, but a demanding model, making specific predictions that may turn out to be right or wrong.

Being eager to know whether or not the value of H is intrinsically meaningful, I sought out the noted and wise economist Pierre Massé, who started his road to fame as the great dam builder of France. I described to him a major finding in M & Wallis 1969b{H}. The parameter H is consistent between different subsamples from a river, and it is systematically higher for French rivers with their source in the Massif Central rather than the Alps. Restated in words, the former are more persistent than the latter. Massé was delighted and astonished that so much of his life-long qualitative experience could be summarized by a single number ... and discovered so directly by a raw non-expert. How would he have reacted to parallel columns of ARIMA coefficients?

One-parameter long-dependence having proved successful, it was combined with ARIMA, yielding "fractional ARIMA" (FARIMA or ARFIMA). As expected, the two ingredients in combination give a better statistical fit than either ingredient by itself. But does long-dependence or ARIMA contribute more significantly to the overall good fit in FARIMA? That is, should the one-parameter FBM dependence be viewed as a last-minute improvement on ARMA, or is it true (as I would expect) that the more significant factor is the exponent H of FBM?

The ARCH representation and its variants. The comments that follow are addressed to readers already acquainted with this common "fix" to Brownian motion. In spirit, the ARCH model is closely related to models that inject a trading time, such as the 1967 model based on subordination (M & Taylor 1967) or the M1972 model. The fit of an ARCH-like model is likely to be good or even excellent, if only because there is no upper bound on the number of parameters. However, all the reservations concerning ARMA extend to ARCH.

To a recently-trained economist who accepts help from the eye, financial data like those in Figure 1 of M 1967j{E15}) are prime material for ARCH modeling. But so does the original output of a multifractal measure, Figure 4 of Chapter E1, or any variant thereof. A short sample output of the fitted ARCH model includes interesting features reminiscent of either figure. But, as I expected but was careful to verify, a longer sample, when seen from a sufficient distance, behaves like the white noise of Figure 1 of Chapter E1. That is, ARCH analysis *fails*, by its very nature, to be faithful to the long-term component that the multifractals involve, and that the eye sees in the financial data.

To be practical, suppose that some statistical test is applied to the long-term components that we all see, and declares them to be in fact statistically non-significant. My first impulse will be to examine the opinion those tests express concerning multifractal samples. I expect that the response will be that the long-term components are non-existent, whereas we know they are present, since they were deliberately built-in.

Interpolation and extrapolation. Yet another deep difference between ARCH and the M 1972 model shows up when models based on daily data have to be interpolated or extrapolated to higher or lower frequencies. In ARCH, analytic procedures are not available. In the M 1972 model, interpolation and extrapolation are both immediate. Moreover, as in the actual data, the distributions over different time spans Δt are different and narrow down as Δt increases.

Clark's "fix" applied to the M 1963 model as a substitute to genuine generalization. M & Taylor 1967 is discussed in Section 8 of Chapter E1, and reproduced in Chapter E21; we shall refer to it as *op. cit.* Once again, the point is that the M 1963 model can be represented as a Brownian motion followed in a trading time defined by using a "subordinator."

Clark 1973 adopted this idea and statistical tests of several alternative subordinator functions concluded that a lognormal fits the daily data better than the scaling suggested in *op. cit.* My response in M 1973c (reproduced in Section 3 of *op. cit*) was a preview of the criticism of Student's distribution presented in Section 4.2. I see nothing special about the time span of one day. If lognormality turns out to also apply to $T > 1$, this would be due to special rules of dependence that no one attempted to describe, much less to compare with the evidence.

The answer to my concerns would be an independent justification of lognormality, but the widely accepted motivation by multiplicative effects is at best arguable, as shown in Chapter E8, and altogether inapplicable in this context. (The case against lognormality is argued in Chapter E9.)

4.4 Touch-ups; scaling versus a patchwork of fixes, within a distinction Chapters E5 and E6 introduce between "states of randomness"

Section 5.3 of Chapter E1 sketched the notion of *mild* randomness, as applied to self-affine models. By contrast, it will be said that fractional Brownian motion is *wildly* random. Thirdly, the words *slow randomness* will describe a range of behavior that is intermediate between mild and wild. In this case, the "long-run" asymptotic behavior is mild, a conclusion that seems reassuring. Unfortunately, this reassuring conclusion is of no observable consequence for slowly random phenomena, because for them the observable "middle-range" transient mimics that of wildly random phenomena.

How does this classification accommodate the financial model that involve fixes? The "Student" model straddles the distinction but at least allows for the possibility that $\alpha < 2$, that is, does not exclude wild behavior. The M 1963 model altered by truncation (Section 4.1) is specifically designed to promise asymptotic mildness, and the difficulties relative to the middle-run are papered over by being assigned to transients. The higher the pyramid of fixes, the longer the transient. The longer the transient, the less significant and worthy the initial model.

This fundamental and delicate issue is discussed in detail in Chapters E5 and E6. It leads me to draw a distinction between fixes and "touch-ups." I think of the former as changing the state of randomness, without clearly acknowledging what is being done, while the latter remain within a single state of randomness and only concern details of fit.

5. "PARADOXES" THAT ILLUSTRATE THE "CREATIVITY" INHERENT IN SCALING

To move into this final section, it is necessary to change gears. The "creativity" that allows very simple fractal and scaling constructions to generate structures of unexpected complexity easily generates paradoxes that deserve being sampled in a jocular vein.

5.1 Paradoxes of expectation in the exponential and the scaling cases

The technical argument presented in Chapter E1, Section 5.2 has extensive concrete implications.

The fact that the exponential distribution is invariant with respect to change of location has a well-known paradoxical consequence. When time intervals between buses are exponential, a passenger's waiting time (in particular, his expected waiting time) is *not* affected by the fact that they barely missed the preceding bus.

This consequence has an obvious counterpart when U is scaling. When $\alpha > 1$, one finds $E(U \mid U > w) = \alpha(\alpha - 1)^{-1} w$. For all α, the conditioned median is $\{U_{1/2} \mid U > w\} = 2^{1/\alpha} w$, and all other quantiles are proportional to w. This is a disconcerting result, and calls for an intuition of randomness that is altogether different from the intuition acquired from exponential queues. This task is best faced through some folklore and related fanciful and paradoxical stories.

5.2 Fanciful but enlightening paradoxical stories

Throughout this chapter runs the assertion that scaling randomness warrants attention. The reason resides in its high probability of generating structures that fit the data well and seem to lie beyond the power of randomness. It is, therefore, proper to end this chapter with fanciful stories that may help this "creativity" become understood. A few were moved here from M 1966b{E19} to increase emphasis. Readers resistant to fancy are urged to forge ahead. The other readers are forewarned: if a story holds up to criticism, I shall claim it as a scientific discovery; if it does not, I shall insist that a parable must not be taken too seriously.

The Lindy Effect. Popular wisdom informs us that "Nothing succeeds like success"; "Advantage brings advantage"; "The greater an active man's past success, the greater further success he may expect in the future"; "The greater the scope an idea or a process has had in the past, the greater the additional scope it may expect to acquire"; "The future career expectation of a television comedian is proportional to the total amount of his past exposure on the medium." The *New York Times Magazine* of March 3, 1968, when quoting this last saying, credited it *The New Republic* of June 13, 1964, which had called it the *Lindy Effect.*

Let us indeed ponder the expectation of the comedian's future career. Clearly, it is the average of (a) the nonexistent future careers of those whose day has already ended (though they may not realize it), (b) the short future careers of most, and (c) the very long future careers of a handful of individuals. Some past performers were extremely durable, therefore the same is generally expected to be the case for present performances, but their identity is usually predicted by no more than a handful of observers who may be exceptionally perceptive or merely lucky. The issue may be discussed interminably, but is seldom determined with certainty.

At this stage of the argument, however, we must stress that the Lindy Effect fails to hold for non-scaling distributions. Thus, as the exposure u increases, expectation remains *constant* if $\Pr\{U \geq u\} = \exp(-u)$. It actually *decreases* if $u > 0$ and $\Pr\{U > u\} = \exp(-u^2)$, which is not too far from the rule that holds for the Gaussian.

Relation between the Lindy Effect and the distribution of personal income. Restated in this book's terminology, Pareto's law claims that the distribution of personal income is scaling. Consider a contemporary U.S. resident, of whom we only know that his annual income is at least w, where w is neither small (a different formula would be needed in that case) nor so large as to identify the resident and hence move beyond the scope of statistics. In the case $\alpha = 2$, it follows that $E(U - w) = w$. In words, the expectation of the unknown portion of income is equal to the ascertained income. The expected value is therefore $2w$, but it is thoroughly misleading. It combines the occasional millionaire with a mass of people whose income barely exceeds w.

"Parable of the Young Poets' Cemetery. In the most melancholy section of the cemetery are the graves of poets and scholars fallen in the flower of their youth, each surmounted by a symbol of loss: one half of a book or

of a column, a tool's handle .... The old groundskeeper, a poet in his youth, urges visitors to take these funereal symbols most literally: 'All who lie here,' he says, 'accomplished enough to be viewed as full of promise, yet fell so young that a limitless future seemed to extend in front of them. Some, indeed, might have challenged the prolific Leonhard Euler, but most were about to be abandoned by their Muses. Having lived long and seen much, I view my charges' lifework before they came here as divided into two exactly equal halves: one half in fulfillment and one half in broken expectation."

"*Where it is shown that a scientist whose work is interrupted when he is still young only realizes half of a promised career.* Most humans never write a scientific article. Of those who do, most write a very small number, but a very few – giants or pigmies – are extraordinarily prolific. According to Alfred Lotka, the distribution of the number of scientific papers signed by an author is scaling with $\alpha = 2$. It follows that however long a person's record, it will on the average continue for an equal additional amount. When it eventually stops, it breaks off half its promise. The only way of avoiding such apparent disappointment is to live to be so old that age corrections must be considered when computing the expected future."

"*The Dean and the Applicant.* A young but already confirmed scientist applies for a position. Legend has it that overcautious deans simply weigh or count publications. Forward-looking deans attempt, through mist and uncertainty, to read the future. To remain "objective" and not misled by imponderables, they must be guided by expectation based on past experience. But after the best possible outcome is averaged with the worst, the Lindy Effect will make forward-looking deans agree with the overcautious ones. A difference is that the overcautious get what they bargained for, while the forward-looking ones worry about "deadwood" they ought to have foreseen."

"*On the exposure and expectation of a whole field of learning.* Whole branches of science also grow, bloom and then wither away, and since there is no human mortality to bound a collective endeavor, it is easier to study a field of science than to study an individual scientist. Assuming that the Lindy Effect holds, the science with the longest uninterrupted successful record should be expected to have the longest future. The sciences that have long been successful also have the greatest "prestige," and warrant – at least according to those who practice them – the greatest social and financial support. But every science will stop sometime, so the greatest disappointment is reserved for individuals who hitch their career to a long-prosperous field just about to run dry."

To continue in the same vein, here is a parable meant to point out that uncertainty is no lesser in the physical than in the social sciences.

"Parable of the Receding Shore. Once upon a time, there was a country called the Land of Ten Thousand Lakes, affectionately known to its inhabitants as Biggest, Second Biggest, ..., N-th Biggest, etc., down to 10,000th Biggest. The widest was an uncharted sea, nay, a wide ocean at least 1600 miles across, the width of N-th Biggest was 1600 $N^{-0.8}$, and in particular the smallest had a width of only 1 mile. (This fits the evidence described in Chapters II and IX of M 1982F{*FGN*}). But each lake was always covered with a haze that made it impossible to see beyond a mile to identify its width. The land was poorly marked, and had few inhabitants to help the traveler. However, the people who lived there, the Lakers, were expert at measuring and great believers in mathematical expectation. As a Laker stood on an unknown shore, he knew he had before him a stretch of water of expected width equal to 5 miles. Having sailed for a few miles m, but failing to reach his goal, our Laker calculates the new expected distance to the next shore, and again obtains 5 miles.

"Could it be that spirits moved the shore away?"

Once again, fanciful stories must not be taken textually. But they must not be spurned either. Suppose, for example, that the lengths of droughts are scaling. Then a drought that already lasted w years will continue for an additional duration proportional to w. This prediction is by no means far-fetched, and is put to good use in the theory of financial bubbles that is presented in M 1966b{E19}.

The Journal of Political Economy: 71, 1963, 421-440. **E3**

New methods in statistical economics

• *Chapter foreword.* An interesting relationship between the methods in this chapter and renormalization as understood by physicists is described in the *Annotation for the physicists* that follows this text. •

✦ **Abstract.** This is an informal presentation of several new mathematical approaches to the study of speculative markets and of other economic phenomena. My principal thesis is that to achieve a workable description of price changes, of the distribution of income, firm sizes, etc., it is necessary to use random variables that have an infinite population variance.

This work should lead to a revival of interest in Pareto's law for the distribution of personal income. The scaling distribution related to this law should dominate economics. ✦

AMONG TODAY'S STATISTICIANS AND ECONOMISTS, Pareto's law for the distribution of income is well-known, but is thoroughly neglected for at least two reasons. It fails to represent the middle range of incomes, and lacks theoretical justification within the context of elementary probability theory. I believe, however, that Pareto's remarkable empirical finding deserves a systematic reexamination, in light of the new methods that I attempt to introduce into statistical economics.

I. INTRODUCTION

Pareto claimed that there exist two constants, a prefactor C and an exponent $\alpha > 0$, such that for large u, the relative number of individuals with an income exceeding u can be written in the form $P(u) \sim Cu^{-\alpha}$.

That is, when the logarithm of the number of incomes greater than u is plotted as a function of the logarithm of u, one obtains for large u a straight line with slope equal to $-\alpha$. Later, the same relation was found to apply to the tails of the distributions of firm and city sizes. In fact, the search for new instances of straight log-log plots has been very popular and quite successful, among others, in Zipf 1941, 1949.

This book reserves the term "law of Pareto" to instances that involve the empirical distribution of personal income. The tail distribution $P(u) \sim Cu^{-\alpha}$ is denoted by the neutral term, *scaling distribution*, that is useable in many sciences and was not available when the paper reproduced in this chapter was published for the first time. The quantity α will be called *scaling exponent*.

Notwithstanding the abundant favorable evidence, Zipf's claims met strong objections from statisticians and economists. Those objections were so strong as to blind the critics to the evidence. In sharp contrast, I propose to show that *the scaling distribution literally cries* for our attention under many circumstances. Those circumstances include (1) taking seriously the simplified models based on maximization or on linear aggregation (2) taking a cautious view of the origin of the economic data or (3) believing that the physical distribution of various scarce mineral resources and of rainfall is important in economics.

In addition, I shall show that, when the "spontaneous activity" of a system is ruled by a scaling rather than a Gaussian process, the causally structural features of the system are more likely to be obscured by noise. They may even be completely "drowned out." This so because scaling noise generates a variety of "patterns;" everyone agrees on their form, but they have no predictive value. Thus, in the presence of a scaling "spontaneous activity, validating a causal relation must assume an unexpectedly heavy burden of proof and must acquire many new and quite perturbing features.

We shall see that the most important feature of the scaling distribution is the length of its tail, not its extreme skewness. In fact, I shall introduce a variant of the scaling distribution, which is two-tailed, and may even be symmetric. Hence, extreme skewness can be viewed as a secondary feature one must expect in variables that have one long tail and are constrained to be positive.

Much of the mathematics that I use as tool have long been available, but viewed as esoteric and of no possible use in the sciences. Nor is this paper primarily an account of empirical findings, even though I was the first to establish some important properties of temporal changes of speculative prices. What I do hope is that the methods to be proposed will constitute workable "keys" to further developments along a long-mired frontier of economics. Their value should depend on (1) the length and number of successful chains of reasoning that they have made possible; (2) the number of seemingly reasonable questions that they may show to be actually "ill-set" and hence without answer; and last, but of course not least, (3) the practical importance of the areas in which all these developments take place.

This paper will not attempt to treat any point exhaustively nor to specify all the conditions of validity of my assertions; the details appear in the publications referenced. Many readers may prefer to read Section VI before Sections II-IV. Section IX examines Frederick Macauley's important and influential critique of Pareto's law.

II. INVARIANCES; "METHOD OF INVARIANT DISTRIBUTIONS"

The approach I use to study the scaling distribution arose from physics. It occurred to me that, before attempting to explain an empirical regularity, it would be a good idea to make sure that this empirical identity is "robust" enough to be actually observed. In other words, one must first examine carefully the conditions under which empirical observation is actually practiced. The scholar observes in order to describe but the entrepreneur observes in order to act. Both know that most economic quantities can hardly ever be observed directly and are usually altered by manipulations. In most practical problems, very little can be done about this difficulty, and one must be content with whatever approximation of the desired data is available. But the analytical formulas that express economic relationships cannot generally be expected to remain unaffected when the data are distorted by the transformations to which we shall turn momentarily. As a result, a relationship will be discovered more rapidly, and established with greater precision, if it "happens" to be invariant with respect to certain observational transformations. A relationship that is noninvariant will be discovered later and remain less firmly established. Three transformations are fundamental to varying extents.

Linear aggregation, or simple addition of various quantities in their common natural scale. The distributions of aggregate incomes are better known than the distributions of each kind of income taken separately. Long-term changes in most economic quantities are known with greater precision than the more interesting medium-term changes. Moreover, the meaning of "medium term" changes from series to series; a distribution that is not invariant under aggregation would be apparent in some series but not in others and, therefore, could not be firmly established. Aggregation also occurs in the context of firm sizes, in particular when "old" firms merge within a "new" one. The most universal type of aggregation occurs in linear models that add the (weighted) contributions of several "causes" or embody more generally linear relationships among variables or between the current and the past values of a single variable (autoregressive schemes). The preference for linear models is of course based on the unfortunate but unquestionable fact that mathematics offers few workable nonlinear tools to the scientist.

There is actually nothing new in my emphasis on invariance under aggregations. It is indeed well known that the sum of two independent Gaussian variables is itself Gaussian, which helps use Gaussian "error terms" in linear models. However, the common belief that only the Gaussian is invariant under aggregation is correct *only* if random variables with infinite population moments are excluded, which I shall *not* do (see Section V). Moreover, the Gaussian distribution is *not* invariant under our next two observational transformations.

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 to the central limit theorems that describe the limits of weighted sums of random variables.

Weighted mixture. In a weighted lottery a preliminary chance drawing selects one of several final drawings in which the gambler acquires the right to participate. This provides a model for other actually observed variables. For example, if one does not know the precise origin of a given set of income data, one may view it as 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 subsidiaries that are almost wholly owned but legally distinct. 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 hits an industry, the distribution of "new" firms may be viewed as a *mixture* of (a) the distribution of companies *not* involved in a merger, (b) the distribution of companies that are the *sum* of two old firms, and perhaps even (c) the *sum* of more than two firms.

Maximizing choice, the selection of the largest or smallest quantity in a set. It may be the case 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 fully reported events are the exceptional ones, such as droughts, floods or famines (and the names of the "bad kings" who reigned in those times) and "good times" (and the names of the "good kings"). Worse, many data are a mixture of full reported data and of data limited to the extreme cases.

Although the above transformations are not the only ones of interest, they are so important that it is important to characterize the distributions that they leave unchanged. It so happens that *invariance-up-to-scale holds asymptotically for all three transformations, as long as the parts themselves are asymptotically scaling.* In the case of infinite aggregation, invariance only holds if the scaling exponent α is less than two. To the contrary (with some qualifications), *invariance does not hold – even asymptotically – in any other case.*

Hence, anyone who believes in the importance of those transformations will attach a special importance to scaling phenomena, at least from a purely pragmatic viewpoint.

This proposition also affects the proper presentation of empirical results. For example, to be precise in the statement of scientific distributions, it is *not* sufficient to say that the distribution of income is scaling; one must list the excluded alternatives. A statistician will want to say that "it is true that incomes (or firm sizes) follow the scaling distribution; it is not true that incomes follow either Gaussian, Poisson, negative binomial or log-normal distributions" But my work suggests that one must rather say: "It is true that incomes (or firm sizes) follow the scaling distribution; it is not true that the distributions of income are very sensitive to the methods of reporting and of observation."

III. INVARIANCE PROPERTIES OF THE SCALING DISTRIBUTION

Of course, the invariance of the asymptotic scaling distribution holds only under additional assumptions; the problem will surely not be exhausted by the present approach. Consider N independent random variables, $U_n (1 \leq n \leq N)$ that follow the weak (asymptotic) form of the scaling distribution with *the same exponent* α. This means 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.

Let me begin with mathematical statements that imply that the scaling behavior of U_n is *sufficient* for the three types of asymptotic invariance-up-to-scale. Short proofs will be given in parentheses, and longer ones in the Appendix. The symbol 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 denote by p_n the probability that U_W is identical to U_n. One can show that this U_w is also asymptotically scaling 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. (*Proof.* It is easy to see that

$$\Pr\{U_W > u\} = \sum p_n \Pr\{U_n > u\} \sim \sum C_n p_n u_n^{-\alpha} = C_w u^{-\alpha}.)$$

Maximizing choice. *Ex-post,* when the values U_n of all the variables U_n are known, let U_M be the largest. One can show that this U_M is also asymptotically scaling, with the scale parameters $C_M = \Sigma C_n$, the *sum* of the separate scale coefficients C_n. (*Proof.* Clearly, in order that $U_M \leq u$, it is both necessary and sufficient that $U_n \leq u$ is valid *for every* n. Hence, Π denoting the product of the terms relative to the N possible values of the index n, we have

$$\Pr\{U_M < u\} = \Pi \Pr\{U_n \leq u\}.$$

It follows that

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

Aggregation. Let U_A be the sum of the random variables U_n. One can show that it is also asymptotically scaling, with a scale parameter that is again the *sum* of the separate weights C_n. Thus, at least asymptotically for $u \rightarrow \infty$, the sum of the U_n behaves exactly like the largest U_n (see M 1960i{E10} for further details). Mixture combined with aggregation is an operation that occurs in the theory of random mergers of industrial firms

(M 1963o). One can show that it also leaves the scaling distribution invariant-up-to-scale.

The converses of the above statements are true only in the first approximation; for the invariance-up-to-scale to hold, the distributions of the U_n need not follow the scaling distribution exactly; but they must be so close to it as to be scaling for many practical purposes.

Strictly invariant distributions that also enter as limits. To introduce two distributions due to Fréchet and Lévy, respectively, and relate them to scaling, let us imitate (with a different interpretation) a principle of invariance that is typical of physics: We shall require that the random variable U_n be strictly invariant up to scale with respect to one of our three transformations.

Let N random variables U_n follow – up to changes of scale – the same distribution 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 the changes of scale – the same distribution as U. This allows one to write U_W (U_M or U_A) in the form a_WU (a_MU or a_AU), where a_w, a_m and a_A are positive functions of the numbers a_n.

As shown in the Appendix, it turns out that the conditions of invariance lead to somewhat similar equations in all three cases; ultimately, one obtains the following results:

Maximization. The invariant distributions must be of the form $F_M(u) = \exp(-u^{-\alpha})$ (Fréchet 1927, Gumbel 1958). These distributions are clearly scaling for large u and correspondingly small $u^{-\alpha}$, since in that range 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 form $N^{-1/\alpha} \max U_n$, where the U_n are asymptotically scaling. There are no other distributions that can be obtained simply by multiplying the mass U_n by an appropriate factor and by having N tend to infinity. But allowing the origin of U to change as $N \rightarrow \infty$, yields the "Fisher-Tippett distribution," which is *not* scaling and not invariant under the other two transformations.

Mixing. In this case, the invariant distributions are $F_W(u) = 1 - Cu^{-\alpha}$, which is the analytical form of the scaling distribution extended down to $u = 0$. This solution corresponds to an infinite total probability, implying that, strictly speaking, it is unacceptable. However, it must not be rejected immediately because in many cases U is further restricted by some relation of the form $0 < a \le u \le b$, leading to a perfectly acceptable conditional probability distribution.

Aggregation. Finally, aggregation leads to random variables that are the "positive" members of the family of "L-stable distributions," other members of which will be encountered =later (Lévy 1925, Gnedenko & Kolmogorov 1954). These distributions depend on several parameters, the principal of which is again denoted by α and must satisfy $0 < \alpha \le 2$. The density $dF_A(u)$ has a closed analytic form in a few cases. The limit case for $\alpha = 2$, is the Gaussian distribution (which, however, is not itself scaling). The density of the positive L-stable distribution is also known in the case $\alpha = 1/2$, which plays a central role in the study of the return to equilibrium in coin tossing. In other cases, no closed analytic expression is known for the stable distribution $F_A(u)$. But Lévy showed that they asymptotically follow the scaling distribution with exponent α, except in the limit case $\alpha = 2$ (for α just below 2, their convergence to their scaling limit is slow).

The L-stable variables yielded by the present argument can take negative values if $1 \le \alpha \le 2$, as is readily seen in the Gaussian case. But there is a very small probability that they take *large* negative values. I have shown how this can be handled in practice by suitably displacing the origin.

L-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 from the viewpoint of ease of mathematical manipulations, they both share the fundamental properties of that distribution from the viewpoint of linear operations. The corresponding forms of the non-classical central limit theorem show that the sum of many additive contributions 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 U itself. This also shows that the log-normal distribution is *not* the only skew distribution that can be explained by addition arguments, thus removing the principal asset of that distribution (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 throughout most of the range of u. However, if $0 < \alpha < 2$, their asymptotic behaviors coincide. Hence, the scaling distribution is also asymptotically invariant with respect to applications of an arbitrary succession of the basic transformations. When 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 distributions attract substantial attention from mathematicians. However, the scaling maximum distributions have few generally known applications and the scaling sum distributions (L-stable distributions) have practically none.

It is true that a celebrated treatise on stable distributions, Gnedenko & Kolmogorov 1954, alludes to forthcoming publications specifically concerned with applications of L-stability. However, when I discussed this allusion with Professor Kolmogorov in 1958 (ten years after 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 is the distribution due to Holtsmark (but often rediscovered,) which rules the Newtonian attraction between randomly distributed stars (see Section 2.8 of M 1960i{E10}). Thus, Gnedenko & Kolmogorov 1954 did not pre-empt my plea that stable distributions should be counted among the most "common" probability distributions.

IV. SIGNIFICANCE OF THE EVIDENCE PROVIDED BY DOUBLY LOGARITHMIC GRAPHS

Limitations on the value of a lead to another quite different aspect of the general problem of observation. It concerns the practical significance of statements having only an asymptotic validity. Indeed, to verify empirically the scaling distribution, the usual first step is to draw a doubly logarithmic graph: a plot 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, look closer at the sampling point of the largest u. Except for the distribution of incomes, one seldom has samples over 1,000 or 2,000 items; therefore, one seldom knows 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 be at best equal to $3/\alpha$ units of the decimal logarithm of u. However, if one wants to estimate reliably the value of the slope a, it is necessary that the width of the graph be close to one unit. In conclusion, one cannot trust any data that suggest that α is larger than 3. Observe that the resulting practical range of α's is wider than in the case of stable distributions.

Looking at the same question from another angle, take doubly logarithmic paper and plot the following distributions: Gaussian, lognormal, negative binomial and exponential. Because all these distributions are very "short tailed," the slope of the graph will become asymptotically infinite. However, in the region of probabilities equal to one-thousandth, the dispersion of sample data is likely to generate – on doubly logarithmic coordinates – the appearance of a straight line having a high but finite slope. In the words of Macaulay 1922 (see Section IX): "The approximate linearity of 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 to describe curve-fitting using special paper. 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 distributions. Any special paper, whether it be log-normal or scaling, should be used only in ranges where the distances that it defines are sensitive to the differences that matter to the particular problem. Hence, the most conservative approach is often to consider several hypotheses, that is, to use several kinds of paper.

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

V. FINITE SAMPLE BEHAVIOR OF RANDOM VARIABLES WITH INFINITE POPULATION MOMENTS

When α is not small (in a sense we shall describe shortly), a scaling distribution is 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 the concept of infinite variance is in no way "improper."

It is of course true that, since observed variables are finite, the sample moments of all orders are themselves finite for finite sample sizes; but this does not exclude the possibility that they tend to infinity with increasing sample size. It may also be true that the asymptotic behavior of the samples is practically irrelevant because the sizes of all empirical samples are by nature finite. For example, one may argue that the history of cotton prices is mostly a set of data from 1816 to 1958, speculation on cotton having been very much decreased by the 1958 acts of the United States Congress. Similarly, when one studies the sizes of United States cities, the statistical populations have a bounded sample size. 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 it actually implies that the only meaningful consequences of infinite population moments are those relative to the sample moments of increasing *subsets* of our various bounded universes.

In Figure 2, the predictions of the mathematical theory are illustrated by computer simulations. Distinct samples of scaling random variables with $\alpha = 1$ were obtained by inverting samples of random variables dis-

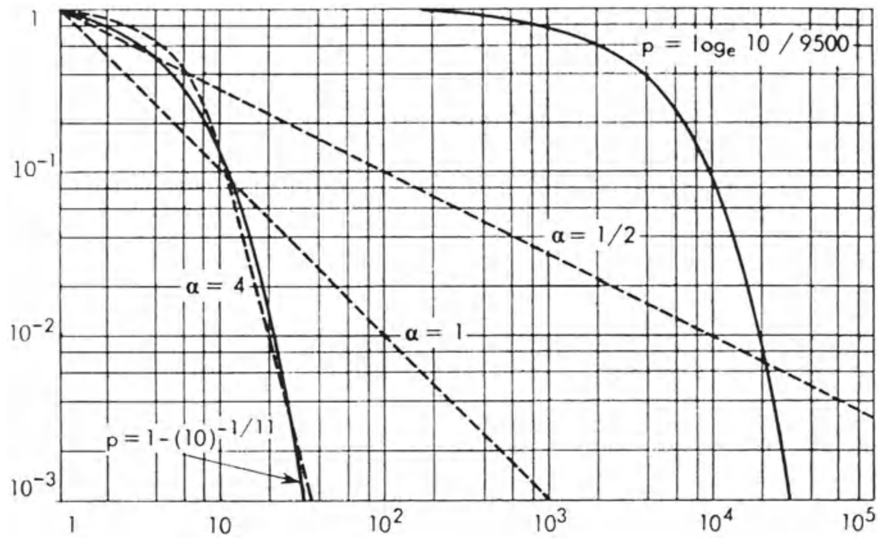


FIGURE E3-1. Five doubly logarithmic plots: (A) Two exponential distributions (*very curved solid lines*) with very different expectations. (B) Two distributions which are uniformly scaling from $u = 1$ and have, respectively, the exponents $\alpha = 1/2$ and $\alpha = 1$. (C) One asymptotically scaling distribution, with the exponent $\alpha = 4$, a large value. The relations between these graphs demonstrate graphically that distributions similar to (C) can readily be confused with the exponential, but small values of the α exponent are reliable.

tributed uniformly over the interval [0, 1]. Plots of the variation of the first and second moments are then created. The sample first moments illustrate what happens when the population moment is given by a barely divergent integral; the sample second moments illustrate what happens when the population moment is given by a rapidly divergent integral. The sample moments do not converge, and – even more impressive – their growth is erratic and very sample-dependent.

Let us now return to experimental data. In some cases, the sample second moment is observed to "stabilize" rapidly around the final value corresponding to the total set. If so, it is unquestionably useful to take this final value as an estimate of the population second moment of a conjectural infinite population from which the sample could have been drawn. But Figure 3 shows that the sample second moments corresponding to increasing subsets may continue to vary widely even when the sample size approaches the maximum imposed by the subject matter. From the

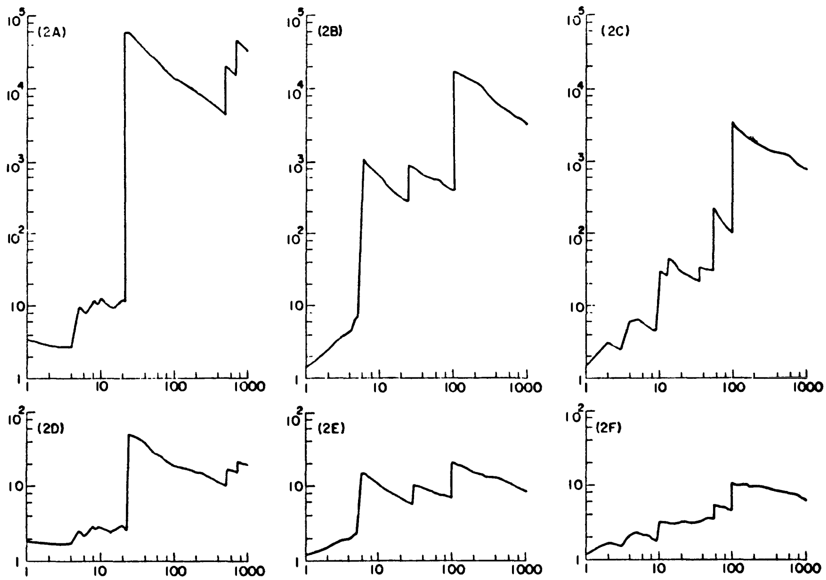


FIGURE E3-2. Monte Carlo runs of the sequential first moment (*lower graphs*) and the sequential second moment (*upper graphs*) of three independent samples from a scaling population of exponent $\alpha = 1$. The term "sequential moment" means that, in each run, the moment is computed for every sample size from 1 to 1,000. This figure suggests the degree to which the sample moments of scaling variables can be erratic and sample-dependent.

viewpoint of sampling, this expresses that even the largest available sample is too small for reliable estimation of the population second moment. In other words, a wide range of values of the population second moment are equally compatible with the data. Now, let us suppose that – as in Figure 3 – the appearance of the sample data recalls Figure 2. Then, the reasonable range of values for the population moment will frequently 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.

To support the alternative that I prefer, let me point out that a realistic scientific model must not depend too critically on 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*, is also by far more cumbersome analytically. The second

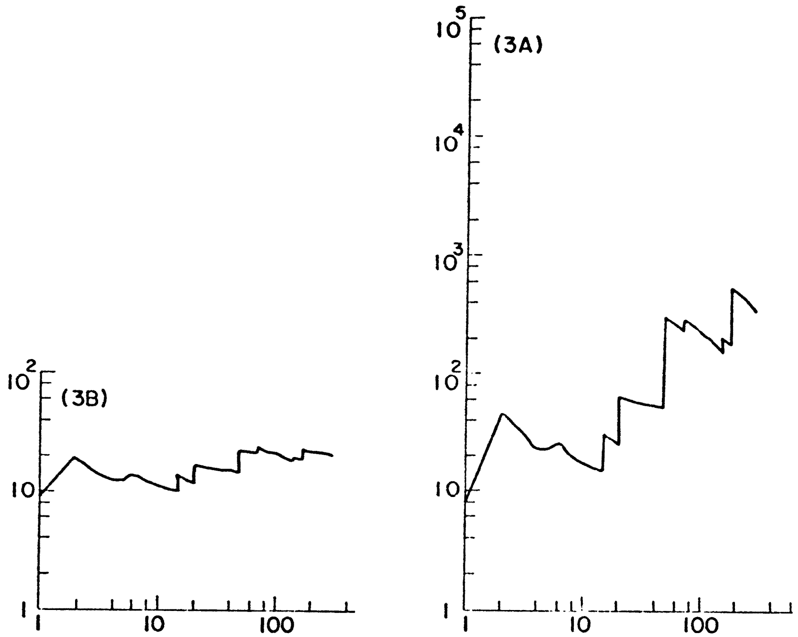


FIGURE E3-3. Sequential first moment (*left*) and the sequential second moment (*right*) of the numbers of inhabitants in United States cities with over 50,000 inhabitants. The cities have been ordered alphabetically. As city sizes have a scaling exponent of about $\alpha = 1.1$, the sample first moment tends – very slowly – to a limit, while the second moment increases less rapidly than in the simulations reported in Figure 2.

assumption, on the contrary, leads to simple analytical developments, and the rapidity of growth of the sample second moment can be modulated to 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 intrinsic bounded variables are drawn at random from infinite populations of unbounded variables having an infinite second moment. But all these infinities are a relative matter, entirely dependent on 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 certain limit, the second moments of some variables may be considered as finite. Conversely, there are variables for which the second moment must be considered finite only if the useful sample size is smaller than some limit.

Actually, this use of infinity is common in statistics, insofar as it concerns the function max $\{u_1, u_2, ..., u_N\}$ of the observations. From this viewpoint, even the use of infinite spans would seem improper. However, it is well known in statistics that little work could be done without using unbounded variables. One even uses the Gaussian distribution to represent the height of adult humans, which is surely positive!

The unusual behavior of the moments of scaling distributions can be used to introduce the least precise interpretation of the validity of the notion of scaling. For example, suppose that the first moment is finite, but the second moment is infinite. Then, as u tends to infinity, the function $1 - F(u)$ must decrease more slowly than $1/u^2$ but more rapidly than $1/u$. In this case, the behavior of $F(u)$ in the tails is very important, and a very useful approximation may be $Cu^{-\alpha}$, with $1 < \alpha < 2$. This approximation is completely harmless as long as one limits oneself to consequences that are not very sensitive to the actual value of a. The situation is very different when the tail is very short, for example, when the population moments are finite up to the fourth order. In that case, the behavior of the function $F(u)$ for large u is far less important than its behavior elsewhere; hence, one will risk little harm with interpolations by the Gaussian or the lognormal distribution.

VI. DIFFICULTIES CONCERNING STATISTICAL INFERENCE AND CONFIRMATION OF SCIENTIFIC DISTRIBUTIONS, WHEN THE ERRORS (THAT IS, THE "BACKGROUND NOISE") ARE SCALING

It is well known that second moments are heavily used in statistical measures of dispersion, or "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 failures have, of course, often been observed empirically and may have contributed to the disrepute in which many writers hold the scaling distribution; but it is clearly unfair to blame a formal expression for the complications made inevitable by the data that it represents. If $2 < \alpha < 3$, second moments exist, but concepts based on third and fourth moments, – for example Pearson's measures of skewness and kurtosis – are meaningless.

I am certain that for practical purposes some of those difficulties eventually will be solved. However, as of today, they are so severe that we must reexamine the meaning of the popular but vague concept of "a structure." It is indeed a truism, especially in fields where actual experimentation is impossible, that one must carefully distinguish between patterns that can only be used for "historical" description of his records and those that are also useful for forecasting some aspect of the future. A useful vocabulary considers the search for distributions a kind of extraction and identification of a "signal" in the presence of "noise." In particular, as we have seen, modern inference theory teaches us always to list both the accepted and the rejected possibilities. The scientist's major problem is frequently to determine whether a conjectured "relation" is statistically 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.

It is not enough, however, that all members of a cultural group agree on the patterns that they read into a historical record. Indeed, although there is unanimity in the interpretation of *certain* Rorschach inkblots, they have no significance from the viewpoint of science as a system of *predictions*. Broadly speaking, a pattern is scientifically significant when it is felt to have a chance of being repeated, meaning that, in some sense, its "likelihood" of having occurred by chance is very small. Unfortunately, the tools of statistics have been mostly designed to deal with Gaussian alternatives and, when the chance alternative is scaling, they are not *at all* conservative or "robust" enough. One will often be able to circumvent this difficulty, but not always. In fields where the background noise is scaling, the burden of proof is closer to that of history and autobiography than of physics.

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

The situation is worse in models known to be very structured (for example, to be autoregressive) with scaling noise. Compared to the case of Gaussian noise, one should expect the data to be *much* more influenced by the noise and *much* less influenced by the structure.

The association between the scaling distribution and "interesting patterns" is nowhere more striking than in the game of tossing a fair coin, which Henry and Thomas have been playing since sometime in the early eighteenth century. When the coin falls on "heads," Henry wins a dollar (or perhaps rather a thaler); when the coin falls on "tails," Thomas wins. We disregard what happened to the game before we break in at time $t = 0$, and we denote by T the time it takes for Henry and Thomas's fortunes to return to the state that they were in at the moment when we broke into the game. For large values t of T, one has the well-known relation: (Feller 1950, Vol. 1).

$$\text{Probability}\ \{\text{that the fortunes return to their initial states after a time greater than } t\} = (\text{constant})\ t^{-1/2}.$$

This relation involves the scaling distribution with exponent $\alpha = 1/2$.

However, gamblers are notorious for seeing an enormous amount of interesting detail in the past records of *accumulated* coin-tossing gains; far more than in the non-cumulative sequences. That is, gamblers are prepared to risk their fortunes on the proposition that these details are not due to mere chance. Several of my papers were based on the idea that very similar phenomena should be expected whenever the scaling distribution 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 sample curves in which both the unskilled and the skilled eye can distinguish the kind of detail that is usually associated with causal relations. In the case of Gaussian processes, such details would be so unlikely that they would surely be considered significant for forecasting; but, this is not true in the scaling case. From the viewpoint of prediction, those structures should be considered *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. A non-structured scaling universe accounts very well for many observations; as a result, it is extremely difficult, at best, to choose between it and an alternative model that postulates causal relations. It is very difficult to challenge someone's belief in the existence of "genuine" structures. But to communicate such a belief to others, with the standards of credibility that are current in physical science, requires *much* more than the statistical tests of significance that social scientists shrug off at the end of a discussion. Such a situation requires a drastic sharpening of the distinction between patterns that – however great the scholar's diligence – can serve only for historical purposes and those patterns that are useable for forecasting.

The question that I have in mind can be well illustrated by the problem of the significance of "cycles." Both the eye and sophisticated methods of Fourier analysis, suggest that almost any record of the past is a sum of periodic components. But the same is also true for a wide variety of artificial series generated by random processes with no built-in cyclic behavior. Furthermore, skilled cycle researchers seldom risk firm, short-term forecasts. Could we then ask two questions that paraphrase Keynes's comments on 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 scaling distribution with respect to various transformations (see Section III), one cannot hope that a simple explanation 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 scaling spontaneous activity differ from the standards of "operationalism" suggested by philosophers. Indeed, 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 realize all of them. Hence, within our lifetime, any given configuration will occur at most once, and one could hardly define a probability on the basis of sample frequency. This conceptual difficulty is common knowledge among physicists, and it is to be regretted that the philosophical discussions of the foundations of probability seldom investigate this point. In a way, the physicists freely indulge in practices that for the historian are mortal sins: to rewrite history as it would have been if Cleopatra's nose had a different shape. My sins are even worse because their actual histories turn out to be very close to some kind of "norm," a property which my models certainly do not possess.

The foregoing argument is best illustrated by two separate reinterpretations of the coin-tossing record plotted in Figure 4. First, forgetting the origin of that figure, 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. Imagine also that this chart has just been brought home by an explorer; the problem is to decide whether it was due to cause or to chance. The naive defense will resort to the Highest Cause. Presenting our graph as fresh evidence that God created Heaven and Earth using a single template, it follows that such concepts as a "continent," an "ocean," an "island," an "archipelago" or a "lake" are precisely adapted to the shape of the Earth. However, a devil's advocate would argue 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 fact that the areas above water happen often to be very short or very long and are rarely of average length.

The preceding example is not as fictitious as it may seem: the distribution of the sizes of actual islands happens to be scaling (M 1962n). Hence, our hypothetical debate emphasizes the two extreme viewpoints 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 used only 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. Moreover, since 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 higher order.

Actual records of the instants when errors occurred do indeed exhibit such clusters in between long periods of flawless transmission. A good idea of the distribution of the errors is provided by yet another look at Figure 4. Consider the sequence of points where the graph crosses the

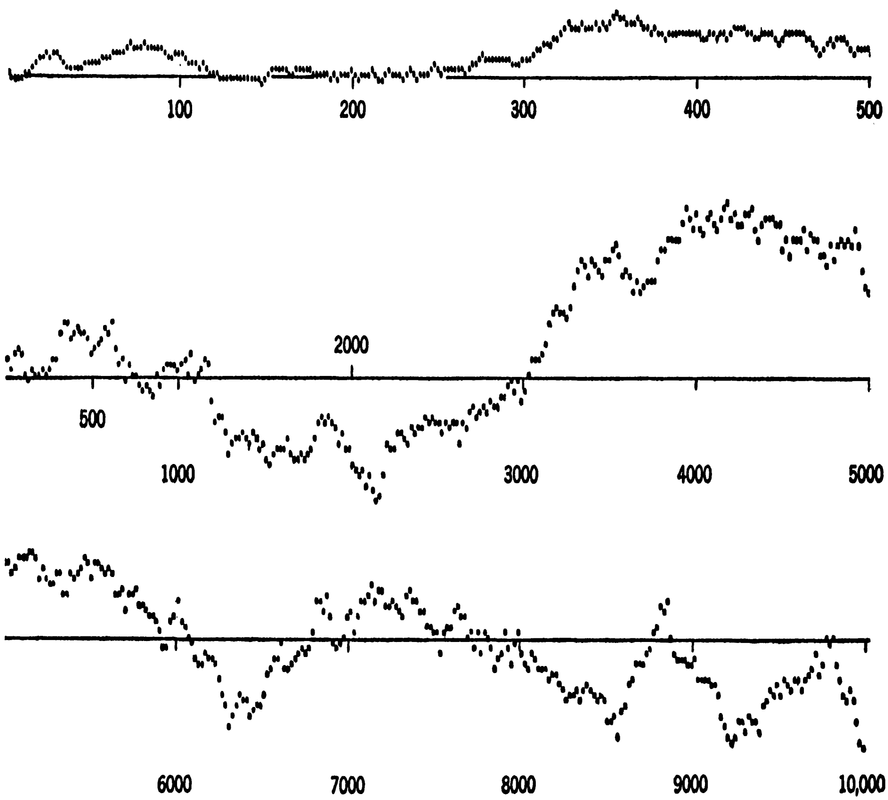


FIGURE E3-4. Record of Henry's winnings in a coin-tossing game, played with a fair coin. Zero-crossings seem to be strongly clustered, although intervals between crossings are obviously statistically independent. This figure is reproduced from Feller 1950 (Volume I).

To appreciate fully the extent of apparent clustering in this figure, note that the unit of time is 2 coin tosses on the first line, and 20 coin tosses on the second and third lines. Hence, the second and third lines lack detail and each apparent zero-crossing is an imperfect representation of a cluster or a cluster of clusters. For example, the details of the cluster centered around the 200th coin toss are clearly separated on line 1.

line that, in an earlier interpretation, had represented sea level. According to the searcher for causes, the precise study of such past records will improve the prediction of errors and will help 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 memory (see Berger & M 1963).

Similar devil's advocates can be heard in many contexts, and someone should take this role 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 a 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 scaling distribution (see Apostel, M & Morf 1957 and M 1961b).

VII. TWO-TAILED AND/OR MULTIDIMENSIONAL STABLE DISTRIBUTIONS

Until now, we have followed tradition by associating the scaling distribution with essentially positive random variables, the distribution of which has a single long tail, making its central portion necessarily quite skew. However, I have discovered important examples in economics of distributions having *two* scaling tails; the most striking example is that of relative changes in the prices of sensitive speculative commodities. The argument of invariance under maximization cannot extend to them. But invariance under mixture simply leads to the combination of the scaling distribution of positive u and the scaling distribution of negative u. Invariance under aggregation is satisfied by every random variable constructed by adding or subtracting two arbitrarily weighted "positive" stable variables of the kind studied earlier in this paper. In particular, these general stable variables can by symmetric; the Cauchy distribution provides a prime example. But their study depends very little on the actual degree of skewness. Hence, the asymmetry of the usual scaling 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 multivariant case, other than the degenerate case of independent coordinates. Very few other distributions (if any) share this property. The reason for this is innately 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 coordinates is a scalar stable variable. This property is essential to the study of multidimensional economic quantities, as well as to the investigation of the dependence between successive values of a one-dimensional quantity, such as income (see M 1961e{E11}).

VIII. THE ROLE OF THE SCALING DISTRIBUTION IN ECONOMICS AND A LINK WITH THE PHYSICAL SCIENCES

The arguments of this paper show that there is a strong pragmatic reason to undertake the study of scaling economic distributions and time series. This category includes prices (M 1963b{E14}), firm sizes (M 1963o) and incomes (M 1960i{E10}, as amended in M 1963p, and also M 1967j{E15}, 1962g), hence making the study of scaling of fundamental importance in economic statistics. Similarly, the example of the distribution of city sizes stresses the importance of the scaling distribution in sociology (M 1965m). Finally, strong indications exist of its importance in psychology, but I shall not even attempt to outbid George Kingsley Zipf in listing all the scaling 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 on aggregation, choice and mixture, one can prove that for a system to be scaling, it *must* be triggered somewhere by essentially scaling "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 of scaling cannot be solved by pushing around the point where such behavior is postulated. Indeed, if it were true, in accordance with "conventional wisdom," that physical phenomena are characterized by the distributions of Gauss and social phenomena are characterized by that of Pareto, we may eventually have to explain the latter using the "microscopic" economics models, such as the "principle" of random proportionate effect, which I prefer not to emphasize in my approach.

I claim, however, that this situation *need not* be the case. Quite to the contrary, the physical world is full of scaling phenomena that one can easily visualize as playing the role of the "triggers" that cause the economic system to be also scaling. For example (M 1962n), I have found that single-tailed scaling 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. This is the case with the areas of oil fields and the sums of their total past production and their currently estimated capacity). The same is true for the valuations of certain gold, uranium and diamond 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 weather data, such as hail records, have a direct influence on important risk phenomena, namely, insurance against hail damage. Other weather data, such as total annual rainfall, obviously influence the sizes of crops and hence, by the distributions of supply and demand, influence the changes of agricultural prices.

If this paper proposed to contribute to "geo-statistics," it should, of course, examine 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 the trigger of a scaling grand economic system *can very well* be found in 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 the scaling exponent α are encountered in both worlds 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 scaling phenomena of physics have also turned out to include some phenomena with no direct relation with economics. For example, Section 3 mentioned that a three-dimensional stable distribution occurs in the theory of Newtonian attraction. Moreover, the distribution of the energies of the primary cosmic rays has long been known to follow a distribution that happens to be identical to that of Pareto with the exponent 1.8 (Fermi's study of this problem includes an unlikely but rather neat generation for the scaling distribution). The same result holds for meteorite energies, an important fact for ionospheric clatter telecommunications. Also, as discussed in Section VI, the intervals between successive errors of transmission on telephone circuits happen to be scaling with a very small exponent, the value of which depends on the physical properties of the circuit.

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

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

Having accumulated so many reasons to view the scaling distribution as extraordinarily important, I am continuously surprised by the attitude described in the first sentence of Section I. I eventually realized that it had deep roots not only in the apparent lack of theoretical motivation for that distribution but also in several seemingly "definitive" criticisms, such as that of Macaulay 1922.

Macaulay's essay is most impressive indeed and – even though I disagree with its conclusions – I strongly recommend it. It disposed of the claim that the α exponent in Pareto's law is the same in all countries and at all times and of the claim that the scaling distribution describes small incomes or the incomes of the lower paid professional categories. Macaulay is also very convincing concerning scaling distributions with a high exponent (see Section V).

I believe, however, that his strictures against "mere curve fitting" have been very harmful. His ideal of a proper mathematical description is so restrictive that he rejects the scaling distribution outright because the sample empirical curves do not "zigzag" around the simple scaling interpolate but rather cross it systematically a few times. This illustrates a basic difference between the care economists bring to statistics and the seeming carelessness of the physicists. For example, when the Boyle law was found to differ from the facts, the physicists simply invented the concept of a "perfect gas," that is, a body that follows Boyle's law *perfectly*. 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, scaling distribution approximations should not even be considered in problems relating to low incomes, but in other investigations they deserve to be the first to be considered.

Therefore I can summarize Macaulay's criticism of the scaling distribution by saying that it only endorses the asymptotic forms. In many cases, however, I believe that it is legitimate to consider more seriously certain "relatives" of the scaling distribution, such as the stable distributions.

APPENDIX: SOME MATHEMATICAL DERIVATIONS

Characterize U by its distribution function $F(u) = \Pr\{U \le u\}$ and its generating function $G(s)$, which is the Laplace transform of $F(u)$, namely

$G(s) = \int_{-\infty}^{\infty} \exp(-us)dF(u)$. In order for $G(s)$ to be finite, it is necessary that $dF \to 0$ very rapidly as $u \to -\infty$. Then, invariance-up-to-scale is expressed by the following conditions:

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

$$\frac{1}{N}\sum F\left(\frac{u}{a_n}\right) = F\left(\frac{u}{a_W}\right).$$

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

$$\sum \log F\left(\frac{u}{a_n}\right) = \log F\left(\frac{u}{a_M}\right).$$

Aggregation. It is necessary that

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

It turns out that the three types of invariance lead to "functional equations" of almost identical form, although they refer to different functions, respectively, F_W, $\log F_M$ and $\log G_A(s)$. Therefore, general solutions of these equations are alike. They assume the following forms

$$F_W(u) = C' - Cu^{-\alpha}, \quad F_M(u) = \exp(-Cu^{-\alpha}), \quad \text{and } G_A(s) = \exp(-Cs^{-\alpha}).$$

One easily verifies that $a_M^\alpha = a_A^\alpha = \Sigma a_{n^\alpha}$ and $a_W^\alpha = (1/N)\Sigma a_n^\alpha$.

I shall now show that the above conditions are 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 such that $F_M(\infty) = 1$. This requires that $C > 0$ and $\alpha > 0$, which leaves us with $F_M(u) = \exp(-Cu^{-\alpha})$.

Mixing. In order that $F_W(u)$ be non-decreasing and satisfy $F_W(\infty) = 1$, it is 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 $C > 0$.

ACKNOWLEDGEMENT

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 – particularly Merton Miller and Lester Telser – for detailed discussions of earlier oral and written versions.

&&&&& POST-PUBLICATION APPENDIX &&&&&

THREE ASPECTS OF THE NOTION OF RENORMALIZATION

1. Footnote 4 in the original, and comment. The many footnotes in the original, except one, were easily integrated in the text. But Footnote 4 did not fit, and it cried out to be emphasized, because it was an early allusion to the theme of self-similarity that came to dominate my life and led to fractals. This footnote 4 read as follows:

"The various criteria of invariance used by physicists are somewhat different in principle from those I propose in economics. For example, the principle of relativity was not introduced to explain a complicated empirical relation, such as scaling. I am indebted to Harrison White for suggesting that I should stress the nuances between my methods and those of physics."

Harrison White is a sociologist with a background in hard science, and his comment was made after a seminar I gave in Cambridge in 1962-3, while I was visiting as professor of economics at Harvard. At that time, little did anyone expect that 1963-4 would still find me at Harvard, having moved over from Littauer Hall to teach applied physics in Pierce Hall. This was in the right place to be reminded of a topic I had studied at Caltech in 1948, namely the 1941 "Kolmogorov" theory of turbulence. The "K41" theory concluded that the spectrum of turbulent velocity should be $k^{-5/3}$. Robert W. Stewart's group at Vancouver had been the first to observe $k^{-5/3}$ in an actual experiment, and Stewart was also visiting Pierce Hall. At this point, it became clear that my version of the method of invariances has far less to do with Einstein than with Kolmogorov.

2. The physicists' concept of "renormalization" and the economists' concept of "aggregation." Section 3 will discuss the relation between the method of invariances used in this chapter, and the physicists' *renormalization*. This last term may be unfamiliar to economists, but is conceptually close to the notion of *economic aggregation*. The latter addresses the question of how, starting with the economic rules that apply to individuals, one can obtain rules relative to families and larger aggregates. It may be, but one cannot be sure, that colleagues' interests in this aggregation helped inspire me to ask how the rules relative to daily price change can be transformed into rules relative to price change over weeks and longer periods.

3. Annotation for the physicists. Only a few years after the events described in Section 1, reporting on the original Footnote 4, a current in the mainstream of "physics" turned very successfully, to the study of the "critical points" of thermodynamics. In the resulting intellectual context, the main themes of this paper are very easy to introduce.

The scaling distribution is known in physics as a "power-law" or "algebraic" distribution.

The operations with respect to which the tail of the scaling distribution is invariant are known in physics as "renormalizations." Three different renormalization are used in this chapter, one linear and two non-linear ones. Each has its own "fixed point," namely, its own "exactly renormalizable" distribution. Therefore, the key fact of this chapter may be described as reporting a property of the asymptotically scaling distribution: it is "asymptotically renormalizable" in three different ways.

Given that this paper was written during the years preceding the original publication in 1963, it could in no way be affected by the later development that introduced renormalization into physics proper.

E4

Sources of inspiration and historical background

✦ **Abstract.** This chapter is written in the style of an acknowledgement of broad intellectual debts. All my scienfitic work fell under the influence of the branch of physics called thermodynamics, and of other independent traditions ranging from deep to very shallow. I came to scaling and renormalization by cross-fertilizing the influences of probability theory (Lévy) and the social sciences (Pareto, Zipf and the economists' idea of aggregation.)

At a point where my views on scaling were already formulated, I became aware that this notion is also fundamental in the study of turbulence (Richardson, Kolmogorov.) The theories of disorder and chaos, which also make extensive use of scaling and renormalization, arose from a different and independent tradition, and did not influence my work until quite late. Furthermore, diverse scaling rules were recorded in geology, but not appreciated, and the biologists' allometry is yet another expression of scaling.

As my study of scaling became increasingly visual and grew into fractal geometry, it became widely agreed that fractal aspects are present in many fields; their importance is limited is some and fundamental in others – including finance. ✦

MY SCIENTIFIC LIFE WAS PERMANENTLY AFFECTED by the fact that I wrote a Ph.D. thesis alone, without an advisor. Very influential, however, were two persons, namely Lévy and Zipf, and a field of inquiry, namely thermodynamics. After the doctorate, I was greatly influenced by John von Neumann (1903-1957), who brought me to the Institute for Advanced Study in Princeton. (Fate made me his last post-doctoral fellow.) His intellectual openness and awe-inspiring versatility became a model to emulate, but there was little technical overlap at that time between his interests and mine.

In any event, fractal geometry had no single source and no actual forerunner. But neither did it materialize out of thin air. It began by answering a few concrete questions that were long left aside, by using a few mathematical tools that were pointedly *not* meant to answer any concrete question. In doing so, it combined two streams of thought that started from widely separate sources, a particularly "pure" one and one that can – to make a point – be called particularly "dirty." In due time, new questions were asked, new tools were developed, and additional sources of scaling were acknowledged, as will be seen in Sections 5 and 6.

This intellectual development can be summarized as follows. It began with a collection of "power-law" statistical distributions (exemplified by Pareto's law for the distribution of personal income) that were collected in Zipf 1949. I was the first to take several steps together: a) to recognize a kinship between those empirical power laws and the theoretical power laws that occur in probability theory; b) to interpret power laws in terms of scaling; c) to interpret the limit theorems of probability theory as involving what is now called "renormalization," (M 1963p{E3}); d) to interpret the limits in those theorems as "fixed points" of renormalization.

The goal of this brief chapter is to elaborate on those steps.

1. Probabilistic "answers without questions"

Lévy. Among those who provided answers without questions, the first was Paul Lévy (1886-1971), whom I met when I was 20, came to know well, and helped reach a high place in the history of his field. In his lifetime, he was ignored or spurned by the Paris mathematical community, often in a harsh way (Lévy 1970, M 1995l.)

Picking up a stream of thought that originated quite explicitly in Cauchy 1853, Lévy made key contributors to the emergence of a tool that plays a central role in this book: the probabilist's version of the notion of scaling, as implemented in the probability distributions that he called *stable*. This term being terribly unfortunate and misleading, I toyed with alternatives. "Pareto-Lévy," "Lévy distribution," and "stable Paretian" did not take root, created confusion, and were withdrawn.

This book is an opportunity to denote my earliest scaling model of price variation, the M 1963 model, by the term "L-stable." This is the word used by Lévy and all probabilists, but preceded by the letter L to warn against misunderstanding and honor my mentor.

Lévy is the hero of one of the biographical sketches in Chapter 40 of *FGN* (see also M 1995l). His utter surprise was forcefully restated each time I reported to him yet another natural phenomenon ruled by what used to be viewed as "mathematical pathology," including his own "pathological" creations. His unwitting contributions to the mathematical toolbook of science went beyond the L-stable distributions. In fact, my fractal forgeries of mountains, which everyone seems to have glimpsed somewhere, began with a suitable generalization of Lévy's Brownian function of several variables.

2. Concrete "questions without answers"

Divisia. To digress briefly, École Polytechnique presented a sharp contrast between the roles of the Professors of Mathematics (Lévy) and Economics. François Divisia (1889-1964) lectured after dinner and the grade he gave hardly mattered to the final ranking that determined the students' later career. This contributed to my having no formal training in economics until the period when the bulk of this book was written, that is, at an age ranging from 36 to 43. No attempt is made to hide my lack of familiarity with economic theory. Not until much later did I acquire close familiarity with economic data, notably those collected by the National Bureau of Economic Research.

Zipf. Among those who provided useful social science questions without answers, the foremost was George Kingsley Zipf (1902-1950), whose name became linked with the rank-size plots discussed in Chapter E7. Zipf 1949 was the work of encyclopedist following an "idée fixe."

He is best remembered for curve-fitting all his data using "power-law" distributions. His zeal was uncontrolled and his assertions should be checked, especially on discoveries of which he was the actual originator. Yet he deserves praise for giving scrupulous credit to others for findings he could have safely claimed for himself, and he played a key role in collecting and preserving knowledge that everyone else ignored.

Unfortunately, an orphan status often befalls empirical discoveries that remain unexplained and not embedded in an over-reaching theory. They remain suspended in an intellectual vacuum, casually disregarded, dismissed, challenged and even ridiculed by professional investigators but attractive to non-professional dabblers of every kind and provincial professors – of which I was one when my career was starting. Most such discoveries rapidly vanish into the dustbin of science. Lacking a theory, scientists did not expect to encounter power-law distributions; therefore, they did not face them, or even failed to see them.

To avoid this fate for his discoveries, Zipf's "idée fixe" was the need for a unifying structure. However, lacking technical background and discipline to draw logical conclusions he took a step familiar in politics and in pseudo-science. He proclaimed that scaling in social sciences follows from a grandiose "principle of least effort" which he did not take the trouble to phrase and study. It evaporates upon examination.

Zipf 1949 received numerous reviews. Those due to social scientists raved about "least effort," but failed to mention the empirical power-law distributions. In sharp contrast stood Joseph L. Walsh (1895-1973), a Harvard mathematician and personal friend of Zipf. Walsh 1949 featured the scaling rule for word frequencies as a puzzle worth serious attention. Walsh's challenge led me to the explanation featured in Section 1.2.4 of Chapter E8; in the longer run, it set a direction to my life.

To avoid being accused of a bias against social science, let me acknowledge an uncomfortable parallelism between Zipf's approach and that of some applied physicists who studied the so-called "$1/f$ noise." That topic is discussed in M 1997N, M 1997H, and elsewhere in my work.

All told, Zipf does deserve a footnote in history, but the sketch in Chapter 40 of M 1982F{*FGN*} ends by noting that "one sees in him, in the clearest fashion - even in caricature - the extraordinary difficulties that surround any interdisciplinary approach."

Pareto. Vilfredo Pareto (1848-1923), being a major Establishment figure in economics, found no place among the maverick heroes of Chapter 40 of *FGN*. But there was a maverick side to him. Besides, while every economist heard of Pareto, a few words may be welcomed by readers of this book who are not economists.

Pareto was an Italian born in Paris, who taught in Lausanne, Switzerland, and left his deepest mark in the countries where he lived. In mathematical economics, his work followed upon Léon Walras (1834-1910) and sought to define and study economic equilibrium. But my interest in Pareto's work was restricted to his empirical law for the distribution of personal income, the topic of several chapters in this book. It is chastening to recall two sets of empirical statistical regularities discovered in the XIXth century.

One set contains the old textbook examples concerned with Army conscripts: their heights were fitted by the normal distribution and the numbers of times they fell off their horses by the Poisson distribution. The statistical fits of each distribution were casual, but the Gaussian and the Poisson became pillars of statistics.

Far greater and more motivated effort was expended by Pareto in fitting personal incomes by power-law distribution. Never forgotten, Pareto's law never flourished either – not unlike the work of Zipf.

Bachelier and belated followers. Louis Bachelier (1870-1946) is the precursor of all statistical approaches to finance. Bachelier 1900 (his Ph.D. dissertation) was eventually translated into English (and issued in Cootner 1964). In 1995, the French original was reprinted in book form, and a French professional group that keeps close to the tenets of the Brownian motion model, without paying attention to its flaws, named itself *Association Louis Bachelier.* It may be that the man's posthumous fame was helped by the biographical sketch found in M 1982F{*FGN*}. His works were not unknown in France, where I first heard of them. But they played a limited role when several distinguished American scholars came, independently of each other, to propose Brownian motion as a model of price behavior. I would like to produce a fair list, but mostly recall Osborne 1959 and Roberts 1959.

Brownian motion would have sufficed as a claim to fame, but Bachelier went further. As is well-known, he originated the notion of *efficient market*, and, to express it mathematically, created the general notion of *martingale.* Among martingales, as true for Bachelier as it is for us today, the special quality of Brownian motion resides in its being by far the easiest to handle analytically.

Incidentally, to answer some revisionist historians, Bachelier stated the notion of martingale correctly on many occasions. The handiest is on pp. 27-28 of the English translation in Cootner 1964 and reads: "The mathematical expectations of the buyer and the seller are zero, [a property that follows from] this fundamental principle: *The mathematical expectation of the speculator is zero.*" For 1900, this was perfectly rigorous.

It is a second but little-known achievement of Bachelier that makes it appropriate to mention him here: he pioneered both in discovering the Gaussian random-walk model and in noting its major weakness. He saw, as noted in Section 1.1 of M 1967j{E156}, that the Brownian model diverges from the evidence in at least two ways: Firstly, the sample variance of $L(t, T)$ varies in time. He observed that if the sample histograms are relative to mixtures of distinct populations, their tails could be expected to be fatter than in the Gaussian case. Second, Bachelier noted that no reasonable mixture of Gaussian distributions could account for the sizes of the very largest price changes, and treated them as "contaminators" or "outliers."

Hurst. The M 1963 model of price variation owes to Lévy, Pareto, and Zipf, but the M 1965 model came to economics via a power-law relation in hydrology. The trigger was a finding concerning persistence in water discharges, namely, an irritating puzzle contributed by Harold E. Hurst (1880-1978) that I recognized as being a symptom of scaling. A biographical sketch is provided in M 1982F{*FGN*}. Fractional Brownian motion strongly links my work in economics and the physical sciences.

Morris. Moving beyond those who passed away, I want to mention that, in the 1960s, my main source of market wisdom and folklore was William S. Morris. When he was riding high, a Princeton University conference on "Operations Research for Top Management" invited him to reveal his secrets; instead he read from Morris 1962, a lavish booklet from which I excerpted in Chapter E17.

The pervasive role the computer now plays on Wall Street did not just happen. Morris has a strong claim to be hailed as the visionary who first used a computer for trading, in addition to accounting. "The timeworn, but little understood, rules of thumb used by jobbers in the practice of their trade can be ... programmed into a computer ... In my company, to underscore our conviction that computerization is feasible now, we have publicly offered to furnish computer, computer program, and the capital to bank the jobbing activity."

The study of aggregation. When I was starting in economics, many of my new colleagues were investigating *aggregation*. This was the search for rules concerning "aggregates" such as families, when one starts from rules concerning individuals. What I was doing was to "aggregate" different sources of income or price changes over different time spans. The physicists' renormalization (Section 6) went very much farther than the economists' aggregation, but started with the same idea.

3. A nemesis of fractals: Gibrat and lognormality

The relation of the lognormal probability distribution to economic inequalities is the work of many hands, but dominated by Gibrat 1932. Robert Gibrat (1904-1980) was a prominent French civil servant.

Thirty years ago, friends and referees were near-unanimous in advising me to give up the scaling distribution and acknowledge the authority of the lognormal. They all pointed out a revealing statement on pp. 101-2 of Aitchison & Brown 1957, which reads that "A number of distributions are given by Zipf, who uses a mathematical description of his own manufacture on which he erects some extensive sociological theory;

in fact, however, it is likely that many of these distributions can be regarded as lognormal, or truncated lognormal, with more prosaic foundations in normal probability theory."

Observe that the "description of Zipf's manufacture" was nothing but the scaling distribution. "Everyone knew" that a theoretical basis existed for the lognormal but not for the scaling. But Chapters E8 and E9 will show that motivations for the lognormal left me unconvinced. For example, right or wrong, everyone expected the Gaussian in additively aggregate phenomena and the lognormal in multiplicative phenomena. But I found that *not all* multiplicative phenomena yield the lognormal and no one has advanced a full explanation of why, to take one example, an oil field's capacity should involve a multiplicative process.

Gibrat's claim that everything of interest is fitted by either the Gaussian or the lognormal distributions owed to a dearly held cliché that goes back to Auguste Comte (1798-1857) and holds that the more perfect fields, like physics, instruct and guide the less perfect ones, like economics, and serve as examples to emulate, while the converse is inconceivable.

In particular, an often stated reason why empirical evidence in favor of scaling distributions was considered suspect was that they were not part of physics. In due time, a few hidden old examples did surface, but in the early 1960s they were not known. Also, there was no awareness of the material to be discussed in Section 6. The examples of scaling collected by Zipf concerned social science and (as mentioned in Section 2) were held in low esteem.

4. The theme of (greatly generalized) statistical thermodynamics

I am a life-long student of thermodynamics for its own sake; thus, M 1964t advanced a variant of its foundation. More important, I treasure thermodynamics for bringing powerful tools and subtle methods of thought that transcend its original applications to gases. The approach to thermodynamics represented by Josiah Willard Gibbs (1839-1903) particularly attracted me. But for the very same reason it was criticized at one point by Ludwig Boltzmann (1844-1906), who spurned and dismissed its generality on the ground that it would only be useful when dealing with "assemblies of typewriters and sewing machines" (I quote from memory.)

In my earliest work on Zipf's law (M 1951; see also Chapter 38 of M 1982F{*FGN*} and Section 4.1 of Chapter E8), one of several variants involved explicit thermodynamics in a phase space of (lexicographic) trees. Later on, referees' pressure led me to abandon gradually all specific references to thermodynamics, therefore I came to scaling and renormalization through probability theory, before their tools became a major part of statistical physics, as will be seen in Section 5. But thermodynamics remains one of the main conceptual threads throughout my work. A novelty concerns the identification of applicable limit theorems of probability. The usual limit theorems characterize the usual state of randomness, which I call "mild", and must be replaced by very different theorems; Chapter E5 describes them as characterizing "wild randomness."

5. The themes of chaos and disorder as defined in physics

As already mentioned, the editor's comments in Cootner 1964 describe M 1963b{E14} as having "marshalled ... evidence of a more complicated and much more disturbing view of the economic world than economists have hitherto endorsed." Read today, these words unavoidably bring physics to mind: "disorder" became attached to materials studied by statistical physics, and "chaos," to phenomena studied by dynamics (also called, for no good reason, "theory of dynamical systems".)

After I drifted away from finance, much of my scientific life was spent in those two fields, beginning with a problem that straddles both, namely, turbulence in fluids. But disorder and chaos became organized too late to influence my use of scaling and renormalization in finance. This is why, as already mentioned in the Preface, economists who took for granted and necessary that physics should lead, went on to mistrust as "premature" a tool that physics had not yet tested.

Disorder in statistical mechanics, scaling and renormalization. Most readers probably know little about physics, therefore even a brief sketch would demand overly extensive preliminaries. But it is essential to mention that yet another source of scaling and renormalization is found in classic investigations by Murray Gell-Mann and Francis Low. Inspired by that work, the same key tools became essential in the theory of critical phenomena and disorderd materials. Major contributors include, listed alphabetically, Michael E. Fisher, Leo P. Kadanoff, Benjamin Widom and Kenneth Wilson. As already said, renormalization went far further than the economists' aggregation, but started with the same idea.

It is worth pointing out that an expository paper, M 1982v included a section on "The Scaling Principle of Economics" (later adapted for *FGN* as Chapter 39.) There, it is predicted that "when the inevitable comes and an economist decides to look at collective phenomena of physics for inspiration in economic modeling, he will find that part of the work has already been done without reference to physics." However, a warning is in order.

Powerful laws (expressed by "Hamiltonians") allow physics to *explain* scaling and exponents with great precision. But those laws have no counterpart in finance. Therefore, to present the use of scaling and renormalization in finance as coming from statistical physics, would contradict history and in addition would ring hollow.

Chaos. Thirty to thirty-five years ago, while I was studying turbulence in prices and physical fluids, a quite separate event occurred when Edward Lorenz wrote the paper that introduced scientists and laymen to the "butterfly effect." The basic idea is now widely known (Gleick 1986) and was already well described in Hadamard 1898. That work came long before its time and became an excellent example of "prematurity" in the sense of Stent 1972. It was filed away in one of those deep vaults that shelter mathematics that is not immediately followed up. While called classical, it remained obscure until its time came.

Lorenz became extraordinarily influential, and his work led to a distinction between *well-behaved* and *chaotic* determinism. It is parallel to the distinction between *mild* and *wild* randomness, which I was beginning to draw at the same time (see M 1987r); it is the topic of Chapter E5. A mutual MIT friend, Erik Mollo-Christensen, had the hunch, brilliantly confirmed by later events, that the intellectual efforts of Lorenz and myself may somehow be related.

As the unavoidable and legitimate free association between fractals and chaos theory became widely known, there were many attempts to improve our understanding of financial fluctuations by invoking the theory of deterministic chaos ... and even the Mandelbrot set! (See Gleick 1986 for a discussion of M 1980n.) Those attempts were *not* a direct descendent of my papers of 1960-1973. In any event, the vague notion that "there is chaos in the market" used to be understood in a statistical fashion. To reinvent it in terms of deterministic chaos would require an algorithm to distinguish between the presence and absence of randomness. Such algorithms were put forward, accompanied by broad claims I always viewed as extravagant; the inadequacy of the Grassberger-Procaccia algorithm is now patent.

The hope to "explain" finance via deterministic chaos is part of a very deep trend. Would-be explanations are welcomed even when they are sketchy, and are less harshly scrutinized than excellent descriptions that are forthright and do not even pretend to explain.

6. Some other sources or uses of scaling

M 1982F{*FGN*} describes how, as the uses of scaling multiplied in my work, I also became aware of independent additional sources of the same idea. They have deep roots in philosophical and poetic discourse (William Blake saw "a world in a grain of salt" and far older sources are quoted in M 1982F{*FGN*}.) More importantly, their occurrences range all over the sciences and the arts. A discussion would not fit in this book, but mentioning a few will serve the cause of full disclosure.

Scaling in the work of Jonathan Swift. Swift 1733, lines 337-340, tells us that

So, Nat'ralists observe, a Flea
Hath smaller Fleas that on him prey,
And these have smaller Fleas to bit 'em
And so proceed ad infinitum.

It is reported that Swift was commenting on the literary society of his time. The following variant form was phrased a hundred years later by Augustus deMorgan:

Great fleas have little fleas upon their backs to bite 'em
And little fleas have less fleas, and so ad infinitum,
And the great fleas themsleves, in turn, have greater fleas to go on,
While these again have greater still, and greater still, and so on.

Implicit in both ditties is the idea that all fleas have the same shape. It is hard not to think of the Ptolemaic planetary system, with its cycles riding on cycles that themselves ride on cycles. Those ditties establish beyond question that the process that led mathematicians to define their *monster curves* did not originate around 1900, but had been widely understood for a far longer time.

Scaling in turbulence. Taking a path-breaking intellectual step, Richardson 1922, p. 66, adapted Swift as follows

Big whorls have little shorls,
Which feed on their velocity
And little whorls have lesser whorls,
And so on to viscosity
(in the molecular sense),

The viscosity "inner cut-off" (the last two lines) has a counterpart in finance, due to the impossibility of trading in strictly continuous time.

The next step after Richardson was taken in Kolmogorov 1941. In a class only with Lévy, Andrei Nikolaievich Kolmogorov (1903-1987) was the greatest probabilist of this century. I barely knew him personally, but greatly admired his extraordinary range of achievement. At the mathematical end of his range of interest in probability theory, Kolmogorov 1933 seemed to me too close for comfort to the work of the ultimate decorator who rearranges existing material. But Kolmogorov's papers on turbulence were filled with novelty and daring.

Kolmogorov 1941 was beginning to be known when I was a student at Caltech (1957-9), but everyone dismissed the so-called "K41" spectrum in $k^{-5/3}$. However, as soon as my main papers in finance were completed, I moved on to the study of noise (Berger & M 1963{N5}). After Robert W. Stewart obtained an experimental vindication of "K41," listening to him made me see strong similarities between Berger & M 1963 and some near-simultaneous papers on turbulence, including Kolmogorov 1962. Errors I found in that great but flawed work spurred me to multifractals (M 1973j{N14}, M 1974f{N15}, and M 1974c{N16}, hence profoundly affected the course of my scientific life. Those events made me realize that scaling, as a fruitful principle in science, has many sources, but I have no recollection of thinking of turbulence when scaling came to my mind in the contexts of income distribution, then finance.

Elliott Waves. This section devoted to miscellanea is as good a place as any to mention Ralph N. Elliott (1871-1948). A former peripatetic accountant and expert on cafeteria management, he studied Fibonacci, the Secrets of the Great Pyramid and the prophecies of Melchi-Zedik, and in 1938 announced a great "discovery," a "Wave Principle" that "really forecasts." A claim that he was a precursor of the use of fractals in finance prompted me to scan Elliott 1994. It is true that some of Elliott's diagrams are qualitatively reminisent of certain self-affine generators of the kind studied in Section 4 of Chapter E6. That is, they embody the wisdom present in Swift's qualitative metaphor quoted earlier in this section, but nothing more. Elliott's work fails the requirements of objectivity and repeatability: in his own words, "considerable experience is required to interpret [it] correctly" and "no interpretation [is] valid unless made by [him or his direct licencees]."

Scaling in biology: allometry. This topic is touched upon in Chapter 17 of M1982F{*FGN*}. Galileo knew that, moving from small to big animals, weight and leg cross-sectional must be roughly proportional, hence by diameter must scale as (overall length)$^{3/2}$. In words, big animals are expected to have thicker legs. Other scaling rules, concerning the branching of trees and rivers, are already mentioned in the *Notebooks* of Leonardo da Vinci. Since his *Notes* were largely records of his wide readings, it may be that those rules were already known to Middle Age engineers. Last point: observed allometric exponents need not be rational numbers.

Scaling in geology. Ohmori's law is an empirical relation concerning earthquakes; it is almost as old as Pareto's law, and used to be even more deeply neglected or spurned, but no longer. The Gutenberg-Richter law, also concerned with earthquakes, was well-known but written in a form that disguised the fact that it expresses scaling. Both are now recognized as being scaling relations, and earth sciences are now rife with fractals.

Comment, a puzzle and a challenge. In most contexts, the idea of scaling failed to be acknowledged (like in geology), remained peripheral (like allometry), or was acknowledged but remained circumscribed (like in the study of turbulence.) Scaling has also many sources in decoration. My own odd combination of the two oldest "academic" sources is described in Sections 1 and 2; it too remained circumscribed as long as it was limited to finance and was wholly accepted by few investigators.

In due time, however, fractal geometry brought together existing flavors of scaling and discovered new examples. The combination became recognized as a new "structure" of pattern, and its fortunes soared. Our century has seen many other "structuralist" syntheses, among them the mathematical school of "Bourbaki."

Was the eye the main unifying factor in the synthesis based on scaling and fractals, or just one of many factors? This is a question for the historian of science.

PART II: MATHEMATICAL PRESENTATIONS

This part, written specially for this book, incorporates the substance of reports and memoranda written over the years. Deliberately, the chapters do not follow each other in strict logical order, and their contents overlap; therefore, they can, to a large extent, be read independently of each other. The topics of Chapters E5 and E6 are largely new but concern themes that long influenced my work. The topic of Chapter E9 is important in practical statistics. The topics of Chapters E7 and E8 have long traditions plagued with casual, questionable, or erroneous writings.

&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&

E5

States of randomness from mild to wild, and concentration from the short to the long run

✦ **Abstract.** An innovative useful metaphor is put forward in this chapter, and described in several increasingly technical stages. Section 1 is informal, but Sections 4 and 5 are specialized beyond the concerns of most readers; in fact, the mathematical results they use are new.

At the core is a careful examination of three well-known distributions: the Gaussian, the lognormal and the scaling with infinite variance ($\alpha < 2$). They differ deeply from one another from the viewpoint of the addition of independent addends in small or large numbers, and this chapter proposes to view them as "prototypes," respectively, of three distinct "states of randomness:" *mild, slow* and *wild*. Slow randomness is a complex inter-

mediate state between two states of greater simplicity. It too splits more finely, and there are probability distributions beyond the wild.

Given N addends, *portioning* concerns the relative contribution of the addends U_n to their sum $\Sigma_1^N U_n$. Mildness and wildness are defined by criteria that distinguish between *even* portioning, meaning that the addends are roughly equal, ex-post, and *concentrated* portioning, meaning that one or a "few" of the addends predominate, ex-post. This issue is especially important in the case of *dependent* random variables (Chapter E6), but this chapter makes a start by tackling the simplest circumstances: it deals with independent and identically distributed addends.

Classical mathematical arguments concerning the long-run ($N \rightarrow \infty$) will suffice to distinguish between the "wild" state of randomness and the remaining states, jointly called "preGaussian."

Novel mathematical arguments will be needed to tackle the short-run ($N = 2$ or "a few"). The resulting criterion will be used to distinguish between a "mild" or "tail-mixing" state of randomness, and the remaining states, jointly called "long-tailed" or "tail-preserving." This discussion of long-tailedness may be of interest even to readers reluctant to follow me in describing the levels of randomness as "states."

In short-run partition, short-run concentration will be defined in two ways. The criterion needed for "concentration in mode" will involve the convexity of $\log p(u)$, where $p(u)$ is the probability density of the addends. The concept of "concentration in probability" is more meaningful but more delicate, and will involve a limit theorem of a new kind. Long-tailed distributions will be defined by the very important "tail-preservation criterion" under addition; it is written in shorthand as $P_N \sim NP$.

Randomness that is "preGaussian" but "tail-preserving" will be called "slow." Its study depends heavily on middle-run arguments ($N =$ "many") that involve delicate transients. ✦

THE NOTION OF CONCENTRATION being central to the study of firm sizes and price changes, this chapter is of direct relevance to economics. It shows that the economics concepts of short, middle and long-run have unsuspected parallels in probability theory: they suggest a distinction between different "states of randomness" that should prove useful in many fields of science, and also involves new mathematical results that may have enough intrinsic interest to be worth developing.

Section 1 is an informal introduction, close in style to Part I of this book. The middle part of the chapter is more technical, yet should interest many readers. Section 2 is devoted to long-run portioning and Section 3 to short-run portioning. The more specialized Section 4 proposes finer states of randomness. Section 5 is even more mathematical: it includes a proof and tackles some problems raised by the moments, and refers to "the moment problem" of classical mathematical analysis. The economic implications of short-run and long-run concentration are explored throughout the book, and serious flaws of the lognormal, in Chapter E9.

Terminology and notation. Once again, a convention is often used in this book. When there is no loss of intelligibility and the context allows, words like "Gaussian," "lognormal," "Bernoulli," "Poisson," and "scaling" will be used as common names, to avoid endless and tiresome repetition of the terms "random variable," "probability distribution," "probability density," or "density." In addition, the tail probabilities and densities will be denoted, respectively, by $P(u) = Pr\{U > u\}$ and its derivative $-P'(u)$, and $P_N(u) = Pr\{\Sigma_{n=1}^{N} U_n, > u\}$ and its derivative $-P'_N(u)$.

1. BACKGROUND AND INFORMAL PRESENTATION

The Gaussian distribution is often called "normal," because of the widespread opinion that it sets a universally applicable "norm." In the case of the phenomena studied throughout my life and described in this book, this opinion is unwarranted. In their case, randomness is highly non-Gaussian, but it is no longer possible to describe it as "pathological," "improper," "anomalous," or "abnormal." Therefore, any occurrence of *normal* in this book, as synonym of *Gaussian,* is the result of oversight, and I try not to think about the second and third syllables of *lognormal.*

1.1. The ageless competition between scaling and lognormal fit, and a motivation for introducing the notion of "states of randomness"

The innovation this chapter puts forward has many roots. One responds to a situation that plagues statistics and is a common reason for its unpopularity and ineffectiveness. All too often, reliable and competent statisticians split into camps that approach the same body of practically relevant data, and sharply disagree in their analysis. An example that concerns random variables is the very old disagreement about the distribution of income. Pareto claimed that it is scaling, and Gibrat, that it is lognormal (see Section 3 of Chapter E4, Chapter E9, and other chapters of this book). Current replays of those disagreements bring in random processes and they concern the records of price changes (Chapter E1 and Parts IV and V of this book.)

Could it be that both camps attempt to prove more than their data allow? Instead of seeking immediately to specify a distribution or a process by an analytic formula with its panoply of parameters, one should perhaps first sort out the possible outcomes into a smaller number of distinct discrete categories. The basic thought behind this classification is that, while the notion of randomness is unified from the viewpoint of mathematical axiomatics, it is of great diversity from the viewpoint of scientific modeling and related statistical tools.

Following this line of thinking, fractals led (first in finance and later in many other fields) to rather bold conclusions. To implement them, it is useful to inject a familiar metaphor and the terminology that comes with it. While a unique theory of physical interactions applies to every form of matter, the detailed consequences of those unique general laws differ sharply, for example, according to temperature and to whether the interactions are short-range or long-range. This is why physics has to distinguish between several *states of matter*, whose traditional number is three.

I propose in this chapter to argue that a similar distinction should be useful in probability theory. In the not-too-distant past, every book of statistics, as well as nearly every scientist engaged in statistical modeling in economics or elsewhere, used to deal with a special form of randomness, which will be characterized as *mild*. It will also be argued that entirely different states of randomness must be distinguished and faced. There is *wild* randomness exemplified by distributions with infinite variance. There is also an intermediate possibility exemplified by the lognormal: it is *slow randomness* – a term deliberately selected to imply what it says.

When faced with a new phenomenon or fresh dataset, the first task is to identify its state of randomness.

The implication is that, instead of ranging continuously, random variables are usefully sorted out in discrete categories exemplified by the Gaussian, the lognormal and the scaling with $\alpha < 2$. When the random variables U are defined by $\Pr(U > u) = P(u)$, the state of randomness differs sharply according to how fast the generalized inverse function P^{-1} decreases as its argument tends to 0, that is, according to how fast the moments U^q increase as $q \rightarrow \infty$. (To define P^{-1} when $P(u)$ is discontinuous, one fills each discontinuity by a vertical interval before the coordinate axes are exchanged.)

The words *mild, slow,* and *wild* were chosen to be short and without competing technical connotations (discounting that everyday usage tends to view all randomness as wild). The word "state" is also carefully chosen. Its existing technical connotations denote gases, solids and liquids; they are strong, but do not compete with the new usage; even some of its ambiguities are helpful, as I propose to argue now.

To begin with mildness, it is characterized by an *absence of structure* and in the case of random processes by a local level of statistical dependence. That is, diverse parts can be modified without much affecting the whole. Remarkably, the same properties also characterize a *gas*. Their importance will be seen in Section 2 of Chapter E8, when discussing the legitimacy of random-walk models of scaling.

Wildness, to the contrary, will be shown throughout this book to be characterized by the opposite qualities: *presence of structure* and long dependence. Remarkably, the same properties characterize a *solid*.

Among the long-recognized states of matter, the third and least-well understood and explained is *liquid*. Characteristic of both physical liquids and slow randomness is a surprising degree of uncertainty in the definition and many technical imperfections. Consider a glass: it behaves from many viewpoints as a solid, but physicists know that in "reality" it is a very viscous liquid. This unresolved problem of physical characterization has a surprising probabilistic counterpart in the distribution of personal income, as seen in several chapters of this book.

Nobody would suggest that income distribution is soft and akin to a gas: it is clearly hard. What remains to be established is whether the better metaphor is a "real solid,"or a "very viscous liquid." Pareto's law presses the claim that income distribution is scaling, therefore like a solid. Gibrat's writings press the counter-claim that it is lognormal, therefore like a very viscous liquid. Chapter E9 will set up a case against the lognormal, and argue that the above disagreement may be of a kind that cannot be settled by inventing better statistical methods.

1.2 The fallacy of transformations that involve "grading on the curve"

Before describing the criteria that distinguish the different states of randomness, it is necessary to dispose of a view that amounts to considering all forms of randomness as effectively equivalent. Indeed, scientists faced with clearly non-Gaussian data are often advised by statisticians to move on to a transformed scale in which everything nicely falls on the Gaussian "bell curve." In schools, the procedure is called "grading on the curve."

When pushed to its logical extreme, the underlying procedure leads to "grading by percentages." This transforms any U into a uniform random variable on [0, 1] defined by $\Pr\{I < x\} = x$. Indeed, a random variable U defined by $\Pr\{U > u\} = P(u)$ is simply the non-decreasing transform of I defined as $P^{-1}(I)$, where P^{-1} is defined in Section 1.1.

Unfortunately, transformation ceases to look attractive as soon as one faces reality. A first complication, beyond the scope of this chapter, concerns sequences of *dependent* variables: when each variable is made uniform, the rules of dependence need not transform into anything simple.

A second complication is this: money is additive, but a transform such as log (money) is not; firm sizes add up to the size of an industry, but a transform like log (firm size) is not additive. In pedantic terms, concrete economics deals with *numerical* variables that can be added, not with *ordinal* variables that can only be ordered.

A third and most important complication is that real-world distributions are not known exactly, but approximately. That is, a random variable does not come up alone, but as part of a natural "neighborhood" that also contains other variables viewed as "nearly identical" to it.

Of enormous significance are the neighborhoods that are automatically implied in every limit theorem of probability theory. For example, to say that a random variable tends to a limit, is to say that it eventually enters a suitably defined neighborhood of the limit. In the usual central limit theorem, the limit is Gaussian, and the neighborhood is defined solely on the basis of the central bell, disregarding the tails. Cramer's large deviations theory splits the neighborhood of the Gaussian in a finer way that does not concern the bell, but the tails. The concrete usefulness of a limit theorem depends initially on whether or not this neighborhood it implies is a "natural" one from the viewpoint of a specific concrete situation.

Now we can describe the major failing of the transformation of U into I : it fails to transform the natural neighborhood of U into the natural neighborhood of I.

Once again, the example of greatest relevance to this book is the notion that for some data the best methods of statistics conclude that $\log X$ is practically Gaussian. This means that the observed deviations from Gaussianity only concern the largest values of X that contribute a few percent of the whole. Faith in the significance of the Gaussian fitted to $\log X$ leads to the recommendation that these exceptional values be neglected or treated as "outliers." The trouble is that in many cases the most interesting data are those in the tail! It follows that differences between alternative notions of neighborhood are *not* matters of mathematical nit-picking.

In the light of these three "complications," the suggestion that any variable can simply be made uniform or Gaussian by transformation is ill-inspired and must be disregarded.

1.3 Portioning on the short or the long-run, and three states of randomness

The proceeding motivation gave one example of each state of randomness. It is now time to define those states. Before we do so, recall that gases, liquids and solids are distinguished through two criteria: flowing versus non-flowing, and having a fixed or a variable volume. Two criteria might define four possibilities, but "non-flowing" is incompatible with "variable volume." Adding in uncanny fashion to the value of our physical metaphor, our three states of randomness are also defined by two mathematical criteria, both deeply rooted in economic thinking. Given a sum of N independent and identically distributed random variables, those criteria hinge on two notions.

Portioning concerns the relative contribution of the addends U_n to the sum $\Sigma_1^N U_n$.

The *concentration ratio* of the largest addend to the sum. Loosely speaking, *concentration* is the situation that prevails when this ratio is high. This idea will, later in this chapter, be implemented in at least two distinct ways. The opposite situation, prevailing when no addend predominates, will be called *evenness.*

The issue must be raised separately on the short- and the long-run, and it will be seen that concentration in the long-run implies concentration in the short-run, but not the other way around. Hence, the contrast between concentration and evenness leads to three principal categories.

- *Mild randomness* corresponds to short- and long-run evenness.
- *Slow randomness* corresponds to short-run concentration and long-run evenness.
- *Wild randomness* corresponds to short- and long-run concentration.

In mild and wild randomness, the short- and long-run behavior are *concordant*; in slow randomness, they are *discordant.*

Here is another bit of natural and useful terminology.

• Taken together, the two non-wild states will be said to define *preGaussian* randomness, the counterpart of *flowing* for the states of matter. An alternative term is "tail-mixing."

• Taken together, the two non-mild states will be said to define *long-tailed* randomness, the counterpart of *fixed-volume* for the states of matter. An alternative term is *tail-preserving.*

Let us now dig deeper, in terms of finance and economics.

Long-run portioning and the distinction between wild and preGaussian randomness. This distinction concerns asymptotics and the long-run. Examples are the relative size of the largest firm in a large industry, the largest city in a large country, or the largest daily price increase over a significantly long period of time. PreGaussian randomness yields approximate equality in the limit, as expressed by the fact that even the largest addend is negligible in relative value. By contrast, wild randomness yields undiminishing concentration, expressed by the property that the largest relative sizes remains non-negligible even in very large aggregates.

The mathematical detail of long-run portioning is delicate and found in standard references, therefore it must and can be summarized. This will be done in Section 2. Additional information is found in Chapter E7.

Short-run portioning, and the distinction between mild and long-tailed randomness. The cleanest contrast to the long-run is the very short-run represented by two items. Given two independent and identically distributed random variables, U' and U'', and knowing the value taken by the sum $U = U' + U''$, what do we know of the distributions of U' and U''? We shall describe U as being "short-run portioned between U' and U'', " and wonder whether those parts are more or less equal, or wildly dissimilar.

As a prelude, consider two homely examples. Suppose you find out that the annual incomes of two strangers on the street add to $2,000,000. It is natural and legitimate to infer that the portioning is concentrated, that is, there is a high probability that the bulk belongs to one or the other stranger. The $2,000,000 total restricts the other person's income to be less than $2,000,000, which says close to nothing. The possibility of each unrelated stranger having an income of about $1,000,000 strikes everyone as extraordinarily unlikely, though perhaps less unlikely that if the total were not known to be $2,000,000.

To the contrary, the total energy of two sub-systems of a gas reservoir is evenly portioned: each molecule has one-half of the energy of the two together, plus a tiny fluctuation. A situation in which most of the energy concentrates in one subsystem can safely be neglected.

Rigorous mathematical argument supports the "intuition" that even portioning is very unlikely in one case, and very likely in the other. Indeed, the above two stories exemplify opposed rules of short-run portioning. Even short-run portioning will define mild randomness, and concentrated short-run portioning will define long-tailed randomness.

Unfortunately, the details of this distinction are not simple. In addition, short-run portioning is not a standard mathematical topic. The question was first raised and discussed heuristically in Section 2.5 of M 1960i{E10} and again in Section V.A of M 1963b{E14}, but, to my knowledge, nowhere else. The first full mathematical treatment, which is new, will be presented in Sections 3 and 5.

The middle-run. Short- and long-run considerations are familiar in economics. They are essential, but, to quote John Maynard Keynes, "in the long-run we shall all be dead." Economic long-run matters only when it approximates the middle-run reasonably, or at least provides a convenient basis for corrective terms leading to a good middle-run description.

Probability theory also favors small and large samples. Samples of a few items are handled by explicit formulas often involving combinatorics. Large samples are handled by limit theorems. Exact distributions for middle-size samples tend to involve complicated and unattractive series or other formulas that can only be handled numerically on the computer. In a way, this chapter proposes to bracket the interesting but untractable probabilistic middle-run between an already known and tractable long-run and a very different tractable short-run.

Digression concerning physics. The model for all sciences, physics, was able for a long time to limit itself to two-body or many-body problems, that is, small or large aggregates. Intermediate ("mesoscopic") phenomena were perceived as hard and only recently did they impose themselves and physics became strong and bold enough to tackle a few of them. In a few examples (some of which occur in my recent work), it is useful and possible to distinguish and describe a distinct *pre-asymptotic* regime of large but finite assemblies.

2. WILD VERSUS PRE-GAUSSIAN RANDOMNESS: CLASSICAL LIMIT THEOREMS DEFINE CONCENTRATION IN THE LONG RUN

This Section is somewhat informal, the technical aspects being available in the literature, and/or taken up in Chapter E7.

2.1 Introduction to long run portioning

Probability theory solved long ago the problems of the typical size of the largest of N addends, relative to their sum, and the problem of distribution around the typical size. The most basic distinction is based is the boundedness of the second moment. Of the many possibilities that are open, the following are the most important.

At one extreme, the addends are bounded, and the concentration is $\sim 1/N$. As $N \rightarrow \infty$, concentration converges to 0. This last conclusion also holds when $EU^2 < \infty$. Since the inequality $EU^2 < \infty$ is generally taken for granted, most scientists view the notion of concentration for large samples as completely solved by probability theory. In particular, one of the justifications of the role of the Gaussian in science is closely patterned after its role in the "theory of errors," as practiced around 1800 by Legendre and Gauss. It is taken for granted that each chance event is the observable outcome of a large number of separate additive contributions. It is also taken for granted that each contribution, even the largest, is negligible compared to the sum, both ex-ante (in terms of distributions) and ex-post (in terms of sample values). The economists' technical term for this premise is "absence of concentration in the long-run," and here it will be called "evenness in the long-run." This premise is widely believed to hold for *all* independent and identically distributed addends. In other words, identity of ex-ante distributions of the parts is believed to lead to evenness of ex-post sample values.

This common wisdom claims to solve one of the problems raised in this chapter. Observed occurrences of concentration are viewed as transients, or possibly the result of strong statistical dependence between addends.

In fact, and this is the main theme of this chapter and of the whole book, the common wisdom is simply *mathematically incorrect*. As sketched in Section 1.3 and discussed in this section, portioning in the long-run can take two distinct forms: *even*, with concentration converging to 0 as $N \rightarrow \infty$ and *concentrated*, with the largest addend remaining of the order of magnitude of the sum.

This distinction largely relies on standard results of probability theory. This book discusses its impact in economics in many places, including in the reprints on income distribution and price variation, and Chapter E13 concerned with firm sizes.

2.2 Alternative criteria of preGaussian behavior

The prototype of mild randomness is provided by the "thermal noise" that marks the difference between the statistical predictions of the theory of gases and the non-statistical predictions of the older thermodynamics. Thermal noise consists in small fluctuations around an equilibrium value. For the "astronomically" large systems that are the (successful) physical analogs of the economic long-run, those fluctuations average out into relative insignificance. If such a system is divided into many equal parts, the energy of the part with the highest energy is negligible compared to the energy of the whole.

Informal statement. More generally, the form of randomness this book calls preGaussian is defined by limit asymptotic properties that are best stated as follows.

• *The fluctuation is averaging,* or *ergodic.* The law of large numbers (LLN) shows that sample averages converge asymptotically to population expectations.

• *The fluctuation is Gaussian.* The central limit theorem (CLT) shows that the fluctuations are asymptotically Gaussian.

• *The fluctuation is Fickian.* The central limit theorem also shows that the fluctuations are proportional to $\sqrt{N}$, when N being the number of addends. For random processes, an alternative, but equivalent statement (less well-known but essential) is that events sufficiently distant in time are asymptotically independent.

More formal questions and answering statements.

Question: Take the sequential sum $\Sigma_{n=1}^{N} U_n$ for a sequence of independent and identically distributed random variables U_n, $1 \le n < \infty$. Is it possible to choose the sequences A_N and B_N and define the notion of "converges to", so that $A_N\{\Sigma_{n=1}^{N} U_n - B_N\}$ converges to a limit?

An answer that defines preGaussian behavior: Under certain conditions described in numerable textbooks, a choice of A_N and B_N is possible in two distinct ways:

The choice of $A_N = 1/N$ and $B_N = 0$ yields the law of large numbers, in which the limit is EU, that is, non-random.

The choice of the Fickian factor $A_N \sim 1/\sqrt{N}$ and $B_N = NEU$ yields the central limit theorem, in which the limit is Gaussian, that is, random. It follows that $\Sigma_{n=1}^{N} U_n$ is asymptotically of the order of $\sqrt{N}$.

The notions of attraction and universality. In many contexts, physicists have no confidence in the details of their models, therefore distrust the models' consequences. An important exception is where the same consequences are shared by a "class of universality," that also includes alternatives that differ (not always slightly) from the original model. Although the word universality is rarely used by probabilists, the basic idea is very familiar to them. For example, few scientists worry about the precise applicability of a Gaussian process, because the Gaussian's domain of attraction is very broad, and slight changes in the assumptions provoke slight deviations in the consequences drawn from the model. The domain of universality of attraction to the Gaussian includes all U satisfying $EU^2 < \infty$, but also some cases when EU^2 diverges slowly enough.

2.3 Exceptions to preGaussian behavior

The preGaussian domain of universality is broad, but *bounded*. The properties of being averaging, Gaussian, or Fickian may fail. The failure of any of these properties defines the *wild* state of randomness.

Failure occurs when the population variance, or even the expectation, is infinite, when the dependence in a random process is not "short-range" or local (contrary to the locality of the Markov process) but "long-range" or global, or when total probability must be taken as infinite. Most notably, the scaling variables with $\alpha < 2$ satisfy $EU^2 < \infty$, and are *not* preGaussian.

Wild randomness and practical statistics. The preface quoted the editor's comments on M 1963b{E14} found in Cootner 1964. They, and countless other quotes by practically-minded investigators, some of them scattered throughout this book, show that non-averaging, non-Gaussian, and/or non-Fickian fluctuations were long resisted and viewed as "improper" or even "pathological." But I realized that many aspects of nature are ruled by this so-called "pathology." Those aspects are not "mental illnesses" that should or could be "healed." To the contrary, they offer science a valuable new instrument. In addition, a few specific tools available in "pure mathematics" were almost ready to handle the new needs. The new developments in science that revealed the need for those tools implied that science was moving on to a *qualitatively different* stage of indeterminism.

The editor's comments in Cootner 1964 also noted that, if it is confirmed that economic randomness is wild, some tools of statistics will be endangered. Indeed, tools developed with mild randomness in mind become questionable in the case of slow randomness. As a rule with many exceptions, they are not even close to being valid for wildly random phenomena, such as those covered by my models of price variation.

Sketch of the domains of universality of attraction to nonGaussian limits. To be outside of the Gaussian domain of attraction or universality is a great complication. In particular, each value $\alpha < 2$ defines its own domain of universality. In addition, in sharp contrast to the width of the domain of the Gaussian, each of those domains is extremely narrow and reduces to the variables for which $Pr\{U > u\} \sim u^{-\alpha}L(u)$, where $L(u)$ is logarithmic or at most slowly varying in the sense that for all h, $\lim_{u \to \infty} L(hu)/L(u) = 1$. The term $L(u)$ is largely a nuisance, and we shall not invoke it unless necessary.

If U_n is in the domain of universality of $\alpha < 2$, the limit is a random variable called *L-stable*, which is widely discussed and used in the papers reprinted in this book.

In the absence of slowly varying term $L(u)$, the choice of A_N is $A_N = N^{1/\alpha}$ for all α, therefore U_N^{Σ} is asymptotically of the order of $N^{1/\alpha}$. The choice of B_N is NEU when $1 < \alpha < 2$, and 0 when $0 < \alpha < 1$.

2.4 Comments on the middle-run and slow randomness

Adding new evidence that the world is not a simple place and science is more difficult than mathematics, the limit theorems of probability do not really matter, unless they also help describe the middle-run. Unfortunately, the middle-run is complicated, hence the existence of a third "middle" state did not fully impress itself on my work until a recent careful look at its foremost example, the lognormal (Chapter E9). Since industrial concentration is incontrovertible (see Chapter E13), the very fact that the lognormal is continually proposed to model industrial concentration means that it cannot really be counted as mild. What is it?

On the long-run, it is indeed averaging, Gaussian and Fickian, therefore, preGaussian. In the middle-run, however, its "nice" asymptotic properties are irrelevant and give no hint of the fact that the strict lognormal yields a "very erratic" sample averages. The statistician who is invited to examine those averages, and not the distribution itself, should conclude that those averages behave "as if" the addends were wild. In other words, a more correct interpolate of the middle-run behavior is obtained if one does not start with the lognormal, but a wild approximation to the lognormal. For actual data that are neither exactly lognormal nor exactly wild, my long-term goal has been to develop viewpoints and techniques that illuminate the middle-run and can be used as the starting point for improvements.

3. MILD VERSUS LONG-TAILED RANDOMNESS: CONCENTRATION IN THE SHORT RUN, CONVEXITY OF $\log p(u)$ AND THE TAIL PRESERVATION RELATION $P_N(u) \sim NP(u)$

Section 2 divides all forms of randomness into wild – defined by concentrated long-run portioning, and preGaussian – defined by even long-run portioning. Our next goal is to subdivide this second category into two categories to be called mild and long-tailed. This will be done in stages.

• A first criterion will be based on *concentration in mode*; it is very simple, but has many flaws.

• A more intrinsic second criterion of wider applicability will be based on asymptotic *concentration in probability.* It will lead to the "tail-preservation" relation $P_N(u) \sim NP(u)$.

The term *slow,* is justified by arguments that cast doubt on the acceptability of slow random models in scientific work. It is best to phrase those arguments in the specific context of the lognormal distribution. This will be done in Chapter E9.

The tail-preservation relation is not, in itself, new to probability theory, since it occurs in the classical "extreme value problem." Indeed, let the random variables $U_j (1 \leq j \leq N)$ be independent and identically distributed, with the tail probability $P(u)$, and let $\tilde{P}_N(u)$ be the tail probability of $\tilde{U}_N = \max (U_j)$. It is well-known that $1 - \tilde{P}_N(u) = \{1 - P(u)\}^N$. In the tail where $P(u) \ll 1$ and $\tilde{P}_N(u) \ll 1$, we find in *all cases* that $\tilde{P}_N \sim NP$.

However, the material that follows *does not concern* the extreme value problem, it merely injects some considerations relative to extreme values into the classical study of *sums.* A striking consequence is that, in this new context, the tail preservation relation for sums holds for *some,* but *not all,* probability distributions. By ceasing to hold universally, it ceases to be a trivial property; instead, it takes up a central role in a fundamental distinction between one state of randomness (mild) and the other states taken together (long-tailed.)

3.1 The doubling convolution and the short-run portioning ratio

As agreed, we denote the common probability density of U' and U'' by $p(u)$. The probability density of $U = U' + U''$, denoted $p_2(u)$, is given by the doubling convolution $p_2(x) = \int p(u)p(x-u)du$. When u is known, the conditional probability density of u' is given by the following expression, to be called "portioning ratio"

$$\frac{p(u')p(u-u')}{p_2(u)}.$$

The denominator is a constant and it remains to study the numerator.

Min (U', U'') and max (U', U'') can be compared in many different ways. The conditional expectation of U', knowing $U = u$, is of no help, since it is always $u/2$, and the conditional expectation of $\min(u', u'')$ is not given by any manageable expansion.

To the contrary, it is often easy to study the location of the most probable values of $\min(u', u'')$ and $\max(u', u'')$, which statisticians call "modes." Those locations lead to a criterion based on the convexity of $\log p(u)$, which will serve to define "concentration versus evenness in mode."

The mode is of little use in probability, but in this instance turns out to be surprisingly close to being satisfactory. Indeed, a more searching stage of this study shows that, under suitable additional assumptions, the integral $\int p(u')p(u-u')du'$ is dominated by values the conditional density $p(u')p(u-u')$ takes in intervals near the modes, while the remaining intervals have a negligible contribution. The underlying mathematical theorem concerns concentration "in probability," but in some cases also holds in the "almost sure" sense.

The proof of this basic theorem also yields the fundamental "tail-preservation criterion" written in shorthand as $P_N \sim NP$. In due time, the assumption of the basic theorem are bound to be improved. Therefore, I propose to define "long-tailedness" as meaning "tail-preserving."

3.2 Sufficient criterion of evenness or concentration "in mode": the graph of $\log p(u)$ is cap- or cup-convex for sufficiently large values of u

In many important cases, the maximum of the product $p(u')p(u-u')$ occurs either near $u' = u/2$, or near $u' = 0$ and $u' = u$. Take logarithms and write

$$\Delta(u) = 2 \log p\left(\frac{u}{2}\right) - [\log p(0) + \log p(u)].$$

When the convexity of $\log p(u)$ is uniform for all u, the sign of $\Delta(u)$ is independent of u.

• The case when the graph of $\log p(u)$ is cap-convex, like the typographical sign $\cap$. In that case, the portioning ratio is *maximum* for $u' = u/2$, and portioning is even in terms of the mode.

• The boundary case when the graph of $\log p(u)$ is straight. In that case, the addends are exponential, and the portioning ratio is a constant.

• The case when the graph of $\log p(u)$ is cup-convex, like the typographical sign $\cup$. In that case, the portioning ratio is *minimum* for $u' = u/2$ and portioning is concentrated in terms of the mode.

Distributions with uniform convexity of $\log p(u)$ suffice to show that the distinction between mild and long-tailed randomness cannot be identified with the distinction between even and concentrated short-run portioning.

3.3 Simple examples of uniform convexity

Every Poisson always yields even *short-run portioning in mode.* When $p(u) = e^{-\gamma}\gamma^u/u!$, the convexity of $\log p(u)$ is that of $\log u!$, which is cap-convex all $u > 0$. The portioning ratio is

$$\frac{p(u')p(u-u')}{p_2(u)} = \frac{e^{-\gamma}e^{-\gamma}}{e^{-2\gamma}} \frac{\gamma^{u'}\gamma^{u-u'}}{(2\gamma)^u} \frac{u!}{u'!(u-u')!} .$$

The non-constant third term is a binomial coefficient that peaks at $u = u/2$ if u is even, and at $(u \pm 1)/2$ if u is odd. At those points, the portioning ratio $p(u')p(u-u')/p(u)$ has a maximum. Even portioning was to be expected: the Poisson distribution rules the number of points of a Poisson process that fall in an interval of given length.

Every Gaussian yields even *short-run portioning in mode.* Here, $\log p(u)$ is essentially $-u^2$, which is cap-convex uniformly for all u. The portioning ratio is

$$\frac{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(u-u')^2}{2}\right\}}{\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{u^2}{2}\right\}} = \frac{1}{\sqrt{\pi}}\exp\left\{-\left(u' - \frac{u}{2}\right)^2\right\}.$$

Thus, a Gaussian is evenly partitioned with a Gaussian "error-term" for which variance is 1/2, that is, does not depend on u.

Every scaling yields concentrated *short-run portioning in mode for all* α. When $p(u)=\alpha u^{-\alpha-1}$ for $u>1$, $\log p(u)$ is cup-convex uniformly for all $u>1$. The same is true of $\log p(u)+\log p(x-u)$, and the portioning ratio is

$$\frac{1}{2}\frac{u'^{-\alpha-1}(u-u')^{-\alpha-1}}{u^{-\alpha-1}}.$$

It is cup-convex and largest for $u=1$ and $u=x-1$.

The family $p(u)=\exp(-u^w)$, *is split, the nature of short-run portioning being dependent on the sign of* $w-1$. The convexity of $\log p(u)$ is, again, uniform for all $u>0$, but here it depends on the sign of $w-1$. Portioning is even for $w>1$ and concentrated for $w<1$. The family $\exp(-u^w)$ is often praised in the literature for the ability of one analytic expression to account for very different behaviors, according to whether $w>1$ or $w<1$. This versatility can also be interpreted in a negative light, as a form of insensitivity to profound differences.

3.4 The lognormal and other examples of non-constant convexity of log $p(u)$; mixed rules of short-run portioning

For many usual distributions, the graph of $\log p(u)$ is cap-convex for all u. But a bell where the graph of $\log p(u)$ is cap-convex is often flanked by one or two tails where the graph is cup-convex. In those mixed cases, portioning depends on u : it is *even for* u *near the mode (i.e., where* $p(u)$ *is largest) and concentrated for large* u. Let us examine a few examples.

The Cauchy. Here,

$$p(u)=\frac{1}{\pi(1+u^2)}\text{ and }p_2(u)=\frac{1}{2\pi(1+u^2/4)}.$$

Here the convexity of log $p(u)$ changes for $u=\pm1$. Hence, portioning is in mode even for $|u|<2$, and concentrated for $|u|>2$.

The "Cournot" (positive L-stable density with $\alpha=1/2$). Here,

$$p(u)=\frac{1}{\sqrt{2}\,\pi}e^{-1/2u}u^{-3/2},\text{ and }p_2(u)=\frac{2}{\sqrt{2}\,\pi}e^{-4/2u}u^{-3/2}.$$

The lognormal. If $EU=1$, there is a single parameter $\sigma^2/2$, and

$$\log p(u) = - \log(\sigma\sqrt{2\pi}) - \log u - \frac{(\log u + \sigma^2/2)^2}{2\sigma^2}.$$

Here, the convexity of $\log p(u)$ changes when $\log u_0 = 1 - 3\sigma^2/2$. Hence, portioning is even when $u < 2u_0$, and concentrated when $u > 2u_0$. When σ^2 is large, so that the lognormal is very skew, the bell lies almost entirely to the left of $EU = 1$ and its total probability is small. Portioning is then most likely to be concentrated and the non-mild character of the lognormal is obvious. When σ^2 is small, so that the lognormal is near-Gaussian with a small tail added, the bell includes $EU = 1$ and its total probability is near 1. Portioning is then most likely to be even, and the lognormal may seem mild.

Note. This example raises an issue of wider applicability. Ostensibly, portioning in the case $N = 2$ is a short-run notion. But in the case of near-Gaussian lognormals, concentration only occurs in very large assemblies.

The log Bernoulli e^B. This is the exponential of a Bernoulli; it has a finite upper bound $\exp(\max B)$, therefore the limit arguments concerning $u \to \infty$ have no meaning for it. As the sum of two addends approaches 2 $\exp(\max B)$, the portioning between the addends necessarily becomes even.

3.5 The problematic gamma family $p(u) = u^{\gamma-1}e^{-u}/\Gamma(\gamma)$; portioning in mode is even for $\gamma > 1$ and concentrated for $\gamma < 1$

The concentration in mode based on the convexity of $\log p(u)$ proves unreasonable in the case of the gamma distribution. In that case, $U' + U''$ is a gamma of parameter 2γ, hence

$$\frac{p(u')p(u-u')}{p_2(u)} = \frac{\Gamma(2\gamma)}{[\Gamma(\gamma)]^2} \frac{u'^{\gamma-1}(u-u')^{\gamma-1}}{u^{2\gamma-1}}.$$

The exponential special case $\gamma = 1$ marks the boundary between two opposite rules of portioning in mode.

When $\gamma > 1$, portioning in mode is even and the maximum at $u/2$ becomes increasingly more accentuated as $\gamma \to \infty$. For integer values of γ, this result was to be expected, because the resulting gamma is the sum of γ independent exponential variables, therefore becomes increasingly close to Gaussian.

When $\gamma < 1$, to the contrary, portioning in mode is concentrated. However, this behavior is not *due* to the tail behavior of the gamma, rather to its behavior near $u = 0$. The tail is *shorter* in the cap-convex case $\gamma < 1$ than in the cup-convex case $\gamma > 1$.

In summary, the gamma shows the need of a criterion of "mildness" that goes beyond the convexity of log $p(u)$.

The multiplicative character of concentration in mode is the gamma case. For the gamma, $W = U'/U$ is independent of u, and has the beta density

$$\Gamma(2\gamma)[\Gamma(\gamma)]^{-2} w^{\gamma - 1}(1 - w)^{\gamma - 1}.$$

Therefore, the fluctuating term can be described as multiplicative. Now apply the same argument formally to the asymptotically scaling. The concentration ratio converges to $w^{-\alpha - 1}(1 - w)^{-\alpha - 1}$. This limit is non-integrable near $w > 0$ and $w = 1$, implying that for the scaling, $w \rightarrow 0$ or 1 as $u \rightarrow \infty$. The underlying reason is that in the scaling case, the distribution of $\min(u', u - u')$ is independent of u for large u, hence the fluctuating term is *not* multiplicative but *additive.*

3.6 Evenness and concentration "in probability," and the criterion $p_2(u) \sim 2p(u)$ of tail preservation under addition as defining long-tailedness

The study of concentration in mode has the virtue of extreme simplicity. The results are surprisingly adequate, but exceptions must be avoided without artificiality. The smallness of the number of exceptions is largely serendipitous, because the criterion based solely on the maxima of $p(u')p(u - u')$ is an extraordinarily crude one. The real question is more searching: is the relative value U'/u likely to lie in a suitably narrow neighborhood of the maximum or maxima?

Definition of short-run concentration in probability. This definition is geared to the case when there is concentration in mode, that is, $p(u')p(u - u')$ is maximum for u' near 0 and u' near u. In that case, given a value of $\tilde{u}$ that satisfies $\tilde{u} < u/2$ and may depend on u, one can split the doubling convolution in three parts, as follows

$$p_2(x) = \int_0^x p(u)p(x - u)du = \left\{ \int_0^{\tilde{x}} + \int_{\tilde{x}}^{x - \tilde{x}} + \int_{x - \tilde{x}}^{x} \right\} p(u)p(x - u)du$$
$$= I_L + I_0 + I_R.$$

I propose to describe $p(u)$ as *short-run concentrated in probability* if it is possible to select $\tilde{u}(u)$ so that the middle interval $(\tilde{u}, u-\tilde{u})$ has the following two properties as $u \rightarrow \infty$.

• The relative probability in the middle interval $I_0/p_2(u)$, *tends* to 0.

• The relative length of the middle interval $(u-2\tilde{u})u$ *does not tend* to 0.

This second requirement opens two sub-possibilities.

• When $p(u)$ is only moderately long-tailed, the relative length of the middle interval tends to 1. The density $p(u)p(\alpha-u)$ concentrated arbitrarily tightly around its mode. Concentration in probability is replaced by a stronger property: almost sure concentration.

• When $p(u)$ is extremely long-tailed, the relative length of the middle interval tends to a limit, or may have a lower bound >0 and an upper bound <1.

The "tail-preservation criterion." Section 5.1 will insure that short-run concentration in probability prevails when log $p(u)$ is smoothly varying, decreasing and cup-convex and its derivative $p'(u)/p(u)$ tends rapidly enough to 0 as $u \rightarrow \infty$.

In addition to concentration in probability, the same proof yields a very perspicuous criterion, namely $p_2(u) \sim 2p(u)$. In terms of the tail probabilities $P(u)$ and $P_2(u)$ of U and $U'+U''$, this criterion reads

$$P_2(u) \sim 2P(u).$$

More generally, writing $P(u)$ and $P_N(u)$ for the tail probabilities of U and of a sum of N variables with the same distribution, one obtains

$$P_N(u) = NP(u).$$

This criterion expresses that the tail behavior of U is preserved under finite addition. The notion of tail preservation, first introduced in M 1960i{E10}, recently turned out to be related to classical and seemingly unrelated considerations in classical "fine mathematical analysis," to be described in Section 5.

Rescaling of tails and a property of scaling distribution. When the tail is asymptotically scaling as in the case of the L-stable distributions, the tail conservation relation acquires a special meaning. It shows that

$$\text{"scale of } U\text{"} = \text{"scale of } U' \times 2^{1/\alpha}\text{."}$$

This result also holds when $p(u) = u^{-\alpha}L(u)$, where $L(u)$ is slowly varying, that is, it satisfies $L(hu)/L(u) \rightarrow 1$ for all $h > 0$) as $u \rightarrow \infty$.

Tail conservation holds for lognormals, but fails to have this special meaning. Lognormality is *not* preserved by addition.

3.7 Mild randomness and mixing behavior when log $p(u)$ is cap-convex

To appreciate the meaning of the criterion $P_N \sim NP$, let us examine cases where it does not hold.

The borderline exponential case. Here, $\Sigma_{n=1}^{N} U$ is a gamma variable; therefore as $u \rightarrow \infty$, P_N/P does not tend to the constant N, but increases like u^{N-1}.

The case of evenness in mode. When the convolution integrand $p(u)p(x-u)$ has a maximum at $u = \alpha/2$, the tail of $p_2(u)$ is little affected by the behavior of $p(u)$ in the tail. But it is greatly affected by its behavior part-way through the tail. The result is that P_N/P can increase very fast. In the Gaussian case with $EU = 0$, when $N \gg 1, P_N/P \sim 1/[\sqrt{N}\,P]$, which grows very fast as $u \rightarrow \infty$. Instead of tail preservation, one encounters an interesting "mixing" behavior whose intensity can be measured by the rate of growth of P_N/P.

3.8 Portioning and the tail-preservation relation $P_N \sim NP$, when N is a small integer above 2

In an equilateral triangle of height u, the distances from a point P to the three sides add up to u, therefore can represent u', u'' and u^0 in the portioning of the sum $u = u' + u'' + u^0$ into its contributing addends u', u' and u^0. When the U are exponential, the conditional distribution of P is uniform within this triangle. When U is mild, the conditional distribution concentrates near the center. When U is short-run concentrated in probability, the conditional distribution concentrates near the corners.

The same distinction holds for $N = 4, 5$ etc.... It is of help in gaining a better understanding of the problematic gamma family. The γ exponent of a sum of N gammas is $N\gamma$, which exceeds 1 as soon as $N > 1/\gamma$. Therefore, U can conceivably be called mild if $\Sigma_{n=1}^{N} U$ has a cap-convex log $p_N(u)$ for all N above some threshold. Starting with $\gamma = 2^k$, where k is a large integer, evenness decreases until $\gamma = 1$. There, a boundary is crossed and portioning becomes increasingly concentrated.

However, as $N \rightarrow \infty$, an altogether different classification takes over, as seen in Section 3.

Next, consider portioning of a sum of four addends $U_1 + U_2 + U_3 + U_4$ into two sums of two addends $U_1 + U_2$ and $U_3 + U_4$. Indeed, even when $\log p(u)$ is cup-convex for all u, one part of the graph of $\log p_2(u)$ is bell shaped. In the scaling case, the bell continues by a cup-convex tail. In the gamma case with $1/2 < \gamma < 1$, the tail is cap-convex.

A seeming paradox of immediate practical importance: when $P_N(u) \sim NP(u)$, the cup-convexity of $\log P_N(u)$ for large u is preserved for all N; this is true both when U is wild and when it is preGaussian. Taking the word "addends" as model, "limitands" is a self-explanatory term for "items" that are made to tend to a limit. The items may be sets, graphs of functions, or analytic expressions. Let $P(L_n)$ and $P(L)$ be properties of each limitand L_n and the limit $L = \lim_{n \to \infty}$, respectively. The "intuition" that $P(L) = \lim_{n \to \infty} P(L_n)$ is often wrong. It used to be that it only failed for artificial mathematical counter-examples, but no longer. Define L_N as the graph of the function $\log p_N(u)$ relative to the sum of N long-tailed random variables U and the property $P(L_N)$ as asserting that, for all N, the graph L_N is cup-convex for large u. Two possibilities are open: When U is wild, this convexity property is indeed preserved in the limit; However, when U is slow, this property fails in the limit, since the limit is the graph of $\log p(u)$ for the Gaussian.

4. A MORE REFINED TENTATIVE SUBDIVISION, YIELDING SEVEN STATES OF RANDOMNESS

The criteria stated in Section 1 and elaborated in Sections 2 and 3 leave open many conceptual and practical "details."

4.1 The boundary between mild and long-tailed and "borderline mild" randomness

Sections 3.3 to 3.6 and Section 4 imply that the exact separation between the mild and the long-tailed states of randomness is not unique, and depends upon the definition selected for the notion of concentration. Within the problematic gamma family (Section 3.6), the convexity of $\log p(u)$ defines a different boundary for each value of N. Granted this fuzziness, one may as well accept the existence of a transitional state between proper mildness and proper long-tailedness.

4.2 Extreme randomness

Wild randomness was characterized by the fact that the largest of many addends is of the same order of magnitude as their sum. But it is possible for concentration to be even more extreme. In the example of the tail probability $P(u) = 1/\log u$, concentration converges to 1 as $N \rightarrow \infty$; asymptotically, it becomes absolute. The same is true whenever $P(u)$ is a "slowly varying function," in the sense that, for all $h > 0, \lim_{u \rightarrow \infty} P(hu)/P(u) \rightarrow 1$. Those $P(u)$ define a state of randomness beyond the wild. I never encountered it in practice.

4.3 The contrast between localized and delocalized moments

Take a hard look at the formula $EU^q = \int_0^\infty u^q p(u) du$. For the scaling, the

integrand is maximum at the trivial values 0 or ∞. But in non-trivial cases, the integrand *may* have a sharp global maximum for some value $\tilde{u}_q$ defined by the equation

$$0 = \frac{d}{du}(q \log u + \log p(u)) = \frac{q}{u} - \left| \frac{d \log p(u)}{du} \right|.$$

The dependence of $\tilde{u}_q$ on q is ruled, once again, by the convexity of $\log p(u)$.

- When $\log p(u)$ is rectilinear, the $\tilde{u}_q$ are uniformly spaced.
- When $\log p(u)$ is cap-convex, $\tilde{u}_q/q$ is decreasing; that is, the $\tilde{u}_q$ are increasingly tightly spaced.
- When $\log p(u)$ is cup-convex, $\tilde{u}_q/q$ is increasing; that is, the $\tilde{u}_q$ are increasingly loosely spaced.

However, knowing $\tilde{u}_q$ is not enough; one must also know $u^q p(u)$ in the neighborhood of $\tilde{u}_q$. The function $u^q p(u)$ often admits a "Gaussian" approximation obtained by the "steepest descents" expansion

$$\log [u^q p(u)] = \log p(u) + qu = \text{constant} - (u - \tilde{u}_q)^2 \tilde{\sigma}_q^{-2/2}.$$

When $u^q p(u)$ is well-approximated by a Gaussian density, the bulk of EU^q originates in the "q-interval" defined as $[\tilde{u}_q - \tilde{\sigma}_q, \tilde{u}_q + \tilde{\sigma}_q]$.

The usual typical examples yield the following results. The Gaussian q-intervals greatly overlap for all values of σ. The Gaussian's moments will be called *delocalized.* The lognormal's q-intervals are uniformly spaced and their width is independent of q; therefore, when the lognormal is sufficiently skew, the q-interval and the $(q+1)$-interval do not overlap. The lognormal's moments will be called *uniformly localized.* In other cases, neighboring q-intervals cease to overlap for sufficiently high q. Such moments will be called *asymptotically localized.*

The notion of localization involves an inherent difficulty. Working in the "natural" scale is essential to problems involving addition, but here it is irrelevant. That is, it suffices to show that $u^q p(u)$ has a good Gaussian approximation in terms of either u or any increasing transform $v = y(u)$.

Example of the density $\exp(-u^w)$. Here, $q = w\tilde{u}_q^w$, hence, $\Delta\tilde{u}_q = \tilde{u}_q - \tilde{u}_{q-1} \sim q^{1/w-1}$; in addition, $\tilde{\sigma}_q^{-2} = wq\tilde{u}_q^{-2}$, hence, $\tilde{\sigma}_q \sim q^{1/w-1/2}$. It follows that $\tilde{\sigma}_q / \Delta\tilde{u}_q \sim \sqrt{q}$. That is, the q-intervals overlap for all values of w. (The same result is obtained using the free variable $v = \log u$.)

Example of the density $\exp[-(\log u)^w]/u$. The expression $u^q p(u)du$, if reexpressed in the variable $V = \log U$, becomes $\exp[-(v^w - qv)]$. One finds

$$\tilde{v} \sim (q/w)^{1/(w-1)} \text{ and } \Delta\tilde{v}_q \sim w^{-1/w} q^{(2-w)/(w-1)},$$

and

$$\tilde{\sigma}_q \sim w - \frac{1}{2(w-1)} (w-1)^{-1/2} q^{(2-w)/2(w-1)}.$$

It follows that $\tilde{\sigma}_q / \Delta\tilde{v}_q \sim q^{(w-2)/2(w-1)}$. When $w > 2$, all the moments of U are delocalized. When $w \leq 2$, they are localized. In the lognormal case $w = 2$, $\tilde{\sigma}_q / \Delta\tilde{v}_q$ is a constant that $\rightarrow 0$ as $w \rightarrow \infty$ and in the case beyond the lognormal, $w < 2$, $\tilde{\sigma}_q / \Delta\tilde{v}_q$ decreases as $q \rightarrow \infty$.

4.4 A tentative list of seven states of randomness

We see that the "slow" state between mild and wild splits into distinct states. Altogether, we shall face seven states of randomness, which we now list, together with examples. Alternative criteria involve the rate of increase as function of q of the moment EU^q or the scale factor $[EU^q]^{1/q}$.)

- *Proper mild randomness.* Short-run portioning is even for $N = 2$. Examples: the Gaussian, the distribution $P(u) = \exp(-u^w)$ with $w > 1$, and the gamma density $-P'(u) = u^{\gamma-1} \exp(-u)/\Gamma(\gamma)$ with $\gamma > 1$.

Mild randomness is loosely characterized, *either* by P^{-1} increasing near $x=0$ no faster than $|\log x|$, *or* by $[EU^q]^{1/q}$ increasing near $q \rightarrow \infty$ no faster than q.

• *Borderline mild randomness.* Short-run portioning is concentrated for $N=2$, but becomes even when N exceeds some finite threshold. Examples: the exponential $P(u)=e^{-u}$, which is the limit case of the preceding non-Gaussian examples for $w=\gamma=1$, and more generally the gamma for $\gamma<1$.

• *Slow randomness with finite and delocalized moments.* It is loosely characterized, *either* by P^{-1} increasing faster than $|\log x|$ but no faster than $|\log x|^{1/w}$, with $w<1$, *or* by $[EU^q]^{1/q}$ increasing faster than q but no faster than a power $q^{1/w}$. Examples: $P(u)=\exp(-u^w)$ with $w<1$, and $P(u)=\exp[-(\log u)^{\lambda}]$ with $\lambda>2$.

• *Slow randomness with finite and localized moments.* It is loosely characterized by *either* P^{-1} increasing faster than any power $|\log x|^{1/2}$ but less rapidly than any function of the form $\exp(|\log x|^{\gamma})$ with $\gamma<1$, *or* by $[EU^q]^{1/q}$ increasing faster than any power of q, but remaining finite. Examples: the lognormal and $P(u)=\exp[-(\log u)^{\lambda}]$ with $\lambda \leq 1$.

• *Pre-wild randomness.* It is loosely characterized *either* by P^{-1} increasing more rapidly than any functions of the form $\exp(|\log x|^{\gamma})$ with $\gamma<1$ but less rapidly than $x^{-1/2}$, or by $[EU^q]^{1/q}$ being infinite when $q \geq \alpha > 2$. Examples: the scaling $P(u)=u^{-\alpha}$ with $\alpha>2$. The power U^q becomes a wild random variable if $q>\alpha/2$.

• *Wild randomness.* It is characterized by $EU^2=\infty$, but $EU^q<\infty$ for some $q>0$, however small. Examples: the scaling $P(u)=u^{-\alpha}$ with $\alpha<2$.

• *Extreme randomness.* It is characterized by $EU^q=\infty$ for all $q>0$. Example: $P(u)=1/\log u$.

4.5 Aside on the medium-run in slow randomness: problems of "sensitivity" and "erratic behavior"

In the slow state of randomness, the middle run poses many problems. The case of the lognormal is investigated in Chapter E9, to which the reader is referred. A more general discussion begins in a straightforward fashion, but is too lengthy to be included here.

5. MATHEMATICAL TREATMENT OF THE TAIL PRESERVATION CRITERION $P_N \sim NP$, AND ROLE OF LONG-TAILEDNESS IN CLASSICAL MATHEMATICAL ANALYSIS

This Section, more mathematical in tone than the rest of this chapter, begins with an important proof and then digresses on some definitions and references.

5.1 Theorem: the tail-preservation criterion $p_2 \sim 2p$ and short-run ($N = 2$) concentration both follow when the function log $p(s)$ is decreasing and cup-convex and has a derivative that tends rapidly to 0 as $s \to \infty$

Let us repeat the definition of I_L, I_0 and I_L:

$$p_2(u) = \int_0^u p(s)p(u-s)ds = \left\{ \int_0^{\tilde{u}} + \int_{\tilde{u}}^{u-\tilde{u}} + \int_{u-\tilde{u}}^{u} \right\} p(s)p(u-s)ds$$
$$= I_L + I_0 + I_R.$$

Bounds on $I_L = I_R$. To establish concentration in probability, it suffices to prove that, as $s \to \infty$, $I_0/I \to 0$ but $1 - 2\tilde{u}/u$ does not tend to 0 . But we shall prove a far stronger result, namely that $I_L = I_R$ can be approximated by $p(s)$, in the sense that, given $\varepsilon > 0$, one can select $\tilde{u}$ so that, for large enough u,

$$(1-\varepsilon)p(u) < I_L = I_R < (1+\varepsilon)p(u).$$

The assumption that $p(u)$ is decreasing yields the following bounds valid for all $\tilde{u}$.

$$p(u)\int_0^{\tilde{u}} p(s)ds \le I_L = I_R \le p(u-\tilde{u})\int_0^{\tilde{u}} p(s)dx \le p(u-\tilde{u}).$$

The desired lower bound of $I_L = I_R$ is achieved if $\int_0^{\tilde{u}} p(s)ds > 1 - \varepsilon$. This inequality will follow automatically from the fact that the upper bound will require that $\tilde{u} \to \infty$ with u.

The desired upper bound is insured if $p(u - \tilde{u})/p(u) \le 1 + \varepsilon$. Assuming $\varepsilon \ll 1$, this reads $\log p(u-\tilde{u}) - \log p(u) < \varepsilon$. Assume that $g(s) = -(d/ds) \log p(s)$ exists and $\to 0$ as $s \to \infty$. Then the desired upper bound requires $\tilde{u} < \varepsilon g(u)$. The condition that $g(s) \to 0$ insures that $\tilde{u} \to \infty$ with u, therefore insures the validity of the lower bound to $I_L = I_R$.

Examples: The scaling cases $p(u) \sim u^{-\alpha-1}$ yields $\tilde{u}/u < \varepsilon/(\alpha+1)$, a constant. The cases $p(u) \sim \exp(-u^w)$ yield $\tilde{u} < \varepsilon u^{1-w}/w$, which increases with u , while $\tilde{u}/u < \varepsilon u^{-w}/w$ decreases. Now assume that $L(u)$ is slowly varying, which means that, for every μ, we have $L(\mu u)/L(u) \rightarrow 1$ as $u \rightarrow \infty$, and consider the density $p(u) \sim \exp(-u/L(u))$; the fact that this density is cup-convex implies that $L(u) \rightarrow \infty$; the resulting densities $p(u)$ yield $\tilde{u} < \varepsilon L(u)$, which again increases while $\tilde{u}/u$ decreases.

Finally, let us check that in the problematic gamma case, the desired upper bound is not available. This case is an example of $p(u) = \exp[-u - L(u)]$; the fact that this density is cup-convex, again implies $L(u) \rightarrow \infty$. Now, $\tilde{u}$ decreases as $u \rightarrow \infty$, albeit slowly. Therefore, the lower bound fails to hold, and the approach is not effective.

Upper bound on I_0 . Because of the cup-convexity of $p(s)p(u-s)$, one has

$$I_0 < (u - 2\tilde{u})p(\tilde{u})p(u - \tilde{u}).$$

The condition $u - 2\tilde{u} \leq u$, and the selection of an upper bound for $I_L = I_R$ have already insured that $p(u - \tilde{u}) \leq p(u)(1 + \varepsilon)$; hence

$$I_0 < (1 + \varepsilon)p(u)[up(\tilde{u})].$$

Return to the example of $p(s)$ considered in discussing the upper bound for $I_L = I_R$. Aside from $p(s) = \exp(-u/L(\underline{s})$, they yield $up(\tilde{u}) \rightarrow 0$, as $u \rightarrow \infty$. The example $p(s) = \exp[-uL(\underline{s})]$ is more complicated and depend on $L(\underline{s})$. Indeed, as $s \rightarrow \infty$, $\log[up(\tilde{u})] \sim \log u - \varepsilon L(u)/L[\varepsilon L(u)]$ behaves like $\log u - \varepsilon L(u)$. This expression may converge to $-\infty$, as for example when $L(u) = (\log u)^2$; in those cases $I_0 \rightarrow 0$. But this expression may also converge to $+\infty$; in those cases, it does not yield, it is a bound of I_0, and more detailed study is needed to tell whether $I_0 \rightarrow 0$. Obviously the issue is far from settled, but this is not the place to pursue the finer study of the domain of validity of the concentration in probability theory.

5.2 A digression: complications concerning the moments, the moment problem, and roles of long-tailedness in classical analysis

Thus far in this chapter, the finiteness of the moments was important, but their actual values and this behavior of EU^q as $q \rightarrow \infty$ were barely mentioned. In the slowly random case with $EU^q < \infty$, this behavior of EU^q is a genuinely hard problem. It is even a topic in what is called "fine (or hard) mathematical analysis" that repeatedly attracted the best minds. Unfortunately, the pure mathematical results are not of direct help to users: the complications that attract the mathematicians' interest prove to be a burden in concrete uses.

Convergence of the Taylor expansion of the characteristic function, and a related alternative definition of long-tailed randomness. It is widely taken for granted that the characteristic function (Fourier transform)

$$\varphi(s) = Ee^{isu} = \int_0^\infty e^{isu} p(u)du$$

always has the Taylor expansion

$$\varphi(s) = Ee^{isu} = \sum i^q s^q = \sum i^q s^q \frac{EU^q}{q!}.$$

When $\lim \sqrt[q]{EU^q/q!}$ exists, this limit is the inverse of the radius of convergence of this Taylor series. (When there is no limit, limit $\sup \sqrt[q]{EU^q/q!}$ always exists and is the inverse of the radius of convergence.)

For the exponential, the series expansion does indeed represent the analytic function $\varphi(s) = \Sigma i^q s^q = 1/(1-is)$. The radius of convergence is 1, and $\gamma(s) = Ee^{isU}$.

For the Gaussian $\varphi(s) = \exp(-2\sigma^2 s^2)$. Here, $EU^q = 0$ if q is odd and $EU^q = q!/2(q/2)!$ if q is even. The formal Taylor expansion has an infinite radius of convergence, defining $\exp(-2\sigma^2 s^2)$ as an "entire function."

But the lognormal yields $\lim \sup \sqrt[q]{EU^q/q!} = \infty$. The function $\gamma(s)Ee^{isU}$ is well-defined, but its formal Taylor series *fails to converge for* $s \neq 0$.

There is a strong temptation to dismiss those properties of the lognormal as meaningless mathematical blips. But they could also provide yet another alternative definition of long-tailed randomness. To do so, it is useful, when $\sqrt[q]{EU^q/q\,!}$ becomes infinite for $q \geq \alpha$, to also write $\lim \sup \sqrt[q]{EU^q/q!} = \infty$. When this is done, the criterion

$$\lim \sup \sqrt[q]{EU^q/q!} < \infty \text{ versus } \lim \sup \sqrt[q]{EUq/q!} = \infty$$

is a criterion of mild versus long-tailed randomness.

The moment problems and additional possible definitions of long-tailed randomness. The following questions were posed by Thomas Stieltjes (1856-1894). Given a sequence M_q, does there exist a measure U (a generalized probability distribution) such that $EU^q = M_q$? If U exists, is it unique? Stieltjes 1894 gave the lognormal as one of several examples where U exists, but is not unique. (See also p. 22 of Shohat & Tamarkin 1943). This property was rediscovered in Heyde 1963, recorded in Feller 1950 (Vol 2, 2nd edition, p. 227), and mentioned in studies of turbulence, including M 1974f{N15}, without suggesting any practical consequence.

The the available partial criteria are either sufficient or necessary, and are not the same on the line and the half-line. Loosely speaking, each known criterion is a way to distinguish between short and long-tailedness the murky border region around mild randomness. The same is true of the criteria encountered in the theory of "quasi-analytic" functions. Some criteria are worth mentioning:

Krein implicitly defines long-tailedness by the convergence of $J = \int_0^\infty \log\, p(u)(1+u^2)^{-1} du$. Koosis 1988-92 is a two-volume treatise that describes many problems where the conditions $J = -\infty$ and $J > -\infty$ are, respectively, the correct ways of expressing that the density $p(u)$ is short or long-tailed. Krein's definition is far more general than the convexity of $\log p(u)$. It is also a little more restrictive, because of the difference it makes between the forms $-u/\log u$ and $-u/(\log u)^2$ for $\log p(u)$.

Carleman implicitly defines long-tailedness by the convergence of $C = \sum(EU^q)^{-1/(2q)}$. For a distribution on the positive half-line to be determined by its moments, a sufficient condition is $C = \infty$. When U is bounded, $EU^q = (u_{\max})^Q$, therefore $C = \infty$. The exponential or the Gaussian also yields $C = \infty$. But $C < \infty$ holds for the scaling and the lognormal.

To conclude, my doubling criterion $P_2 = 2P$ is a new addition to an already overflowing collection. Who knows, perhaps this newcomer may add fresh spice to an aging mathematical game, or conversely.

E6

Self-similarity and panorama of self-affinity

✦ **Abstract.** This long and essential chapter provides this book with two of its multiple alternative introductions. The mathematically ambitious reader who will enter here will simply glance through Section 1, which distinguishes between self-similarity and self-affinity, and Section 2, which is addressed to the reader new to fractals and takes an easy and very brief look at self-similarity. Later sections approach subtle and diverse facets of self-affine scaling from two distinct directions, each with its own significant assets and liabilities.

Section 3 begins with WBM, the Wiener Brownian motion. In strict adherence to the *scaling principle* of *economics* described in Chapter E2, WBM is self-affine in a statistical sense. This is true with respect to an arbitrary reduction ratio r, and there is no underlying grid, hence WBM can be called the *grid-free*. Repeating in more formal terms some material in Sections 6 to 8 of Chapter E1, Section 3 discusses generalizations that share the scaling properties of WBM, namely, Wiener or fractional Brownian motion of fractal or multifractal time.

Section 4 works within grids, hence limits the reduction ratio r to certain particular values. Being *grid-bound* weakens the *scaling principle of economics*, but this is the price to pay in exchange for a significant benefit, namely the availability of a class of self-affine non-random functions whose patterns of variability include and exceed those of Section 3. Yet, those functions fall within a unified overall master structure. They are simplified to such an extent that they can be called "toy models" or "cartoons."

The cartoons are grid-bound because they are constructed by recursive multiplicative interpolation, proceeding in a self-affine grid that is the simplest case prescribed in advance. The value of grid-bound non-random fractality is that it proves for many purposes to be an excellent surrogate for randomness. The properties of the models in Section 3 can be

reproduced with differences that may be viewed as elements of either indeterminacy or increased versatility. Both the close relations and the differences between the cartoons could have been baffling, but they are pinpointed immediately by the enveloping master structure. At some cost, that structure can be randomly shuffled or more deeply randomized. Its overall philosophy also suggests additional implementations, of which some are dead-ends, but others deserve being explored.

Wiener Brownian motion and its cartoons belong to the *mild* state of variability or noisiness, while the variability or noisiness of other functions of Section 3 and cartoons of Section 4 are *wild*. The notions of states of mild and wild randomness, as put forward in Chapter E5, are generalized in Section 5 from independent random variables to dependent random processes and non-random cartoons. Section 5.4 ends by describing an ominous scenario of extraordinary wildness.

Being constrained to scaling functions, this chapter leaves no room for slow variability. ✦

WHEN DISCUSSING THE ORGANIZATION OF THIS BOOK, the Preface mentions several welcoming entrances. This and the preceding chapters are the entrances most suited for those who do not fear mathematics. (This chapter grew to become too long, and may be best viewed as several chapters bound together.

While Chapter E5 restricted itself to independent random variables, this chapter allows dependence, either deterministic or statistical, but restricts itself to self-affine scaling. This allows for mild and wild randomness, but not for slow randomness. That is, this chapter describes dependent functions or processes in continuing time that generalize a special family considered in Chapter E5, namely sequences of L-stable variables, with their Gaussian limit case.

The term *Panorama* in the title is meant to underline that, beyond the specific needs of this book on finance, this chapter also opens vistas that involve many other fields. Indeed, self-affine random variation is by no means restricted to economics. It is also often encountered in physics, for example, in $1/f$ *noises*. Different examples of those noises involve several of the variants in this *Panorama*, but there is no field in which all variants have been fully implemented. Fuller versions of the same text, updated and with different biases, are scheduled for M 1997N and M 1997H, which mention $1/f$ noise in the title. Those versions will be more technical and perhaps more practical.

The text can first be skimmed and later read in increasing detail. Students of finance who do not favor mathematics may be satisfied to examine the illustrations, and to be aware that this chapter helps organize and relate the Bachelier "B 1900 model," and my successive models in finance, M 1963, M 1965, M 1967 and M 1972, as sketched in Sections 6 to 8 of Chapter E1.

The strong term "cartoons" used to describe the "grid-bound" implementations of self-affinity collected in Section 4. A political cartoon's effectiveness hinges on its being highly simplified, yet preserving the essentials of what it refers to. In the same spirit, as known to readers familiar with elementary fractals and sketched below in Section 2 for the sake of other readers, the non-random Koch islands were mathematical curios until I injected them as "cartoons" of realistic fractal models of coastlines; in turn, those random models are cartoons of real coastlines. Some of the non-random self-affine constructions in Section 4 are cartoons of the random self-affine process in Section 3; the latter, in turn, are cartoons of real price records. Cartoons being unavoidable, the user should learn to like them, and the provider must develop ways to make them simple yet instantly recognizable.

Disclaimers. This *Panorama* is by no means the last word on its topic, in part, because the field of fractals has not yet become unified. Some studies grow from the top down: they first set general principles and then proceed to the consequences. To the contrary, fractal geometry grows from the bottom up. It continues to draw new substance from a succession of explorations with focussed ambitions. In parallel, it continues an effort to rethink the available substance in fashions that are increasingly organized, and suggest new explorations.

Use of the cartoons to resolve a widespread confusion between the M 1963 and M 1965 models. Between the L-stable motion behind the M 1963 model, and the fractional Brownian motion behind the M 1963 model, mathematicians see a number of parallelisms often described as "mysterious." Fortunately, Section 4 suggests that self-affinity may be one of those cases for which order and simplicity are restored, and confusion vanishes, when *a*) the standard models are made *more*, rather than *less*, general, and *b*) the resulting wider family of possibilities is presented in very graphic fashion.

In particular, very simple arguments relative to the cartoons suffice to eliminate a confusing complication that concerns the value of the fractal dimension. Depending on which feature is being singled out, the dimension is as follows:

• Either $D_G = 2 - H$ or $D_T = 1/H$ for the graphs or trails of fractional Brownian motions (M 1965 model; see Section 3.3).

• Either $D = 2 - 1/\alpha$ or $D_T = \alpha$ for the graphs or trails of L-stable processes (M 1963 model).

• Moreover the M 1972 model (see Section 3.13) yields two values, D_T and $D_G > D_T$ that are not functionally related to each other. Other dimensions also enter into contention.

A multiplicity of binary splits. It is useful to underline the versatility of self-affine constructions by describing how they split in several overlapping ways. The following list uses terms that will not be defined until later in this chapter, therefore should be viewed as merely suggestive.

• Between *grid-free* and *grid-bound.*

• Between *mildly* and *wildly variable.*

• Between *continuous* and *discontinuous.*

• Between *monotone,* either non-decreasing or non-increasing, and *oscillating* up and down.

• Between *non-intermittent,* that is, allowing no interval of clock time when motion stops, and *intermittent,* with variation concentrated on a fractal trading time. Variation can also be *relatively intermittent,* if it is concentrated on a new construct: a multifractal trading time.

• Between *unifractal,* characterized by a single exponent *H*, *mesofractal,* which also includes other values of *H* restricted to be 0 and/or infinity, and *multifractal,* characterized by a distributed exponent *H*.

• When *H* is single-valued, between the case $H = 1/2$ and the cases $H \neq 1/2$.

• Finally (but this will not be discussed in this book), between constructions that are *stable* or *unstable* under wild randomization.

1. CONTRAST BETWEEN SELF-SIMILARITY AND SELF-AFFINITY

This chapter concerns scaling behavior in the graph of a function, more precisely, linearly scaling behavior. Before seeking examples, one must know that this scaling has two principal geometric implementations: self-similar fractals, and the more general self-affine fractals. Self-similarity, the narrowest and simplest, is the most standard topic of fractal geometry, and it is good to begin by briefly considering it in Section 2. But the remainder of this chapter and this book are limited to functions whose graphs are *self-affine fractals*.

This distinction is essential, and it is most unfortunate that many authors use one word, *self-similar*, to denote two concepts. I gave a bad example, but only until M 1977F, when I found it necessary to introduce the term *self-affine*. This term is now accepted by physicists, engineers, and mathematicians who study non-random constructs. Unfortunately, many probabilists persist in using *self-similar* when they really mean *self-affine*; this is the case in the book by Baran 1994 and Somordnitsky & Taqq 1994.

Let us elaborate. Many geometric shapes are approximately isotropic. For example, no single direction plays a special role when coastlines are viewed as curves on a plane. In first-approximation fractal models of a coastline, small pieces are obtained from large pieces by a similarity, that is, an isotropic reduction (homothety) followed by a rotation and a translation. This property defines the fractal notion of self-similarity. Self-similar constructions make free use of angles, and distances can be taken along arbitrary directions in the plane.

But this book deals mostly with geometric shapes of a different kind, namely, financial charts that show the abscissa as the axis of time and the ordinate as the axis of price. The scale of each coordinate can be changed freely with no regard to the other. This freedom does not prevent a distance from being defined along the coordinate axes. But for all other directions, the Pythagorean definition,

$$\text{distance} = \sqrt{(\text{time increment})^2 + (\text{price increment})^2}\ ,$$

makes no sense whatsoever. It follows immediately that circles are not defined. Rectangles must have sides parallel to the axes. Squares are not defined, since – even when their sides are meant to be parallel to the axes – there is no sense in saying that time increments = price increment.

There is a linear operation that applies different reduction ratios along the time and price axes. It generalizes similarity, and Leonhard Euler called it an *affinity*. More precisely, it is a *diagonal affinity*, because its matrix is diagonal. It follows that for graphs of functions in time, like price records, the relevant comparison of price charts over different time spans involves the scaling notion of self-affinity. Self-affinity is more complicated and by far less familiar than self-similarity, therefore this chapter begins by surveying the latter. Readers already acquainted with fractals may proceed to Section 3.

On the measurement of texture, irregularity or roughness. The very irregular and rough shapes often encountered in Nature never tire of exciting the layman's imagination, but science long failed to tackle them. Thus, no serious attempt was made to define and measure numerically the irregularity of a coastline or a price record.

Topology provides no answer, even through its name seems to promise one. For example, consider a chart or time record of prices when filled-in to be continuous; this curve can be obtained from the line without a tear, using a one-to-one continuous transformation. Disappointedly, this property defines *all* price charts as being topological straight lines!

Nor does statistics provide a useful answer. For example, examine the perennial and objective problem of measuring the roughness of physical surfaces. Statistics suggests following a procedure familiar in other fields: first fit a trend-like plane (or perhaps a surface of second or third degree), then evaluate the root-mean-square (r.m.s.) of the deviation from this trend. What is unfortunate is that this r.m.s., when evaluated in different portions of a seemingly homogeneous surface, yields conflicting values.

Does the inappropriateness of topology and statistics imply that irregularity and roughness must remain intuitive notions, inaccessible to mathematical description and quantitative measurement? Fractal geometry is a geomety of roughness, and it answers with a resounding no. It shows that in many cases, roughness can be faced and overcome to a useful extent, thanks to scaling exponents that underlie the scaling principles of mathematical and natural geometry.

For example, coastlines are nearly self-similar, and the most obvious aspect of their roughness is measured by a quantity called *fractal dimension,* which is described in Section 2. Many irregular physical surfaces are self-affine, and their roughness is measured reliably by two numbers. One is the exponent H introduced in Section 3; engineers have already come around to call it simply the "roughness exponent," but mathematically, it is a Hurst-Hölder exponent and a fractal co-dimension. The second characteristic number is a scale factor similar to a root-mean-square, but more appropriately defined. (This topic is treated in M 1997H). The study of roughness in terms of self affinity has become a significant topic in physics; see Family & Vicsek 1991.

2. EXAMPLES OF SELF-SIMILAR RECURSIVE CONSTRUCTIONS

2.1 Getting answers without questions to work on questions without answers

Of the five diagrams in Figure 1, the largest and most complicated is the composite of two wondrous and many-sided broken lines. One of them is violently folded upon itself, and gives the impression of attempting a monstrous task for a curve: to fill without self-contact the domain bounded by the less violently folded second curve. This impression was intended, since we witness an advanced stage of the construction of a variant "space-filling curve," an object discovered by Giuseppe Peano (1858-1932).

Actually, "space-filling curve" is an oxymoron. An improved substitute that I proposed is "space-filling (or Peano) *motion*." Thus, Figure 1 illustrates a variant of the original Peano motion, bounded by a less violently folded fractal "wrapping." By construction, both curves are precisely as complicated in the small as in the large. The wrapping, introduced in M 1982F{*FGN*}, Chapter 6, is patterned after one that Helge von Koch used in a celebrated shape called "snowflake curve." The filling was introduced in M 1982t. Never mind that Koch's motivation was purely mathematical: he was seeking a curve without tangent anywhere, meaning that the direction of a cord joining any two points has no limit as these points converge to each other. To achieve this goal, the simplest was to demand that this cord fluctuate exactly as much in the small as in the large.

Fractal geometry preserved this demand, but changed its motivation from purely mathematical to very practical. When irregularity is present at all scales, it is simplest when, whatever the magnification, the fine details seen under the microscope are the same (scale aside) as the gross features seen by the naked eye. Using the vocabulary of geography, the fine details seen on a very precise map are the same as the gross features seen on a rough map. Concrete reinterpretations of Koch's recursive procedure continually inspire me in empirical work. In summary, one can state two guiding principles.

A) *Scaling principle of natural geometry*. Shapes whose small and large features are largely identical except for scale, are useful approximations in many areas of science.

B) *Scaling principle of mathematical geometry.* Sets wherein small and large features are identical except for scale are interesting objects of study in geometry.

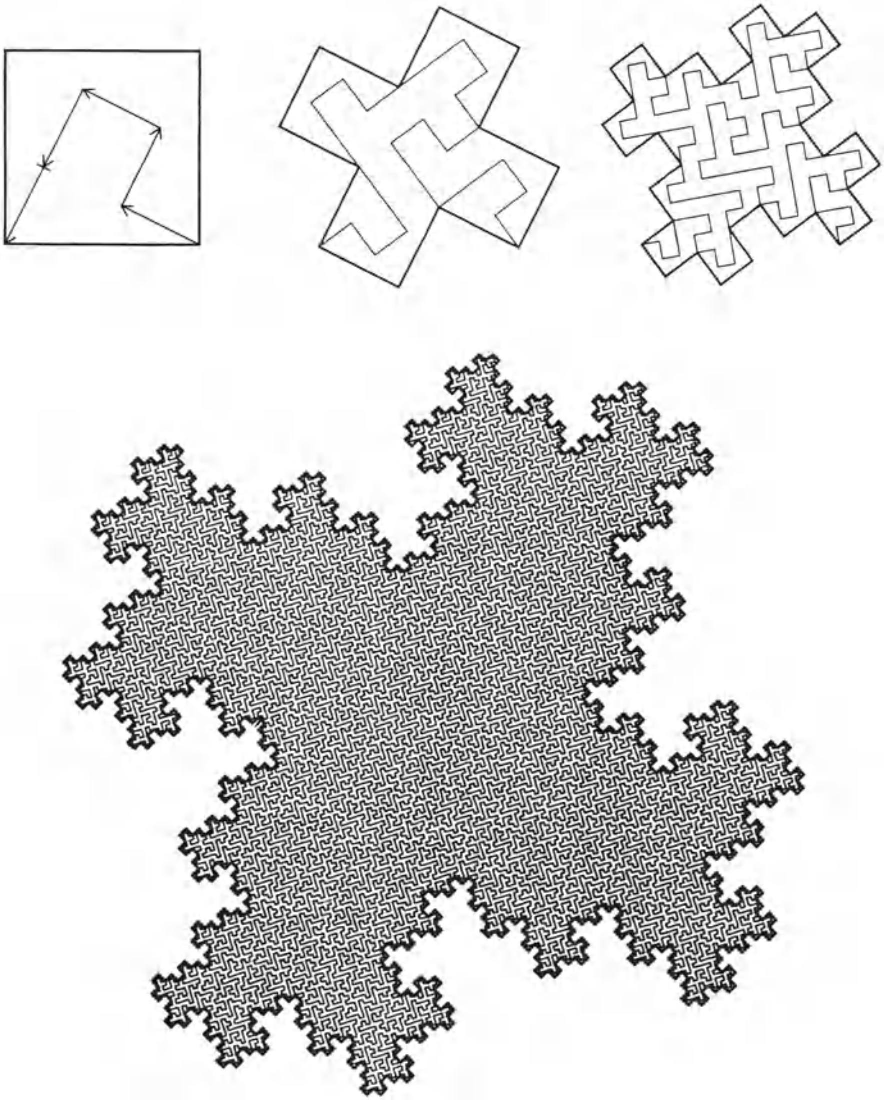


FIGURE E6-1. Construction of a Peano motion "wrapped" is a squared Koch curve that it fills.

Part of my life-work consists in viewing B as providing a collection of answers without question, and setting them to work on the questions without answers summarized under A.

2.2 Examples of self-similar fractal shapes

To implement the goal that Koch stated in his way before I restated it in mine, the easiest is to proceed step by step. Select an *initiator*, often an interval, and a *generator*, also a broken line. The first construction stage replaces each side of the initiator by an appropriately rescaled, translated and rotated version of the generator. Then a second stage repeats the same construction with the more broken line obtained at the first stage, and so on.

The early stages of the constructions shown on Figure 1 are illustrated by four small diagrams to be followed clockwise from left center, in order of increasing complication. The initiators are the four sides of a unit square for each of four repeats of the wrapping of Peano motion and one side of this square for the Peano motion itself. The generator of the motion is an irregular open pentagon that does its best to fill the square, using sides equal to $1/\sqrt{5}$. One perceives an underlying square lattice of lines $1/\sqrt{5}$ apart, and the Peano generator crosses every lattice vertex contained in the original wrapping. The wrapping generator has $N=3$ sides of length $r=1/\sqrt{5}$.

In the next stage of the construction, each side of the pentagon is replaced by an image of its whole reduced in the ratio of $1/\sqrt{5}$, and suitably rotated. The result no longer fits within the square, but fills uniformly the cross-like shape obtained by replacing each side of the square by the wrapping generator. The same two constructions are then repeated ad infinitum in parallel. Zooming in as the construction proceeds, one will constantly witness the same density of filling; watching without zooming in, one sees a curve that fills increasingly uniformly a wrapping whose complexity keeps increasing.

The Peano motions which mathematicians designed during the heroic period from 1890 to 1922 filled a square or a triangle, but the present boundaries are more imaginative.

Figure 2 carries the construction of a curve of Figure 1 one step further and the filling is interpreted as the cumulative shoreline of several juxtaposed river networks; the wrapping is the combination of a drainage divide surrounding these networks and of a portion of seashore. To build up the network, one proceeds step by step: (1) Each dead-end square in the basic underlying lattice – meaning that three sides belong to the filling – is replaced by its fourth side, plus a short "stream" with its source at the center of the dead-end, and its end at the center of the square beyond the newly filled-in side. (2) One proceeds in the same fashion with the polygons left in after the processed dead-ends are deleted. (3) And so on until the filling is changed from a broken line with no self-contact to a collection of "rivers" forming a tree. At this point, the wrapping becomes reinterpreted as the river network's external drainage divide.

To use an old sophomoric line, after you think of it imaginatively, carefully, and at great length, it becomes obvious that a plane-filling motion fails at its assigned task of being a mathematical monster. I proved it to be nothing but a river network's cumulative shore. The converse is also true. Much better-looking river networks are given in my book, M 1982F{*FGN*}, but the basic idea is present here. There is not much else to Peano motions. Thus, the mathematicians who used to tell us that Peano motions are totally nonintuitive had deluded themselves and misinformed the scientists.

2.3 The notion of fractal dimension of a self-similar geometric shape

Each stage of a Koch construction replaces an interval of length 1 by N intervals of length r, therefore multiplies a polygon's length by a fixed factor $Nr > 1$. It follows that the limit curves obtained by pursuing the recursions ad infinitum are of infinite length. Furthermore, it is tempting to say that the filling is "much more infinite" than its wrapping, because its length tends to infinity more rapidly. This intuitive feeling is quantified mathematically by the notion of fractal dimension. The original form was introduced by Hausdorff and perfected by Besicovitch. It is inapplicable to empirical science, and had to be replaced by a variety of alternative definitions.

The explanation of the underlying idea begins with the very simplest shapes: line segments, rectangles in the plane, and the like. Because a straight line's Euclidian dimension is 1, it follows, for every integer $\gamma > 1$, that the "whole" made up of the segment of straight line $0 \leq x < X$ may be "paved over" (each point being covered once and only once) by $N = \gamma$ segments of the form $(k-1)X/\gamma \leq x < kX/\gamma$, where k goes from 1 to γ. Each of these "parts" can be deduced from the whole by a similarity of ratio $r(N) = 1/N$. Likewise, because a plane's Euclidian dimension is 2, it follows that, whatever the value of γ, the "whole" made up of a rectangle $0 \leq x < X; 0 \leq \gamma < Y$ can be "paved over" exactly by $N = \gamma^2$ rectangles defined by $(k-1)X/\gamma \leq x < kX/\gamma$ and $(h-1)X/\gamma \leq y < hY/\gamma$, where k and h

go from 1 to γ. Each part can now be deduced from the whole by a similarity of ratio $r(N) = 1/\gamma = 1/N^{1/2}$. Finally, in a Euclidian space whose dimension is $E > 3$, a D-dimensional parallelepiped can be defined for any $D \le E$. All those classical cases satisfy the identity

$$D = \frac{-\log N}{\log r(N)} = \frac{\log N}{\log(1/r)}.$$

This expression is the self-similarity dimension. Its value lies in the ease with which it can be generalized. Indeed, the fact that it was first used for a segment or a square is not essential for its definition. The critical requirement is scaling, meaning that the whole can be split into N parts deducible from it by a self-similarity of ratio r (followed by translation, rotation, or symmetry). Such is precisely the case in Figure 1. For the wrapping, $N = 3$ and $r = 1/\sqrt{5}$, hence

$$D = \log 3/\log\sqrt{5} = \log 9/\log 5 = 1.3652.$$

For the filling, $N = 5$ and $r = 1/\sqrt{5}$, hence

$$D = \log 5/\log\sqrt{5} = 2.$$

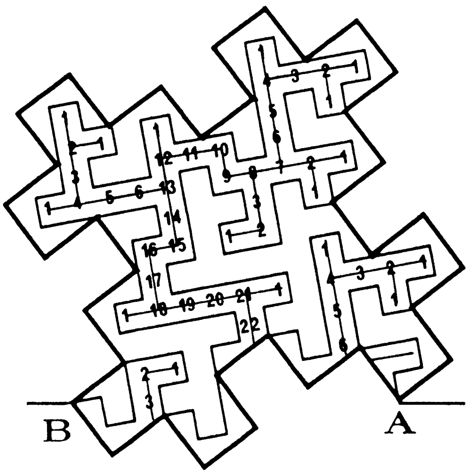


FIGURE E6-2. The third diagram of Figure 1, reinterpreted in terms of a river network. This interpretation led M 1982F{*FGN*} to boast of having "harnessed the Peano Monster Curves."

Thus, the impression that the filling is "more infinite" than its wrapping is both confirmed, and quantified by the inequality between their dimensions. The impression that the filling really fills a plane domain is confirmed and quantified by its dimension being $D=2$.

The preceding argument may seem overly specialized, so it may be comforting to know (a) that fractal dimension can be defined using alternative methods of greater generality and full rigor and (b) that the result behaves in many other ways like the old-fashioned integer-valued dimension. For example, consider the notion of *measure.* If a set is self-similar and measure is taken properly, then the portion of this set that is contained in a sphere of radius R is of measure proportional to R^D.

3. SELF-AFFINE FRACTAL MODELS IN FINANCE

Joined by readers who knew about fractals and skipped section 2, we now turn from self-similarity to self-affinity and to a collection of possible models of price variation that follow the scaling principle of economics. Sections 3 and 4 cover roughly the same material in two very different ways. There is enough overlap to allow Sections 3 and 4 to be read in either sequence.

The bare facts were already sketched in Sections 6 to 8 of Chapter E1, but *one need not* read these sketches before this Section. Furthermore, this chapter has no room for a full treatment of L-stable motion, fractional Brownian motion and multifractals; L-stable motion is the topic of much of the second half of this book; fractional Brownian motion is the topic of M 1997H, and multifractals are the topic of M 1997N.

3.1 The 1900 model of Bachelier, Brownian motion

$B(t)$ is defined as being a random process with Gaussian increments that satisfies the following "Fickian" diffusion rule:

$$\text{for all } t \text{ and } T,\ E\{B(t+T)-B(t)\}=0 \text{ and } E\{B(t+T)-B(t)\}^2=T.$$

A Fickian variance is an automatic consequence if the increments are assumed independent. Conversely, Fickian variance guarantees the orthogonality of the increments. Adding the Gaussian assumption, it guarantees independence.

3.2 Tail-driven variability: the M 1963 model and the L-stable processes

Reference. The concept of L-stability is discussed throughout the second half of this book, and it would be pointless to repeat here the definition due to Paul Lévy. It is enough to say that L-stability means that the sum of N *independent* L-stable variables is itself L-stable. The Gaussian shares this property and indeed it is a limit case of the L-stable variables, when the parameters α tends to 2. Moreover, consider a weighted index of *independent* variables $\Sigma W_g X_g$. When the weights are *not* random, the variables X and the weighted index are L-stable.

The "ruin problem" for the L-stable processes. Suppose a speculator is called ruined if his holdings fall beyond a prescribed level called "threshold." What is the probability that ruin occurs before a time t_{max}? Questions of this type are thoroughly explored for Wiener Brownian motion. For L-stable processes, the literature is limited, but includes Darling 1956 and Ray 1958.

Invariance under non-randomly and randomly weighted forms of addition. This digression is addressed to readers who know the concept of fixed point of a (semi-)group of transformations. L-stable variables are *fixed points* in the operation that consists in transforming independent random variables by taking a *non-randomly weighted average.* A distribution invariant under addition of *independent* addends used to be thought as necessarily Gaussian until M 1960i{E10} injected L-stable addends in a down-to-earth concrete situation. The M 1972 model, to be presented in Section 3.8, involves an actual generalization of L-stability, and it is good to mention how this generalization relates to Lévy stability. The considerations in M 1974f{N15} and M 1974c{N16} also involve a weighted index $\Sigma W_g X_g$; but there is the important innovation that the weights W are not constants, but independent values of the same random variable W.

Given W and N, I investigated the variable Y such that $\Sigma W_g Y_g$ has, up to scale, the same distribution as Y. Using the terminology already applied to L-stable variables, my variables Y are *fixed points* in the operation that consists in taking *randomly weighted averages of independent* random variables. The variables Y range from being close to L-stable (a limit case) to being very different indeed.

3.3 Dependence-driven variability: the 1965 model and fractional Brownian motion

The fractional Brownian motion (FBM) $B_H(t)$ is the random process with Gaussian increments that satisfies the following diffusion rule

$$\text{for all } t \text{ and } T,\ E[B_H(t+T) - B_H(t)] = 0 \text{ and } E[B_H(t+T) - B_H(t)]^2 = T^{2H}.$$

The value $H = 1/2$ yields the Wiener Brownian motion, whose diffusion is called "Fickian." However, the exponent H is only constrained to $0 < H < 1$. For $H \neq 1/2$, the diffusion of FBM is widely called "non-Fickian." In a different terminology, a mysterious but widely used one, H is called "strength of singularity" at time t.

This process was introduced in M 1965h{H} and fully described in M & Van Ness 1968{H} as a model of diverse phenomena that exhibit cyclic non-periodic variability at all time scales. The oldest recorded example concerned the annual discharge of the Nile River and is associated with the Biblical story of Joseph, the son of Jacob. Therefore, I refer to non-periodic cyclicity as the *Joseph Effect*. The use of FBM is economics was pioneered in M 1970e, M 1971n, M 1971q, M 1972c and M 1973j. Recent mathematical references are Baran 1994 and Samorodnitsky & Taqqu 1994 (Section 7.2). Unfortunately, as already mentioned, both books use the word self-similarity where the correct concept, hence the correct term, is self-affinity). A recent book for engineers is Bras & Rodriguez-Iturbe 1993 (pages 210-261).

The property of uniscaling. The above definition implies that the scale factors based on moments satisfy

$$\{E[\,|B_H(t+T) - B_H(t)|^q\}^{1/q} = \text{(a constant) } T^H$$

for all powers $q > -1$. (For $q \leq -1$, this expression becomes infinite.) That is, q-th order scale factor defined by the left hand side, is independent of q. (For $q < -1$, the left hand side is infinite, therefore the equality holds trivially, with an infinite constant). This obvious corollary is said to express *uniscaling*. It will become important in Section 3.8 and 3.9, and the cases when the q-th scale factor depends on q will be called *multiscaling*. By contrast, $B_H(t)$ will be called *uniscaling*, and no complication can arise from writing $\Delta B_H \sim \Delta t^H$.

Hurst, Hölder, and an exponent that bridges mathematics and concrete needs. With a suitable definition of the symbol ~, the functions $B_H(t)$, including $B(t) = B_{1/2}(t)$, satisfy

$$\frac{\log |\Delta B_H|}{\log \Delta t} \sim H.$$

Observe that, in this definition, Δt and ΔB_H are increments over a non-vanishing interval, *not* infinitesimal quantities. One says that, as defined here, H is a *coarse* quantity, not a *fine* or *local* one. In addition, H is defined for all values of t.

The idea behind the exponent H has two thoroughly disparate historic roots. When introduced in M 1965h{H}, $B_H(t)$ was motivated by a difficult problem from civil engineering, and referred to the initial letter of the hydrologist H. E. Hurst (1880-1978), briefly mentioned in Chapter E. But H also has a second set of deep roots in pure mathematics, namely, in the work of L. O. Hölder (1859-1937). Serendipitously, the names of Hurst and Hölder shared the same initial letter. However, Hölder's original definition had to be very much generalized. In Section 4, the underlying idea will split further.

$B_H(t)$ and the phenomenon of long-run statistical dependence. The most striking single property of $B_u(t)$ concerns the quantities $[B_H(0) - B_H(-T)]/T$, called *past* average and $[B_H(T) - B_H(0)/T$, called *future* average. Both are Gaussian random variables, and their correlation is easily seen to be

$$C = \frac{1}{2}\frac{(2T)^{2H} - T^{2H} - T^{2H}}{(T^H)^2} = 2^{2H-1} - 1.$$

That is, C is independent of T. This fact could be called "intuitive" because it follows from self-affine scaling. But an older form of "intuition" of the nature of randomness is more demanding, and insists that a distant past and a distant future "should" become statistically independent. This second intuition is correct in the Wiener Brownian case, where $C = 0$, and in other cases of mild randomness. But it must be "unlearned" in all other cases.

More precisely, $C > 0$ in the "persistent" case $1/2 < H < 1$ and $C < 0$ in the "anti-persistent" case $0 < H < 1/2$. In both cases, $B_H(t)$ *is* definitely *neither* a martingale *nor* a Markov process.

Spectral properties: the fractional Gaussian noise $B'_H(t)$ as a "continuing" or "humming" form of " $1/f$ noise." $B_H(t)$ is continuous but not differentiable. However, one can define for it a "generalized derivative" $B'_H(t)$. The spectral density of $B'_H(t)$ is $\propto f^{-B}$, with the exponent $B = 2H - 1$ ranging between 1 and − 1. Physicists denote such phenomena by the curious term of "$1/f$ noises." When $1/2 < H < 1$, the spectral density diverges at $f = 0$. This is one token of long-run statistical dependence. When $0 < H < 1/2$, the integral of the spectral density is 0, which is a different (and far less "robust") token of long-run statistical dependence.

The multiplicity of co-existing fractal dimensions for $B_H(t)$, including the value $D_G = 2 - H$, and the larger value $D_T = 1/H$. Section 2.3 describes how the irregularity of a self-similar fractal curve is in large part measured by a number called its "fractal dimension" D. Self-affine curves are significantly more complicated, as M 1997H will show from several distinct viewpoints.

A first complication is this. While self-similar fractals have a unique fractal dimension, I showed that self-affine fractals demand several, depending on which aspect is being considered. In the case of $B_H(t)$, some careful authors only quote the value $D_G = 2 - H$, while other careful authors only quote $D_T = 1/H > 2 - H$.

Those two sets of authors report different answers because the values $2 - H$ and $1/H$ refer to different geometric objects.

The value $2 - H$ can be shown to be the box dimension of the graph of $X(t)$, hence the suffix G.

The value $1/H$ can be shown to be the box dimension of a different but related geometric object, namely a "trail," hence the suffix T.

The distinction between graph and trail is developed in M 1982F{*FGN*}, but the main facts can be summarized here. First consider a Wiener Brownian motion in the plane. Its coordinates $X(t)$ and $Y(t)$ are independent Brownian motions. Therefore, if a 1-dimensional Brownian motion $X(t)$ is combined with another independent 1-dimensional Brownian motion $Y(t)$, the process $X(t)$ becomes "embedded" into a 2-dimensional Brownian motion $\{X(t), Y(t)\}$. The value $D_T = 2 = 1/H$ is the fractal dimension of the three dimensional graph of coordinates t, $X(t)$ and $Y(t)$, and the projected "trail" of coordinates $X(t)$ and $Y(t)$. However, the dimension $D_G = 2 - H$ applies to the projected graphs of coordinates t and $X(t)$ or t and $Y(t)$. A heuristic derivation of this value is best postponed to Section 3.13, where it will be generalized.

An FBM with $H \neq 1/2$ can only be embedded in a space of dimension $E \geq \max(2, 1/H)$.

We are done now with explaining how two values of the dimension coexist peacefully in the *unifractal* case of the FBM $B_H(t)$. Thinking ahead, Section 3.13 tackles the next case, to be called *multifractal*, and shows that D_T and D_G cease to be related functionally.

There is a second complication (but it is beyond the scope of this book): in the self-affine case, the notion of fractal dimension splits into local and global forms. The above-mentioned values are local, and the global values are different.

3.4 Trading time, compound processes, and a fundamental fact: preservation of the trail dimension D_T under compounding

The variation of most prices is neither tail- nor dependence-dominated, but ruled by both contributions in combination. To model such combinations, one must go beyond the M 1963 and M 1965 models. This is a task I first attacked piecemeal, by seeking suitable random functions and later attacked systematically, by introducing a flexible general family of random functions. (Actually, several options were considered, but the present discussion will be limited to one.)

Trading time and compound processes. The processes in this family are "compound," "decomposable," or "separable" in the following sense: by construction, their variation is "separated" into the combination of two distinct contributions. The first is a trading time θ, a random non-decreasing function of clock time t. In the terminology in Feller 1950 (Vol. II, p. 347), $\theta(t)$ is called *directing function*. The second, which yields price as function of trading time, $X(\theta)$, will be called *compounding function*.

In the absence of further restrictions, the notion of compounding is useless. Indeed, given a function $P(t)$, an arbitrary choice of $\theta(t)$ automatically defines also a function $X(\theta)$ such that $X[\theta(t)] = P(t)$. Our attention will be restricted to the case when the two components are statistically independent.

Furthermore, we wish to insure that the compound process is self-affine, that is, follows the scaling principle of economics. The easiest is to demand that *both* $\theta(t)$ and $X(\theta)$ be self-affine functions. In addition, the directed functions will be WBM and FBM, thus preserving something of the Bachelier model and the M 1965 model. The hope, of course, is that the outcome provides a sensible approximation to interesting data that are driven by a combination of tail and dependence.

Preservation of the trail dimension $1/H_T$ *under continuous compounding.* Section 3.3 distinguishes between the graph of $X(t)$ and an embedded trail of coordinates $X(t)$ and $Y(t)$. Compounding can be continuous or discontinuous, as will be seen momentarily. When it is continuous, it modifies the graph of $X(t)$, but leaves unchanged the trail of coordinates $X(t)$ and $Y(t)$. In particular, the trail dimension remains $1/H$. When compounding is discontinuous, it modifies both the graph and the trail.

Comment. In this section, trading time is a notion that is borrowed from our historical, and therefore intuitive, knowledge of how markets operate. In Section 4, trading time will enter in a far more intrinsic fashion.

3.5 A major but unrecognized "blind spot" of spectral analysis: spectral whiteness is insensitive to change of trading time, therefore misleading

To engineers, successive increments ΔB of Wiener Brownian motion define a *white noise.* They are independent, therefore uncorrelated ("orthogonal"), and their spectral density is a constant, defining a white spectrum. Now, let us follow Brownian motion in a trading time chosen at will (self-affine, or not). The increments of the compound motion are very strongly dependant. However, most remarkably, they are uncorrelated, therefore *they remain spectrally white.* In other words, spectra as applied to a compound process are only sensitive to the whiteness of the directed function, and completely blind to the properties of the directing function.

Indeed, given two non-overlapping time increments $d't$ and $d''t$, the corresponding increments $d'B(t)$ and $d''B(t)$ are, by definition, independent. It is obvious that this property continues to hold when B is followed in a trading time θ that is in a non-linear non-decreasing function of t, and $B(t)$ is replaced by $B^*(\theta)=B[t(\theta)]$. The increments of B^* exhibit very strong dependence, yet they are white, that is, uncorrelated.

Remark concerning statistical method. When interpreting spectra in a non-Gaussian and non-Brownian context, this dangerous possibility must be kept in mind. This serious "blind spot" was noted, but not developed, in my papers on noise of the 1960s, to be collected in M 1997H. It constitutes a fundamental limitation of spectral analysis that statistics must face.

Remark concerning the spectral whiteness of financial data. During the 1960s spectral analysis was introduced into economics with fanfare, but never lived to its promise. The "blind spot" of spectra suffices to account for many puzzling observations reported in the literature. Indeed, Voss 1992 and the contributors to Olsen 1996 are neither the first nor the only authors to report on such whiteness. Both parties also examined records of absolute price change, or of price change squared. The spectrum is no longer white but instead takes the "$1/f$" form characteristic of FBM. It will be shown at the end of Section 3.9 that this apparent contradiction is characteristic of the M 1972 model, namely, of the Brownian motion in multifractal time.

Remark concerning R and R/S analysis. This form of analysis is mentioned and referenced in Section 7.4 of Chapter E1 and discussed in M 1997H. Changes in tradii.g time leaves the range unchanged, but removal of the trend (as it is practiced in R/S) does modify the range. This topic must be withheld for consideration in M 1997H.

3.6 A special form of discontinuous compounding, "subordination;" the notion of fractal time

Definitions. The simplest directing function $\theta(t)$ are functions with non-negative statistically independent increments. This form of compounding is denoted by the term *subordination,* which is due to S. Bochner. The most general implementation is a non-decreasing random function with infinitely divisible increments. The topic is discussed in Feller 1950 (Vol. II, p. 347). When the compounding function is Markovian, so is the compounded function.

Self-affine subordination and the fractal devil staircases. When the directing function is self-affine, it must be an L-stable non-decreasing function, sometimes called "stable subordinator." This is a non-decreasing function of trading time whose graph is an inverse Lévy devil staircase, the latter being a Cantor devil staircase made random.

M 1982F{*FGN*} discusses Lévy staircases in Chapter 31 and illustrates them on Plate 286 and 287. It discusses Cantor staircases and illustrates one in Plate 83 of Chapter 8. The term "staircase" is motivated by the presence of flat steps. The steps are infinitely numerous, and most are infinitesimally small. Between its steps, a fractal staircase moves up by infinitesimal amounts. The values of θ where steps end form a "Cantor dust" or a "Lévy dust." The latter is fully characterized by a single exponent α which is a fractal dimension. Conversely, trading time followed as function of physical time, reduces to a series of jumps of widely varying size. The idea of subordination is that, a fleeting instant of clock time allows trading time to change by a positive amount, generating the price jumps to be considered in Section 3.7.

M 1977F proposed that a trading time ruled by a devil staircase be called a *fractal time.*

Subordination came to play an important role in many aspects of fractal geometry, therefore is discussed in detail in Chapter 32 of M 1982F{*FGN*}, where it is illustrated and interpreted in a variety of contexts.

3.7 Fractal compounding: LSM is identical to WBM, as followed in a trading time defined by a fractal devil staircase

A representation of L-stable motion. The original and simplest form of subordination was used in M & Taylor 1967{Sections 1 and 2 of E21}, to which the reader is referred. It takes price to be a Wiener Brownian motion of fractal trading time. The interesting fact is that the procedure happens to reproduce exactly the L-stable process that M 1963b{E14} proposed for the Noah Effect. The exponent α is "fed in" by the Lévy staircase.

A generalization that calls for detailed exploration: fractional Brownian motion of fractal time. This obvious generalization has two parameters: the α exponent of the Lévy staircase, which is a fractal dimension, and the exponent of the compounding function. $B_H(t)$, the Hölder exponent of the observed process, depends on α and H as we shall see in Section 3.9.

As mentioned in Section 6 of Chapter E1 and Section 3 and Annotations in Chapter E21, Clark 1973 proposed to preserve subordination, while replacing fractal time by a lognormal time, which is non-fractal. M 1973c{E21, Section 3} argued against Clark's substitute. But I never implied that the M 1963 model, as restated in M & Taylor 1967{E21}, said the last word, quite to the contrary. However, instead of "patching up" the subordinator, I propose to replace subordination itself by a suitable more general form of compounding.

3.8 A form of continuous compounding, called multifractal, and a form of variability driven by tail and serial dependence acting together

A direct introduction of dependence into LSM had proven difficult, but compounding beyond subordination opened the gates to diverse possibilities, to which we now proceed. Observe that LSM, FBM and subordination were part of the mathematical literature, but what follows is new, even from the mathematical viewpoint.

Multifractality. The key step in moving beyond subordination consists in changing trading time from fractal to a more richly structured (and more complicated) form called *multifractal.* This step is explained in Chapter ix of M 1975o and in a section on "relative intermittency" on p.

375 of M 1982F{*FGN*}: both argue that many patterns that seem fractal in a first approximation prove on a second look to be multifractal. This step is now taken near-automatically in many fields. It was first taken in M 1969b, a paper concerned with turbulence, and my first full publication in that field, M 1972j{N14} ends (p. 345 of the original) as follows:

"The interplay ... between multiplicative perturbations and the lognormal and [scaling] distributions has incidental applications in other fields of science where very skew probability distributions are encountered, notably in economics. Having mentioned the fact, I shall leave its elaboration to a more appropriate occasion."

Multifractal measures and functions. The concept introduced in M 1972j{N14} and developed in 1974f{N15} and M 1974c{N16} involves nondecreasing multifractal random functions with an infinite number of parameters. Their increments are called *multifractal measures.* The original example introduced in M 1972j{N14} is the "limit lognormal multifractal measure;" it remains after all those years the main example that is "homogeneous" in time. Most explicitly constructed multifractals are grid-bound "cartoons;" they are defined and studied in Section 4. (In the same vein, the main example of fractal trading time with strong homogeneity remains the Lévy staircase used in Section 3.6. Figure 4 of Chapter E1 is a plot of the measures contained within successive intervals of the abscissa, and was originally simulated on a computer in order to model the gustiness of the wind and other aspects of the intermittency of turbulence. But the resulting pattern reminded me instantly of something entirely different, namely, Figure 1 of M 1967j{E15} which represents the variance of cotton price increments over successive time spans.

The limit lognormal multifractal measures are *singular*, and the same is true of all the examples invoked in the early literature – but not of some more recent ones. Being "singular", the integral $M(t)$ of the plot in Figure 5 of Chapter E1 is monotone increasing and continuous, yet non-differentiable anywhere. There is no trace of the step-like intervals corresponding to vanishing variation that characterize the Cantor and Lévy devil staircases. My immediate thought in 1972 was to use this function $M(t)$ as graph of a multifractal trading time $\theta(t)$. In the simplest cases, the inverse function $t(\theta)$ is also multifractal. This thought was not elaborated until recently and is published for the first time here and in three papers by M, Fisher & Calvet in different permutations. Tests delayed for twenty-five years suggest that my 1972 hunch led to a surprisingly good approximation, as will be seen in Section 3.15. I heard rumors of other investigations of multifractals in finance; after all, once again, this is the next obvious step after fractals. But it is also an extremely delicate one.

A remarkable novelty: Multifractals allow concentration to occur with or without actual discontinuity. The fact that the typical early functions $M(t)$ are continuous is linked to the subtitle of this book and the topic of Section 1.3 of Chapter E2. Indeed, WBM and FBM of multifractal time are capable of achieving an arbitrarily high level of concentration *without* the actual discontinuity that is characteristic of LSM.

As a matter of fact, LSM can be viewed as a limit case. If one looks very closely, this limit is atypical and the convergence to it is singular. But this book need not look close enough to be concerned.

3.9 Characterization of multiscaling: "tau" functions that describe the moments' behavior for the directing and the compound functions

Except for scale, FBM is characterized by one parameter, LSM by two, and the major properties of a self-similar fractal follow from one parameter, its fractal dimension. Multifractals are more complicated: the closer one investigates them, the larger the number of parameters. This is because multifractals are characterized by a plethora of scaling relations, with correspondingly many exponents. The list of principal exponents defines a function "tau" which will now be described in two forms. (While this function is fundamental, it does *not* uniquely describe a multifractal.)

The moment exponent function $\tau_D(q)$ of the directing function. In a multifractal measure, as first shown in M 1974f{N15} and M 1974c{N16}, the moments of ΔM typically take the form

$$E[(\Delta M)^q] = \Delta t^{\tau_D(q)+1}.$$

(*Digression.* Some readers may be surprised by the equality sign, because other writers define a function τ as a limit. The technical reason is that the original method I used to define multifractals focuses on "fixed points" for which equality prevails.)

Moment-based scaling exponents. The q-th root of the q-th moment is a scale factor. For multifractals,

$$\{E(\Delta M)^q\}^{1/q} = \Delta t^{\sigma_D(q)}, \text{ where } \sigma_D(q) = \frac{1+\tau_D(q)}{q}\ .$$

(*A warning.* The literature also uses the notation $D(q) = \tau(q)/(q-1)$; except in the unifractal case, $D(q) \neq \tau(q)$).

The uniscaling cases. When multifractal time reduces to clock time, $\tau_D(q)+1=q$, implying uniscaling, since $\sigma_D(q)$ is independent of q.

The multiscaling cases. In the cases to be considered in this chapter, $\tau_D(q)$ satisfies two conditions: a) $E[(\Delta M)^0]=1$, that is, $\tau_D(0)=-1$, and b) $E(\Delta M)=\Delta t$, that is, $\tau_D(1)=0$. However, the graph of $\tau_D(q)+1$ is not a straight line. It follows that $\sigma_D(q)$ decreases as $q \rightarrow \infty$.

Interpretation of the quantity $\tau'_D(1)=D_1$. This quantity has a very important concrete interpretation, as the fractal dimension of the set of values of $\theta(t)$ where the bulk of the variation of θ occurs. It is often denoted as D_1, and will be needed momentarily.

The power exponent function $\tau_C(q)=\tau_D(qH)$ *of the compound function.* Since $\Delta X = G(\Delta\theta)^H$, where G is a reduced Gaussian,

$$E[|\Delta X|^q]=E[|G|^q]E[(\Delta\theta)^{qH}]=(\text{a numerical constant})(\Delta t)^{1+\tau_D(qH)}.$$

This important new result defines an additional "tau" function, namely,

$$\tau_C(q)=\tau_D(qH).$$

"Multifractal formalism." This is the accepted term for the study of the functions $\tau(q)$ and associated functions customarily devoted by $f(\alpha)$. The latter are often called "singularity spectra," but they are best understood by generalizing to oscillating function, the original approach pioneered in M 1974c{N16}: they are limits of probability densities of ΔX, but plotted in a special way. The general idea can be inferred from the discussion in Section 8.4 of Chapter E1, where it is pointed out that linear transformation *cannot* collapse the densities, but *can* collapse the quantities $\rho_\sigma(u)$. Calvet, Fisher & M 1997 sketches the role of the function $f(\alpha)$ in the context of economics, and numerous chapters of M 1997N will fully describe my approach to multifractal measures and functions, and compare it to alternative approaches.

3.10 The FBM of multifractal time accounts for two facts about the tails that constitute "anomalies" with respect to the M 1963 model

It was mentioned repeatedly that reports came out very early that some price records disagree with the M 1963 model. Some authors report tails that follow the scaling distribution but with an exponent α that exceeds

Lévy's upper bound $\alpha = 2$. Other authors report distributions that fail to collapse when superposed with the proper scaling exponent.

We shall now show that multifractals provide a framework compatible with either or both observations.

The critical tail exponent q_{crit}. The equation $\tau_D(q) = 0$ has always the root $q = 1$. In addition, the function τ_D being cap convex, the equation $\tau_D(q) = 0$ may also have a finite second root; when it exists, it is denoted by q_{crit}. The limit lognormal case always yields $q_{\text{crit}} < \infty$. An important and suprising discovery is reported in M 1972j{N14}, M 1974{N15} and M 1974c{N16}: when a second root q_{crit} exists, the distribution of ΔM has an asymptotically scaling tail of the form

$$\Pr\{M > u\} \sim u^{-q_{\text{crit}}}.$$

Thus, from the viewpoint of the tail, q_{crit} is a "critical tail exponent." It plays the same role as the Lévy exponent α, namely, $E(\Delta M)^q < \infty$ if, and only if, $q < q_{\text{crit}}$. The essential novelty is that the range of q_{crit} is *no longer* $0 < q_{\text{crit}} < 2$; instead it becomes $1 < q_{\text{crit}} < \infty$.

A way to obtain a tail exponent of price change that exceeds the upper bound 2 that is characteristic of L-stability. Now return to compounding, namely to a fractional Brownian function $B_H(t)$ of a limit lognormal trading time. Its increments will satisfy $E(\Delta B_H)^q < \infty$ if, and only if, $q < q_{\text{crit}}/H = \alpha$. In the Brownian case $H = 1/2$, q_{crit} can range over $[1, \infty]$, hence α can range over $[2, \infty]$, which conveniently extends the L-stable range $[1, 2]$ of α. Furthermore, choosing H in the range $[1/2, 1]$ extends the range of α to $[1, \infty]$, which is the maximum conceivable in the case where expectations are finite.

Nevertheless, q_{crit} need not exist, that is, a multifractal ΔM *need not* have a scaling tail. This may sound confusing, but only means that not every property of every multifractal is scaling.

It is nice that multifractal trading time makes it possible to extend the range of the asymptotic exponent $\alpha > 2$, but this result is not achieved without major changes. Indeed, the multifractal increments ΔM are *not* scaling in the sense that applies to the L-stable variables. They have more than one characteristic exponent, hence a structure, called *multiscaling,* that is far richer and has many distinct aspects.

Multiscaling implies that the tails of the compound process become increasingly shorter as T increases. This is because the scale factors $[E[\Delta X]^q]^{1/q}$ are scaling and their exponent $\sigma_D(q) = [1 + q_D(\tau H)]/q$ decreases as $q \to \infty$. To illustrate the importance of this fact, consider the renormalized increments $\Delta X/[E\Delta X^2]^{1/2}$. Contrary to the L-stable increments of the M 1963 model, those multiscaling increments do *not* collapse; instead, their distributions' tails become shorter and shorter as T increases. In other words, the multifractality of trading time is a sufficient explanation of the two anomalies described in Section 3.8.

3.11 Fourier spectral properties before and after rectification

The spectral exponent B_C of the increments of the compound process. The behavior of $E[\Delta X^2]$ for $\Delta t \to 0$ determines the behavior for $f \to \infty$ of the spectral density of the increments of X. That density takes the "$1/f$" form:

$$\text{spectral density} \sim f^{-B_C}, \quad \text{where } B_C = \tau_C(2) = \tau_D(2H).$$

The WBM case, $H = 1/2$, yields $B_C = \tau_D(1) = 0$, as we already know from Section 3.5. When $H \neq 1/2$ but is close to $1/2$, we have

$$B_C = \tau(2H) \sim \tau_D(1) + (2H-1)\tau'_D(1) = (2H-1)\tau'(1) = (2H-1)D_1.$$

Conclusion. In the white case $H = 1/2$, we encounter once again the very important blind spot of spectral analysis noted in Section 3.5. For $H \neq 1/2$, compounding changes the spectral exponent. However, the nearly white cases exhibit an extraordinary and very welcome simplification: the exponent B_C of the compound process "separates" into a product. In the case $D_1 = 1$, which corresponds to FBM in clock time, it is confirmed that the spectral exponent and sole parameter of the increments of the compounding function is $(2H-1)$. As to the directing function, it is not represented by its full function $\tau_D(q)$, only by a single parameter, the dimension D_1. Additional structural details of the directing function, which may be complicated, do not matter.

Value of the spectral exponent, after the increments of the compound process have been "rectified", in the sense of having their absolute values raised to the power $1/H$. Electrical engineers and applied physicists know (more accurately perhaps, used to know) that to understand a "noise," it is good to study it in two steps at least: first in its natural scale, then after it has been "rectified," which usually means taking the absolute value or squaring. This approach motivated the tests carried out in Voss 1992 and mentioned at the end of Section 3.5, and perhaps also the tests in Olsen 1996. In the present context, let us show that a particularly appropriate rectification consists in "taking the power $1/H$." When $H=1/2$, this reduces to squaring.

Indeed, take two non-overlapping intervals of duration Δt separated by a time span T, and form the "covariance" of the compounding increments $(\Delta' X)^{1/H}$and $(\Delta'' t)^{1/H}$. We know that G' and G'' are independent Gaussian variables $\Delta' X = G'(\Delta'\theta)^H$ and $\Delta'' X = G''(\Delta''\theta)^H$. Hence,

$$E\{(\Delta' X)^{1/H}(\Delta''(X))^{1/H}\} = E(G')^{1/H}E(G'')^{1/H}E(\Delta'\theta\Delta''\theta).$$

The numerical prefactor $E(G')^{1/H}E(G'')^{1/H}$ depends on H, but otherwise this last expression solely reflects the properties of the directing function. The covariance and the spectral density of ΔX can be shown to be proportional, respectively, to $s^{-\tau_D(2)}$ and $f^{-1-\tau_D(2)}$. For this reason, $\tau_D(2)$ acquired the strange name of "correlation dimension."

Summary. In the WBM case $H=1/2$, the appropriate rectification boils down to $(\Delta X)^2$. In the FBM case where $H \neq 1/2$ but H is close to $1/2$, one needs corrective factors, but reporting them here would delay us too much.

The spectrum reflects the form of dependence, but only in a limited fashion; it is distinct from, and only distantly related to, the features of $\tau_D(q)$ that affect the shape of the tails. A striking feature of the multifractals is this: scaling may, but need not, be present in the tails, but is always present in the dependence. A Brownian or fractional Brownian function of a multifractal trading time follows the same scaling rule of long-run statistical dependence as found in fractional Brownian motion.

3.12 The notions of partition function, or q-variation, for the directing multifractal time and the compound process.

Take the length of the available sample as time unit, divide it into non-overlapping intervals of lengths $\Delta_i t$, and consider the expression

$$\chi_D(q) = \sum |\Delta_i X|^q.$$

To statisticians, this is a non-normalized "sample estimate" of the moment $\Sigma|\Delta_i X|^q$. To physicists who follow a thermodynamical analogy, $\chi_D(q)$ is a "partition function." To mathematicians who follow N. Wiener, $\chi_D(q)$ is a "q-variation." Extraneous difficulties are avoided by choosing the unit of X so that $\Delta_i X < 1$ for all i.

Unequal $\Delta_i t$. For some purposes, as when we compare the q-variations taken along several alternative "times," it is important to allow the $\Delta_i t$ to be unequal. One takes the infimum of $\chi_D(q)$ for all subdivisions such that $\Delta_i t < \Delta t$, then one lets $\Delta t \rightarrow 0$. The values of q such that $\chi_D(q) \rightarrow 0$ and $\chi_D(q) \rightarrow \infty$, respectively, are separated by a critical value that will be denoted as $1/H_T$.

A very important observation concerning the contribution of the discontinuities to the value of χ. The case of direct interest is when $H_T < 1$. If so, χ_D can be divided into the contribution of the discontinuities and the rest. In the limit, the discontinuities contribute 0 if $q > 1$, therefore if $q > 1/H_T$. As a result, it makes no difference whether or not the discontinuities are included.

Equal $\Delta_i t = \Delta t$. For other purposes, however, one assumes that the $\Delta_i t$ are equal to Δt. This makes it possible to follow $\chi(q)$ as function of Δt, and one finds

$$\chi_D(q, \Delta t) = (\Delta t)^{\tau_D(q)}, \text{ with } \tau_D(q) = \log \Delta\chi(q, dt) / \log \Delta t.$$

The same argument can be carried out when the increments of trading time are replaced by the increments of the compounded process. It yields a new partition function

$$\chi_C(q, \Delta t) = \Delta t^{\tau_C(q)}.$$

3.13 A record's trail and graph have different fractal dimensions

This topic is best approached by a roundabout path.

The special case of FBM in clock time. The function $\tau(q)$ is associated with multifractals, but can also be evaluated for $B_H(t)$. Its value is found to be yielding $X_d(q, \Delta t) = (\Delta t)^{Hq-1}$, hence $\tau(q) = Hq - 1$.

We know from Section 3.4 that the trail dimension is $D_T = 1/H$ with or without compounding. Now let us sketch a standard argument **BUG** that begins with the fact that $\tau(1) = H - 1$, and concludes for the graph dimension with the value $D_G = 1 - \tau(1) = 2 - H$. This argument consists in covering the graph with square boxes of side Δt. Each Δt and the corresponding Δx contribute a stack of $|\Delta x| / \Delta t$ boxes. (Actually, one needs the smallest integer greater than the ratio $|\Delta x| / \Delta t$, but this ratio is $\sim (\Delta t)^{-1/2}$, hence is large when Δt is small.) Denote by $N(\Delta t)$ the total number of boxes in all the stacks and by D_G the box dimension. One has

$$N(\Delta t) = \sum |\Delta x| / \Delta t = (\Delta t)^{\tau(1)-1}, \text{ hence } D_G = \frac{\log N(\Delta q)}{\log(1/\Delta q)} = 1 - \tau(1) = 2 - H.$$

The general case of FBM of a multifractal trading time. The values obtained for $\tau(1)$ and D_G are specific to FBM, but the bulk of the preceding argument is of wider applicability. The total number of boxes of side Δt needed to cover the graph is $\sim \sum |\Delta x| / \Delta t = (\Delta t)^{\tau_C(1)-1}$. Taking a ratio of logarithms, this heuristic argument yields for the dimension of the graph of $X(t)$ the value

$$D_G = 1 - \tau_C(1) = 1 - \tau_D(qH).$$

From $\tau_C(1) < 0$, it follows that $D_G \geq 1$, as is the case for every curve, hence for every graph of a function.

Under multifractal compounding, there is no *functional relation between* D_T and D_G. The unifractal functions FBM are specified by a single parameter H, hence the values of D_T and D_G are necessarily functionally related. Indeed,

$$H = \frac{1}{D_T} = 2 - D_G.$$

A compound process is more complicated, since its specification includes both H and the function $\tau_D(q)$. Hence the values of D_T and D_G *cease to be* functionally related. The best one can say is that an inequality established in M 1974f{N14} implies $D_G < D_T = 1/H$; in fact, $D_G \leq 2 - H$, which we know to be the value relative to FBM.

3.14 Statistical estimation for multifractals, beginning with H, and continuing with the $\tau_C(q)$ function of the multifractal time

The preceding title includes two statistical problems. The good news is that they can be faced separately. This is so because the asymptotic behavior of $\chi(q)$ has the remarkable property of separating the properties of the compounding function $X(\theta)$ from those of the directing function $\theta(t)$.

The estimation of H. It suffices to identify the value of q for which $\tau_C(qH) = 0$. Actually, H can be defined without injecting equal Δt's and the resulting function $\tau(q)$. Indeed, $1/H$ is a "critical value" of the exponent > 1 such that $\chi(q) \to 0$ for $q > 1/H$ and $\chi(q) \to \infty$ for $q < 1/H$.

The estimation of the directing function, once H is known. It suffices to plug H into τ_C to obtain τ_D. When one only wishes to obtain D_1, one can estimate the spectral exponent B and write $D_1 = B/(2H-1)$. Unfortunately, this is the ratio of two factors that may be small simultaneously, therefore, is not very reliable.

The special case of WBM or FBM in clock time. The critical value is $1/H$. Consequently, the behavior of $\chi_C(q)$ suggests a new method of estimating H, to be added to the standard correlation or spectral analysis and the less standard R (range) or R/S methods (see M 1997H).

The case when the trading time θ is multifractal and a continuous function of the clock time t. Once again, the test of whether $\chi(q) \to 0$ or $\chi(q) \to \infty$ does not require the $\Delta_i t$ to be identical, only that they all tend to 0. When $\theta(t)$ is a continuous function, the same critical value φ is obtained by using uniform intervals of θ and uniform intervals of t. Uniform intervals of θ bring us back to the compounding FBM function $B_H(t)$, but trading time is not observable directly, and investigation of actual samples imposes uniform intervals of t.

The discordant case of $B_\sigma\{\theta(t)\}$, when the trading time θ is a discontinuous function of the clock time. This case occurs in the M 1967 representation of the M 1963 model, when compounding reduces to subordination. In some way, it is the limit of the case of continuous directing functions. However, this limit is extremely atypical, the reason being that the Δt can be made increasingly small, but not the $\Delta\theta$. The illuminating behavior of $\chi(q)$ when the $\Delta\theta$ are equal and tend to 0 is inaccessible and not reflected in the behavior of $\chi(q)$ when the Δt are equal and tend to 0.

In particular, recall that the L-stable process of exponent α is the WBM of a fractal time and is twice the exponent of the Lévy devil staircase. In this case, the correct value $H = 1/2$ is *not* revealed by the critical exponent of $\chi(q)$ evaluated with constant Δt. (Digression: this subtle point can be better understood be examining Plate 298 of M 1982F{*FGN*}.)

The converse problems. Now suppose the preceding statistical analysis is carried out on a process that is *not* a FBM of a multifractal trading. The q-variation exponent is defined for *every* function, therefore the algorithm to estimate H_C yields a value in every case.

3.15 The experimental evidence

As mentioned in the Preface, empirical testing of the M 1972 model was slow and could not be as broad and complete as I wished. But we studied the changes in the dollar/deutschmark and other foreign exchange rates obtained from Olsen Associates in Zürich; the results, which are extremely promising, will be published in three papers by M, Fisher and Calvet in different permutations. The principal figure of Fisher, Calvet & M 1997 is reproduced here as Figure 3. It is a log-log plot of the variation of $\chi_D(q, \Delta t)$ as function of Δt, for several values of q close to 2. Two distinct datasets were matched, namely, daily and high frequency data.

The first observation is that the diagrams are remarkably straight, as postulated by multifactality.

The scaling range is very broad, three and a half decades wide, from Δt of the order of the hour to Δt of more than a hundred days (at least.)

The second observation concerns the value of q for which this graph is horizontal, meaning that $\tau_D(q) = 0$. This value of q defines the trail dimension D_T, and the data show that it is close to the Wiener Brownian value $D_T = 2$. This value was implied when Voss 1992 and Olsen 1996 described the spectrum of the rate changes as being white.

At closer look, however, D_T seems a bit smaller than 2, suggesting $H_T > 1/2$. If confirmed, this inequality would be a token of persistent fractional Borwnian motion in multifractal time.

Increments Δt below one hour seem to exhibit a different scaling, with D_T clearly different from 2. Once again, full detail is to be found in Fisher, Calvet & M 1997.

3.16 Possible directions for future work

A major limitation of the fractional Brownian motion of time was acknowledged in Section 9 of Chapter E1: the resulting marginal distributions are symmetric. A possible way out was also referenced, namely, the "fractal sums of pulses."

This section's context instantly suggests an alternative way out: to replace $B_H(t)$ by an asymetric form of the L-stable process that underlies the M 1963 model. The presence of two parameters (an exponent α and a skewness parameter β) can only help improve the fit of the data. But the resulting process remains unexplored and may prove to be unmanageable.

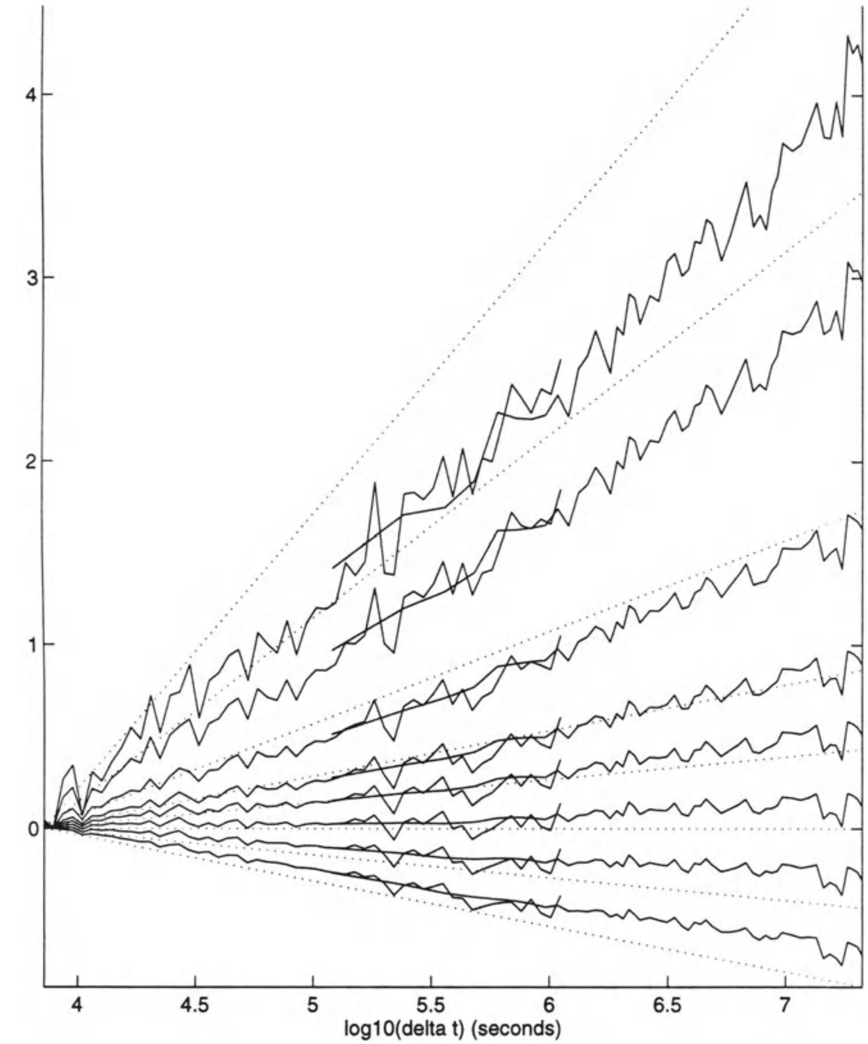


FIGURE E6-3. Doubly logarithmic plot of $\chi_D(q, \Delta t)$, as function of Δt in the case of the Olsen data for the US dollar/Deutschmark exchange rate. The main observations are a) the fact that the plots are straight from Δt of the order of one hour to the end of the data, which corresponds to Δt of more than a hundred days; the slopes of the plots define the function $\tau_D(q)$; b) the fact that the value of $q = D_T$ for which $\tau_D(q) = 0$ is close to 2.

Observation a) is a symptom of multifractality and observation b) is a symptom that the process is close to being a Wiener Brownian motion that is followed in multifractal time. The true value of D_T is a bit smaller than 2, suggesting the inequality $H_T > 1/2$. If confirmed, this would be a token of persistent fractal Brownian motion in multifractal time.

4. DIAGONAL-AXIAL SELF-AFFINE CARTOON SURROGATES

This section covers roughly the same material as Section 3, but in entirely different style. Section 3 concerned constructions that first arose in limit theorems of probability theory (some of them classical and/or difficult). Those constructions are of great independent interest and of "universal" rather than arbitrary character. But their origins are not coordinated, so that, they did not fit comfortably together. To the contrary, this Section proceeds in tightly coordinated fashion and the parts fit well together. The new feature is that those parts are non-universal and to a large extent arbitrary. Their baroque wealth of structure is a loss from the viewpoint of simplicity and esthetics; but it may be a gain from the viewpoint of apprehending the baroque wealth of structure found in nature.

Let me elaborate. By abundantly illustrating self-similarity, M 1982F{*FGN*}, demonstrated that the principle of recursive construction exemplified in Section 2 is very versatile. That is, it is not sharply restrictive but leaves room for many varied implementations. To an even larger extent, self-affinity is versatile almost to excess, hence insufficient by itself for any concrete purpose in science. The goal of this section is to transform the 1900, M 1963, M 1967 and M 1972 models of price variation into constructions that fit together as special examples in a broader, well-organized but diverse collection. The implementation of this goal is distantly inspired by a construction due to Bernard Bolzano (1781-1848). In a terminology that may be familiar to some readers, this implementation is "multiplicative." The more familiar "additive" constructions (patterned on the non-differentiable functions due to K. Wierstrass) proved to be of insufficient versatility.

4.1 Grid-bound versus grid-free, and fractal versus random constructions

The role of grids in providing simplified surrogates. Fractal construction are simplest when they proceed within a grid. Grids are not part of either physics or economics. But suitable grid-based constructs can act as "surrogates" to the grid-free random process, like the 1900, M 1963, M 1965, M 1967, and M 1972 models. When this is possible, the study is easier when carried out on the grid-based cartoons. Besides, the cartoons in this chapter fit as special cases of an overall "master structure" which relates them to one another, is enlightening and is "creative" in that it suggests a stream of additional variants. I came to rely increasingly on this master structure in the search for additional models to be tried out for new or old problems. Striking parallelisms were mysterious when first observed in the grid-free context, but became natural and obvious in this master structure.

Fractality versus randomness from the viewpoint of variability. The main cartoon constructions in this chapter are non-random. To the contrary, the 1900, M 1963, M 1965, M 1967 and M 1972 models in Section 3 are random, for reasons explained in Section 1 of Chapter E1.

However, an important lesson emerges from near every study of random fractals. At the stage when intuition is being trained, and even beyond that stage, *being random is for many purposes less significant than being fractal.* It helps if the rules of construction are not overly conspicuous, for example, if no two intervals in the generator are of equal length. That is, the non-random counterparts of random fractals exhibit analogous features, and also have the following useful virtue: they avoid, postpone, or otherwise mitigate some of the notorious difficulties inherent to randomness. Those non-random fractals for which acceptable randomizations are absent or limited, are also of high educational value.

Distinction between the contributions of Wiener and Khinchin. In the spirit of the preceding remarks, the readers acquainted with the Wiener-Khinchin theory of covariance and spectrum may recall that a single theory arose simultaneously from two sources: Wiener studied non-random but harmonizable functions, and Khinchin studied second-order stationary random functions. The two approaches yield identical formulas.

A drawback: grid-bound constructions tend to "look "creased" or "artificial." This drawback decreases at small cost in added complication, when the grid is preserved, but the construction is randomized to the limited extent of choosing the generator among several variants. This will be done in Figures 4 and 5. Nevertheless, the underlying grid is never totally erased. To a trained eye, it leaves continuing traces even after several stages of recursion. (A particularly visible "creasing" effect is present in computer-generalized fractal landscapes, when the algorithm is grid-based. Around 1984, this issue was a serious one in the back-offices of Hollywood involved in computer graphics.)

Addition versus multiplication. Specialists know that fractals are usually introduced through additive operations, and multifractals, through multiplicative operations. The reason for selecting the fractal examples that follow is that they can be viewed as either additive or multiplicative, making it unnecessary to change gears in the middle of the section.

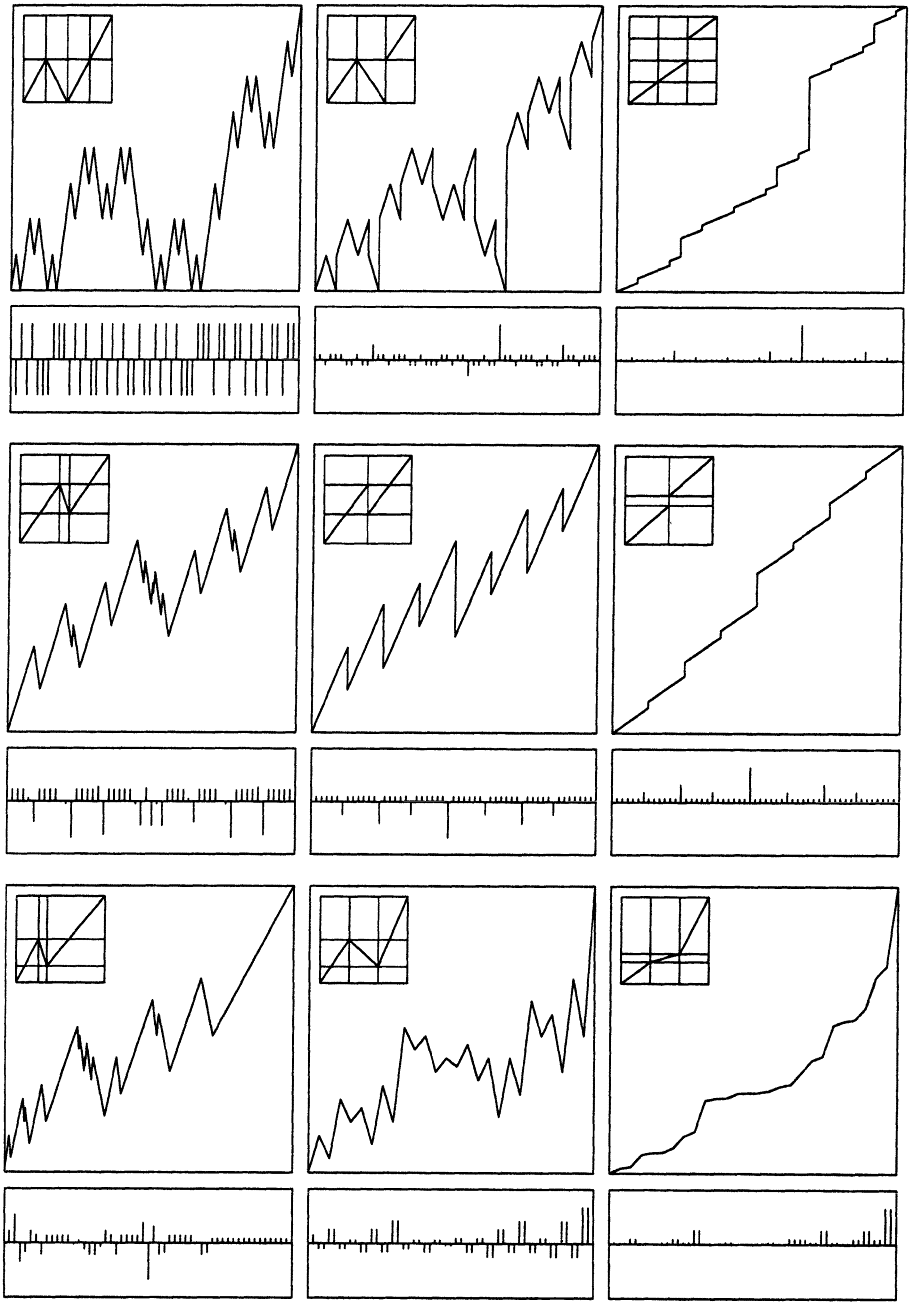


FIGURE E6-4. Six alternative cartoon constructions explained on the next page.

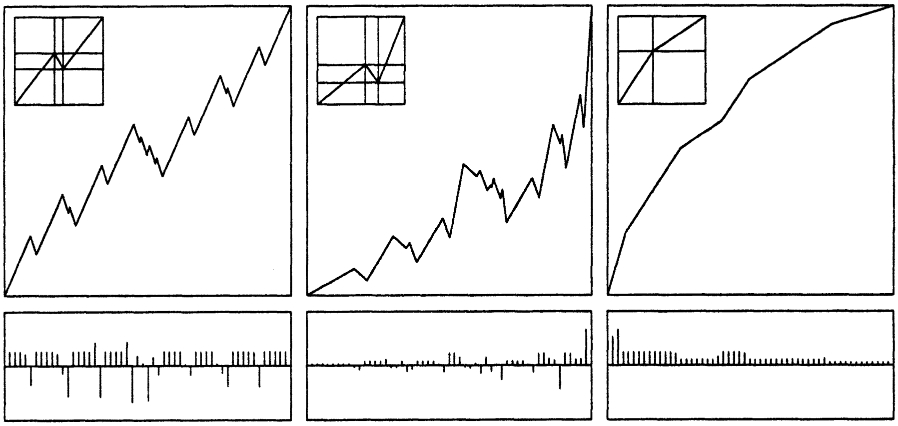


FIGURE E6-5. Three additional cartoons. The following explanation applies to this figure and the preceding one.

Each construction is illustrated by a square diagram and a longitudinal one. The generator is shown in a small window within the square diagram: it is either diagonal or diagonal-and-axial. The square diagram shows the level-2 approximation and the corresponding longitudinal diagram shows the increments of the level-2 approximation, taken over equal time increments.

The diagrams juxtaposed on a horizontal row in Figure 4 are intimately related, as described in Section 4.9.

The power and limitations of the eye. The eye discriminates better between records of *changes* of a function than between records of the function *itself.* Therefore, the results of many constructions that follow will be illustrated in both ways.

Recursiveness also allows many other possibilities that lie beyond the scope of this book: they include the "fractal sums of pulses" (M 1995n and several papers for which the co-author listed first is Cioczek-Georges).

4.2 Description of the grid-bound recursive constructions with prescribed initiator and generator

To start from scratch, each diagram in Figures 4 and 5 is drawn within a *level-0 box,* whose sides are parallel to the coordinate axes of t and x. The "initiator," is an ordered interval (an "arrow") acting as a "hidden string" that crosses the level-0 box from bottom left to top right. Therefore, it is useful to think of this and other boxes as "beads." The width and height of the level-0 box are chosen as units of t and x, making the box a square. (The question of what is meant by a *square* in the affine plane is a subtle issue, to be tackled below, after the definition of H.)

In addition, each diagram contains a *string generator* that joins the bottom left of the initiator to its top right. Alternative descriptions for it are "string of arrows," "broken line," and "continuous piecewise linear curve." The number of intervals in the generator, b, is called *"generator base"*. When the generator is increasing, $b \geq 2$; when the generator is oscillating, the lower bound becomes $b \geq 3$. The larger b becomes, the greater the arbitrariness of the construction. Hence, the illustrations in the chapters use the smallest acceptable values of b.

Axial and diagonal generator intervals. To insure that the recursive construction generates the graph of a function of time, the string generator must be the "filled-in graph" of a function $x = G(t)$, to be called *generator function.* To each t, the ordinary graph attaches a single value of x. To each t where $G(t)$ is discontinuous, the filled-in graph attaches a vertical oriented interval of values of x that spans the discontinuity. The resulting interval in the generator is called *axial.* (The general case, mentioned later but not used in this book, also allows for horizontal intervals.) A non-axial interval is called *diagonal,* and the rectangle that it crosses diagonally from left to right defines a *level-1 box.* In some cases the level-1 boxes can be superposed by translation or symmetry, in other cases they cannot.

Recursive construction of a self-affine curve joining bottom left to top right, using successive refinements within a prescribed self-affine grid. Step 0 is to draw the diagonal of the initiator. Step 1 is to replace the diagonal of the initiator by the filled-in graph of the generator function $G(t)$. Step 2 is to create a line broken twice, as follows. Each diagonal interval within the generator is broken by being replaced by a downsized form of the whole generator. To "downsize" means to reduce linearly in the horizontal and vertical directions. In the self-affine case, the two ratios are distinct. In some cases, one must also apply symmetries with respect to a coordinate axis. As to the generator's axial intervals, they are left alone. One may also say that they are downsized in the sense that the ratio of linear reduction in one direction is 0, collapsing the generator into an interval.

The "prefractal" approximations of self-affine graphs can take one of two forms. They may consist of increasingly broken lines. Figures 6 and 7 take up important generators and draw corresponding approximations as boundaries between two domains, white and black. This graphic device brings out finer detail, and helps the eye learn to discriminate between the various possibilities. Alternatively, each diagonal interval may be replaced by a rectangular axial box of which it is the diagonal. If so, the prefractal approximation consists in nested "necklaces" made of increasingly fine boxes, linked by axial pieces of string.

As the recursive construction of an oscillating cartoon proceeds, its increments Δu over increasingly small intervals Δt tend to become symmetrically distributed. That is, the ratio of the numbers of positive and negative increments tends to 1. {Proof: After k stages, each increment is the product of k factors, each of the form sign $(\Delta_i x)$. But Π sign $(\Delta_i x)$ is > 0 if Σ sign $(\Delta_i x)$ is even, and is < 0 if Σ is odd. The distribution of Σ sign $(\Delta_i x)$ is binominal and smoothly varying, therefore even and odd values are equally frequent asymptotically.}

4.3 The H exponents of the boxes of the generator

This and the next sections show how the fundamental scaling exponent H of fractional Brownian motion splits into a number of significantly different aspects.

Diagonal boxes and their finite and positive H exponents. Given a diagonal box β_i of sides $\Delta_i t$ and $\Delta_i x$, an essential characteristic is

$$H_i = \frac{\log \Delta_i x}{\log \Delta_i t} = \frac{\text{log of the absolute height of the box } \beta_i}{\text{log of the width of the box } \beta_i}.$$

In other words,

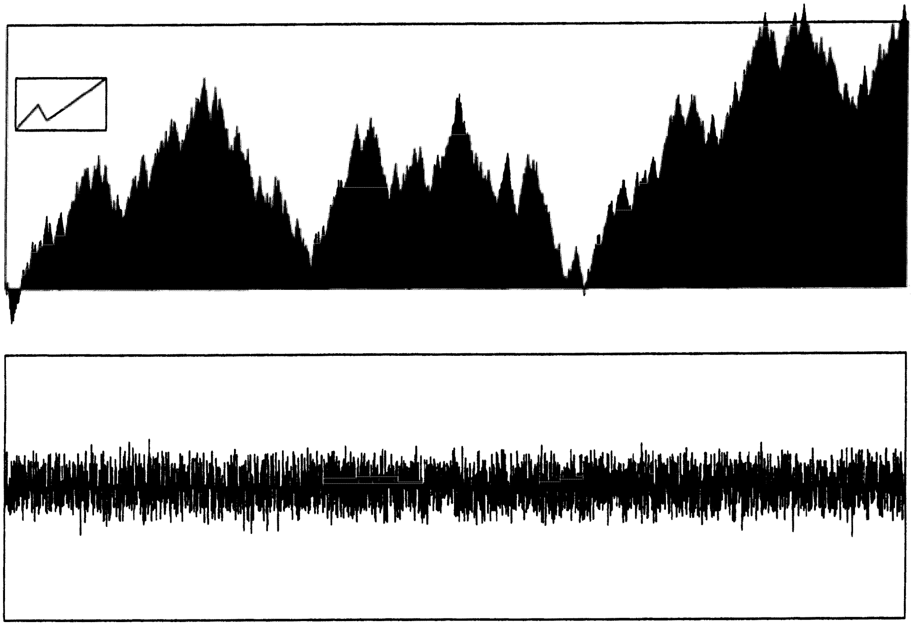


FIGURE E6-6. The top line illustrates a cartoon of Wiener Brownian motion carried to many recursion steps. The generator, shown in a small window, is identical to the generator A2 of Figure 2. At each step, the three intervals of the generator are shuffled at random; it follows that, after a few stages, no trace of a grid remains visible to the naked eye.

The second line shows the corresponding increments over successive small intervals of time. This is for all practical purposes a diagram of Gaussian "white noise" as shown in Figure 3 of Chapter E1.

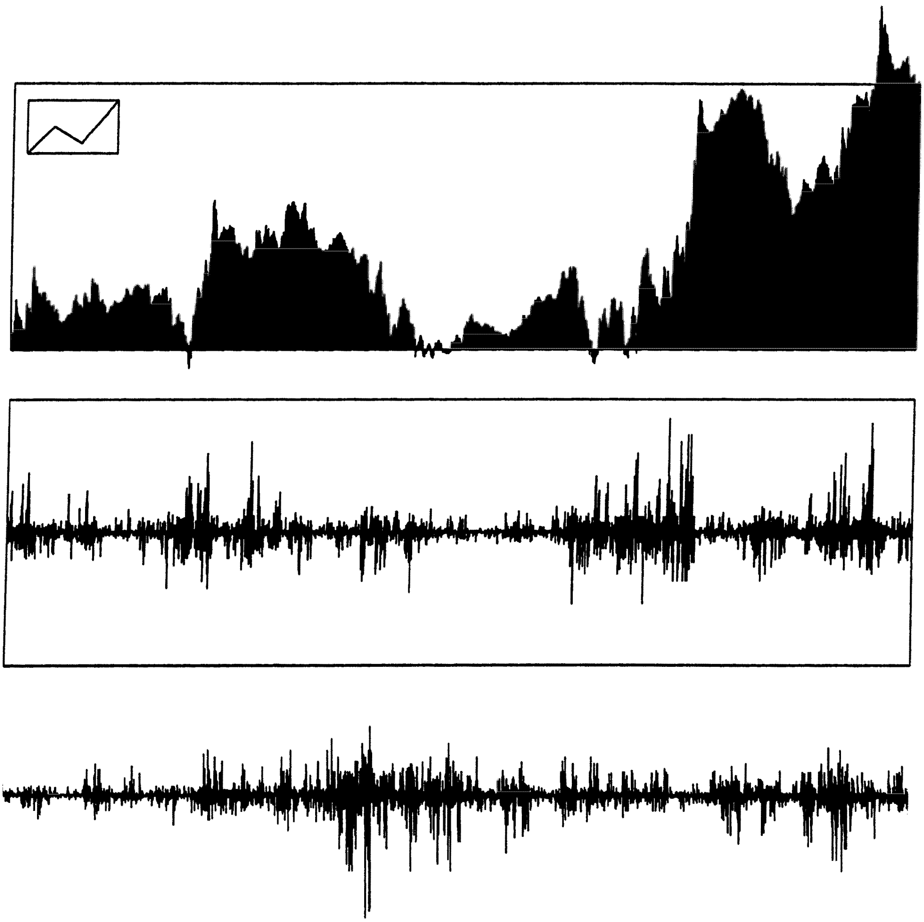


FIGURE E6-7. This figure reveals – at long last – the construction of Figure 2 of Chapter E1. The top line illustrates a cartoon of Wiener Brownian motion followed in a multifractal trading time. Starting with the three-box generator used in Figure 6, the box heights are preserved, so that D_T is left unchanged at $D_T = 2$ (a signature of Brownian motion), but the box widths are modified. (Unfortunately, the seed is not the same as in Figure 6.)

The middle line shows the corresponding increments. Very surprisingly, this sequence is a "white noise," but it is extremely far from being Gaussian. In fact, serial dependence is conspicuously high. The bottom line repeats the middle one, but with a different "pseudo-random" seed. The goal is to demonstrate once again the very high level of sample variability that is characteristic of wildly varying functions.

The resemblence to actual records exemplified by Figure 1 of Chapter E1 can be improved by "fine-tuning" the generator.

$$\Delta_i x = (\Delta_i t)^{H_i}.$$

This identity concerns non-infinitesimal boxes, therefore H_i is a *coarse* coefficient. Now proceed to the limit $\Delta t \to 0$. If $x(t)$ had a well-defined derivative x', one would have $\Delta x \sim x' \Delta t$. Therefore, differentiable functions yield $\alpha = 1$ (but the converse is not true).

Hurst, Hölder, and a way to conciliate mathematical and concrete needs. Section 3.2 mentions that H has roots in the works of both the hydrologist H. E. Hurst (1880-1978), and the mathematician L. O. Hölder (1859-1937). However, these concrete and mathematical contexts require a special effort before they fit comfortably together. For example, assume that all boxes of the same level are equal, with $\Delta t = b^{-1}$ and $\Delta x = b'^{-1}$ for level-1, therefore $\Delta t = b^{-k}$ and $\Delta x = b'^{-k}$ for level-k. It follows that $H_i = \log b' / \log b = H$ for all boxes at all levels; level 0 yields $\log 1/\log 1 = 0/0$, which can be interpreted as equal to H. However, if the level-0 box had sides 1 and B, all the level-k boxes would yield

$$H = \frac{\log b' + \log B/k}{\log b}.$$

In the pure mathematical interpretation due to Hölder, H is a local concept that concerns the limit $k \to \infty$. Its value is not affected by B. By contrast, the concrete interpretation of H that I pioneered do not concern local asymptotics but concrete facts, therefore applies uniformly to all sizes. If the resulting "coarse" H is to serve a purpose, it must be independent of all units of length; this is achieved by setting $B = 1$.

Axial intervals and the values H=0 and $H = \infty$. One may say that horizontal intervals yield $H = \infty$, and discontinuities yield $H = 0$. The value $H = \infty$ can occur almost everywhere, and the value $H = 0$ can occur at most on a denumerable set, therefore, on a set of dimension 0.

Comments on the examples on Figure 4. The columns are denoted by letters (A, B, and C) from the left, and the rows by numbers (1, 2, 3, 4) from the top. Each example will first be listed by itself, then different examples will be shown to be related to one another. Figure 5 carries further the constructions based on generations A3 and B3, adding randomization for increased realism.

4.4 Unifractality for cartoons, and selected major examples

This section proceeds beyond general statements on grid-bound cartoons. It introduces (by definitions and examples) some key distinctions that explain how those cartoons can serve as "surrogates" of the grid-free self-affine processes described in Section 3. When a process is denoted by XYZ, its cartoon surrogate will be denoted by C(XYZ).

Unifractality. Each box β_i of the generator has its own exponent H_i. If all the H_i are identical, > 0 and $< \infty$, the cartoon construction will be called *unifractal.* The conditions $H > 0$ and $H < \infty$ exclude axial generator intervals. This case is very special, but of fundamental importance, because it includes cartoons of WBM and FBM.

Unibox versus multibox constructions. In the unifractal case, the generator boxes can be either identical, defining the *unibox* case or not, defining the *multibox* case. All unibox constructions are unifractal. Many of their properties depend only on H, but other properties depend on the boxes' two sides, and some properties also depend on the details of the arrangement of the boxes. Multibox constructions depend on a larger number of parameters; they are less regular, hence less "artificial-looking," therefore their fractality is a better surrogate for randomness.

- **C_1(WBM).** (Generator A1). This cartoon of base $b = 4$ is a unibox (hence unifractal) surrogate for Wiener Brownian Motion. It has become widely used in physics (M 1986l, M 1986t, Family & Vicsek 1991). To find where it comes from, consider the Peano-Cesaro motion illustrated (in approximation) on Plate 65 of M 1982F{*FGN*}. Follow this motion as it proceeds from the lower left to the upper right corner, moving through the upper left corner. The projection of this motion on the x-axis will have A1 as its generator. Variants are described in M 1986t{H}, and in various articles collected in M 1997N and M 1997H.

- **C_2(WBM).** (Generator A2). The generator of this multibox cartoon contains $b = 3$ intervals, which is the smallest value that allows oscillations. Denoting the side intervals by x, the middle interval is of height $2x - 1$. Since $H = 1/2$ for WBM, the generating identity becomes $2x^2 + (2x - 1)^2 = 1$, yielding $x = 2/3$.

- **C_3(WBM).** (Generator A3). The novelty is that all three intervals were made unequal, to add realism to the construction. A form of it is carried over many stages in Figure 6.

- **C(FBM).** (Generator A4). Cartoon C_2 (WBM) is readily generalized to $H \neq 1/2$. It suffices to take for x the positive root x_0 of the equation $2x/H + (2x - 1)/H = 1$. For $H > 1/2$, $x_0 < 2/3$; for $H < 1/2$, $x_0 > 2/3$.

Two steps beyond unifractality . The usual contrast to unifractality is provided immediately by *multifractality*, but the present context makes it necessary to single out an intermediate case that did not, until now, warrant a special name. Since "in between" is denoted by the Greek root *meso*, this case will be called *mesofractal*, a word that is used here for the first time.

4.5 Mesofractality for cartoons, and selected major examples

Mesofractality. This term will denote cartoons whose generator includes vertical intervals with $H=0$, in addition to diagonal intervals sharing a unique H satisfying $0<H<\infty$. (The definition extends painlessly to allow horizontal intervals with $H=\infty$; such cases are not needed in this book, but will be discussed in detail in the introductory material of M 1997N.)

Like unifractality, mesofractality is a special case, but it too is of fundamental importance because it characterizes the several important cartoons that follow. The first and the second can be described as surrogates of LSM, even though the first only includes negative jumps. The third is a surrogate of fractal trading time, a notion to be defined in Section 4.5.

- **C_1(LSM).** (Generator B2). Begin with C_2(WBM) and modify the generator's boxes by the following transformation: keep the heights constant, expand the first and third box to be of width 1/2, and reduce the second box to be of width 0, hence $H_2=0$. All the jumps are negative.

- **C_2(LSM).** (Generator B1). A more realistic surrogate of LSM must have both positive and negative jumps. To achieve this goal, it is necessary to use a generator containing at least $b=4$ intervals. Begin with C_1(WBM), and modify the generator's boxes by the following transformation: keep the heights constant, expand the first, second and fourth box to be of width 1/3, and reduce the third box to be of width 0. The new H values are $H_1=H_2=\log_3 2$, $H_3=0$, and $H_4=\log_3 2$.

- **C_1(FTT) and C_2(FTT).** (Generators C1 and C2). These are inverse functions of variants of the classical devil staircase.

4.6 Multifractality for cartoons, and selected major examples

Multifractality. The most general category of cartoons allows the generator to include diagonal boxes with different values of H_i, ranging from $H_{min}>0$ to a maximum satisfying $0<H_{max}<\infty$. Those cartoons are necessarily multibox. Boxes created at the k-th stage of recursion are characterized by H distributed over the interval $[H_{min}, H_{max}]$ in increasingly tight fashion. The limit of this distribution is described by the multifracatal formalism $f(\alpha)$ mentioned at the end of Section 3.9.

As defined above, multifractality allows some generator intervals to be axial, hence includes cartoons that combine continuous variation with jumps. However, jumps are absent, both from the bulk of the abundant literature on multifractals, and from the single grid-free multifractals of Section 3, like those in M 1972j{N14}, M 1974f{N15} and M 1974c{N16}. They will be discussed in M 1997N and M 1997H, but not in this book.

- **C_1(MFM).** (Generator B3). This example of oscillatory multifractal motion is a much simplified version of a construction due to Bernard Bolzano (1781-1848). Begin with C_3(WBM thru FBM), and modify the generator's boxes by the following transformations: keep the heights constant and change the width to 1/3. (These linear transformations are invertible, therefore called *affinities*. The linear transformations used to define C_1(LSM) and C_2(LSM) cannot be inverted.)

- **C_2(MFM).** (Generator B4). A three interval oscillating generator was chosen haphazardly. A form of this case is carried over to many stages in Figure 7.

- **C_1(MTT).** (Generator C4). A point was chosen in the unit square, and joined to the lower left and upper right corners by a non-decreasing broken line.

- **C_2(MTT).** (Generator C3). This is a three-interval generator yielding a multifractal trading time.

4.7 "First step towards a compound cartoon" representation of a general cartoon: definition of the trail dimension D_T and the trail exponent $H_T = 1/D_T$; spectral density of the form f^{-B} where $B = 1 - 2/D_T$

The major examples in Sections 4.4 to 4.6 include cartoons of WBM, LSM and FBM, which concern M 1900, M 1963 and M 1965 models, and other grid-bound self-affine functions that combine long tails and long memory. The generating functions are of great diversity, and innumerable additional examples immediately come to mind. Their very multiplicity might have been a source of disorder and confusion.

Fortunately, it is not, thanks to a very strong result that will be established in Section 4: every oscillating cartoon construction can be rephrased as a compound function, namely in the form of C(WBM) or C(FBM) as followed in suitable "multifractal" trading time that is a monotone non-decreasing function of clock time.

Recall that the grid-free constructions sketched in Section 3 consist in unifractal Wiener or fractional Brownian motions in a trading time that is either linear, or fractal, or multifractal. That collection of models grew step-by-step from 1963 to 1972, and lacked intrinsic cohesion or legitimacy, except as a rather abstract companion of the functions $\tau_C(q)$ and $\tau_D(q)$.

The grid-based cartoons are much simpler in that respect. In their case, most remarkably, the notion of trading time is intrinsic, compelling and inevitable. We proceed by steps, and begin by defining D_T and $H_T = 1/D_T$.

A simple identity that characterizes unifractal generators. In all cartoons, the box widths satisfy $\sum \Delta_i t = 1$. In the unifractal case where $H_i = H$ for all i, *define* D_T as $1/H$. It follows that the box heights satisfy

$$\sum |\Delta_i x|^{D_T} = 1, \text{ with } D_T = 1/H.$$

A simple identity that characterizes monotone multifractal generators. When $\Delta_i x > 0$ for all i, the equality $\sum \Delta_i t = 1$ trivially implies

$$\sum |\Delta_i x|^{D_T} = 1, \text{ with } D_T = 1.$$

The dimension-generating equation of a cartoon construction. This term will denote the equation $\sum |\Delta_i x|^{\sigma} = 1$, the unknown being σ. We know two cases already: in the unifractal case the only positive root is $\sigma = D_T = 1/H$, and in the monotone case the only positive root is $\sigma = 1$. We now proceed to the remaining possibility.

A new and highly significant concept: generalized values of D_T and $H_T = 1/D_T$, as defined for oscillating multifractal cartoons. In the oscillating multifractal case, the quantities H_i cease to be identical. The generating equation ceases to be a restatement of a mildly relevant identity. However, it remains meaningful and becomes highly significant. Its only positive root D_T satisfies $D_T > 1$ and is a *fundamentally* important characteristic of the construction.

Geometric interpretation of D_T by embedding, as the trail dimension of a closely related vectorial process. As we know from Section 2, a set is a self-similar fractal, when the whole is made of parts that are obtained from the whole by reductions. The generating equation for σ is formally identical to the classical Moran equation that gives the dimension of such a set, where the reduction ratios, $r_i = |\Delta_i x|$, are not equal, therefore, the simplest formula $D = \log N/\log(1/r)$, is inapplicable.

The procedure that gives substance to this analogy is embedding. It was already used when Sections 3.3 and 3.4 interpreted a scalar FBM $X(t)$ as a projection of a vectorial FBM of at least $1/H$ coordinates. The argument had to refer to a known theorem, that $1/H$ is the dimension of that vectorial FBM. In the present context, embedding is even simpler and requires no delicate reference.

The argument is simplest when $H = 1/2$ and $b = \max i \geq 4$. Consider, in a b-dimensional space, the point P of coordinates $\sqrt{b}\,\Delta_i x$. The squared distance from O to P is $\sum |\Delta_i x|^2 = 1$. Now consider projections on the main diagonal of our b-dimension space. The vector OP projects on an interval of length $1/\sqrt{b}$, and the vector of length $\Delta_i x$ along the i-th coordinate axis projects on an interval of length $\Delta_i x/\sqrt{b}$.

We are now ready to construct a self-similar curve in b-dimensional space, by taking OP as the initiator and the sequence of coordinate vectors of length $\Delta_i x$ as the generator. A classical theorem due to Moran tells us that the fractal dimension of that curve is the root D_T of the dimension-generating equation. This interprets D_T as a fractal dimension, with no reference to the values of the $\Delta_i t$. (The reason for postulating $b \geq 4$ is to some extent esthetic: in a space of $b > 3$ dimensions, one can obtain a spatial curve without double points.)

The preceding construction relies on the Pythagoras theorem, which is why the case $D_T = 2$ is the simplest possible, but the same result can be obtained for all $D_T > 1$.

Spectral density of the embedding vectorial motion. It can be shown to be of the " $1/f$ " form f^{-B}, with $B = 1 + 2H_T$ for the motion itself and $B = -1 + 2H_T$ for the "derivative" of the motion, which is a white noise when $H_T = 1/2$.

4.8 The graph dimension D_G of a cartoon; it is not functionally related to the trail dimension D_T

Having generalized D_T beyond the value $1/H$ relative to FBM, the next step is to generalize D_G beyond the corresponding value $2 - H$.

The special case where $\Delta_i t = 1/b$ for all i. The derivation of D_G for FBM was sketched in Section 3.13. The idea is to cover the graph with stacks of square boxes of side Δt. When $\Delta_i t = 1/b$ for all i , take square boxes of side b^{-k}. One defines $\tau_C(1)$ by writing

$$\chi(1, b^{-k}) = \sum |\Delta X| = (\sum b^{-H_i})^k = (b^{-k})^{\tau_C(1)}.$$

One can define H_G by writing

$$\chi(1, b^{-k}) = (b^{-k})^{1-H_G}, \text{ that is }, b^{-H_G} = \frac{1}{b} \sum b^{-H_i}.$$

For oscillating functions, $b \geq 3$, and the two exponents H_G and H_T are distinct functions of $b-1 \geq 2$ independent parameters H_i. Therefore, they are *not* linked by a functional relation.

The general case. Multifractal formalism. Section 3.9 defines the functions $\tau_C(q)$ and $\tau_D(q)$. The same definitions apply to the cartoon constructions. The details will be described in M 1997N.

4.9 Constructions of an intrinsic "compound cartoon" representation of a general cartoon; trading time is fractal for mesofractal cartoons and multifractal for multifractal cartoons

Before we complete the task of demonstrating that the cartoon constructions are surrogates of the compound functions examined in Section 3, one last step is needed. Examine the three sets of cartoons that are shown in Rows 1, 2 and 3 of Figure 2, and mark the coordinates as follows: θ and X in Column A, t and X in Column B and t and θ in Column C. Seen in this light, each cartoon in Column B is reinterpreted as a compound cartoon involving its neighbors in the same row. It is obtained from its neighbor in Column A, by replacing the clock time by the fractal or multifractal time defined by its neighbor in Column C. Let us now show here that such a representation can be achieved for every cartoon.

The intrinsic duration of an interval in the generator. We start with a recursive construction of fractal dimension as defined early in this section. To make it over into a cartoon of FBM with the exponent $H_T = 1/D_T$. we must apply the inverse of the linear transformation that led from C_1(WBM) to C_1(LSM) and from C(FBM) to C(MFM). Starting from an arbitrary generator box, the recipe is in two steps:

- keep the height $\Delta_i x$ constant,
- by definition of $\Delta_i \theta$, change the width from $\Delta_i t$ to $|\Delta_i x|^{D_T} = \Delta_i \theta$.

Intrinsic definition of a cartoon's trading time. We are now ready to take a last and basic step. We shall show that an oscillating, but otherwise arbitrary cartoon can be represented as a unifractal oscillating cartoon of exponent $H_T = 1/D_T$ of a multifractal (possibly fractal) trading time. In Section 3, the notion of trading time entered as a model of our historical and intuitive knowledge of competitive markets. Now, it enters through an inevitable mathematical representation.

The idea is to construct trading time using a cartoon generator defined by the quantities $\Delta_i t$ and $\Delta_i \theta$. Each recursion stage ends with an "approximate trading time that becomes increasingly "wrinkled" as the interpolation proceeds. The limit trading time may, but need not, involve discontinuity, but in all cases, its variation becomes increasingly concentrated in increasingly at each stage of interpolation, and in the limit manifests a high degree of concentration in very short periods of time. How this process builds up is a very delicate topic that cannot be discussed here in detail, but constitutes a core topic of M 1997N.

4.10 The experimental evidence

The visual resemblence between Figure 7 and Figure 1 of Chapter E1 deserves to be viewed as impressive, because of the extraordinary (in fact, seemingly "silly") simplicity of the underlying algorithm. However, multifractal analysis, using $\tau_c(q)$, as in Figure 1, or using the equivalent technique of $f(\alpha)$, shows that the resemblence is not complete. Indeed, the simulated data of Figure 5 yield slopes $\tau_C(q)$ that disagree with Figure 1. This is as it sould be: indeed, in order to simplify the construction to the maximum, Figure 5 uses a single generator, except that the three intervals are randomly shuffled at each iteration. A closer agreement requires the fully random algorithm of M 1972J{N14}, or at least a "canonical" algorithm of in the sence M 1974f{N15}, with lognormal weights. These topics are delicate and must be postponed to M 1997N.

4.11 Why should price variation be multifractal, and would multifractality have significant consequences?

Possibly explanatory power of multiplicative effects. This book is eager to study the consequences of scaling, but reluctant to look for its roots; in particular, Chapter E8 expresses doubts about explanations that involve "proportional effects".

Nevertheless, such an argument underlies multifractals, and is worth sketching.

The structure is especially clear in a generating method that is an alternative to the cartoons described in this section. It concerns the variance of price movements and proceeds as follows. The originator is a uniform intensity, and it is pertubed by pulses independent of one another and random in every way. Metaphorically, the variability of the variance is attributed to an infinity of "causes," the effects of one cause being described by one pulse. The main feature is that the distribution of pulse lengths must be scaling; this means that the effects are very short-lived for most causes and very long-lived for some.

There is little doubt that, ex-post, such "pulses" could be read into the data, but this is not the proper place to discuss the pulses' reality.

Extrapolation of multifractals and an ominous and possibly inconceivable implication. This brief paragraph simply draws attention to Section 5.4.

4.12 Possible directions for future work

A sketch of directions cannot be comprehensive and comprehensible, without detailed acquaintance will further developments of the theory which are necessarily postponed to M 1997N,H.

5. THE DISTINCTION BETWEEN MILD AND WILD VARIABILITY EXTENDS FROM RANDOM VARIABLES TO RANDOM OR NON-RANDOM SELF-AFFINE FUNCTIONS

As applied to discrete-time sequences of independent random variables, the notions of "mild" and "wild" were discussed in Chapter E5. This section moves on, to establish the same distinction in two additional classes of functions: the continuous time grid-free *random* self-affine functions discussed in Section 3 and the grid-bound *non-random* self-affine cartoon functions discussed in Section 4. Both classes involve numerous "either-or" criteria that sort out diverse possibilities: continuous or not, unifractal or not, and, in the cartoon case, unibox or not. But none of these "either-or" criteria is more important than the distinction between mild and wild.

The fact that one can extrapolate those notions to a non-random context is a special case of the general and important fact, already mentioned in Section 4.1, that fractality is often an excellent surrogate for randomness. The fact that this section is restricted to self-affine functions brings a major simplification, as compared to Chapter E5: there will be no counterpart as slow randomness.

Here is a summary of this section's conclusions. True WBM and its cartoons C(WBM) exhibit *mild variability*. The remaining processes described in Section 3 and other cartoons described in Section 4 contradict mildness in diverse ways, alone or in combination. Those contradictions exemplify different possible forms of *wild variability*. Thus, "Noah wild" recursive functions are cartoons of discontinuous wildly random processes whose jumps are scaling with $\alpha < 2$. "Joseph wild" recursive functions are cartoons of (continuous) Gaussian processes called fractional Brownian motions. The "sporadic wild" recursive functions are cartoons of wildly random processes that I called sporadic because they are constant almost everywhere and supported by Lévy dusts (random versions of the Cantor sets.)

5.1 The notions of mild and wild in the case of random functions

The basic limit theorems. Given that self-affinity forbids slow randomness, it suffices to show that the limit theorems that define mildness remain meaningful beyond random sequences of independent identically distributed variables. For many purposes – including the present one – those theorems are best split into three parts, each concerned with the existence of a renormalizing sequence $A(T)$, such that $\sum_{t=1}^{T} X(t)/A(T) - B(t)$ has a non-degenerate limit as $T \rightarrow \infty$.

LLN. When such a limit exists for $B = 0$ and $A(T) = T$, X satisfies the *law of large numbers.*

GLT. When there exists two functions, $A(T)$ and $B(T)$, such that $X\Sigma(T)A(T) - B(T)$ converges to the Gaussian, X satisfies the *central limit theorem with Gaussian limit.*

FLD. When $A(T) = T^{1/2}$, $X(t)$ satisfies the *Fickian law of diffusion,* which says that diffusion is proportional to $\sqrt{T}$.

Those theorems hold for independent Gaussian random variables such as the Gaussian, which deserved in Chapter E5 to be called *mildly* random. But all three theorems fail for independent Cauchy variables. And wild variables are those for which one or more of those three theorems fail.

5.2 Extrapolation of recursive cartoon constructions as an "echo" of interpolation

To show that LLN, GLT and FLD have exact counterparts for the extrapolated non-random "cartoons" in Figure 1, and that those counterparts statements may be true or false, we must first extrapolate the recursive construction of our cartoons. The conclusion will be that mild variability is found only in the cartoon of Wiener Brownian motion. Every other cartoon exemplifies a form of wild behavior, with one added possibility to be mentioned momentarily.

A complete self-affine fractal shape is not only infinitely detailed, but also infinitely large. This infinitely fine detail is absent in the case of sequences of random variables in discrete time and has no significance in either physics or finance. The sole reason why Section 4 did not use extrapolation is because interpolation is far easier to describe, study, and graph. The k-th level of extrapolation can also be called level-$(-k)$ of the construction.

First examine the case where each level-1 box is obtained from the original level-0 box by reduction of ratio b_t horizontally and b_x vertically. A straightforward procedure achieves both k levels of interpolation and k levels of extrapolation: it suffices to start with the prefractal interpolate pushed to the $2k$-th level of interpolation and dilate it in the ratios b_t^k horizontally and b_x^k vertically. This dilation transforms each level-$(-k)$ box into a unit square.

While interpolation is a uniquely specified procedure, extrapolation is not. In order to specify it, it is necessary to first select the fixed point of the dilation. The simplest is the origin 0 of the axes of t and x. However, this is not the only possible choice: the fixed point can be the bottom left or upper right corner of any box in the generator (or the limit of a sequence of such points). The most natural procedure is to select the fixed point at random; in this sense, *all* extrapolated self-affine shapes are intrinsically random. (Interpolated sets do not become random until all selects an origin, but this is an optional step that one may not need to face.)

Observe, however, that the fixed point *cannot* be located on a vertical interval of the generator. If extrapolation is attempted around such a point, the interval it contains will lengthen without bound, into an infinitely large discontinuity characterized by $H=0$. (It is also impossible to select the fixed point on horizontal interval of the generator. The extrapolation will lengthen this interval without bound, into an infinitely long gap characterized by $H=\infty$.)

5.3 The notions of mild and wild in the case of extrapolated cartoons

We shall examine the law of large numbers and the Fickian law of diffusion.

Counterparts of the law of large numbers (LLN). Several cases must be distinguished.

• *The unifractal case.* The function varies in clock time, H is unique with $H \leq 1$, and $\Delta X \sim (\Delta t)^H$ for every t. If so, the sample average is $\Delta X / \Delta t \sim (\Delta t)^{H-1}$. It follows that the LLN holds and the limit is 0.

• *The mesofractal case.* The function varies in fractal time, and H takes a unique non-degenerate value $H < 1$, except that in interpolation, $H = 0$ for points in a discontinuity. In extrapolation, the fixed point is never in a discontinuity; therefore, the LLN holds.

• *The multifractal case.* The function varies in multifractal time, and in interpolation replaces the single H by a collection of H_i, some of them > 1 while others < 1. One can show that if the fixed point is chosen at random, LLN *fails* with probability 1. The set of fixed points for which LLN holds is very small (of zero measure). We shall return to this issue in Section 5.3.

Absence of counterpart of the central limit theorem (CLT). No choice of $A(T)$ makes $X(T)/A(T)$ converge to a non-trivial limit. This is part of the price one has to pay for the replacement of randomness by non-random fractality.

Counterparts of the Fickian law of diffusion (FLD). One tends to view the Fickian form $A(T) = \sqrt{T}$ as a simple corollary of the Gaussianity of the limit. But it is not. When $H_i = H$ for all i, only $A(T) = T^H$ makes $X(T)/A(T)$ oscillate without end, rather than collapse to 0 or ∞. The Fickian law of diffusion is satisfied when $H = 1/2$. This requirement allows the cartoon of WBM, of course. Mesofractal cartoons also allow $H = 1/2$, but a careful study (which must be postponed to M 1997N and 1997H) suggests that this case is of limited interest.

5.4 An ominous and possibly inconceivable implication of the extrapolation of multifractals

Thus far, the passage from fractals to multifractals deliberately avoided conceptual roadblocks and proceeded in as low a key as possible, but for the next topic a low-key tone would be hard to adopt. In the Gauss-Markov universe and related processes, the effects of large excursions are well-known to be short-lived and to regress exponentially towards the mean. The multifractal world is altogether different and the following inference gives a fresh meaning to the term, "wildness".

Consider a function $x(t)$ drawn as a multifractal cartoon of the kind examined in Section 4, with both t and x varying from 0 to 1. Now hold Δt constant and extrapolate to a time interval of length 1 placed at a distance T away from the original [0, 1]. A striking result concerns the increments Δx of $x(t)$ over that increasingly distant interval. As $T \rightarrow \infty$, those increments *will not regress at all.* For example (but we cannot stop here for a proof), the average of $|\Delta x|^{1/H}$ will *increase* without bound, like T to the power $-\tau_D(-1) > 0$, if that power is finite, and even faster otherwise.

In words, exponential regression to the mean is replaced by a power law "explosion." We already know that tail lengths explode as one interpolates, and now find that the same is true as one extrapolates.

To understand intuitively the explosion that accompanies extrapolation, the easiest is to reinterpret Figure 5, by imagining that it relates to time span much longer than 1 and that the unit time interval from which one will wish to extrapolate is chosen at random. We know that it is in the nature of interpolation for multifractals that a randomly chosen short interval will with high likelihood fall within a region of *low* variation (and a very short interval will fall within a region of *very low* variation.) A corollary is that the variation is likely to be wilder outside the unit interval than it is inside.

The most likely response to this wildly "unstable" scenario is the usual one: to argue that, well before any explosion occurs, the process is bound to "cross over" to another process obeying different rules. Be that as it may, the consequences of this scenario are fascinating, and will be explored in a more suitable context.

E7

Rank-size plots, Zipf's law, and scaling

✦ **Abstract.** Rank-size plots, also called Zipf plots, have a role to play in representing statistical data. The method is somewhat peculiar, but throws light on one aspect of the notions of concentration. This chapter's first goals are to define those plots and show that they are of two kinds. Some are simply an analytic restatement of standard tail distributions but other cases stand by themselves. For example, in the context of word frequencies in natural discourse, rank-size plots provide the most natural and most direct way of expressing scaling.

Of greatest interest are the rank-size plots that are rectilinear in log-log coordinates. In most cases, this rectilinearity is shown to simply rephrase an underlying scaling distribution, by exchanging its coordinate axes. This rephrasing would hardly seem to deserve attention, but continually proves its attractiveness. Unfortunately, it is all too often misinterpreted and viewed as significant beyond the scaling distribution drawn in the usual axes. These are negative but strong reasons why rank-size plots deserve to be discussed in some detail. They throw fresh light on the meaning and the pitfalls of infinite expectation, and occasionally help understand upper and lower cutoffs to scaling. ✦

THIS LARGELY SELF-CONTAINED CHAPTER covers a topic that goes well beyond finance and economics and splits into two distinct parts. Hence, the points to be made are best expressed in terms of two definite and concrete contexts. The bulk is written in terms of "firm sizes," as measured by sales or number of employees, but would be unchanged if firm sizes were replaced by such quantities as city populations. The second context to be invoked, word frequencies, warrants a digression from this book's thrust , if only because the straightness of a log-log rank-size plot is explained most readily and simply in that context.

Restatement of the probabilists' notation. A capital letter, say U, denotes a quantity whose value is random, for example the height of man or the size of an oil reservoir selected at random on the listing of the data. The corresponding lower case letter, say u, denotes the sample value, as measured in numbers of inches or in millions of barrels.

1. INTRODUCTION

1.1 Rank-size plots for concrete quantities

A concrete random variable is a quantity that is measured on an "extrinsic" or "physical" scale. Humans are measured by *height*, firms by sales or numbers of employees, and cities by *numbers* of inhabitants. More generally, statistical quantities such as "height" and "number of inhabitants" are originally defined in a non-stochastic context. Their physical scale serves to rank those random variables by increasing or decreasing value, through either $F(u) = \Pr\{U \leq u\}$ or the tail distribution $P(u) = 1 - F(u)$.

The concrete reality that underlies the notions of $F(u)$ and $P(u)$ can be also represented in the following alternative fashion. The first step is to *rank* the elements under investigation by decreasing height, size, and number. The largest item will be indexed as being of rank $r = 1$; the largest of the remaining items will be of rank $r = 2$, and so on. The second step is to specify size, or any other suitable quantity Q, as a function of rank. One way to specify the distribution of a random quantity is to specify the corresponding function $Q(r)$.

By definition, $Q(r)$ varies inversely with r : it decreases as r increases. Granted the possibility of more than one item of equal size, $Q(r)$ must be non-increasing. This is the counterpart of the fact that $F(u)$ and $P(u)$ are non-decreasing and non-increasing, respectively.

Special interest attaches to the positive scaling case, when the assertion that Q varies *inversely* with r can be strengthened to the assertion that Q is *proportional to the inverse* of r, or perhaps that log Q varies *linearly* with log r. Unfortunately, some careless rank-size studies confuse different meanings of "inverse variation."

1.2 "Static" rank-frequency plots in the absence of an extrinsic scale

The occurrence of a word in a long text is not accompanied by anything like "a human's height" or "a city's number of inhabitants". But there is a simple and beautiful way out. Even when extrinsic "physical" quantities are not present, every random event involves at least one intrinsic quantity: it is the event's own probability.

Thus, in the case of word frequencies, rank-size does not involve the usual functions $F(u)$ and $P(u)$, but begins with a function $Q(r)$ that gives the probability of the word whose rank is r in the order of decreasing probabilities. To some authors, this looks like a snake biting it's tail, but the paradox is only apparent and the procedure is quite proper. In the scaling case, $\log Q$ varies linearly with $\log r$.

Furthermore, this ranking happens to be justified a posteriori in the theory of word frequencies introduced in M 1951, sketched in Section 1.2.4 of Chapter E8, and developed in M 1961b. That theory introduces a quantity that is always defined and often has desirable additivity properties similar to those of "numbers of inhabitants;" it is the function $-\log p$, where p is a word's probability. By introducing $-\log p$, the ranking based on frequency is reinterpreted as conventional ranking based on $-\log p$ viewed as an intrinsic random variable. In practice, of course, one does not know the probability itself, only an estimate based upon a sample frequency.

There are strong reasons to draw attention to a wide generalization of my derivation of the law of word frequencies. One reason is that it may bear on the problem of city population via a reinterpretation of the central place theory. A second reason is that this generalization involves a phenomenon described in the next section, namely a built-in crossover for low ranks, that is, frequent words and large city population. The reader interested in the derivation is referred to M 1995f, and the reader prepared to face an even more general but old presentation is referred to M 1955b.

1.3 Distinction between the terms, Zipf distribution and Zipf law

The term "Zipf law" is used indiscriminately, but the concepts behind this terms *distribution* and *law* are best kept apart. The fairest terminology seems to be the following one.

Zipf distribution will denote all instances of rank-size relation $Q(r)$ such that, with a suitable "prefactor" Φ, the expression

$$Q(r) \sim \Phi r^{-1/\alpha}$$

is valid over an intermediate range of values of r, to be called *scaling range.* This range may be bounded by one or two crossovers, $r_{\min} \geq 1$ to $r_{\max} \leq \infty$, to which we return in Section 1.7. Allowing crossovers automatically allows all values of $\alpha > 0$. When $\alpha < 1$, the scaling range need not, but can, extend to $r \rightarrow \infty$ with no crossovers; when $\alpha \geq 1$, the scaling range is necessarily bounded from above.

Zipf emphasized the special case $\alpha = 1$. If so, $Q(r)$ does not only vary inversely with r but varies in inverse proportion to r. In the special case of word frequencies, Zipf asserted $\alpha = 1$ and $\Phi = 1/10$, which are very peculiar values that demand $r_{max} < \infty$.

Zipf law will denote all empirical cases when the Zipf distribution is found to hold.

1.4 Zeta and truncated zeta distributions

"Zeta" and "truncated zeta" distributions are the terms to be used to denote exact statements valid for all values of r.

The zeta distribution. When $\alpha < 1$, hence $1/\alpha > 1$, the function

$$\zeta(1/\alpha) = \sum_{s=1}^{\infty} s^{-1/\alpha}.$$

is the mathematicians' Riemann zeta function. This suggests "zeta distribution" to denote the one-parameter discrete probability distribution

$$p(r) = \frac{r^{-1/\alpha}}{\sum_{s=1} \text{to}\infty s^{-1/\alpha}} = \frac{r^{-1/\alpha}}{\zeta(1/\alpha)} = \Phi r^{-1/\alpha}.$$

In the coordinates $\log r$ and $\log p(r)$, the zeta distribution plots as an exact straight line of slope $-1/\alpha$. Clearly,

$$\int_1^{\infty} s^{-1/\alpha} ds = \frac{\alpha}{1-\alpha} < \sum_{s=1}^{\infty} s^{-1/\alpha} < 1 + \int_1^{\infty} s^{-1/\alpha} ds = \frac{1}{1-\alpha}.$$

When α is near 1, the two bounds are close to each other.

Under the zeta distribution, the relative size of the largest firm is $\zeta^{-1}(1/\alpha)$. The joint share of the r largest firms is

$$\zeta^{-1}(1/\alpha)\sum_{s=1}^{r} s^{-1/\alpha}.$$

The ratio: sum of sizes of firms of rank strictly greater than r, divided by the size of the r-th firm, is

$$\left\{\sum_{s=r+1}^{\infty} s^{-1/\alpha}\right\} r^{1/\alpha}.$$

As r increases, the sum in braces becomes increasingly closer to the integral $\int_r^\infty x^{-1/\alpha}dx$, and the preceding ratio becomes

$$\frac{r^{(1-1/\alpha)}r^{1/\alpha}}{(1/\alpha - 1)} = \frac{r\alpha}{(1-\alpha)}.$$

Truncated zeta distribution. When $\alpha < 1$ and $V > -1$, define

$$\zeta(1/\alpha, V) = \sum_{s=V+1}^{\infty} s^{-1/\alpha} = \sum_{1}^{\infty}(r+V)^{-1/\alpha}.$$

I use the term "truncated zeta distribution" to denote the two-parameter discrete probability distribution

$$p(r) = \frac{(r+V)^{-1/\alpha}}{\zeta(1/\alpha, V)} = \Phi(r+V)^{-1/\alpha}.$$

Plotted in the coordinates $\log r$ and $\log p(r)$, the tail is straight, of slope $-1/\alpha$, as in the truncated zeta distribution, but this tail is preceded, for small ranks, by an appreciable flattening that extends to values of r equal to a few times V.

1.5 Dynamic evolution of a rank-size plot as the sample-size increases

The considerations in Sections 1.1 and 1.2 are called static because they concern a fixed sample. Some sort of dynamics enters if the rank-size plot is continually updated as data are drawn from this sample. Let us show that the examples of Section 1.1 and 1.2 behave very differently from that viewpoint. In other words, the commonality of structure that seems to be implied by the term Zipf law is misleading.

Firms. Create an increasing sample of firms from an industry by picking them at random. One approximation is to follow a list ordered lexicographically. As the sample develops, the largest firm will repeatedly change, and a given firm's rank will increase as new firms flow in. The rank-size plot will grow by its *low rank* end. Furthermore, however long a list of prices may be, it is certainly finite. Therefore, as the sample size increases, the straightness of the rank-size plot must eventually break down at the high-rank end. Additional reasons for breakdown will be examined in the next sub-section.

Words. By way of contrast, increase a sample of words by reading a scrambled text, or perhaps a book by James Joyce. The most probable word will soon establish and maintain itself and other words' rank will gradually settle down to those words' probabilities. Experience suggests that in most cases the number of distinct words is so extremely high, that fresh words keep being added as the sample increases. Therefore, the rank-size plot will grow at its *high rank* end.

1.6 Large estimated values of α are not reliable, hence not significant

The scaling range from (r_{min}, Q_{max}) to (r_{max}, Q_{min}). might be reported in the form of a "number of decades," defined as the decimal logarithm of either of two ratios, namely $\log_{10}(Q_{max}/Q_{min})$ or $\log_{10}(r_{max}/r_{min})$. When $\alpha \sim 1$, the two ratios are close to each other. When α is large and $1/\alpha$ is small, the two ratios differ significantly. One is tempted to report the larger of the two values, $\log_{10}(r_{max}/r_{min})$, but the proper value is the smaller. The reason is that the intrinsic quantity is not r but Q. The issue is discussed in Chapter E3.

For example, consider the reports of phenomena for which $1/\alpha = 1/4$ holds over a seemingly convincing range of 2 decades in terms of r. Restated in terms of Q, this range reduces to an unconvincing one-half decade.

1.7 The many forms taken by the crossovers

The difference Section 1.5 draws between the cases of firms and words is essential from the viewpoint of the width of the scaling ranges from (r_{min}, Q_{max}) to (r_{max}, Q_{min}). Let us run through a few examples.

Personal income. Scaling was observed by Pareto and is discussed in several chapters of this book. But scaling breaks down for large values of the rank, because small incomes *do not* follow a scaling distribution. There is also an operational reason for breakdown: small incomes are neither defined nor reported with accuracy. As a result, the log-log rank-size plot is expected to cross-over for high values of r into a near-vertical portion. Once again, however, and this is important to the discussion of incomes in this book, the evidence suggests that scaling holds for unboundedly large incomes, implying a straight log-log plot for small ranks. An exception is that the straightness is not expected to hold for $r = 1$, because, as Section 3.2 will show, the largest value $U(1, N)$ is expected to have extraordinarily high sample scatter.

Firms. This notion breaks down into artificiality and irrelevance for very small sizes, because of legal reasons to register or not to register.

City sizes. Both ends of the graph are affected by artificiality, for example by political boundaries that represent nothing worth studying quantitatively.

Word frequencies. As already mentioned, Section 1.2.4 of Chapter E8 describes my reasons why one should expect word frequencies to follow Zipf's law in the form $Q(r) \sim \Phi r^{-1/\alpha}$. But those reasons rely on limit theorem of probability and say nothing about small values of r. In general, the model yields unrelated values of α and Φ, which fail to satisfy the equality $\Phi^{-1} = \zeta(1/\alpha)$ that is characteristic of the zeta distribution. When such is the case, a crossover is inevitable. One can *define* a correction factor V by the relation

$$\Phi^{-1} = \zeta(1/\alpha, V),$$

and use as approximation the truncated zeta expression

$$Q(r) = \Phi(r + V)^{-1/\alpha}.$$

In the context of word frequencies, this relation is often referred to as the *Zipf-Mandelbrot law.*

Summary. All told, the expectation that one or both ends of the curve will cross over implies that the estimation of α must often neglect the values of very low or very high rank.

Analytic expressions for the behavior of a non-scaling distribution beyond the scaling interval: limitations to their usefulness. Many specialists in curve-fitting insist that one can account for crossovers by replacing a linear log, log plot, by the plot of a second-order polynomial. When the second order is not enough, one moves to a polynomial of higher order.

A different approach is suggested by a different tradition that is very strong both in physics and in economics (where it goes back to Pareto; see (Chapter E2, Section 3.3). In that tradition, one multiplies $r^{-1/\alpha}$ by $r^{-1/\alpha}\exp(-\beta r)$, or perhaps by a factor that varies more slowly than an exponential, such as $r^{-1/\alpha}/\log r$.

Those "all-purpose" traditional corrective terms may improve the fit, or broaden the range in which a single formula prevails. But they are not useful, in my judgement, and draw attention away from the impact of approximate straightness. The only corrective terms I find valuable are not those imitated from physics, but those suggested by theory.

1.8. The power of words and pictures

Words are powerful. Probabilists who now speak of *distribution* used to speak of *law*, which sounds or "feels" more impressive. "Scaling" distribution and "power-law distribution" are neutral terms that do not seek mystery and do not promise much in common between the various occurrences of scaling. By contrast, experience shows that "Zipf's law" is a repulsive magnet to professional students of randomness, but an attractive magnet for non-professional dabblers of all kind. The same is true of "$1/f$ noise," a term that necessity often forces me to both use and fight. Its near-synonym "self-affine function" makes no ringing statement, but experience proves that "$1/f$ noise" suggests a single underlying phenomenon, which happens to be very far off the mark.

Zipf's law as attractor. Zipf 1949 put forward the bold claim that scaling is the "norm" for all social phenomena, while for physical phenomena the "norm" is the Gaussian. His claims created quite a stir when I was a post-doc at MIT, in search for unusual facts to investigate.

In 1953, I gained durable praise from linguists for having shown that a straight rank-size plot for word frequencies is devoid of meaning for linguistics; there is nothing in it for syntax or semantics. However, Zipf's law proved interesting in probabilistic terms and (as told in Chapter 42 of M 1982F{*FGN*}) somehow started me on a path that led, first, to finance and economics, and eventually to fractals.

Zipf's law as repeller. Very different is the conventional conclusion, already mentioned in Chapter E4, that is recorded in Aitchison & Brown 1957. On pp. 101-2, we read that "A number of distributions are given by Zipf, who uses a mathematical description of his own manufacture on which he erects some extensive sociological theory; in fact, however, it is likely that many of these distributions can be regarded as lognormal, or truncated lognormal, with more prosaic foundations in normal probability theory." This statement proves two things: a) Aitchinson and Brown did not feel it necessary to check; b) they did not know what they were talking about. Few other technically competent persons knew.

As I write in 1997, the "bad vibes" that overselling had created in the nineteen fifties are forgotten, and Zipf's law is again oversold as a fresh and mysterious key to complexity or to a "linguistic" analysis of DNA structure. Those old dreams should crawl back in some hole.

2. FAST TRACK FROM A SCALING DISTRIBUTION TO A STRAIGHT RANK-SIZE PLOT

The themes of this section will be discussed again rigorously in Section 3.

2.1 From scaling to straight rank-size plots

The quantity U is called *scaling* when one has the relation

$$\Pr\{U \geq u\} = \text{ probability that } U \geq u = P(u) \sim Fu^{-\alpha}.$$

α is called *scaling exponent*, and F is a *numerical prefactor* that includes a *scale factor*. The sign ~ expresses that the relation is valid only for large values of u. Scaling does not exclude negative values of u, but this chapter does not dwell on them.

Assimilating the relative number of cases to a probability, a sample made of N independent drawings from a scaling distribution yields

$$\text{Nr}\ \{U \geq u\} \sim NFu^{-\alpha}.$$

The quantity Nr $\{U \geq u\}$ becomes the rank r of an item in the ordering by decreasing frequency, population or income. Once again, the biggest firm has rank $r = 1$ and size $U(1, N)$, the second biggest has rank $r = 2$ and size $U(2, N)$, etc..

Plotting this expression on transparent paper and turning the sheet around the main diagonal of the axes will yield u as function of r,

$$u(r, N) = F^{-1/\alpha} r^{-1/\alpha} N^{1/\alpha}.$$

Diagrams are not neutral, and different presentations of the same set of data emphasize one thing or another. The eye tends to be drawn to the top of a figure. In the plot of $\Pr\{U > u\} \sim u^{-\alpha}$, this position contains the many cases where u is small, while the other cases hide in the tail. In the plot of $u \sim r^{-1/\alpha}$, the opposite is true. When the values of u that matter most are the few largest ones, they are seen best in rank-size plots.

2.2 Relative size, the prefactor and criticality of the exponent $\alpha = 1$

Careful discussions of the rank-size relation consider the relative size

$$u_R(r, N) = \frac{\overline{u}(r, N)}{\sum_{s=1}^{N} \overline{u}(s, N)}.$$

We shall write

$$U_R(r, N) = \Phi r^{-1/\alpha}.$$

This formula involves a new prefactor Φ for which a numerical value is often reported with no comment. This implies the belief that Φ is independent of N. This strong statement is not obvious at all, in fact, it expresses a specific and unusual property. An essential role of this chapter is to tackle the case where the scaling range continues to $u_{\max} = \infty$, and to give necessary conditions for the prefactor Φ to be independent of N. One condition is that the exponent should satisfy $\alpha < 1$. Another condition is $1 < \alpha < 2$ combines with $EU = 0$. In all other scaling cases, we shall see that Φ is a decreasing function of N.

3. CAREFUL DERIVATION: FROM A SCALING DISTRIBUTION TO A STRAIGHT RANK-SIZE PLOT

This Section begins informally and becomes rigorous thereafter.

3.1 Typical absolute size as function of rank

A) Select the unit of "firm size" so that the tail distribution is

$$\Pr\{U \geq u\} = P(u) \sim u^{-\alpha},$$

and take a random sample of N firms. A "typical value" of the number of firms larger than u is the expectation

$$r(u, N) = Nu^{-\alpha}.$$

B) Exchange the role of variable and function, rank the firms in decreasing order of size, and define $U(r, N)$ as the size of the r-th largest firm in this ranking. Inverting the preceding function for a group of N firms, the number of those of size $U(r, N)$ or larger, will "typically" be

$$r \sim NU(r, N)^{-\alpha}.$$

C) Draw N firms independently from the same scaling distribution and rank them as in B). For given r, a "typical value" of $U(r, N)$ will be

$$\bar{u}(r, N) = (N/r)^{1/\alpha} = N^{1/\alpha} r^{-1/\alpha}.$$

3.2 Rigorous results replacing the "typical" values in Section 3.1

The standard statistical theory of extreme values confirms that, as the number of firms in an industry increases, the size of the largest increases proportionately to $N^{1/\alpha}$. The precise results are as follows.

Theorem concerning weighting by $N^{1/\alpha}$. (Many references, including Arov & Bobrov 1960, Formula 19). As $N \to \infty$, the sampling distribution of the ratio $U(r, N)N^{-1/\alpha}$ converges to the truncated gamma distribution

$$\lim_{N\to\infty} \Pr\left\{ \frac{U(r, N)}{N^{1/\alpha}} < x \right\} = \frac{1}{\Gamma(k)} \int_{x^{-\alpha}}^{\infty} z^{r-1} e^{-z} dz.$$

The most probable value of $U^{-\alpha}(r, N)/N$ is $r-1$, giving some legitimacy to $U \sim (r-1)^{-1/\alpha}$. More importantly,

$$\lim_{N\to\infty} E\left\{ \frac{[U(r, N)]^q}{N^{q/\alpha}} \right\} = \frac{\Gamma(r - q/\alpha)}{\Gamma(r)} \text{ when } r > \frac{q}{\alpha}, \text{ and } = \infty \text{ otherwise .}$$

Double asymptotics. As $N \to \infty$ and $r \to \infty$, the Stirling formula yields

$$E\{[U(r, N)]^q N^{-q/\alpha}\} \sim r^{-q/\alpha}.$$

For $q=1$, this Stirling approximation for $r \to \infty$ agrees with the "typical value" $\bar{u}(r, N)$. Moreover, as $N \to \infty$ and $r \to \infty$, the variability factor $EU^2/(EU)^2 - 1$ tends to 0; more generally, $U(r, N)N^{-1/\alpha}$ becomes for all practical purposes non-random. This was implicitly taken for granted in the heuristic argument of Section 2, and has now been justified.

Preasymptotic behavior. The Stirling formula is not a good approximation until very large values of r are reached.

When α is estimated from the low range portion of the plot, there is a clear statistical bias. It is due to averaging of $U(r, N)$ for fixed r, therefore represents a self-inflicted complication due to the use of rank-size plots.

The fact that $E\{U(r, N)\} = \infty$ for $r < 1/\alpha$ is unfortunate; it is avoided using the following result.

Theorem concerning weighting by $U(1, N)$. (Arov & Bobrov 1960, formula 21) As $N \to \infty$, the sampling distribution of the ratio $U(r, N)/U(1, N)$ converges to

$$\lim_{N\to\infty} \Pr\left\{ \frac{U(r, N)}{U(1, N)} < x \right\} = 1 - (1 - x^{\alpha})^{r-1}.$$

It follows that

$$\lim_{N\to\infty} E\left\{ \left[\frac{U(r, N)}{U(1, N)} \right]^q \right\} = \frac{\Gamma(1 + q/\alpha)\Gamma(r)}{\Gamma(r + q/\alpha)}$$

As $r \to \infty$,

$$\lim_{N\to\infty} E\left\{ \left[\frac{U(r, N)}{U(1, N)} \right]^q \right\} \sim \Gamma\left(1 + \frac{q}{\alpha}\right) r^{-q/\alpha}.$$

For $q=1$, this formula agrees with the ratio of "typical values" $\bar{u}(r, N)/\bar{u}(1, N)$, except for the prefactor $\Gamma(1 + 1/\alpha)$ which is greater than 1, and implies that $U(r, N)/U(1, N)$ remains scattered even when it is large. Since the variability of $U(r, N)$ tends to 0 as $r \to \infty$, the variability of $U(r, N)/U(1, N)$ solely reflects the scatter of $N^{1/\alpha}/U(1, N)$. Of course, the moments of $N^{1/\alpha}/U(1,N)$ follow from the fact that $N^{-1/\alpha}U(1, N)$ follows the Fréchet distribution $\Pr\{X < x\} = \exp(-x^{\alpha})$.

Clearly,

$$\lim_{N\to\infty}\Pr\left\{\left[\frac{U(r,N)}{U(1,N)}\right]^{\alpha} > y\right\} = (1-y)^{r-1},$$

hence,

$$\lim_{N\to\infty}E\left\{\frac{U(r,N)}{U(1,N)}\right\}^{\alpha} = \frac{1}{r}.$$

This means that $E\{U(1,N)^{-\alpha}\}$ is near 1. The sizes of firms of low rank are very sample dependent, hence are not necessarily close to their typical values. To avoid this variability, it is best to take a different point of comparison.

Weighting $U(r,N)$ by the cumulative size of the firms of rank higher than r. From the rank-size argument, the ratio of the sizes of the r' largest firms and the r'' largest firms is approximately equal to

$$\frac{1+\cdots s^{-1/\alpha}+\cdots r'^{-1/\alpha}}{1+\cdots s^{-1/\alpha}+\cdots r''^{-1/\alpha}}.$$

This expression is the same as for the zeta distribution. It varies continuously with α; for α near one, and large r' and r'', its order of magnitude is $\log r'/\log r''$.

3.3 Additional considerations

Logarithmic plots. Log-log plots involve the expectation of $\log[U(r,N)]/\log[U(1,N)]$ rather than of $[U(r,N)/U(1,N)]^{\alpha}$. This change brings no difficulty as long as r is not too small: $U(r,N)$ clusters tightly around its own expectation, which validates the approximation

$$E\left\{\frac{\log U(1,N)}{\log U(r,N)}\right\} \sim \frac{E[\log U(1,N)]}{E[\log U(r,N)]} = \frac{EV(1,N)}{EV(r,N)},$$

where $V(r,N)$ is r-th largest among N exponential variables $V=\log_e U$.

Visual estimation of α from the rank-size plot on doubly logarithmic paper. Despite the encouraging values of the various expected values reported in this section, the small r sampling distributions of $U(r,N)/U(1,N)$ are usually too scattered for complete statistical comfort. As the term of comparison, it is better *not* to use the size of the largest firm but rather a firm of rank as large as practical; the larger, the safer. This feature bears on the problem of the estimation of α. The usual procedure is to fit a straight line to the tail of $\log U(r, N)$ considered as a function of $\log r$, and to measure the slope of that line. When this is done, the points of rank 1, 2 or 3 are too sample-dependent, and should be given little weight. The resulting informal procedure can be approximated in several stages.

The *first* approximation would be to choose two values of r (say $r''=5$ and $r'=20$), and draw a line through the corresponding points on a doubly logarithmic graph; the sampling distribution of this estimator of alpha could be derived from the second theorem of this section.

A *second* approximation is to choose two couples (r', r'') and fit a straight line to 4 points. The sampling distribution would no longer be known exactly because the $U(r, N)$ are so defined that they do not provide independent information about α, but the precision of estimation naturally increases with the number of sampling points. The commonly practiced visual fitting amounts to weighting the estimates corresponding to various couples (r', r''), thus eliminating automatically the outlying estimates and averaging the others. It would be desirable to formalize this procedure and informal visual fitting should be studied more carefully, but it does not deserve its shady reputation.

3.4 Total industry size when $U>0$: contrast between the cases $\alpha>1$ (hence $EU<\infty$) and $\alpha<1$ (hence $EU=\infty$)

The size of the industry is the sum of the sizes of the N firms it contains, $\sum_{s=1}^{N} U_s$. While the arguments in Sections 3.1 and 3.2 hold for all α, it is now necessary to distinguish between $\alpha>1$ and $\alpha<1$.

D1) *The case when* $\alpha>1$, *hence* $EU<\infty$. Firm size being positive, $EU>0$, and "common sense" and the tradition of practical statistics take it for granted that the law of large numbers hold, so that the *total industry size is approximately N times the expected size of a randomly selected firm.*

D2) *The case when* $\alpha<1$, *hence* $EU=\infty$. The inequality $EU<\infty$ cannot and must not be taken for granted: it fails when the random variable U is scaling with $\alpha<1$. Many authors describe this feature as being "improper," and failed to face it. But *it is not* improper, and *must* be faced.

Applied blindly to the case $EU=\infty$, the law of large numbers claims that the total industry size is approximately infinite. This ridiculous result shows that one can no longer rely on common sense that is based on expectations.

Heuristically, if expectation is replaced by a different "typical" value, the total industry size is the sum of the above-written typical values $\overline{u}(r, N)$

$$\sum_{s=1}^{N} \overline{u}(s, N) = N^{1/\alpha} \sum_{s=1}^{N} s^{-1/\alpha}.$$

The most important feature is that the customary proportionality to N has disappeared. For very large N, it must be replaced by proportionality to $N^{1/\alpha}$. For moderately large N,

$$\tilde{u}N^{1/\alpha} \sum_{s=1}^{N} s^{-1/\alpha} \sim \tilde{u}N^{1/\alpha}[\zeta(1/\alpha) - \frac{N^{1-1/\alpha}}{1/\alpha - 1}] = \tilde{u}[\zeta(1/\alpha)N^{1/\alpha} - \alpha(1-\alpha)^{-1}N].$$

Because of $\alpha < 1$, the factor in $N^{1/\alpha}$ grows faster than the factor in N.

3.5 Relative shares when $U > 0$: contrast between $\alpha > 1$ and $\alpha < 1$; when $\alpha < 1$ and $N \rightarrow \infty$, Φ has a limit and the largest addend does not become relatively negligible

The two paths started in Section 3.3 continue in profoundly different fashions.

E1) *The case* $\alpha > 1$. As $N \rightarrow \infty$, (due to point C), the r -th largest firm increases proportionately to the power $N^{1/\alpha}$, and (due to point D1)) the sum of all firm sizes increases proportionately to N.

$$\overline{U}_R(r, N) \sim \frac{N^{1/\alpha} r^{-1/\alpha}}{NEU} = N^{-1+1/\alpha} r^{-1/\alpha}.$$

As $N \rightarrow \infty$, this ratio tends to zero. This is a familiar and widely used property; for example, the relative size of the largest of N Gaussian, exponential, Poisson, Gamma, or lognormal variables becomes negligible.

E2) *The case* $\alpha < 1$. *Heuristics.* As $N \rightarrow \infty$, both the r-th largest firm (due to point C)) and the sum of all firm sizes (due to point D2)) increase proportionately to the power $N^{1/\alpha}$. It follows that the relative share of the r-th largest firm behaves roughly like

$$\bar{u}_R(r, N) = \frac{r^{-1/\alpha}}{\sum_{s=1}^{N} s^{-1/\alpha}}.$$

When size is measured by the work force, the preceding relation gives an estimate of the probability that a worker chosen at random is an employee of the r-th firm.

E3) *The case* $\alpha < 1$, *continued. Rigorous results.* Given its significance, the argument yielding $\bar{u}_R(r, N)$ must be scrutinized carefully. This assumption that the numerator and denominator are statistically independent as $N \rightarrow \infty$ is false, but the conclusion is correct. Darling 1952 shows that $U_R(1, N)$ indeed has a distribution that is asymptotically independent of N. The formulas look forbidding and are not needed here, therefore, were put in the Appendix.

3.6 Comments

Chapter E9 will study the lognormal distribution, and show that this chapter's uncomfortable conclusion can be "papered over" by asserting that the observed facts concern an ill-defined "transient", but it is better to face it squarely. Against the background of the usual practical statistics, the fact that it is possible for Φ to be independent of N is astounding. The usual inference, once again, is that when an expression is the sum of many contributions, each due to a different cause, then the relative contribution of each cause is negligible. Here, we find, not only that the predominant cause is not negligible, but that it is independent of N.

The reader may be reminded of the distinction that Chapter E5 makes between *mild, slow,* and *wild* fluctuations. Most scientists' intuition having been nourished by examples of *mild* randomness, the preceding conclusion is *wild* and "counter-intuitive," but it will not go away.

APPENDIX A: THEOREMS CONCERNING LONG-RUN CONCENTRATION FOR THE WILD SCALING DISTRIBUTIONS

Theorem (Darling 1952). There exists a family of distribution functions, $G(1, \alpha, y)$, a special case of the distributions $G(r, \alpha, y)$ which will be examined later, such that

(A) if $0 < \alpha < 1$, $\lim_{N \to \infty} \Pr\left\{ \frac{\sum_{n=1}^{N} U_n - U(1, N)}{U(1, N)} \le y \right\} = G(1, \alpha, y).$

(B) if $1 < \alpha < 2$, $\lim_{N \to \infty} \Pr\left\{ \frac{\sum_{n=1}^{N} U_n - NE(U) - U(1, N)}{U(1, N)} \le y \right\} = G(1, \alpha, y).$

(C) if $1 < \alpha < 2$ and $EU \neq 0$, one has, in addition

$$\lim_{N \to \infty} \Pr\left\{ \frac{\sum_{n=1}^{N} U_n}{U(1, N)} \le y\tilde{u}N^{-1-1/\alpha} \right\} = \exp\{-[y/E(U)]^{\alpha}\}.$$

The distribution $G(1, \alpha, y)$ cannot be written as a simple analytic expression but its characteristic function $\hat{G}(\alpha, z)$ is known. It is as follows:

$$\text{If } 0 < \alpha < 1,\ \hat{G}(\alpha, z) = \frac{1}{1 - \alpha \int_0^1 (e^{izs} - 1)s^{-(\alpha+1)}ds} = \frac{1}{e^{iz} \int_0^1 e^{isz} s^{-\alpha} ds}$$

$$\text{If } 1 < \alpha < 2,\ \hat{G}(\alpha, z) = \frac{1}{-1 + \frac{iz\alpha}{(\alpha - 1)} - \alpha \int_0^1 (e^{isz} - 1 - isz)s^{-(\alpha+1)}ds}.$$

The essential thing about G is that it does not reduce to the degenerate value 0 as is the case in the distributions cited in Section 5, but has finite and non-vanishing moments of all orders. It is important to note the following: when $1 < \alpha < 2$, then $NE(U)$ must be subtracted from ΣU_n in order to make its expectation even to zero. If $0 < \alpha < 1$, one finds

$$E\left\{ \frac{\sum_{n=1}^{N} U_n - U(1, N)}{U(1, N)} \right\} = \frac{\alpha}{1 - \alpha}.$$

Theorems (Arov & Bobrov 1960). These theorems generalize the results in Darling 1952 to firms of ranks 2, 3, etc.. We have the following:

$$\text{If } 0 < \alpha < 1, \lim_{N \to \infty} \Pr\left\{ \frac{\sum_{n=1}^{N} U_n - U(1, N) - \cdots - U(r, N)}{U(r, N)} \leq y \right\} = G(r, \alpha, y),$$

where the distribution $G(r, \alpha, y)$ relates to the sum of r independent random variables, each following the law of distribution $G(1, \alpha, y)$; in particular, its expected value is $r\alpha/(1 - \alpha)$,

$$\text{If } 1 < \alpha < 2, \lim_{N \to \infty} \Pr\left\{ \frac{\sum_{n=1}^{N} U_n - NE(U) - U(1, N) \ldots U(r, N)}{U(r, N)} \leq y \right\} = G(r, \alpha, y).$$

APPENDIX B: TWO MEASURES OF CONCENTRATION AND THEIR DEPENDENCE ON THE FINITENESS OF *EU* AND *EU*²

To establish the usefulness of the rank-size rule and of the preceding heuristics, it is good to examine their bearing on existing techniques of statistical economics.

B.1 An index that measures inequality by a second moment

Herfindahl proposed the following statistical index of inequality

$$H = \sum_{s=1}^{N} \left\{ \frac{U(s, N)}{\sum U_n} \right\}^2 \leq 1.$$

This index has no independent motivation, and we shall see that its behavior is very peculiar. It is odd that it should ever be mentioned in the literature, even solely to be criticized because it is an example of inconsiderate injection of a sample second moment in a context where even the existence of expectation is controversial. Three cases must be distinguished.

The case where $EU^2 < \infty$. For large N, the law of large numbers applies to both U and U^2 and yields

$$H \sim \frac{NEU^2}{N^2(EU)^2} = \frac{1}{N}\frac{EU^2}{(EU)^2}.$$

The ratio $EU^2/(EU)^2$ is a normalized second moment, and H is expected to depend inordinately on the sample size N, in a way that is inextricably intertwined with its dependence on the distribution.

The case where $EU = \infty$, *in particular where* U *is scaling with* $0 < \alpha < 1$. For large N, the law of large numbers applies to neither U nor U^2. To obtain a first order of magnitude of H, one can take the heuristic step that uses the rank-size argument. This yields

$$H \sim \tilde{H} = (\text{ a constant })\sum_1^N s^{-2/\alpha}\left\{\sum_{s=1}^N s^{-1/\alpha}\right\}^{-2}.$$

As $N \rightarrow \infty$, this ratio tends to the positive and finite limit

$$\tilde{H}(\alpha) = (\text{ a constant })\zeta(2/\alpha)\zeta^{-2}(1/\alpha).$$

When α is close to 1, which is the value claimed for firm sizes,

$$\tilde{H}_\infty(\alpha) \sim (\text{a constant })\zeta(2/\alpha)(1-\alpha)^2.$$

The values of H do not depend much on N, but it amplifies the statistical fluctuations around the rank-size typical value.

The case when $EU < \infty$ *but* $EU^2 = \infty$, *in particular when* U *is scaling and* $1 < \alpha < 2$. According to the rank-size argument, Herfindahl's index is of the order of $N^{-2+2/\alpha}$ and tends to 0 as $N \rightarrow \infty$.

According to reports, Herfindahl's index is taken seriously in some publications. This is hard to believe.

B.2 Lorenz curves

As a measure of concentration, Lorenz proposed the function

$$L(x) = \frac{\sum_{s=1}^{xN} U(s, N)}{\sum_{s=1}^{N} U(s, N)}.$$

This function yields the proportion $L(x)$ of the total size as function of the proportion x of the number of firms, starting from the largest. It is taken for granted that the function $L(x)$ is obtained by a simple transformation from the size distribution $F(u) = \Pr\{U \le u\}$, and that the graph of $L(x)$, to be denoted by $\mathcal{L}$, is visually "more telling" than either the graph of $F(u)$ or the corresponding rank frequency graph $Q(r)$.

Skeptics respond that Lorenz curves emphasize a concept of inequality that involves the whole distribution and may be very misleading because the data in the bell of the distribution are frequently very incomplete. However, Lorenz curves also encounter a more serious theoretical objection. Indeed, it seems to be implicitly assumed that $\mathcal{L}$ only depends as the degree of concentration within a sample, not on the sample size N. Let us show that this implicit assumption is correct when $EU < \infty$, but *not* when $EU = \infty$. For scaling distribution, the implicit assumption is correct for $\alpha > 1$ but not for $\alpha < 1$.

A) For distributions with moments of all orders, Lorenz curves are theoretically unobjectionable. But their reputation for being visually telling is undeserved.

Indeed, in terms of $P(u) = \Pr\{U > u\}$, the number of firms of size greater than u is $NP(u)$ and their share of the sum of all firm sizes is $-N\int_u^\infty s dP(s)$. One can therefore write relative numbers and relative shares as functions of u as follows:

$$x = P(u) \text{ and } L(x) = -\int_u^\infty s dP(s).$$

This means that both x and $L(x)$ are independent of N, and define a curve $\mathcal{L}$. When N is small, the sample Lorenz curve will be made up of segments of straight line; but it will tend towards the well-defined limit $\mathcal{L}$ as $N \to \infty$.

For example, if U is scaling with $\alpha > 1, P(u) \sim u^{-\alpha}$ and the curve behaves as follows near the point $L = 0; x = 0$:

$$x(L) \sim u^{-\alpha} \sim [(\alpha - 1)\alpha^{-1}L]^{\alpha/(\alpha - 1)}; \text{ or } L(x) = \alpha(\alpha - 1)^{-1}x^{(\alpha - 1)/\alpha}.$$

This behavior has an unfortunate by-product: $\mathcal{L}$ may be well-defined independently of L, yet fails to deliver on its promise of being "visually telling." Indeed, if $\alpha - 1$ is small, a large number of derivatives of the function $x(L)$ vanish at the point $x = 0$, meaning that the curve $\mathcal{L}$ has a contact of very high order with the axis $x = 0$. On computer-drawn Lorenz curves, the exact order of this contact is not at all clear to the eye. As to old-time draftsmen, they near-invariably destroyed any evidence of contact by drawing Lorenz curves with a slope that is neither zero nor infinite near $x = L = 0$.

B) When U is scaling with $\alpha < 1$, sample Lorenz curves are unjustified and misleading, because they are greatly dependent on sample size. Indeed, we know that the relative share of the r largest firms is independent of N. Therefore, any prescribed ordinate $L(x)$ will be achieved for an abscissa x that tends towards zero as $N \rightarrow \infty$. This means that for $0 < \alpha < 1$, the sample Lorenz curve will tend towards the "degenerate" limit made up of the lower edge and the right edge of the unit square. Hence, the sample curves for finite N will not be representative of anything at all. In particular, sample Lorenz curves will depend even more critically upon the thoroughness with which small firms have been tabulated.

When U is scaling near the borderline value $\alpha \sim 1$, the convergence of $\mathcal{L}$ to its degenerate limit is very slow, which makes $\mathcal{L}$ especially misleading.

Early draft dated 5/8/63

E8

Proportional growth with or without diffusion, and other explanations of scaling

✦ **Abstract.** However useful and "creative" scaling may be, it is not accepted as an irreducible scientific principle. Several isolated instances of scaling are both unquestioned and easy to reduce to more fundamental principles, as will be seen in Section 1. There also exists a broad class of would-be universal explanations, many of them variants of proportional growth of U, with or without diffusion of $\log U$. This chapter shows why, countering widely accepted opinion, I view those explanations as unconvincing and unacceptable.

The models to be surveyed and criticized in this expository text were scattered in esoteric and repetitive references. Those I quote are the earliest I know. Many were rephrased in terms of the distribution of the sizes of firms. They are easily translated into terms of other scaling random variables that are positive. The two-tailed scaling variables that represent change of speculative prices (M 1963b{E14}) pose a different problem, since the logarithm of a negative change has no meaning. ✦

THIS CHAPTER MEETS HEAD-ON the legitimate and widespread wish to explain the prevalence of scaling in finance and other fields.

Section 1 describes scattered instances in which scaling is explained fully by a brief and mathematically straightforward argument: by eliminating an intrinsic variable between two intrinsically meaningful exponentials. Because there are so few of them, those instances acquire an otherwise undeserved standing.

Section 2 provides a careful analysis and critique of a typical attempt to explain scaling by multiplicative diffusion of U, that is, ordinary ("Fickian") diffusion of $\log U$. Deep reservations about this approach motivated a preference for viewing scaling as a postulate that brings economies of thought, good fit and a useful basis for practical work. The scaling distribution must be thoroughly understood, and consequences explored, without waiting for explanation.

Section 3 describes a classic first-order model, in which randomness is limited to a transient, and models involving permanent randomness.

Reasons why this text came to be. In the late 1950s and early 1960s, much of my research concerned scaling in economics. But questions I found fascinating kept being described by others as not worth detailed study. Stated in today's vocabulary, the reason was that scaling was viewed as haging been explained by very simple arguments involving a "principle of proportional effect." Therefore, scaling "should be expected," and there was not much to it. M 1982F{*FGN*} alludes to those events in Chapter 42, titled "The Path to Fractals."

My constantly restated response involved several separate points.

A) A successful explanation could be described as proceeding "upstream" from scaling, while the consequences scaling proceed "downstream." In the case of scaling, upstream considerations happen not to affect my downstream investigations and the latter prove to be surprising or even shocking, therefore, extremely worthwhile.

This first response rarely convinced my prospective audience, making additional responses necessary.

B) A careful look shows that, with a few exceptions, the existing upstream explanations are oversold in one or more of several ways.

B1) Nearly all lean heavily on probability limit theorems that concern the state of a system in a "long-run when we shall all be dead" (to quote J. M. Keynes once again). Arbitrarily set initial conditions do not affect the limit, but the pre-asymptotic behavior may be poorly approximated by the theoretical asymptotics. Chapter E9 shows that this "defect" is especially nefarious when the limit is lognormal.

B2) In physics, the long-run is attainable, and "universality" often prevails. This grand word means that the same result is obtained from seemingly different assumptions, therefore details do not matter much. In particular, the interaction between particles leads to the same equilibrium distribution whether the total energy of a system is fixed, or allowed to fluctuate. To the contrary, would-be models of scaling outside of physics are overly sensitive to the choice between these last two assumptions, as will be argued in Section 2.

B3) In addition to clearly stated assumptions, many explanations of scaling use additional hypotheses that are rarely stated and by no means compelling. Seemingly imperceptible changes of unstated assumptions often yield a final outcome that is *completely different* and *usually non-scaling*. For example, they generate the lognormal distribution, instead of the scaling distribution that one wishes to explain.

B4) Some would-be "explanations" are circular and/or mathematically incorrect. It will be seen that diffusion demands *second-order* differential equations, yet some authors inplicitly believe that the same effects can be obtained by using a *first-order* difference equation. To achieve a scaling output, such models must begin with a scaling input.

The above-listed features were either not noticed or not appreciated. Reluctantly, I wrote M 1959s, which overwhelmed many readers, then M 1963g, which took a broader view. But at that point, the sudden wide interest in M 1963b{E14} seemed to eliminate the need for M 1963g, and that text was left unpublished. Unfortunately, the hopes that led to questionable models prove to be durable. This motivates the present text, loosely based on M 1963g, with two short papers added in appendices. This survey leaves aside a very novel generation of scaling via multifractals (M 1972j{N14}); Chapters E1 and E6 mention and use them to model price variation.

Aside from Section 1.2.4, and the criticism of diffusion models, this chapter surveys the work of others. It does not pretend to be either exhaustive, or entirely accurate in terms of historical credit. But it hopes to discourage the tedious process of piecemeal and independent discovery of models that are essentially equivalent and equally unconvincing. This text should provide the reader with background to tackle other questions of interest.

1. EXCEPTIONAL SCALING DISTRIBUTIONS THAT ARE COMPLETELY EXPLAINED IN A FEW LINES

Some models sketched in this Section are properly random, others obtain the scaling distribution by straightforward elimination of a common variable between two exponential relations.

1.1 Properly random models

Section 1.1.1. and 1.1.2. describe a few examples that take only a moment and yield scaling with $\alpha = 1$. Section 1.1.3 refers to a classical example.

1.1.1. Ratios of independent random variables and the effects of small denominators. *Ratios of Gaussian variables.* The ratio $R = Y/X$ of two independent Gaussian variables X and Y with $EX = EY = 0$ and $EX^2 = EY^2 = 1$ is well-known to be a Cauchy variable, with the density $1/[\pi(1+z^2)]$. Here is a perspicuous geometric proof. It follows from the assumptions on X and Y that $Z = X + iY$ is an isotropic random variable. Hence, its curves of equal probability density are circles, and $\theta = \tan^{-1}(Y/X)$ is uniformly distributed from 0 to 2π. The density of $R = Y/X$ follows immediately. The Cauchy density is only asymptotically scaling of exponent $\alpha = 1$ (it is L-stable).

Ratios of exponential variables. When X and Y are both exponential with $EX = EY = 1$, the curves of equal probability density of $Z = X + iY$ are no longer circles (Z is not isotropic), but lines with an equation of the form $X + Y =$ constant. For every given $X + Y$, hence also unconditionally, it follows that the ratio $(Y - X)/(Y + X)$ is uniformly distributed from –1 to 1. That is, U being defined as uniform from 0 to 1, R satisfies

$$\frac{Y-X}{Y+X} = \frac{R-1}{R+1} = 1 - \frac{2}{R+1} = 1 - 2U, \text{ hence } R = \frac{1}{U} - 1.$$

Conclusion, $R + 1$ follows *exactly* the scaling distribution of exponent $\alpha = 1$. No reference was found for this unbeatably simple result, but it would be surprising if it were new.

The scaling character of ratios whose denominators are often enough very small. In the 1960s, interest was aroused by econometric techniques that conclude with ratios having a small denominator and infinite moments. Those moments' divergence was viewed as an anomaly to be avoided, but anyone interested in explaining scaling should hold the precisely opposite view. Indeed, "it would suffice" to reexpress scaling quantities as ratios in which the denominator can be small. Students of mechanics know, at least since Poincaré, that small denominators lead to chaotic behavior.

1.1.2. Car queues on a one-lane road. Starting with many cars trying to maintain a constant randomly selected speed, the system will evolve to one in which cars queue behind a slow car. What will be the steady-state distribution of the length of this queue? In the simplest case, the intended speeds are independent and identically distributed random variables U_m with $\Pr\{U < u\} = F(u)$. The length of the queue is the first value of m such that $U_m < U_1$. When u_1 is known, $\Pr\{M = m \mid U = u_1\} = [1 - F(u)]^{m-1} F(u_1)$. Assume that there is a zero probability of anyone trying to drive at some minimum speed. Thus,

$$\Pr\{M=m\}=\int dF(u)[1-F(u)]^{m-1}F(u)=\frac{1}{m}-\frac{1}{m-1};\ \Pr\{M\geq m\}=\frac{1}{m}.$$

The queue length follows *exactly* the scaling distribution of exponent $\alpha=1$.

1.1.3. Number of tosses of a coin before a prescribed high gain level is first reached. It is well-known that this number's tail probability is $\sim u^{-1/2}$, which is asymptotically scaling with $\alpha=1/2$. A concrete application of this distribution is inserted at the end of this Chapter as Appendix I.

1.2 Straightforward elimination of an intrinsic variable between two intrinsically meaningful exponentials

This section proceeds in more or less chronologic order, tackling mutations, gravitation cosmic rays and words in discourse.

1.2.1. Attraction from within a very thin cone. In a Euclidean space of dimension E, create a cloud of unit point masses called stars, pick one star as the origin Ω and draw a very thin one-sided cone with apex at Ω and height R. If fluctuations are neglected, the proportion of stars found in this cone and also in a sphere of radius ρ is $(\rho/R)^E$. Assume that attraction follows the generalized Newton's law $u=\rho^{-N}$. Eliminating ρ, we find

$$\text{Prob }\{\text{attraction}>u\}=R^{-E}(u^{-1/N})^E=R^{-E}u^{-E/N}.$$

This is a scaling distribution. Three arbitrary features are the cutoff at a finite R, the restriction to a very thin cone, and the fact that we only consider the attraction from one star other than Ω.

Holtsmark's problem. Appendix III of M 1960i{E10} carries out a full argument for the physical case $E=3$ and $N=2$, but the same method generalizes without problem to all cases where $N>E/2$.

1.2.2. Bacterial mutations (Luria & Delbrück 1943). Disregarding the fact that numbers of bacteria are random and are integers, choose the time unit so that bacteria multiply and mutate deterministically and exponentially as follows: In the absence of mutation, a culture that contains b_0 bacteria at time $t=0$ contains b_0e^t bacteria at time t. In the presence of mutations at the rate m, the size of the culture becomes $b_0\exp[t(1-m)]$. Also suppose that a clone that descends from a mutation attains the size $\exp(gx)$ at age x. The number of clones will increase without end, the biggest clone being the oldest one.

If a mutant clone is picked at random among those present after a long time t, $\Pr\{\text{age} \geq x\} = \exp[-x(1-m)]$, hence $\Pr\{U \geq u\} = \exp(-gx)$. Eliminating x, we find in the limit the scaling distribution

$$\Pr\{U \geq u\} = u^{-D} \text{ with } D = (1-m)/g.$$

This argument helped molecular biology take off. A full and explicit treatment taking account of fluctuations was first provided in M 1974d, which is incorporated in this chapter as Pre-Publication Appendix II.

1.2.3. The energy of incoming cosmic rays (Fermi 1949.) The energy of primary cosmic rays is found to follow a scaling distribution with $\alpha \sim 1.7$ over 11 units of $\log_{10}$ (energy). The breadth of this range lies beyond economists wildest dreams; even in physics it is rarely encountered.

Fermi's assumptions easily translate into terms of firm growth, disappearance and replacement: all firms that exceed the size $\tilde{u}$ grow exponentially, until they die, but the population is replenished at a uniform rate to insure a steady-state distribution of sizes. We choose the unit of time so that T units of time after the size $\tilde{u}$ was exceeded

$$\text{firm size } = u(T) = \tilde{u}e^{T}.$$

Independently of its size $u \geq \tilde{u}$, a firm is given the probability αdT of disappearing during the time increment dT. It follows that the average "lifetime" of a firm is $1/\alpha$, and that

$$\Pr \{ \text{ a firm survives for the time } > T\} = \exp(-\alpha T).$$

Under these assumptions,

$$\Pr \{U > u\} = \Pr \{T > \log (u/\tilde{u})\} = \exp(-\log(u/\tilde{u}) = (u/\tilde{u})^{-\alpha}.$$

This means that $\Pr\{\text{size} > u\}$ is scaling with the exponent α. Q. E. D.

The expected change of a firm's size. The probability of dying out in the next unit of time is αdt, and the firms that do not die out will grow by the factor $(1 + dt)$. Hence, neglecting second-degree terms

$$E[U(t+dt) \mid u(t)] = u(t)(1-\alpha dt)(1+dt) = [1 + (1-\alpha)dt - (dt)^2]u(t),$$

hence

$$E[U(t+dt) \mid u(t)] - u(t) = (1-\alpha)dt.$$

Thus, an overall steady state will be established. But a given firm will, on the average, increase in size if $\alpha < 1$, and decrease in size if $\alpha > 1$. If $\alpha = 1$, the expected change of $U(t)$ vanishes; $U(t)$ is a martingale (see Chapter E1 or M 1966b{E19}).

A Fermi-like model with variable immigration rate $\varphi(t)$. If $\varphi(t) = \exp(\beta t)$, the expected number of firms of size u is

$$\int_0^{\infty} \exp(-\alpha T)\exp[\beta(t-T)]dt = \lambda \exp(at)\left(\frac{u}{\tilde{u}}\right)^{-(\alpha+\beta)} u^{-1}du.$$

If $\alpha + \beta > 0$, the number of firms surviving to time t will have a finite expected value equal to

$$\lambda\int_0^{\infty} \exp(-\alpha T)\exp[(t-T)]dt = \frac{\lambda \exp(\beta t)}{(\alpha+\beta)}.$$

Contrary to Fermi's original process, the number of firms now increases without bound, and there is no steady-state. But the expected *relative* number of firms of size u is scaling with exponent $\alpha + \beta$.

More generally, the distribution of firm sizes is the Laplace transform of the immigration rate $\varphi(t)$. This model may yield *any distribution of firm sizes,* as long as the inverse Laplace transform is positive.

1.2.4. A different example of straightforward elimination between two exponentials: word frequencies, lexical trees, and "Zipf's" law (M 1951, M 1982F{FGN}, Chapter 38.) Take an alphabet of $M+1$ letters L_m, with L_0 denoting the improper letter "space". Let "typing monkeys" use this alphabet to produce a random text in which L_0 is used with the probability p_0, and each of the other letters, with the probability $(1-p_0)/M$. There will be M^k distinct words made of k proper letters followed by space, each with the probability

$$p = p_0[(1-p_0)/M]^k = p_0 e^{-k \log B}, \text{ by definition of } B = (1-p_0)/M.$$

The most probable word, corresponding to $k=0$, is the shortest one. Now rank the other words by decreasing probability. In this ordering, a k -letter word will have a rank satisfying

$$r \propto M^k; \text{ that is } k = \frac{\log r}{\log M}.$$

Eliminating k between r and p yields the scaling distribution

$$p \propto p_0 \exp(-\ \log r \frac{\log B}{\log M}) = P_0 r^{-1/\alpha},$$

with

$$\frac{1}{\alpha} = \frac{\log B}{\log M} = \frac{-\log(1-p_0) + \ \log M}{\log M} = 1 + |\log_M (1-p_0)| > 1.$$

This scaling expression was found to be obeyed by words in large but homogenous samples of homogenous natural discourse, such as some long books. In the (unattainable!) limit $\alpha = 1$, this expression is called *Zipf's law for word frequencies.*

There is *nothing more* to Zipf's law with $\alpha > 1$. The derivation merely relies on compensation between two exponentials.

Markovian discourse and other generalizations yield the same result for $r \rightarrow \infty$ (M 1954b) But the probability distribution of "m-grams" formed by m letters *is not expected to be scaling*. Scaling does not take over until after the m-grams for all values of m have been sorted out in order of decreasing probability.

For word frequencies, the compensation between two exponentials can be rephrased in several ways. There is a "thermodynamical" or "information-theoretical" restatement (M 1982F{*FGN*}, Chapter 38); it looks learned and is enlightening to the specialist, but brings nothing new for most readers.

(In an amusing tongue-in-cheek etymology, Lee Sallows described the number of letters in a word as being its logorithm, from *logos*=word and *arithmos*=number. Thus, the simplest model described above assigns to a word a *logorithm* that is proportioned to the *logarithm* of its inverse probability.)

2. "RANDOM PROPORTIONATE EFFECT" AND ITS FLAWS

Improvements that inject randomness in the models of Luria-Delbruck and Fermi will be examined in Sections 3.2 and 3.3. But we begin by considering a widely popular approach that recreates the conditions of Section 1.2, by introducing an artificial variable that enters into two exponentials, and can be eliminated to yield the scaling.

2.1 Introduction to models in which the logarithm of the firm size performs a discrete random walk, a Brownian motion, or a diffusion

The overall scheme is familiar: when U follows the scaling distribution $\Pr\{U > u\} \sim (u/\tilde{u})^{-\alpha}$, the auxiliary variable $V = \log_e(U/\tilde{u})$ satisfies the exponential distribution $\Pr\{V \geq v\} = \exp(-\alpha v)$. To explain U by explaining V, one must a) motivate the transformation from U to V, and b) explain why V should be exponential. Task a) is difficult, but task b) seems formally very easy, because the exponential distribution plays a central and well-understood role in physics. It is not a surprise that a number of models of scaling are more or less obvious and/or conscious economic translations of various classical models of statistical thermodynamics.

Section 3.1 will classify those models as being of either the first or the second order and, therefore, as leading to a flow or a diffusion. But certain issues must first be faced, and this is done best by focusing on a bare bones diffusion model.

2.1 Tempting dynamic explanation of the exponential distribution for V by diffusion contained by a reflecting barrier

We begin with a random walk that is the simplest example of second order and diffusion. The physics background is a collection of particles that a) are subjected to a uniform downward gravity force, b) form a gas at a uniform temperature and density, and c) are constrained to remain in a semi-infinite vertical tube with a closed bottom and an open top. These particles' final equilibrium distribution will be a compromise: a) *gravity* alone would pull them down, b) *heat motion* alone would diffuse them to infinity, and c) the tube's *bottom* prevents downward diffusion. The result is classical in physics: acting together, these three tendencies create at the height z an exponential density distribution of the form $\exp(-\alpha z)$, where $1/\alpha$ increases with the temperature. Therefore, scaling could be explained by any model which (consciously or not) will re-interpret the above three physical forces in terms of economic variables.

2.2. A surrogate for diffusion of log U, based on a biased random walk

Let time t be an integer and $\log_e U$ be of the form kc, where k is an integer and $c > 0$. Between times t and $t+1$, allow the following possibilities:

a) $\log_e U$ can increase by c, the probability being p;

b) $\log_e U$ can decrease by c, the probability being $1-p$.

A necessary condition for equilibrium. For a distribution of $\log U$ to be invariant by the above transformations, it is *necessary* that the expected number of firms growing from size e^{kc} to size $e^{(k+1)c}$ be equal to the expected number of firms declining from size $e^{(k+1)c}$ to size e^{kc}. This can be written

$$\frac{\Pr\{\log_e U > (k+1)c\} - \Pr\{\log_e U > kc\}}{\Pr\{\log_e U > kc\} - \Pr\{\log_e U > (k-1)c\}} = \frac{p}{1-p}.$$

A steady-state solution in which $P(\log U > v) = \exp(-\alpha v)$ requires $\exp(-\alpha c) = p/(1-p)$. The condition $\alpha > 0$ requires $p < 1-p$ or $p < 1/2$ (this is an economic counterpart of the "force of gravity" referred to in section 2.1). Taken by themselves, the conditions a) and b) make firm sizes *decrease* on the average. Hence, to insure that the number of firms remains time-invariant above a lowest value $\tilde{u}$, one needs an additional factor, the counterpart of the closed bottom in Section 2.1.

c) A reflecting barrier can indifferently be interpreted in either of two principal ways. The firms that go below $\log_U = \log_e \tilde{u} - (1/2)c$ are given a new chance to start in life at the level $\tilde{u}$, or become lost but replaced by a steady influx of new firms starting at the level $\tilde{u}$.

Altogether, $P(k,t) = \Pr\{\log_e U = kc$ at time $t\}$ satisfies the following system of equations

$$P(k, t+1) = pP(k-1, t) + (1-p)P(k+1, t) \quad \text{if } k > k_0 \log_e \tilde{u}/c,$$
$$P(\log_e(\tilde{u}/c), t+1) = (1-p)P(\log \tilde{u}/c, t) + (1-p)P(1 + \log_e(\tilde{u}/c), t).$$

Due to the "diffusive" character of these equations, there is a steady-state limit function $P(k,t)$, independent of the initial conditions imposed at a preassigned starting time $\tilde{t}$. That limit is the scaling distribution. (*Proof:* Under the steady-state condition $P(k, t+1) = P(k,t)$, the second equation yields $pP(k_0, t) = (1-p)P(1+k_0, t)$, then the first equation gives the same identity by induction on $k_0 + 1, k_0 + 2$, and so on.) Therefore, conditions a), b), and c) provide a possible generating model of the exponential distribution for log U.

Champernowne's formally generalized random walk of log U. Champernowne 1953 offers a more general model. It seems sophisticated, but I came to view it as violating a cardinal rule of model-making, namely, that the assumptions must not be heavier and less transparent than the outcome. Details are found in M 1961e{E11}.

2.3 Aside on diffusion *without* a reflecting barrier: as argued by Gibrat, it yields the lognormal distribution, but no steady-state

As mentioned in Chapter E, the scaling and lognormal are continually suggested as alternative representations of the same phenomena, and the lognormal is believed by its proponents (Gibrat 1932) to result from a proportional effect diffusion. This is correct, but only up to a point. It is true that *in the absence of a reflecting barrier*, diffusion of log U does not yield an exponential, but a Gaussian, hence the lognormal for U. Indeed, after T tosses of a coin, $\log_e U(t+T) - \log_e U(t)$ is the sum of T independent variables taking the values c or $-c$ with the respective probabilities p and $1-p$. By the central limit theorem,

$$\frac{\log U(t+T) - \log U(t) - T\,c(2p-1)}{[2\,Tp\,(1-p)]^{1/2}}$$

will tend towards a reduced Gaussian variable. Hence, if at $T=0$ all firms have equal sizes, $\log_e U(t+T)$ will become lognormally distributed as $T \rightarrow \infty$. The same holds for other diffusion models without reflecting barrier.

While the above argument is widely accepted, it has a lethal drawback: the lognormal describes an instantaneous state, not a steady state distribution; for example, in time, its variance increases without bound.

2.4 Misgivings concerning the relevance to economics of the model of scaling based on the diffusion of log $U = V$

In the models for V that lead to a proper steady-state (Sections 2.1 and 2.2), the transformation $U = \exp V$ seems to work a miracle of alchemy: the metamorphosis of a mild variable V into the wild variable U, in the sense described in Chapter E5. But I propose to argue that no metamorphosis took place, because the conclusions reached ex-post destroy the intuition that justified ex-ante the diffusion of V. This contradiction between the ex-post and the ex-ante comes on top of the limitations stressed under B1) in the second page of this chapter.

Reminder of why a diffusion of V is a reasonable idea in physics. Physics is fortunate to have a good and simple reason why diffusion models are good at handling the exchanges of energy between gas molecules. In gases, the energy of even the most energetic molecule is negligible with respect to the total energy of a reservoir. Hence, all kinds of things can happen to one molecule, while hardly anything changes elsewhere. This enormous source of simplicity is a major historical reason why the theory of gases proved relatively easy to develop. For example, the wonderful property of universality (described under B2 at the beginning of this chapter) implies that it makes no difference whether the total "energy" ΣV in a gas is fixed (as it is in a "microcanonical" system) or allowed to fluctuate (as it is in a "canonical" system.) This is why the diffusion model's conclusions do not contradict its premises.

Reasons why diffusion of $\log U$ *seems* not to be *a reasonable idea in economics.* The transformation $V = \log U$ may seem innocuous, but it introduces a major change when $\alpha < 2$. The original U (firm size, income and the like) is additive, but *not* the logarithm $V = \log U$, and U is "wildly random." The concentration characteristic of wild randomness has very concrete consequences.

Firm and city sizes are arguably scaling with $\alpha \sim 1$ (Chapter E13 and Auerbach 1913), and it is typical for a country's largest city to include 15% of the total population. It surely will matter, both intuitively and technically, whether ΣU is fixed or allowed to fluctuate. When the total employment or population are kept fixed, it seems far-fetched to assume that the largest firm or city could grow or wane without influencing a whole industry or country. I cannot even imagine which type of observation could confirm that such is the case. Ex-post, this would be an interesting prediction about economics. Ex-ante, however, this presumed property is surely no more obvious than scaling itself, hence, cannot be safely inserted in an explanatory model of scaling.

Conclusion: a baffling embarrassment of apparent riches. The economic predictions yielded by the diffusion of $\log U$ are baffling. The model does yield a scaling distribution for U, but the model conclusion makes its premises highly questionable. Additionally, the argument ceases to be grounded in thermodynamics, because the latter does not handle situations where canonical and microcanonical models do not coincide. It is true that neither the success nor the failure of a model in physics can guarantee its success or failure in economics.

All told, the models of U based on the diffusion of $\log U$ leaves an embarrassment of riches. The user's response is a matter of psychology.

The *utter pessimist* will say that flawed models are not worth further discussion.

For the *moderate pessimist,* the diffusion of $\log U$ is questionable but worth playing with. I favor this attitude and M 1961e{E11} is an example of its active application.

The *moderate optimist* will not want to look too closely at a gift horse; even in physics, it is common that stochastic models yield results that contradict their assumptions. For example, the search for the "most likely state" applies Stirling's approximation to the factorial $N!$; but this argument eventually yields Ns equal 0 or 1, so that Stirling's approximation is unjustified even though it is successful. This is true, but physics also knows how to read the same results without assumption that contradict the conclusions; economics does not know how to do it. The feature that saves many stochastic models is that errors somehow seem to "cancel out." But in the presence of concentration the counterparts of "the contributing errors" are neither absolutely nor relatively small. For this reason, I view the moderately optimistic position as very hard to adopt.

The widespread *very optimistic* and unquestioning view of diffusion as explaining scaling is altogether indefensible. Models based on the diffusion of $\log U$ do not make me change my policy of viewing scaling as a postulate that brings economies of thought, good fit and a useful basis for practical work. It remains my strongly held belief that scaling must be thoroughly understood, and its consequences explored, without waiting to explain them. Nevertheless, after a short digression, we shall devote Section 3 to an examination of variant models.

2.4. A static characterization of the exponential: it is the "most-likely, "and the "expected" steady-state of Brownian motion with a barrier

Among the other models proposed to account for an exponential V in finance and economics, most are also dynamic, if only because "static" is a derogatory word and "dynamic" a compliment.

It is very well-known in thermodynamics that the state which a system attains as the final result of random intermolecular shocks can also be characterized as being "the most likely." The exponential distribution $P(v) = \exp[-\alpha(v - \tilde{v})]$ is indeed obtained by maximizing the expression $-\sum P(v) \log P(v)$ under the two constraints $v \geq \tilde{v}$ and $\sum vP(v) = \text{constant}$. A longer proof shows that the exponential is also "the average state."

As applied to the distribution of income, this approach goes back at least to Cantelli 1921. I see no good reason for retaining it: in physics, the expressions $\sum vP(v)$ and $\sum -P(v)\log P(v)$ have independent roles, the former as an energy and the latter as an entropy (communication theory calls it quantity of information.) But those roles do not give those expressions any status in the present context.

Many variants seem to avoid the transformation $V=\log_e U$. Castellani 1951 assumes without motivation the (equivalent) postulate that u is submitted to a "downward force" proportional to $1/u$.

3. PROPORTIONATE GROWTH AND SCALING WITH EITHER TRANSIENT OR PERMANENTLY DIFFUSIVE RANDOMNESS

Let us start from scratch. The basic idea of *random proportionate effect* is that when t increases by dt, the increment du is proportional to the present value of u, hence the change of $v=\log u$ is independent of v. However, those words are far from completely specifying the variation of V in time. Many distinct behaviors are obtained by varying the boundary conditions and the underlying differential equation itself. To end on a warning: due to time pressure, the algebra in this section has not been checked through and misprints may have evaded attention.

3.1 First and second-order random proportionate effects

Most important is the fact that the underlying equation can be of either first or second order.

Flow models ruled by a first-order equation. Solution of first-order differential equations largely preserve their initial conditions and essentially correspond to a nonrandom proportionate growth of U, in which the value of u only depends upon "age." This assumption must be recognized and may or may not be realistic. The model found in Yule 1924, corresponds to a first-order finite-difference equation. It predicts a random transient in youth, followed by effectively nonrandom, mature growth. This feature makes it inapplicable to economics, even though it is defensible in the original context of the theory of evolution, and its restatement in terms of bacterial mutation (see Appendix II) helped provide the first full solution of a classic problem in Luria & Delbrück 1943. The analysis in Section 3.2 applies with insignificant changes to a slight variant of Yule's process advanced in Simon 1954. After a short transient, the growth postulated in that paper becomes non-random.

Diffusion models ruled by a second-order equation. An actual "diffusion" embodies a permanently random proportionate effect. The initial conditions may become thoroughly mixed up, but lateral conditions (such as the nature of a side-barrier, the kind of immigration, etc.) have a permanent influence, even on the qualitative structure of the solution.

In the model in Aström 1962, the boundary is "natural," in the terminology of Feller, and no lateral conditions need or can be imposed. However, the equation possesses certain special features that amount to additional hypotheses in the model that the equations embody.

In any event, many assumptions are necessary if the diffusion model is to lead to the desired scaling solution and/or to a steady-state solution. One must not artificially stress some assumptions, while not acknowledging others of special importance in the existence of the reflecting barrier that entered in Section 2.1, but not Section 2.3.

3.2 The first-order finite difference model in Yule 1924

Except for an initial genuinely stochastic transient, the model in Yule 1924 is undistinguishable from the model in Fermi 1949 changed by the addition of exponential immigration, and restriction to a vanishing decay constant $\alpha = 0$. The calculations are much more involved than Fermi's, but it is useful to analyze Yule's model fully and concisely because there are occasions where it is applicable.

The fundamental assumption is that firm size is quantized and that during each time increment dt, the size of a firm has a known probability dt of increasing by unity. It is well-known that a firm growth from size 1 to size u during the elapsed time T has the geometric distribution for the probability that a firm grew from size 1 to size u during the elapsed time T. To this assumption, Yule adds the further hypothesis that new firms immigrate into the system with a probability dt during the time increment from t to $t+dt$. Then, at the calendar time t, the probability that one of these firms is of size u is given by

$$e^{-T}(1-e^{-T})^{u-1}.$$

To this, Yule adds another hypothesis: new firms immigrate into the system with a probability $\varphi(t)dt$ during the time-increment from t to $t+dt$. Then, at the calendar time t, the probability that one of these firms be of size u is given by

$$f(u, t) = \int_{-\infty}^{t} e^{-(t-\tilde{t})}[1 - e^{-(t-\tilde{t})}]^{u-1}\varphi(\tilde{t})d\tilde{t}.$$

As to the number of firms at time t, its expected value is given by

$$n(t) = \int_{-\infty}^{t} \varphi(\tilde{t})d\tilde{t},$$

and the expected value of the total size of all firms is

$$k(t) = \int_{-\infty}^{t} e^{t-\tilde{t}}\varphi(\tilde{t})d\tilde{t}.$$

In Yule's case,

$$\varphi(\tilde{t}) = \alpha e^{\alpha\tilde{t}} \text{ for } \tilde{t} > 0.$$

Writing $\exp(\tilde{t} - t) = y$, one obtains

$$f(u, t) = \alpha\int_{0}^{t} e^{-(t-\tilde{t})}[1 - e^{-(t-\tilde{t})}]^{u-1}e^{\alpha\tilde{t}}d\tilde{t} = \alpha e^{\alpha t}\int_{0}^{e^{-t}} y^{\alpha}(1-y)^{u-1}dy.$$

This is Euler's incomplete Beta integral. For e^{-t} close to 1,

$$f(u, t) \sim \alpha e^{\alpha t}\frac{\Gamma(\alpha+1)\Gamma(u)}{\Gamma(\alpha+u+1)}.$$

For large u,

$$f(u, t) \sim \alpha e^{\alpha t}\Gamma(\alpha+1)u^{-(\alpha+1)}.$$

Yule's model predicts the birth rate

$$k(t) = \alpha\int_{0}^{t} d^{t-\tilde{t}}e^{\alpha\tilde{t}}d\tilde{t} = \frac{\alpha}{\alpha-1}(e^{\alpha t} - e^{t}).$$

If $\alpha < 1$, $k(t) \sim \dfrac{\alpha}{1-\alpha} e^{t}$, and $f(u, k) \sim (1-\alpha)^{\alpha}\alpha^{1-\alpha}\Gamma(\alpha+1)k^{\alpha}u^{-(\alpha+1)}$.

If $\alpha > 1$, $k(t) \sim \dfrac{\alpha}{\alpha-1} e^{\alpha t}$, and $f(u, k) \sim \dfrac{\Gamma(\alpha+1)}{\alpha-1} k u^{-(\alpha+1)}$.

The relative rate of addition of new firms equals $(1-\alpha)^{-1}$. In the case of firm sizes, the mechanism of proportional growth cannot be presumed to start until the size has exceeded some sizable quantity $\tilde{u}$. Therefore, the Yule model is *not applicable*, since the correspondingly modified model will be based upon the probability that a population grows from size $\tilde{u}$ to size u in the time T, which is well-known to be given by the negative binomial distribution (Feller 1950). Therefore,

$$f(u, t) = \int_0^t \frac{\Gamma(u)\varphi(\tilde{t})d\tilde{t}}{\Gamma(u-\tilde{u}+1)\Gamma(\tilde{u})} e^{-\tilde{u}(t-\tilde{t})}[1-e^{-(t-\tilde{t})}]^{u-\tilde{u}}.$$

Using Yule's rate of addition of new population,

$$f(u, t) = \alpha e^{\alpha t} \frac{\Gamma(u)}{\Gamma(u-\tilde{u}+1)\Gamma(\tilde{u})} \int_0^t e^{-\tilde{u}T}[l-e^{-(t-\tilde{t})}]^{u-\tilde{u}}$$

$$= \alpha e^{\alpha t} \frac{\Gamma(u)}{\Gamma(u-\tilde{u}+1)\Gamma(\tilde{u})} \int_0^{e^{-t}} y^{\tilde{u}-1+\alpha}(1-y)^{u-\tilde{u}}dy .$$

If t is large,

$$f(u, t) = \alpha e^{\alpha t} \frac{\Gamma(u)}{\Gamma(u-\tilde{u}+1)\Gamma(\tilde{u})} \frac{\Gamma(\tilde{u}+\alpha)\Gamma(u-\tilde{u}+1)}{\Gamma(\alpha+u+1)}$$

$$= \alpha e^{\alpha t} \frac{\Gamma(\tilde{u}+\alpha)}{\Gamma(\tilde{u})} \frac{\Gamma(u)}{\Gamma(\alpha-u+1)} .$$

As expected, $f(u,t)$ is the Yule distribution truncated to $u \geq \tilde{u}$.

Now examine non-exponential rates $\varphi(\tilde{t})$. If $\tilde{u}$ is large (e.g, if it exceeds 100), the kernel function $e^{-ux}(1-e^{-x})^{u-\tilde{u}}$ becomes *extremely* peaked near its maximum for $e^{x} = u/\tilde{u}$. To approximate this kernel, expand $\log[e^{-\tilde{u}T}(1-e^{-T})^{u-\tilde{u}}]$ in Taylor series up to terms of second-order. The exponential of the result is

$$e^{-\tilde{u}T}(1-2^{-T})^{u-\tilde{u}} \sim \left\{ \left(\frac{\tilde{u}}{u} \right)^{(\tilde{u}/u)} \left(1 - \frac{\tilde{u}}{u} \right)^{(1-\tilde{u}/u)} \right\}^{u} \exp \left\{ - \frac{[T - \log(u/\tilde{u})]^2}{2(1/\tilde{u} - 1/u)} \right\}.$$

From this representation, it follows that the full integral that gives $f(u,t)$ can be replaced by the integral carried over a "four-standard-deviation" interval around $\log(u/\tilde{u})$, namely

$$\log\left(\frac{u}{\tilde{u}} \right) - 2\sqrt{\frac{1}{\tilde{u}} - \frac{1}{u}} \le t \le \log\left(\frac{u}{\tilde{u}} \right) + 2\sqrt{\frac{1}{\tilde{u}} - \frac{1}{u}}\,.$$

That is, for all practical purposes, $f(u,t)$, will not depend upon the firms that started at any other times.

As a consequence, suppose that the size u has exceeded a sizeable threshold, such as $\tilde{u} = 10{,}000$. Yule's model predicts that, from then on, $u(t)/10{,}000$ will differ from $\exp(t - \tilde{t})$ by, at most, a quantity of the order of $10{,}000^{-1/2} = 1\%$. That is, the growth of sizes *will for all practical purposes proceed with little additional randomness.*

To summarize, except for insignificant "noise," Yule's model makes the prediction that the size of the larger firms increases exponentially. Among the firms larger than "Smith and Co.," almost all will have reached the threshold size 10,000 before Smith and Co., and all the ranking by size of most larger firms is identical to their ranking by date of achievement of the threshold size 10,000.

It is easy to ascertain that the same conclusion will be reached if the definition of Yule's process is modified, as long as the modification does not affect the fundamental differential equations of that process. In all cases, U exhibits a stochastic transient during take-off, lasting as long as a firm is small. However, if and when its size becomes considerable, all sources of new randomness will have disappeared. This makes Yule's model indistinguishable in practice from Fermi's with zero death-rate, and there is little gain from Yule's far heavier mathematics. In Simon 1955, the independent variable is the actual total population k at time t, but – after a short transient – growth is non-random.

3.3 A "Fermi-like" model involving permanent diffusion

To eliminate the principal drawback of the model in Fermi 1949, as applied to economics, exponential lifetimes can be combined with growth that allows *random* increments of $\log_e U$. When this argument was presented to my IBM colleague M. S. Watanabe, he noted it could be applied to the cosmic-ray problem, but my model turned out to be simply a variant of Fermi's approach, as defined by the following assumptions.

Let dt and $\sigma\sqrt{dt}$ be the mean value and the standard deviation of the change of V per time increment dt. Let t be the time elapsed since the moment when a firm exceeds the size $\tilde{u}$, i.e., when V first exceeds the value 0. The distribution of V after the time t will be Gaussian, i. e., the probability that $v \leq V \leq v + dv$ will be

$$\frac{1}{\sigma\sqrt{2\pi t}} \exp[-\frac{(v-t)^2}{2t\sigma^2}]dv.$$

Now, the probability that $v < V \leq v + dv$, regardless of "age," will be

$$\frac{\lambda}{\sigma\sqrt{2}\,\pi}\left\{\int_0^\infty t^{-1/2} \exp[-bt - \frac{(v-t)^2}{2t\sigma^2}]dt\right\}dv.$$

A result in p. 146 of Bateman 1954, reduces this probability to

$$\frac{\lambda}{\sqrt{2\sigma}}(b + \frac{1}{2\sigma^2})^{-1/2} \exp\left\{\frac{v}{\sigma^2} - \frac{v\sqrt{2}}{\sigma}\sqrt{b + \frac{1}{2\sigma^2}}\right\},$$

which is exponential distribution with

$$\alpha = -\frac{1}{\sigma^2} + \frac{\sqrt{2}}{\sigma}\sqrt{b + \frac{1}{2\sigma^2}}.$$

To check that $\alpha > 0$, multiply the positive expression by

$$\tilde{\alpha} = \frac{1}{\sigma^2} + \frac{\sqrt{2}}{2^\sigma}\sqrt{b + \frac{1}{2\sigma^2}},$$

The product is $\alpha\tilde{\alpha} = 2b/\sigma^2 > 0$, therefore $\alpha > 0$, as it should.

The value $\alpha = 1$ corresponds to

$$\sqrt{2b+1}\,/\sigma^2 = \sigma(1 + \frac{1}{\sigma^2}) = \sigma + \frac{1}{\sigma} \text{ or } \sigma^2 = 2(b-1).$$

Again $\alpha = 1$ corresponds to the martingale relation for U. Indeed,

$$\exp\,[dt + (1/2)\sigma^2 dt] \sim 1 + [1 + \sigma^2/2]dt,$$

when Z is the expected value of a variable whose logarithm is Gaussian with mean dt and standard deviation $\sigma\sqrt{dt}$. Hence, whichever b,

$$E[U(t+1)\,|\,u(t)] \sim u(t)(1 - bd)[\,+ dt(1 + \sigma^2/2)]$$
$$\sim u(t)[1 + (\sigma^{-2}/2 + 1 - b)dt],$$

and the martingale relation indeed requires $\sigma^2 = 2(b-1)$.

3.4 Aström's diffusion model

While studying a problem of control engineering, Aström 1962 introduced the continuous time variant of the difference equation

$$U(t+1) = (1 - m' + G)U(t) + m''.$$

Here $m', m'' > 0$ and $\sigma > 0$ are constants, and G is a Gaussian variable of mean zero and unit standard deviation. For large u, this equation is another variant of random proportionate effect. However, it presents the originality that the reflecting barrier is replaced by the correction expressed by the positive "drift" m'', which is independent of u and negligible for large u. This gives interest to the equation. Since we expect the density to be asymptotically scaling, it is reasonable to try as a solution a density of the form

$$f(u) = \varphi(u)u^{-(\alpha+1)},$$

where $\varphi(u)$ rapidly tends to a limit as u increases.

To avoid advanced mathematics, let us first determine α by requiring $u(t+1)$ and $(1 - m'G\sigma)u(t)$ to have the same distribution for large u. This can be written as

$$u^{-(\alpha+1)} = \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty w^{-1} \exp\left\{ - \frac{[u - (1 - m')w]^2}{2\sigma^2 w^2} \right\} w^{-(\alpha+1)} dw;$$

defining y as u/w, this requirement becomes

$$u^{-(\alpha+1)} = u^{-(\alpha+1)} \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty y^\alpha \exp\left\{ - \frac{[v-(1-m')]^2}{2\sigma^2} \right\} dy.$$

In the limit where the basic equation changes from difference to differential, m' and σ^2 are small. Letting $\psi(y) = \log_2 v - [v-(1-m')]^2(2\sigma^2)^{-1}$, we see that $\exp[\psi(y)]$ is non-negligible only in the neighborhood of its maximum, which occurs approximately for $y = \tilde{y} = 1 - m' + \alpha\sigma^2$. Near $\tilde{y}$, $\psi(y) \sim \psi(\tilde{y}) + (y-\tilde{y})\psi'(\tilde{v}) + (1/2)(v-\tilde{v})\psi'(\tilde{v})$, hence the steady-state condition:

$$1 = \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty \exp[\psi(v)]dy = \exp[\psi(\tilde{v})]\sigma^{-1}[\psi''(\tilde{v})]^{-1/2}.$$

Easy algebra yields

$$\alpha = 1 + \frac{2m'}{\sigma^2}.$$

We again encounter the increasingly familiar relation between the sign of $\alpha - 1$ and that of the regression of $u(t+1)$ on $u(t)$. In particular, $\alpha = 1$ corresponds to the "martingale" case $m' = 0$.

In the present case, however, $\varphi(u)$ is determined by the equation itself, therefore, the model does not need special boundary conditions at $u = 0$. Suffices to say that the Fokker-Planck equation yields

$$\varphi(u) = \exp(-2m''/\sigma^2) = \exp(-\beta/u),$$

so that, normalizing the density $f(u)$ to add to one, we obtain

$$f(u) = \frac{\beta^\alpha}{\Gamma(\alpha)} u^{-(\alpha+1)} \exp(-\beta/u).$$

a) U_A is the inverse of a Gamma-variable of exponent α. Hence, $\Sigma 1/u_n$ is a "sufficient statistic" for the estimation of α from a set of sample values u_n. Unfortunately, this statistic is unusable, being overly dependent upon values that are small and fall in the range where data do not exactly follow the inverse of a gamma distribution. This exemplifies the difficulties which are encountered in the statistical theory of the estimation of the scaling exponent.

b) For $\alpha = 1$, U_A reduces to the Fréchet distribution that rules the largest of N identical scaling variables of exponent 1.

c) For the special value $\alpha = 1/2$, U_A happens to be identical to the L-stable distribution of exponent $\alpha = 1/2$ and of maximum skewness, in other words identical to the limit of the distribution of $(1/N^2)$-th of the sum of N identical scaling variables of exponent $1/2$.

&&&&& PRE-PUBLICATION APPENDICES &&&&&

APPENDIX I (M 1964o): RANDOM WALKS, FIRE DAMAGE, & RISK

Actuarial science accumulated a substantial store of knowledge. This paper discusses the risk due to fire. Very similar mechanisms apply in many other problems, hence a more general goal is to illustrate why it is said that the scaling distribution constitutes a "source of anxiety for the risk theory of insurance."

I.1 Introduction

In somewhat simplified form, the following statement summarizes an empirical law established by Benckert & Sternberg 1957.

"The damage a fire causes to a house follows the scaling distribution."

That is, suppose that the damage is greater than a minimum threshold $m = \$20$ and smaller than M, defined as the maximum destroyable amount of the building. Then

$$\text{for } m < x < M, \qquad \Pr\{\text{damage } \geq x\} = x^{-\alpha} m^{\alpha};$$
$$\text{but} \qquad \Pr\{\text{damage } = M\} = M^{-\alpha} m^{\alpha}.$$

This law applies to all classes of Swedish houses outside of Stockholm. The reported values of α range between 0.45 and 0.56; we shall take $\alpha = 1/2$ and investigate the consequences.

I.2 A model of fire damage amount

The scaling distribution of exponent 0.5 plays a central role in probability theory: it is the distribution of the returns to equilibrium in the game of tossing a fair coin. This theory is developed in most textbooks of probability (such as Feller 1950), and can be translated into insurance terms.

Our first assumption is that the intensity of a fire is characterized by a single number, designated by U, which can take only integer values: there is no fire when $U=0$; a fire starts when U becomes equal to 1, and it ends when either U becomes equal to 0 again, or when all that can possibly burn has already burned out.

We also assume that, at any instant of time, there is a probability $p=1/2$ that the fire encounters new material so that its intensity increases by 1, and there is a probability $q=1/2$ that the absence of new materials or the action of fire-fighters decreases the intensity by 1. In the preceding statement, 'time' is to be measured by the extent of damage. If there is no finite maximum extent of damage, and no lower bound to recorded damages, the duration of a fire will be an even number given by a classical result concerning the return to equilibrium in coin-tossing:

$$\Pr\{\text{duration of a fire} = x\} = 2\binom{1/2}{x/2}(-1)^{x/2-1}.$$

Except for the first few values of x, this expression is proportional to $x^{-3/2}$. Damages smaller than the minimum threshold m are not properly recorded. Hence, the duration of a fire (i.e., the extent of damage) will be given by the scaling distribution with exponent $1/2$:

$$\Pr\{\text{duration of a fire } > x > m\} = (x/m)^{-1/2}.$$

Finally, take account of the fact that the fire *must* end if and when the whole house has burnt out. We see that the prediction of the above argument coincides precisely with the Benckert-Sternberg finding.

I.3 Relations involving the size of the property and the expected amount of the damage due to fire

It is easy to compute the expected value of the random variable considered in Section 1. One finds

$$\text{Expected fire damage } = \int_m^M x(1/2)x^{-(1/2+1)}m^{1/2}dx + M(M/m)^{-1/2}$$
$$= 2\sqrt{Mm} - m.$$

This value tends to infinity as $M \rightarrow \infty$.

On the other hand, according to von Savitsch & Benktander 1953, the expected number of fires per house in the course of a year is a linear function of M. If this is indeed so, it would imply that, for large values of M, the rate of insurance should be proportional to $\sqrt{M}$.

When the distribution of property sizes M is known, a simple argument yields

$$\Pr\{\text{damage per fire} > x\} = \Pr\ \{M > m\}(x/m)^{-1/2}.$$

Moreover, if von Savitsch & Benktander 1953 is correct, one has

$$d \Pr\{\text{fire damage per year } > x\} = Cxd \Pr\ \{\text{ damage } > x\}$$
$$= Cxd\left[\Pr\ \{M > m\}(x/m)^{-1/2}\right].$$

Let the distribution of M be itself scaling with the exponent α^*. This is the case in all kinds of liability amounts. The distribution of damage in a single fire will then be scaling with exponent $\alpha^* + 1/2$; if $\alpha^* > 1/2$, the distribution of damage per year will be scaling with exponent $\alpha^* - 1/2$. This demonstrates that the mathematical manipulations based on the scaling distribution are especially convenient.

I.4 Generalization

The random walk with $p = q = 1/2$ represents a kind of equilibrium state between the fire and the fire-fighters. If the quantity of combustible property were unbounded, such a random walk will surely die out, although its expected duration would be infinite.

To the contrary, if $p > q$, the fire-fighting efforts would be inadequate, and there would be a nonvanishing probability that the fire continue forever.

If $p < q$, the probability of the fire running forever would again be zero, and the expected duration of the fire would be finite. The law giving the duration of the fire would then take the form:

$$d\,\Pr\{\text{duration of a fire } \geq x > m\} \sim c^* \exp(-cx)x^{-3/2}dx,$$

where c^* and c are two constants depending upon m, p, and q. If x is actually bounded and $q-p$ (and hence c) is small, the above formula will be indistinguishable in practice from a scaling distribution with an exponent slightly greater than 1/2; this is perhaps an explanation of the more precise experimental results of Benckert and Sternberg.

Generalizations of the random walk provided by many classical models of diffusion may be rephrased in terms of fire risks. A random walk model in which the intensity of the fire goes up, down, or remains unchanged would only change the time scale.

I.5 Another application

The results of Benckert & Sternberg 1957 strongly recall those of Richardson 1960 and Weiss 1963, and the model that has been sketched above is easy to translate in terms of Richardson's problem. It would be fascinating to ponder how relevant that translation may be.

Annotation to Appendix I: The importance of the scaling distribution in the theory of risk. The topic of Appendix I may seem narrowly specialized, but many risks against which one seeks insurance or reinsurance do follow scaling distributions. The sole reason for focussing on fire damage was the abundance and quality of the data on Swedish wooden houses.

Until recently, however, no one faced such risks squarely, quite to the contrary. Once, an insurance representative visited to report that his industry was bound one day to face risks following the scaling distributions, but he never called again. More significantly, a (short-lived) manager I once had at IBM pointedly described the present paper as childish and insignificant in comparison with the bulk of my work. Events suggest, to the contrary, that this paper was a quiet long-term investment. Important occurrences of very long-tailed risk distribution that are being investigated include Zajdenweber 1995ab.

APPENDIX II (M 1974d): THE NUMBER OF MUTANTS IN AN OLD CULTURE OF BACTERIA

Luria & Delbrück 1943 observed that, in old cultures of bacteria that have mutated at random, the distribution of the number of mutants is extremely long-tailed. Here, this distribution is derived (for the first time) exactly and explicitly. The rates of mutation will be allowed to be either positive or infinitesimal, and the rate of growth for mutants will be allowed to be either equal, greater or smaller than for nonmutants. Under the realistic limit condition of a very low mutation rate, the number of mutants is shown to be a L-stable random variable, of maximum skewness, β, whose exponent α is essentially the ratio of the growth rates of nonmutants and of mutants. Thus, the probability of the number of mutants exceeding the very large value m is proportional to $m^{-\alpha-1}$; it is "asymptotically scaling." The unequal growth rate cases $\alpha \neq 1$ are solved for the first time. In the $\alpha = 1$ case, a result of Lea & Coulson is rederived, interpreted, and generalized. Various paradoxes involving divergent moments that were encountered in earlier approaches are either absent or fully explainable. The mathematical techniques used are standard and will not he described in detail; this paper is primarily a collection of results.

II.1 Introduction

Let the bacteria in a culture grow, and sometimes mutate, at random, for a long time. In an occasional culture, the number of mutants will be enormous, which means that "typical values," such as the moments or the most probable value, give a very incomplete description of the overall distribution. Also, when the same mutation experiment is replicated many times, the number of mutants in the most active replica may exceed by orders of magnitude the sum of the numbers of mutants in all other replicas taken together. Luria & Delbrück 1943 first observed the above fact, and also outlined an explanation that played a critical role in the birth of molecular biology: The advantage of primogeniture is so great that the clone to which it gives rise has time to grow to a very much larger size than other clones in the same replica, or than the largest clone grown in a more typical replica in which no early mutation happened to be included.

Interest in expressing this explanation quantitatively, by describing the full distribution of the numbers of mutants, first peaked with Lea & Coulson 1949, Kendall 1952, Armitage 1952, 1953 and Bartlett 1966, but the solutions advanced were not definitive. Several investigators only calculated moments. Also, rates of growth were always assumed to be the same for mutants and non-mutants, and the rate of mutation to be very small. Kendall 1950 was very general; it may include in principle the results to be described, but not explicitly.

This short paper describes the whole distribution (for the first time) under assumptions that seem both sufficiently general to be realistic and sufficiently special for the solution to be exact and near explicit – in the sense that the Laplace transform of the distribution is given in closed analytic form.

The extreme statistical variability characteristic of the Luria & Delbrück experiment is also found in other biological experiments in progress. One may therefore hope that the present careful study, which settles the earliest and simplest such problem, would provide guidance in the future. It may help deal with some new cases when very erratic behavior is unavoidable, and in other instances, it may help avoid very erratic random behavior and thus achieve better estimates of such quantities as rates of mutation.

II.2 Background material: assumptions and some known distributions

Assumptions. (A) At time $t=0$ the culture includes no mutant, but includes a large number b_0 of nonmutants.

(B) Between times t and $t+dt$, a bacterium has the probability mdt of mutating.

(C) Back mutation is possible.

(D) Neither the mutants nor the nonmutants die.

(E) The rate of mutation m is so small that one can view each mutation as statistically independent of all others.

(F) Mutants and nonmutants multiply at rates that may be different. The scale of time is so selected that, between the instants t and $t+dt$, the probability of division is gdt for a mutant and dt for a nonmutant.

Non-mutants. A bacterium that mutates may be considered by its non-mutant brethren as having died. Therefore, $N(t,m)$, defined as the number of non-mutant bacteria at the instant t, follows the well-known "simple birth and death process" (see, e.g., Feller 1950, Vol. I, 3rd ed., p. 454). When $b_0 \gg 1$, the variation of $N(t,m)$ is to a good approximation deterministic

$$N(t,m) \sim EN(t,m) \sim b_0 e^{t(1-m)}.$$

Non-random clones. A clone being all the progeny of one mutation, denote by $K(t,m)$ the number of clones at the instant t. From Assumption (E), $K(t,m)$ is so small relative to $N(t,m)$ that different mutations can be considered statistically independent. It follows that $K(t, m)$ is a Poisson random variable of expectation

$$m\int_0^t fN(s, m)ds = b_0 m(l - m)^{-1}[e^{t(1-m)} - 1].$$

Random clones. Denote by $Y(t, m, g)$ the number of mutants in a clone selected at random (each possibility having the same probability) among the clones that have developed from mutations that occurred between the instants 0 and t. The distribution of $Y(t, m, g)$ will be seen to depend on its parameters through the combinations e^{gt} and $\alpha = (1-m)/g$. Since eventually we shall let $m \to 0$, α nearly reduces to the ratio of growth rates, $1/g$. One can prove that, after a finite t,

$$\Pr\{Y(t,m,g) \geq y\} = \alpha[1 - e^{-t(1-m)}]^{-1}\int_1^{\exp(gt)} fv^{\alpha - y}(v-1)^{y-1}dv.$$

In the case $\alpha = 1$, this yields explicitly

$$\Pr\{Y(t, m, g) \geq y\} = y^{-1}[1 - e^{-gt}]^{y-1}.$$

The generating function (g.f.) of Y, denoted $\tilde{Y}$, equals

$$\tilde{Y}(b, t, m, g) = \alpha[1 - e^{-t(1-m)}]\int_1^{\exp(gt)} fv^{-\alpha-1}\{[v(e^b - 1) + 1]^{-1}dv\}.$$

As $t \to \infty$, while m and g are kept constant, Y tends to a limit random variable $Y(\alpha)$ that depends only on α. When $\alpha = 1$,

$$\Pr\{Y(\alpha) = y\} = \int_0^1 fv(1-v)^{y-1}dv = \frac{1}{y(y+1)},$$

a result known to Lea & Coulson 1949. For all α,

$$\Pr\{Y(\alpha) \geq y\} = \alpha\frac{\Gamma(\alpha)\Gamma(y)}{\Gamma(\alpha + y)}.$$

For large y,

$$\Pr\{Y(\alpha) \geq y\} \sim \Gamma(\alpha+1)y^{-l-\alpha}.$$

The $Y(\alpha)$ thus constitutes a form of asymptotically "hyperbolic" or "Pareto" random variable of exponent α. The population moment $EY^h(\alpha)$ is finite if $h < \alpha$ but infinite if $h \geq \alpha$. For example, the expectation of $Y(\alpha)$ is finite if, and only if, $\alpha > 1$ and the variance is finite if, and only if, $\alpha > 2$. Infinite moments are a vital part of the present problem.

II.3 The total number of mutants

$M(t, m, g)$ will denote the number of mutant bacteria at the instant t. Thus, $M(0, m, g) = 0$, and

$$M(t, m, g) = \sum_{k=1}^{K(t,m,g)} Y_k(t, m, g).$$

Denote its g.f. by $\tilde{M}(b, t, m, g)$. Since K is a Poisson random variable of expectation EK, $\log \tilde{M}(b, t, m, g) = EK[\tilde{Y}(b, t, m, g) - 1]$.

The distributions of K and Y both depend on t (and are therefore interrelated). For this reason, the standard theorems concerning the limit behavior of sums (Gnedenko & Kolmogorov 1954, Feller 1950, Vol. II) are not applicable here. Fortunately, the special analysis that is required is straightforward. An approximate formal application of the standard theorems – letting the Y converge to the $Y(\alpha)$ and then adding K of them – would be unjustified, but some of its results nevertheless remain applicable. (Some of the paradoxes encountered in the analyses circa 1950 are related to cases where inversion of limit procedures is unjustified.)

One correct formal inference concerns the choices of a scale factor $S(K)$ and location factor $L(K)$, so as to ensure that the probability distribution of $R = S(K)[M - L(K)]$ tends to a nondegenerate limit as $K \to \infty$. Setting

$$\delta = [b_0 m(l-m)^{-1}]^{-1/\alpha},$$

the scale factors are as follows:

$$\begin{array}{rll} \alpha > 2 : & L(K) = EY & ; \quad S(K) = (EK)^{-1/2} \\ 1 < \alpha < 2 : & L(K) = EY & ; \quad S(K) = (EK)^{-1/\alpha} \\ \alpha = 1 : & L(K) = \log EK & ; \quad S(K) = (EK)^{-1/\alpha} = \delta e^{-gt} \\ \alpha < 1 : & L(K) = 0 & ; \quad S(K) = (EK)^{-1/\alpha} \end{array}$$

With these scale factors, the limits depend on α, as follows.

The case $\alpha > 2$**.** Here, $\lim_{t \to 0} (EK)^{-1/2}(M - EM)$ can be shown to be Gaussian. Nothing original!

The case $\alpha < 1$**.** Here, $\lim_{t \to 0} (EK)^{-1/\alpha}\Sigma_{k=1}^{K} Y_k$ can be shown to have a g.f. equal to

$$\tilde{R}(b, \infty, m, g) = \exp\left\{\alpha \int_0^{\delta} fbw^{-\alpha}(bw + 1)^{-1} dw\right\}.$$

The corresponding limit r.v., call it $R(\infty, \alpha, \Delta)$, appears for the first time (to the best of my knowledge) in the present context. The fact that it is nondegenerate (not reduced to either 0 or ∞) confirms that the above standardization was well chosen. Moreover, near $b = 0$, $\tilde{R}(b, \infty, m, g)$ has a good expansion in Taylor series, so all moments of $R(\infty, \alpha, \delta)$ converge. However, this convergence has limited significance because, in actual practice, m is extremely small and Δ is extremely large, so the moments of $R(\infty, \alpha, \delta)$ are themselves enormous and tell us very little about the distribution of $R(\infty, \alpha, \delta)$. On the other hand, as was shown by Luria & Delbrück, the birth and mutation process is illuminated by a sort of "diagonal" procedure. In this procedure, while t increases, m and/or b_0 change in such a way that $\Delta \to \infty$ while $g > 1$, while α remains between 0 and 1. As a result, the function $\tilde{R}$ tends towards

$$\exp\left\{-\alpha b^{\alpha}\int_0^{\infty} fz^{-\alpha}(1 + z)^{-1} dz\right\} = \exp\left[-b^{\alpha}\alpha\pi / \sin(\alpha\pi)\right].$$

This is an unfamiliar expression for a well-known function, namely the g.f. a positive Lévy stable random variable of maximal skewness, $\beta = 1$ (Gnedenko & Kolmogorov 1954, Feller 1950, Volume II). It is also the limit one would have obtained for $K \to \infty$ if we let $Y \to Y(\alpha)$ and then consider the similarly standardized sum of K independent random variables of the form $Y(\alpha)$. In the limit, all the moments of order $h > \alpha$ (including all integer moments) diverge. As a practical consequence, the statistical estimation of m and g from values of M is both complicated and unreliable. Traditionally, statistics has relied heavily on sample averages, but when the population averages are infinite, the behavior of the sample averages is extremely erratic, and any method that involves them must be avoided.

***The case* $1 < \alpha < 2$.** Here, $\lim_{t \to \infty} (EK)^{-1/\alpha} \sum_{k=1}^{K} [Y_k - EY_k]$ can be shown to have the g.f.

$$\exp \left\{ \alpha \int_0^\delta f b^2 w^{-\alpha+1} (bw+1)^{-1} dw \right\}.$$

As $\Delta \to \infty$, this function tends towards

$$\exp [- b^\alpha \alpha \pi / \sin (\alpha \pi)],$$

which is again the g.f. of a stable random variable of exponent α and maximal skewness, i.e., of the limit of a similarly standardized sum of K independent random variables of the form $Y(\alpha)$. The theory of these limits is well known, but their shape is not; see M 1960i{E10}, M & Zarnfaller 1959.

***The case* $\alpha = 1$.** Here, $\lim_{t \to \infty} (EK)^{-1} \sum_{k=1}^{K} [Y_k - \log EK]$ can be shown to have the g.f.

$$\exp [b \log b + b \log (1 + 1/b\Delta)].$$

As $\Delta \to \infty$, this function tends towards $\exp [b \log b]$, corresponding to the stable density of exponent $\alpha = 1$ and maximal skewness, $\beta = 1$, sometimes called the "asymmetric Cauchy" density. It was derived (but not identified) in Lea & Coulson 1949, which concerns the case when the mutation rate m is small, and the growth rates for the mutants and the nonmutants are equal, so that $\alpha \sim 1$.

II.4 The total number of bacteria and the degree of concentration

Designate by $B(t, m, g) = N(t, m) + M(t, m, g)$ the number of bacteria of either kind at the instant t. In the straightforward special case $g = 1$, the function $B(t, m, g)$ follows a "simple birth process" or "Yule process"; see Feller

1968. When $b_0 \gg 1$, the growth of B is for all practical purposes deterministic and exponential, meaning that $B(t) \sim b_0 e^t$.

In the cases $g \neq 1$, things are much more complex, but much of the story is told by the orders of magnitude for large t: $M(t, m, g) \sim e^{gt}$ and $N(t, m, g) \sim e^{(l-m)t}$.

When $\alpha < 1$, $B(t, m, g) \sim M(t, m, g)$, meaning that the mutants – which we know are subject to very large fluctuations – become predominant.

When $\alpha > 1$, $B(t, m, g) \sim e^{(1-m)t}$ with little relative fluctuation, since the random factor that multiplies t is nearly the same as it would be if there had been no mutation. Thus, the dependence of B upon g is asymptotically eliminated.

Now examine the "degree of concentration" of the mutants, namely the ratio ρ of the number of mutants in the largest of the K clones in a replication, divided by the total number of mutants in the other clones of this replication.

Luria & Delbrück discovered that an alternative ratio can be quite large. This ratio is the number of mutants in the largest among H replications, divided by the sum of the number of mutants in the other of the replications. It can be shown that the above two ratios follow the same distribution, so it will suffice to study the first, beginning with two extreme cases.

Consider the case where a mutation brings enough competitive disadvantage and enough decrease in the growth rate to result in $\alpha \gg 2$. Then, the number of young and small clones increases much faster than the size of the single oldest clone in an experiment. Therefore, it is conceivable that a negligible proportion of mutants will be descended from either this oldest clone or any other single clone. This expectation is indeed confirmed. We know that if $\alpha > 2$ the quantity $M(t, m, g)$ tends towards a Gaussian limit, so the contribution of any individual addend Y_k to their sum is indeed negligible.

Now consider the opposite extreme case, where a mutation brings enough competitive advantage and enough increase in the growth rate to result in $\alpha \ll 1$. Then, the size of the oldest clone in an experiment (corresponding to the earliest mutation) grows much faster than the number of fresh clones. It is conceivable, therefore, that the largest clone in an experiment is comparable in size to the sum of all the other clones. An appreciable proportion of the mutants could descend from the single largest clone. This expectation is indeed confirmed in two different ways. First, it has been shown by Darling 1952 (see also Feller 1950, Volume II, p. 439, problem 20), that if $\alpha < 1$, the ratio ρ does not tend to zero as $K \rightarrow \infty$. Rather, its distribution tends to a nondegenerate limit, and $E(1/\rho)$ has the nondegenerate limit $\alpha/(1-\alpha)$. As α varies from 0 to 1, this limit varies from 0 to ∞. That is, when mutation causes an enormous increase in growth rate so that the value of α is very small, $1/\rho$ is nearly 0 on the average, and the limit value of ρ for large K is often very large. When, on the contrary, mutation brings very slight advantage, so that α is very nearly 1, $1/\rho$ is very large on the average and ρ tends to be small. But its values can be seen to be widely scattered, and large values are not unlikely.

The limits described by the preceding theorem are approached rapidly when α is small, but very slowly when α is near 1. Thus, in the Lea & Coulson case corresponding to $\alpha = 1$, the value of K must be very large for ρ to become negligible. For ordinary values of K, the typical value of ρ is non-negligible, and the dispersion of ρ around this typical value is very wide, so that the original argument of Luria & Delbrück is justified.

Note: The formulae on random clones described in Section 2 restate some results obtained in Yule 1924. Yule's paper is known to have introduced the birth process, but has otherwise been neglected. It treated a nominally different problem: our "growth" was his "increase in the number of species in one genus," our "mutation" was his "starting of a new genus." Simon 1954 attempted to modify Yule's argument to obtain diffusion from less strong first-order assumptions. This attempt unavoidably failed.

Editorial comment. This reprint corrects a horrendous typographical error. In the original, both in the abstract and at the end of Section 2, the exponent $-\alpha$ was printed as $-\alpha-1$.

E9

A case against the lognormal distribution

✦ **Abstract.** The lognormal distribution is, in some respects, of great simplicity. This is one reason why, next to the Gaussian, it is widely viewed as the practical statistician's best friend. From the viewpoint described in Chapter E5, it is short-run concentrated and long-run even. This makes it the prototype of the state of slow randomness, the difficult middle ground between the wild and mild state of randomness. Metaphorically, every lognormal resembles a liquid, and a very skew lognormal resembles a glass, which physicists view as a very viscous liquid.

A hard look at the lognormal reveals a new phenomenon of delocalized moments. This feature implies several drawbacks, each of which suffices to make the lognormal dangerous to use in scientific research. Population moments depend overly on *exact* lognormality. Small sample sequential moments oscillate to excess as the sample size increases. A non-negligible concentration rate can only represent a transient that vanishes for large samples. ✦

AFTER LÉVY, ZIPF AND PARETO were described as providing inspiration to scaling and fractal geometry, Chapter E4 also listed a widely-followed nemesis. Robert Gibrat, the author of *Les inégalités économiques* (Gibrat 1932), remains foremost among the many who claim that economic inequalities (presumably all of them) can be described and explained by the lognormal. As is well-known, Λ is called lognormal when $G = \log\Lambda$ is Gaussian. Section 1 recalls the basic facts about the lognormal, and describes in parallel several reasons why it is liked, and counterbalancing reasons why its assets are misleading. In a word: this distribution should be avoided. A major reason, elaborated in Section 2, is that a near-lognormal's *population* moments are overly sensitive to departures from exact lognormalities. A second major reason, elaborated in Section 3, is that the *sample moments* are not to be trusted, because the sequential sample moments oscillate with sample size in erratic and unmanagable manner.

Once again, the preceding paragraph and the rest of this book avoid endless and tiresome repetition of the terms "density," "distribution," "random variable," and the like. It is better to deal with such words as "Gaussian," "lognormal," "Bernoulli," "Poisson," and "scaling" as common names. For example, if there is no loss of intelligibility and the context allows, "lognormal" will be a synonym either of "lognormal distribution," or of "lognormal random variable." Only a slip of the pen can make me use the word "normal" as synonym of "Gaussian." The reason is that in this book the norm is randomness that used to be called "anomalous" and that Chapter E5 describes as "wild." Since the word "lognormal" will not change, I try not to think about its undesirable root.

Some statisticians tell practicing scientists that there is no need to deal with many different random variables, because every variable can be transformed into a Gaussian ... or even a uniform variable. This transformation is discussed and dismissed in Chapter E5.

The lognormal claims to represent both the bell and the tails in distribution of personal income, though only roughly. The scaling is concerned with the tail only, but claims to represent that part in more precise, more enlightening and more useful fashion. The L-stable is claimed in Chapter E10 to represent the tails well and the bell, reasonably. More generally, the lognormal, the scaling and other narrower-purpose distributions continually compete in the many fields of science where skew long-tailed histograms are a fact of life and concentration ratios are not small. My research life began by facing the conflict between the lognormal and the scaling in the study of word frequencies.

The endless conflict between the lognormal and the scaling is illustrated on Figure 1. It is annoying and boring, and its very existence is irritating and implies that the two distributions differ less than their vastly different analytic forms would suggest. Section 3 will show that such is indeed the case: many lognormals can be approximated over wide spans of values of the variable by judiciously chosen scaling, and conversely.

This assertion does *not* endorse the claim by statisticians who despise log-log plots, that "everyone knows that *every* log-log plot is straight; therefore, a straight log-log plot cannot mean anything." If this were true, the scaling distribution could not be conceivably proved wrong ("falsified" in Popper's terminology.) But it would not be a candidate for serious scientific discourse. Be that as it may, all log-log plots *are not* straight.

The lognormal's properties helped Chapter E5 draw a deep difference between mild, slow, and wild "states of randomness." The Gaussian is mildly random. The scaling thrives on its own wildness: it faces the many difficulties due to skewness and long-tailedness, and this is why it is usable and realistic. The lognormal lies between the mild and the wild, in the state of "slow randomness;" it even provides an excellent illustration of this intermediate state and its pitfalls. It is beloved because it passes as mild: moments are easy to calculate and it is easy to take for granted that

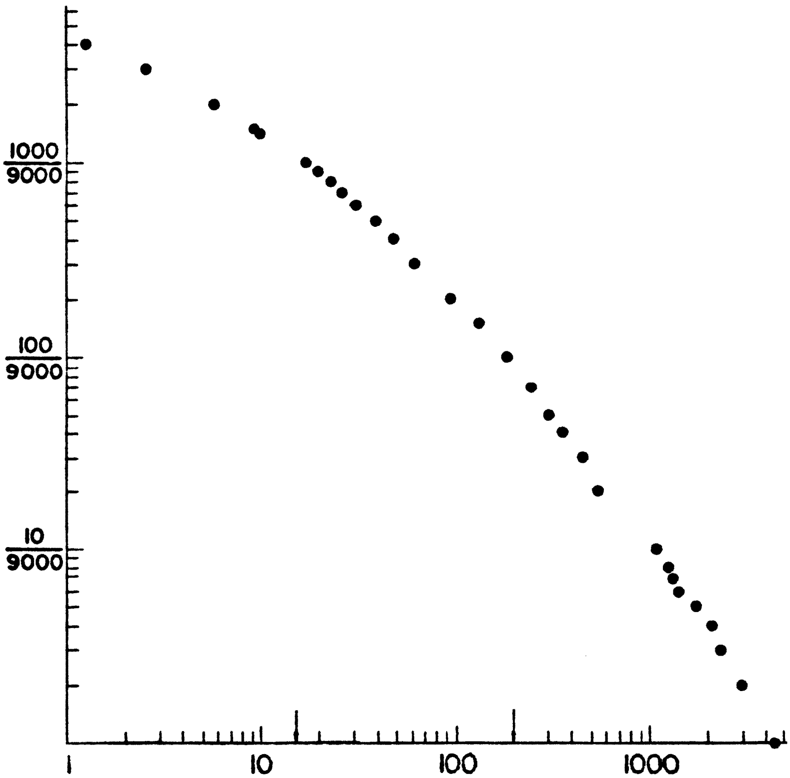


FIGURE E9-1. Illustration of how a sample of a very skew lognormal random variable can "pass" as being from a scaling. The abscissa is $\log x$, the ordinate is $\log \text{Fr}(X > x)$, with Fr the frequency in a sample. This is the plot of the distribution (cumulated from the tail) of a sample of 9 000 lognormal variables X, where $\log X$ has zero mean and a standard deviation equal to $\log_e 10$. The graph "passes" as straight. The arrow near $x = 12$ marks the mean, and the arrow near $x = 150$, the mean plus one standard deviation.

they play the same role as for the Gaussian. But they do not. They hide the difficulties due to skewness and long-tailedness behind limits that are overly sensitive and overly slowly attained.

In the metaphor of "states of randomness," the contrast between liquid and solid leaves room for glasses. These hard objects used to be viewed as solids, but their properties are *not* explained by the theory of solids (as a matter of fact, they remain poorly explained). In time, strong physical reasons arose for viewing glasses as being very viscous fluids. The glassy state is a convenient metaphor to characterize the lognormal, but also a challenge that will be taken up in this chapter. Therefore, the lognormal's wondrous properties are irrelevant and thoroughly misleading; it *is not* the statisticians' best friend, perhaps even their worst one. For those reasons, and because of the importance of the topic, this chapter was added to bring together some points also made in other chapters.

Given the serious flaws of the lognormal, there are strong *practical* reasons to prefer the scaling. But scientists learn to live with practical difficulties, when there are solid *theoretical* reasons for doing so. The scaling has diverse strong theoretical points in its favor, while Chapter E8 shows that the usual theoretical argument in favor of lognormality is weak, incomplete and unconvincing. Unfortunately, the fields where the lognormal and the scaling compete lack convincing explanations.

Helpful metaphors. There are many issues that the scaling distribution faces straight on, but the lognormal distribution disguises under a veneer. The lognormal distribution is a wolf in sheep's skin, while the scaling density is a wolf in its own skin; when living among wolves, one must face them on their own terms.

References. The literature on the theory and occurrences of the lognormal is immense and I do not follow it systematically. Aitchison & Brown 1957 was up-to-date when I took up this topic, and I marvelled even then at the length of the mathematical developments built on foundations I viewed as flimsy. See also Johnson, Kotz & Balakrishnan 1994.

A warning against a confusion between "lognormal" and "logBrownian." To my continuing surprise, "lognormal" is also applied here and there to the "logBrownian" model according to which log (price) performs a Brownian motion, à la Bachelier 1900. The only feature common to those two models (not counting the evidence against both) is slight: the logBrownian model asserts that where a price is known at time $t=0$, its value at time t is a lognormal random variable.

1. INTRODUCTION

1.1 The lognormal's density and its population moments

Let $V = \log \Lambda$ be Gaussian, that is, of probability density

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{(v-\mu)^2}{2\sigma^2} \right\}.$$

Then the probability density of the lognormal $\Lambda = \exp V$ is

$$p(\lambda) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{\lambda} \exp\left\{ -\frac{(\log\lambda-\mu)^2}{2\sigma^2} \right\}.$$

The simplest distinction between states of randomness (Chapter E5) involves the convexity of $\log p(\lambda)$ and the finiteness of the variance.

***The cup-convexity of the tail of* $\log p(\lambda)$.** For the lognormal, there is a "bell" where $\log p(\lambda)$ is cap-convex, and a tail where $\log p(\lambda)$ is cup-convex. Most of the probability is in the bell when σ^2 small, and in the tail when σ^2 is large. If generalized to other distributions, this definition sensibly states that the Gaussian has no tail. Because of the cup-convexity of $\log p(\lambda)$ in the tail, Chapter E5 calls the lognormal "long-tailed."

Finiteness of the moments. An easy classical calculation of $E\Lambda^q$ needed in the sequel yields

$$\begin{aligned} E\Lambda^q &= \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty \lambda^{q-1} \exp\left\{ -\frac{(\log\lambda-\mu)^2}{2\sigma^2} \right\} d\lambda \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty \exp\left\{ -\frac{(\log\lambda-\mu)^2}{2\sigma^2} + (q-1)\log\lambda \right\} d\lambda \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^\infty \exp\left\{ -\frac{(v-\mu)^2}{2\sigma^2} + qv \right\} dv. \\ &= \frac{1}{\sigma\sqrt{2\sigma}} \exp\left\{ \mu q + \frac{\sigma^2 q^2}{2} \right\} \int_{-\infty}^\infty \exp\left\{ -\frac{[v-(\mu+\sigma^2 q)]^2}{2\sigma^2} \right\} dv. \end{aligned}$$

Hence the following result, valid for all $q(-\infty < q < \infty)$

$$E\Lambda^q = \exp\left\{\mu q + \frac{\sigma^2 q^2}{2}\right\} < \infty \text{ and } [E\Lambda^q]^{1/q} = \exp\left\{\mu + \frac{\sigma^2 q}{2}\right\} < \infty.$$

Ways of normalizing Λ. One can set $\mu = 0$ by choosing the unit in which λ is measured. To achieve $E\Lambda = 1$, it suffices to set $\mu = -\sigma^2/2$. Using the notations $\mu = -m$ and $\sigma^2 = 2m$.

$$E\Lambda^q = \exp[-qm + q^2 m] = \exp[mq(q-1)].$$

In particular, $E\Lambda^2 = \exp(2m) = \exp\sigma^2$, and the variance is $v^2 = \exp(2m) - 1$.

Skewness and long-tailedness. The lognormal's skewness and kurtosis confirm that, as $m \rightarrow \infty$, the distribution becomes increasingly skew and long-tailed. But skewness and kurtosis are less telling than the above-mentioned notions of "bell" and "tail".

The reader is encouraged to draw several lognormal densities, normalized to $E\Lambda = 1$ and parametrized by the standard deviation v. On both sides of the point of coordinates 1 and $p(1)$, include an interval of length $2v$. As soon as $m > (\log 2)/2 \sim 0.35$, this interval extends to the left of the ordinate axis. This fact underlines the unrepresentative nature of the standard deviation, even in cases of moderate skewness.

This fact also brings to mind one of the deep differences that exist between physics and economics. In physics, moments of low order have a clear theoretical interpretation. For example, the population variance is often an energy that must be finite. In economics, to the contrary, the population variance is nothing but a tool of statistical analysis. Therefore, the only real interest is restricted to the insights that population moments can yield, concerning phenomena ruled by *sample moments*. This chapter will show that the predictions drawn from the lognormal are too confused to be useful while those drawn from the scaling are clear-cut.

1.2 Three main reasons why the lognormal is liked, and more-than-counterbalancing reasons why it should be avoided

An asset: the lognormal density and the formulas for its moments are very simple analytically. So are *products* of lognormals.

A more-than-counterbalancing drawback: the distributions of sums are unmanageably complicated. Dollars and firm sizes do not multiply; they add and subtract. But sums of lognormals are not lognormal and their analytic expressions are unmanageable. That is, the lognormal has invariance properties, but not useful ones.

This is a severe handicap from the viewpoint of the philosophy of invariances described in Chapter E1 and throughout this book. Once again, each scientific or engineering problem involves many fluctuating quantities, linked by a host of necessary relations. A pure curve-fitting doctrine proposes for each quantity the best-fitting theoretical expression, chosen in a long list of all-purpose candidates. But there is no guarantee at all that the separately best fitting expressions are linked by the relations the data must satisfy. For example, take the best fit to one-day and two-day price changes. The distribution of the sum of one day fit need not be analytically manageable, and, even if it is, need not be identical to the distribution of a two-day fit.

Major further drawback: Section 2 shows that the population moments of the lognormal are not at all *robust with respect to small deviations from* absolutely precise *lognormality.* Because of this lack of robustness, X being approximately Gaussian is not good enough from the viewpoint of the population moments of $\exp X$. The known simple values of $E\Lambda^q$ are destroyed by seemingly insignificant deviations. The technical reason behind this feature will be described and called "localization of the moments." Hence, unless lognormality is verified with absolute precision, the moments' values are effectively arbitrary.

The deep differences between the lognormal as an exact or an approximate distribution were unexpected and led to confusions even under the pen of eminent scientists. Few are the flaws in the *Collected Works* of Andrei N. Kolmogorov (1903-1987), but his influential papers on lognormality (especially in the context of turbulence) are deeply flawed. Hard work to correct those flaws led M 1972j{N14} and M 1974f{N15} to results on multifractals that overlap several fields of inquiry and greatly contributed to fractal geometry and the present discussion.

Another major drawback: Section 3 shows that the sequential sample moments of the lognormal behave very erratically. This additional drawback tends to prevent the first one from actually manifesting itself. The population moments of a lognormal or approximate lognormal will eventually be approached, but how rapidly? The answer is: "slowly."

When the lognormal Λ is very skew, sample size increases, the answer is that the sequential sample average undergoes very rough fluctuations, and does not reach the expectation until an irrelevant long-run (corresponding to asymptotically vanishing concentration). In the middle-run, the sample and population averages are largely unrelated and the formulas that give the scatter of the sequential sample moments of the lognormal are impossibly complicated and effectively useless. This behavior is best explained graphically, the Figure captions being an integral part of the text. Figure 2 uses simulated lognormal random variables. while Figure 3 uses data.

Powers of the lognormal being themselves lognormal, all sample moments are averages of lognormals. Their small, and medium sample variability is exteeme and *not* represented by simple rules deduced from lognormality. By contrast, the scaling interpolations of the same data yields simple rules for the very erratic sample variability. Erratically behaving sample moments and diverse other difficulties that the scaling distribution faces straight on, are characteristic of wild randomness.

A widely assumed asset: it is believed that the lognormal is "explained" by a random "proportional effect" argument. Aside from its formal simplicity, the greatest single asset of the Gaussian is that it is the limit in the most important central limit theorem. That theorem's limit is not affected by small changes in the assumptions, more precisely, limit Gaussianity defines a "domain of "universality," within which details do not count. Similarly, the lognormal is ordinarily viewed as being justified via so-called "proportionate effect" models. They represent log X as the sum of independent proportionate effects, then invoke the central limit theorem to conclude that log Z must be approximately Gaussian.

A more-than-counterbalancing drawback: the random proportional effect models yield the Gaussian character of log Λ *as an approximation and the conclusions concerning* Λ *cannot be trusted.* In most scientific problems, the lack of exactitude of central limit approximations makes little difference. The number of conceivable multiplicative terms of proportionate effect is not only finite (as always in science) but small. Therefore, the Gaussian involved in the limit theorem is *at best a distant* asymptotic approximation to a preasymptotic reality. When John Maynard Keynes observed that in the long-run we shall be all dead, he implied that asymptotics is fine, but economists should be concerned with what will happen in some middle run. Unfortunately, we deal with one of those cases where, because of the already-mentioned sensitivity, approximations are not sufficient.

Under the lognormal assumption, the basic phenomenon of industrial concentration must be interpreted as a transient that can occur in a small sample, but vanishes asymptotically. In an industry including N firms of lognormally distributed size, how does the relative size of the largest depend on N? This topic is discussed in Chapter E7 and E8.

In the long-run regime $N \to \infty$, the relative size of the largest of N lognormal addends decreases and soon becomes negligible. Hence, a sizeable relative size of the largest, could only be a transient and could only be observed when there are few firms. Furthermore, the formulas that deduce the degree of concentration in this transient are complicated, evade intuition, and must be obtained without any assistance from probability limit theorems.

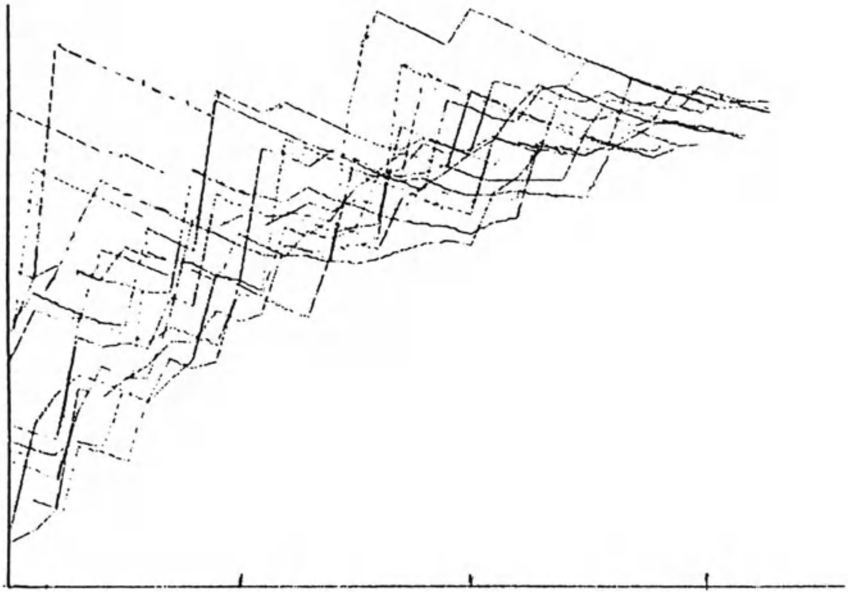


FIGURE E9-2. My oldest illustration of the erratic behavior of the sample averages of very skew approximately lognormal random variables. Several samples were generated, each containing over 10,000 values. Then the sample average $N^{-1}\sum_{\lambda=1}^{N} X_\lambda$ was computed for each sample, and plotted as a line.

Both coordinates are logarithmic. In an initial "transient" zone, the averages scatter over several orders of magnitude. The largest average is often so far removed from the others, that one is tempted to call it an outlier and to disregard it. The approximate limit behavior guaranteed by the law of large numbers is far from being approached. The expectation EX is far larger than the bulk of sample values X_n, which is why huge sample sizes are required for the law of large numbers to apply.

In addition, the limit depends markedly on the Gaussian generator. In this instance, $\log X_n = \sum I_n - 6$, where the I_n are 12 independent pseudorandom variables with uniform distribution. With a different approximation, the limit would be different, but the convergence, equally slow and erratic.

To the contrary, in an industry in which firm size is scaling, the relative size of the largest firm will depend little on the number of firms. Furthermore, the asymptotic result relative to a large number of firms remains a workable first-order approximation where the number of firms is not very large.

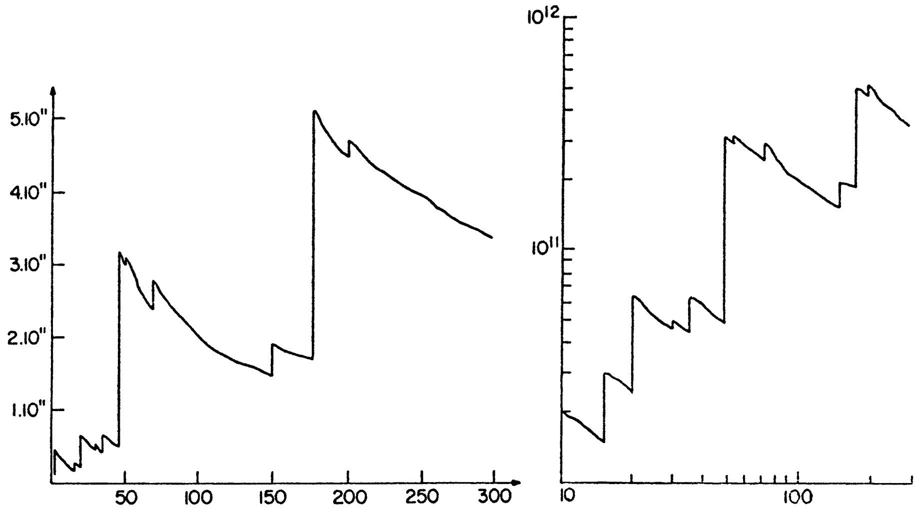


FIGURE E9-3. Illustration of the erratic behavior of the sample mean square of a set of very skew natural data, namely the populations of the 300 largest cities in the USA. This old graph was hand-drawn in 1986. The alphabetical order was picked as approximately random and $N^{-1}\Sigma X_n^2$ was computed for every value of N. The curve to the left uses linear coordinates in units of 10^{11}; the curve to the right uses log-log coordinates.

There is not even a hint of convergence.

In light of this Figure, examine two conflicting claims. Gibrat 1932 claims that the distribution of city populations is lognormal, and Auerbach 1913 that this distribution is scaling. It may well be that both expressions fit the histograms. But it is clear that the fitted lognormal only describes the asymptotic behavior of the sample mean square and gives no information until the sample size enters the asymptotic range. However, the sample of city sizes is exhaustive, and cannot be increased any further, hence the notion of asymptotic behavior is a figment of the imagination. To the contrary, the fitted scaling distribution does predict the general shape of this Figure.

Conclusion. Even in the study of the transients, it is better to work with the scaling approximation to the lognormal than with the lognormal itself. This scaling approximation makes one expect a range of sizes in which the concentration depends little on N.

The 3 and 4-parameter generalized lognormals. They will not be discussed here. To all the defects of the 2-parameter original, the generalizations add defects of their own. Simplicity is destroyed, the moments are equally meaningless and Gibrat's purported justifications, already shaky for the lognormal, lose all credibility when parameters are added.

2. THE POPULATION MOMENTS OF A NEAR-LOGNORMAL ARE LOCALIZED, THEREFORE OVERLY SENSITIVE TO DEPARTURES FROM EXACT LOGNORMALITY

2.1 Summary of a first argument against the lognormal

The expressions obtained in the Section 1.1 prove to be of little consequence unless the lognormal holds with exactitude beyond anything that any scientist or engineer can reasonably postulate for a statistical distribution. Otherwise, the classical and easily evaluated population moments are *devoid of practical relevance.*

2.2 Even when G is an acceptable Gaussian approximation of Z, the moments of e^G may drastically differ from the moments of e^Z

This sensitivity is a very serious failing. When a theoretical probability distribution is characterized by only a few parameters, a host of properties are intimately tuned to each other. It suffices to verify a few to predict the other. Moving on from a theoretical distribution to one obtained by fitting, one hopes that "small" errors of fitting yield small prediction errors. Such is, indeed, the case for the Gaussian G, but not for the lognormal $\Lambda = e^G$. The trouble is that the practical use of the lognormal consists of predictions that are very sensitive to departure of G from *exact* Gaussianity.

The sensitivity of the lognormal will not be proved theoretically, but will instead be illustrated by comparing a) the Gaussian, and the following near-Gaussian examples: b) a Bernoulli random variable B obtained as sum of K binomial variables bounded by $\max B$, c) a Poisson random variable, P, and d) a gamma random variable Γ obtained as the sum of γ exponentials. Textbooks prove that B, P and Γ can be made "nearly identical" to a normal G. The underlying concept of "near identity" is critical; for sound reasons, it is called "weak" or "vague." Let us show that it allows the moments of the "approximations" e^P, e^B and e^Γ to depend on q in ways that vary with the approximation, and do not match the patterns that is characteristic of e^G.

a) *The lognormal.* To match the Poisson's property that $EP = EP^2 = \rho$, we set $EG = \mu = \rho$ and $\sigma^2 = \rho$. It follows that $[E(e^G)^q]^{1/q} = \exp[\rho(1 + q/2)]$. Thus, $[E(e^G)^q]^{1/q}$ is finite for all q, and increases exponentially.

b) *The logBernoulli.* Here, $[E(e^B)^q]^{1/q} \leq \exp(K \max B)$. Thus, $[E(e^B)^q]^{1/q}$ is bounded; in the vocabulary of states of randomness expounded in Chapters E5, e^B is mildly random, irrespective of K, but this property is especially devoid of contents from the viewpoint of the small- and the middle-run.

c) *The logPoisson.* $[E(e^P)^q]^{1/q} = \exp[\rho(e^q - 1)/q]$. Thus, $[E(e^P)^q]^{1/q}$ is finite but increases more rapidly than any exponential. Like U_e, the lognormal e^P belongs to the state of slow randomness

d) *The log-gamma.* $E(e^{\Gamma/\alpha})^q = \infty$ when $q > \alpha$. Thus, $e^{\Gamma/\alpha}$ is of the third level of slow randomness when $\alpha > 2$, and is wildly random when $\alpha < 2$.

Expectations. By design, the bells of G and P are very close when ρ is large, but $E(e^G) = \exp(1.5\rho)$ and $E(e^P) = \exp(1.7\rho)$ are very different; this shows that the expectation is not only affected by the bell, which is roughly the same for G and P, but also by their tails, which prove to be very different.

The coefficients of variation. They are

$$\frac{E[(e^G)^2]}{[E(e^G)]^2} - 1 = e^\rho - 1 \text{ and } \frac{E[(e^P)^2]}{[E(e^P)]^2} - 1 = \exp[(e-1)^2\rho] - 1 \sim e^{3\rho} - 1.$$

The dependence on the tails is even greater for e^P than it is for $E\Lambda$.

Higher order moments differ even more strikingly. In short, it does not matter that a large ρ insures that B and P are nearly normal from the usual viewpoint of the "weak-vague" topology. The "predictive error" $E(e^P)^k - E(e^G)^k$ *is not* small. Less good approximations Z yield values of the moments $E(e^Z)^k$ that differ even more from $E(e^G)^k$.

Illustration of the appropriateness of the term "weak topology." In a case beyond wild randomness that is (thankfully) without application but serves as warning, consider $Z = \exp\tilde{V}_N$ where $\tilde{V}_N$ is a normalized sum of scaling addends with $\alpha \geq 2$. By choosing N large enough, the bells of $\tilde{V}_N$ and G are made to coincide as closely as desired. Moreover, $EV^2 < \infty$, hence the central limit theorem tells us that $\tilde{V}_N$ converges to a Gaussian G, that is, comes "close" to G in the "weak", "vague" sense. The underlying topology is powerful enough for the central limit theorem, but for $q > \alpha$ moments *cannot* be matched, since $E\tilde{V}_N^q = \infty$ while $EG^q < \infty$, and *all* positive moments $E \exp(q\tilde{V}_N)$ are infinite, due to the extraordinarily large values of some events that are so extraordinarily rare that they do not matter.

2.3 The moments of the lognormal are sensitive because they are localized, while those of the Gaussian are delocalized

The formula $E\Lambda^q = \exp(\mu q + \sigma^2 q^2/2)$ reduces all the moments of the lognormal to two parameters that describe the middle bell. However, let us consider a general U and take a close look at the integral

$$EU^q = \int u^q p(u)du.$$

For many cases, including the lognormal and the Gaussian, the integrand $u^q p(u)$ has a maximum for $u = \tilde{u}_q$, and one can approximate $q \log u + \log p(u)$ near its maximum by a parabola of the form $-(u - \tilde{\mu}_q)/2\tilde{\sigma}_q^2$, and the integral is little changed if integration is restricted to a "leading interval" of the form $[-\tilde{\sigma}_q + \tilde{\mu}_q, \tilde{\mu}_q + \tilde{\sigma}_q]$, where $\tilde{\sigma}_q$ is the width of the maximum of $u^q p(u)$. When q' is allowed to vary continuously instead of being integer-valued and close to q, the corresponding leading intervals always overlap. We shall now examine what happens as q' moves away from q. There is continuing overlap in the Gaussian, but not in the lognormal case. It follows that different moments of the lognormal are determined by different portions of the density $p(u)$; therefore, it is natural to describe them as *localized*. By small changes in the tail of $p(u)$, one can strongly modify the moments, not independently of each other, to be sure, but fairly independently. This fact will help explain the observations in Section 2.1.

2.3.1 The Gaussian's moments are thoroughly delocalized. Here,

$$\log [u^q p(u)] = \text{ a constant } + q \log u - \frac{u^2}{2\sigma^2}.$$

At its maximum, which is $\tilde{\mu}_q = \sigma\sqrt{q}$, the second derivative is $2/\sigma^2$, hence $\tilde{\sigma}_q = \sigma/\sqrt{2}$. Successive leading intervals overlap increasingly as q increases. Numerically, the second percentile of $|G|$ is given in the tables as roughly equal to 2.33. Values around the second percentile greatly affect moments of order 5 or 6. The value $|G|=3$ is encountered with probability 0.0026, and its greatest effect is on the moment of order $q=9$. Therefore, samples of only a few thousand are expected to yield nice estimates of moments up to a fairly high order.

*2.3.2 **The lognormal's moments are localized.*** For the lognormal, a good choice variable of integration, $v=\log u$ yields the formula in Section 1.1

$$E\Lambda^q = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\mu q + \frac{\sigma^2 q^2}{2}\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{[v-(\mu+\sigma^2 q)]^2}{2\sigma^2}\right\} dv.$$

When $\mu=-m$ and $\sigma^2=2m$, so that $E\Lambda=1$,

$$\tilde{\mu}_q = \mu + \sigma^2 q = m(2q-1), \quad \text{while} \quad \tilde{\sigma}_q = \sigma = \sqrt{2m} \text{ is independent of } q.$$

Consequences of the dependence of $\tilde{\mu}_q$ and $\tilde{\sigma}_q$ on q, *concerning the localized character of the moments of the lognormal.* The midpoints of the leading intervals corresponding to q and $q+\Delta q$ differ by $\sigma^2\Delta q$. When σ is small, the leading intervals overlap only with neighbors. When $\sigma^2 > 2\sigma$, integer qs yield non-overlapping leading intervals.

Consequences of the values of $\tilde{\mu}_q$ and $\tilde{\sigma}_q$ concerning the direct estimation of the moments $E\Lambda^q$ from the data on a lognormal Λ. One can evaluate $E\Lambda^q$ from the mean and variance as estimated from the distribution of $\log \Lambda$, or from Λ itself. The latter method shows that the population moments of the lognormal are delocalized and overly dependent on separate intervals of rare values.

The moment $E\Lambda$. As soon as $m=2.33$, $\log \tilde{\lambda}_1$ lies on the distribution's first percentile to the right. That is, the estimation of $E\Lambda$ from the λ data is dominated by one percent of the data. As soon as $m=3.10$, $\log \tilde{\lambda}_1$ corresponds to the first per-mil to the right. That is, the estimation of $E\Lambda$ from the λ data is dominated by one-thousandth of the data.

The moment $E\Lambda^2$. Its estimation is dominated by $\log \tilde{\lambda}_2 \sim \mu + 2\sigma^2 = 3m$. The percent and per-mil thresholds now occur, respectively, for $m=0.77$ and $m=1.03$. Therefore, the empirical variance makes no sense, except for very small m and/or a very large sample size.

The moment $E\Lambda^3$. Its estimation, hence the value of the empirical skewness, is dominated by $\log \tilde{\lambda}_3 \sim \mu + 3\sigma^2 = 5m$.

The rest of the argument is obvious and the practical meaninglessness of its estimate is increasingly accentuated as q increases.

Implications of the sensitivity of the population moments to the confidence a scientist may place in them. For both the Gaussian and the lognormal, a standard formula extrapolates all the $E\Lambda^q$ and the tail's shape from two characteristics of the bell, namely, μ and σ^2. For the Gaussian, the extrapolation is safe. For the lognormal, the extrapolated high moments cannot be trusted, unless the underlying distribution is known in full mathematical precision, allowing no approximation. But *absolute exactitude for all* λ is not of this world. For example, Section 1.2 mentioned that the statisticians' attachment to the lognormal is rationalized via the Central Limit Theorem, but this theorem says *nothing* of the tails. Moreover, due to the localization of the lognormal, high order moments largely depend on a tail that is effectively unrelated to the bell.

Prediction in economics demands such as extrapolation from the fitted distribution to larger samples and corresponding larger values. From this viewpoint, data for which one may hesitate between the lognormal and the scaling distributions are genuinely difficult to handle. By fitting the scaling distribution the difficulties are made apparent and can be faced. By contrast, lognormal fitting hides them and prevents them from being recognized, because it fails to be sensitive in the regions in which sensitivity matters. The decision between lognormal or the scaling cannot be helped by the development of better statistical techniques. When data are such that the scaling and lognormal representations are equally defensible, and the limited goal is compression of data for the purpose of filing them away, one may just as well flip a coin. But we must move beyond that limited goal.

3. THE POPULATION MOMENTS OF THE LOGNORMAL BEING LOCALIZED, THE FINITE SAMPLE MOMENTS OSCILLATE IN ERRATIC AND UNMANAGEABLE MANNER

3.1 Summary of the second argument against the lognormal

Population moments can be evaluated in two ways: by *theory*, starting from a known distribution function, or by *statistics*, starting from sample moments in a sufficiently large sample. For the lognormal, Section 2 took up the first method. We now propose to show that the sensitivity of the population moments to rare events has another unfortunate consequence:

the second method to estimate the population moments is no better than the first.

3.2 From exhaustive to sequential sample moments

Every form of science used to depend heavily on the possibility of reducing long lists of data to short lists of "index numbers," such as the moments. But Section 5.1 of Chapter E5 argues that computer graphics decreases this dependence. Moreover, the heavy reliance on moments seems, perhaps unconsciously, related to the notion of statistical sufficiency. As is well-known, the sample average is sufficient for the expectation of a Gaussian, meaning that added knowledge about the individual values in the sample brings no additional information. This is true for estimating the expectations of the Gaussian but not in general. I always believed, in fact, that sample moments pushed concision to excess. This is why my old papers, beginning with M 1963b{E14}, did not simply evaluate a q-th moment, but made sure to record a whole distribution.

3.3 The lognormal's sequential sample moment

Given a set of $N =$ max n data and an integer q, the sequential qth sample moment is defined by

$$S_q(n) = \frac{1}{n} \sum_{m=1}^{n} U_m^q.$$

The question is how $S_q(n)$ varies as n increases from 1 to $N =$ max n.

For the lognormal Λ and near-lognormals with $EU^q < \infty$, we know that $S_q(n)$ does converge to a limit as $n \to \infty$. But Section 3.4 will show that the sample sizes needed for reliable estimation of the population moments may be colossal, hence impractical. For reasonable sample sizes, convergence is erratic. With a significant or even high probability, the sample moments will seem to vary aimlessly, except that, overall, they appear to increase.

The key fact is that, for large enough q, the event that $U_m^q < EU^q$ has a very high probability, hence also the event that $S_q(n) < EU^q$. Colossal sample sizes are needed to allow $S_q(n)$ to reach up to EU^q.

The nature and intensity of those difficulties depends on skewness. In the limit $\sigma \ll 1$ and $E\Lambda = 1$, one has $\mu = \sigma^2/2$, hence $|\mu| \ll 1$ and $\Lambda = \exp[\sigma(G - \mu)] \sim 1 + \sigma(G - \mu)$. That is, Λ is near Gaussian, and one anticipates sample moments converging quickly. Low-order moments confirm this anticipation. However, Λ^q being also lognormal, the qth moment of one lognormal is the sample average of a less skewed one. Since a large enough q makes the parameters $\sigma_q = q\sigma$ and $\mu = -q^2\sigma^2/2$ as large as desired, Λ^q become arbitrarily far from being Gaussian.

That is, every lognormal's sufficiently high moments eventually misbehave irrespective of the value of σ. Since the moments' behavior does not depend on q and σ separately, but through their product $q\sigma$, we set $q = 1$, and study averages for a lognormal having the single parameter σ.

3.4 The growing sequence of variable "effective scaling exponents" that controls the behavior of the sequential moments of the lognormal

The scaling and lognormal distributions are best compared on log-log plots of the tail densities, but those plots are complicated. To the contrary, the log-log plot of the density are very simple and give roughly the same result. An effective α exponent $\tilde{\alpha}(\lambda)$ is defined by writing

$$\frac{d}{d\lambda} \log p(\lambda) = \frac{d}{d\lambda}\left\{ -\log(\sigma\sqrt{2}\,) - \log \lambda - \frac{(\log \lambda - \mu)}{2\sigma^2} \right\}$$
$$= -1 - \frac{\log \lambda - \mu}{\sigma^2} = -\tilde{\alpha}(\lambda) - 1.$$

After reduction to $E\Lambda = 1$,

$$\tilde{\alpha}(\lambda) = \frac{\log \lambda - \mu}{\sigma^2} = \frac{1}{2} + \frac{\log\lambda}{\sigma^2}.$$

From Section 2.2.2, the values of λ that contribute most to $E\Lambda^q$ satisfy $\log \lambda \sim \tilde{\mu}_q = \sigma^2(q - 1/2)$, hence yield an effective $\tilde{\alpha}(q) \sim q$. For example, the range corresponding to $q = 1$ yields an effective $\tilde{\alpha}(q) \sim 1$. Within a sample of finite size $N = \max n$, one can say that the behavior of the sequential $S_q(n)$ is not affected by the tail of the density, only by a finite portion, and for the lognormal that finite portion corresponds to an effective $\tilde{\alpha}$ that grows, but slowly.

The existence of an effective $\tilde{\alpha}$ follows from the localization of moments. An effective $\tilde{\alpha}$ is not defined for the Gaussian, or can be said to increase so rapidly that small samples suffice to make it effectively infinite. By contrast, the scaling distribution has a constant $\tilde{\alpha}$, which is the true α.

We know that a scaling X makes specific predictions concerning the distribution of sequential sample moments, and those predictions are simple and identical in the middle and the long-run. When the q-th population moment diverges for $q > \alpha$, the sequential moment $S_q(n)$ has no limit for $n \rightarrow \infty$, but the renormalized form $N^{-q/\alpha}\Sigma X_n^q$ tends in distribution to a L-stable variable of exponent α/q and maximal skewness. In particular, median $[N^{-1}\Sigma X_n^q]$ is finite and proportional to $N^{-1+q/\alpha}$, and the scatter of the sample q-th moment, as represented by the ratio of X_n^q to its median, also tends in distribution to a L-stable random variable.

Why inject the wildness of infinite population moments into a discussion in which all moments are actually safe and finite? Because the very same behavior that some authors used to describe as "improper" is needed to predict about how the sequential moment of the lognormal varies with sample size. While this behavior is practically impossible to obtain from direct analytic derivations, it is readily described from a representative "effective" sequence of scaling distributions.

For small N, the sample $S_q(n)$ will behave as if the lognormal "pretended" to be scaling with a very low α, that is, to be wild with an infinite $E\Lambda$, suggesting that it will *never* converge to a limit value. For larger samples, the lognormal mimics a scaling distribution with $1 < \tilde{\alpha} < 2$, which has a finite $E\Lambda$, but an infinite $E\Lambda^2$. As the sample increases, so does the effective $\tilde{\alpha}(\lambda)$ and the sample variability of the average decreases. It is only as $\lambda \rightarrow \infty$, therefore $\tilde{\alpha}(\lambda) \rightarrow \infty$, that the lognormal distribution eventually acknowledges the truth: it has finite moments of all orders, and $S_q(n)$ ultimately converges. Those successive ranges of values of λ are narrow and overlap when σq is small, but are arbitrarily wide and nonoverlapping when σq is large.

But where will the convergence lead? Suppose that Λ is not exactly, only nearly lognormal. The qualitative argument will be the same, but the function $\tilde{\alpha}(\lambda)$ will be different and the ultimate convergence will end up with different asymptotics.

Sequential sample moments that behave erratically throughout a sample are often observed in data analysis, and must be considered a fact of life.

PART III: PERSONAL INCOMES AND FIRMS

The idea of shortening those papers on income distribution came to mind, but a certain level of repetition may help the reader understand scaling and L-stability.

&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&

International Economic Review: 1, 1960, 79-106 & 4, 1963, 111-115. **E10**

L-stable model for the distribution of income

✦ **Abstract.** This paper introduces a new model for the distribution of income, and hopes to draw attention to the great potential of the "L-stable" family of nonGaussian probability distributions. This new tool may be as important as the specific application to be discussed. In other words, the same approach translates immediately to analogous quantities for which it may be more reasonable, or give a better fit. One might even paraphrase Einstein's cagey comment about Brownian motion: it is possible that the properties studied in that paper are identical to those of income; however, the information available regarding incomes is so lacking in precision that one cannot really form a judgement on the matter. ✦

THE THEME OF THIS PAPER is the empirical expression proposed by Vilfredo Pareto for the distribution of high personal incomes. Pareto 1896 asserts that the social-economic structure of a community and the precise definition chosen for "income" at most influence the parameters of a universal expression. Otherwise, or so he claimed, the same distribution represents the incomes of a few hundred "burghers" of a Renaissance city-state and of all USA taxpayers.

Section 1 carefully distinguishes between the terms "Pareto law" and "scaling distribution," each of which can be either "uniform" or "asymptotic." Section 1.5 comments on several existing theories of income distribution.

Section 2 introduces the *"positive L-stable distribution,"* which is not restricted to high incomes but also partially explains the data relative to the middle-income range. This new alternative was suggested by two arguments. *First,* among all possible interpolations of the asymptotic version of the scaling distribution, only the L-stable distributions strictly satisfy a certain (strong) form of invariance relative to the definition of income. *Second,* under certain conditions, the L-stable distributions are possible limits for the distribution of renormalized sums of random variables. This implies, of course, that *the Gaussian distribution is not the only possible limit,* contrary to what is generally assumed. It is unnecessary (as well as insufficient) to try to save the limit argument by applying it to $\log U$ instead of to U, as is done in some theories leading to the lognormal distribution for U. The L-stable distributions also have other very desirable properties, which will be discussed.

1. INTRODUCTION

Let $P(u)$ be the percentage of individuals whose yearly income U exceeds u; income is assumed to be a continuous variable. The empirically observed values of $P(u)$ are of course percentages relative to finite populations, but we follow Pareto in considering them as sample frequencies in a random sample drawn from an infinite population. That is, U will be treated as a random variable with values u, and the curve $U(t)$ describing the variation of U in time t will be treated as a random function.

1.1. The uniform scaling distribution and uniform Pareto law

The *uniform Pareto law* asserts that personal income follows the uniform scaling distribution. This term means that there exist two "state variables" $\tilde{u}$ and α such that

$$P(u) = \begin{cases} (u/\tilde{u})^{-\alpha} & \text{when } u > \tilde{u} \\ 1 & \text{when } u < \tilde{u}. \end{cases}$$

Here, $\tilde{u}$ is a minimum income; it is a scale factor that depends on the currency. The exponent $\alpha > 0$ will be used to quantify the notion of the inequality of distribution. Graphically, the uniform distribution implies that the doubly logarithmic graph of $y = \log P$ as a function of $v = \log u$, is a straight line. The corresponding density $p(u) = -dP(u)/du$ is

$$p(u) = \begin{cases} \alpha(\tilde{u})^{\alpha} u^{-(\alpha+1)} & \text{when } u > \tilde{u} \\ 0 & \text{when } u < \tilde{u}. \end{cases}$$

In this statement of the uniform scaling distribution, the value of α is only constrained to satisfy $\alpha > 0$. Pareto also made the even stronger statement that $\alpha = 3/2$, which is clearly invalid.

1.2 Asymptotically scaling distribution and asymptotic Pareto law

The uniform scaling distribution is empirically unjustified, and it should be noted that many purported "disproofs" of "the" Pareto law apply to this variant only. On the contrary, there is little question of the validity of the scaling distribution if sufficiently large values of u are concerned. The *asymptotic scaling distribution* asserts that personal incomes are asymptotically scaling. This term means that

$$P(u) \sim (u/\tilde{u})^{-\alpha}, \text{ as } u \rightarrow \infty.$$

This statement is useful only if the exact definition of the sign $\sim$, "behaves like," conforms to the empirical evidence and – taking advantage of the margin of error in such evidence – lends itself to easy mathematical manipulation. The following definition is usually adequate:

$$\frac{P(u)}{(u/\tilde{u})^{-\alpha}} \rightarrow 1, \text{ as } u \rightarrow \infty,$$

that is,

$$P(u) = \{1 + e(u)\}(u/\tilde{u})^{-\alpha}, \text{ where } e(u) \rightarrow 0, \text{ as } u \rightarrow \infty.$$

The sign ~ expresses that the graph of log$P$ versus log$u$ is asymptotic for large u to the straight line characteristic of uniform scaling.

1.3 Interpolation of asymptotic scaling to the bell of the distribution

Only over a restricted range of values of u does $P(u)$ behave according to scaling. Elsewhere, the "density" $-dP(u)/du$ is represented by a curve that is quite irregular and whose shape depends in particular on the breadth of coverage of the data considered. This has been emphasized in Miller 1955: if one includes all individuals, even those with no income and part-time workers, and if one combines the incomes of men and women, then the income distribution is skew and presents several maxima. However, for most of the different occupational categories, as distinguished by the census, the separate income distributions are both regular and fairly symmetric. Thus, the main source of skewness in the overall distribution can be traced to the inclusion of self-employed persons and managers together with all other wage earners. One may also note that the method of reporting income differs by occupational categories; as a result, the corresponding data are not equally reliable.

The above reasons make it unlikely that a single theory could ever explain all the features of the income distribution or that a single empirical formula could ever represent all the data. As a result, it has been frequently suggested that several different models may be required to explain the empirical $P(u)$. Unfortunately, we know of no empirical attempts to verify this conjecture. In any case, the present paper will be devoted almost exclusively to a theory of high income data and the asymptotic scaling distribution. It is unlikely that the interpolation of the results of our model will be able to explain all middle-income data and we shall not examine this point in great detail.

1.4 Two distributions of income distribution which contradict the asymptotic scaling distribution

1.4.1. Crossover to exponential asymptotic decay. Pareto himself suggested

$$p(u) = -dP/du = ku^{-(\alpha+1)}\exp(-bu), \text{ where } b > 0.$$

However, the asymptotic scaling distribution must be at least approximately correct for large u. Hence, the parameter b must be very small, and there is little evidence against the hypothesis that $b = 0$. Therefore, the choice between the hypotheses $b=0$ and $b \neq 0$ may be influenced legitimately by the ease of mathematical manipulation and explanation. We shall see that the foundation for a theoretical argument resides in some crucial properties of the asymptotic Pareto distribution that correspond to $b=0$. We shall therefore disregard the possibility that $b \neq 0$.

1.4.2. The lognormal distribution. That distribution (Gibrat 1932, Aitchison & Brown 1957) claims that the variable $\log U$, or perhaps the variable $\log(U-\tilde{u})$ (where $\tilde{u}>0$) is well represented by the Gaussian distribution. The empirical evidence for this claim is that the graph of (P, u) on log-normal paper seems to be straight. However, such a graph emphasizes one range of values of u, while the doubly logarithmic graph emphasizes a different range. Thus, the graphical evidence for the two distributions is not contradictory. Moreover, the motivation for the lognormal distribution is largely not empirical but theoretical, as we shall now proceed to show.

1.5. "Thermodynamic" models of income distribution

There is a great temptation to consider the exchanges of money that occur in economic interaction as analogous to the exchanges of energy that occur in physical shocks between gas molecules. In the loosest possible terms, it is felt that both kinds of interactions "should" lead to "similar" states of equilibrium. Driven by this belief, many authors tried to explain the distribution of income distribution by a model similar to that used in statistical thermodynamics. Other authors took the same path unknowingly.

Unfortunately, the scaling $P(u)$ decreases much more slowly than the usual distributions of physics. To apply the physical theory mechanically, one must give up the additivity properties of U and argue that U is a less intrinsic variable than some slowly increasing function $V(U)$. The universal choice seems to be $V=\log U$ or perhaps $V'=\log(U-\tilde{u})$, with $\tilde{u}$ a positive minimum income. This choice is suggested by the fact that empirical records use income brackets with a width that increases with u. In addition, $\log V$ can be traced back to the "moral wealth" of Bernoulli, and it has been argued that $\log V$ can be justified by some distribution of proportionate effect, a counterpart in economics of the Weber-Fechner distribution of psychophysiology.

Even if this choice of V is granted, one has to explain why it seems that the normal distribution applies in the middle zone of v's and the exponential distribution $P(v)=\exp\{-\alpha(v-\tilde{v})\}$ applies for large v's.

Many existing models of the scaling distribution are reducible to the observation that $\exp(-\alpha v)$ is the barometric density distribution in the atmosphere. Alternatively, consider Brownian particles floating in a gas that has a uniform temperature and density and that is enclosed in a semi-infinite tube with a closed bottom and an open top. Assume further that the gravitational field is uniform. Then, the equilibrium density distribution of the Brownian particles, and their limit distribution, is exponential. This limit results from a balance between two forces, which are both uniform along the tube; gravity alone pull all particles to the bottom, and heat motion alone would diffuse them to infinity. Clearly, the models of income distribution that we are now considering involve interpretations of the forces of diffusion and gravity.

Unfortunately, the counterpart of the bottom of the tube is essential: *removing the bottom changes everything.* There is no longer any steady limit state because all the Brownian particles diffuse down to infinity. It is true that, if all the particles start from the same point, the *conditional* distribution for V is Gaussian, but this fact cannot be an acceptable basis for a model of the log-normal distribution.

The boundary conditions already matter when diffusion is replaced by a *random walk.* In this approximation, time is an integer and V is an integer multiple of a unit $\tilde{v}$. This model was explicitly introduced into economics in Champernowne 1953. It assumes (1) that the variation of V is Markovian, that is, $V(t+1)$ is a function only of $V(t)$ and of chance, and (2) that the probability that $V(t+1) - v(t) = k\tilde{v}$, which *a priori* could be a function of k and of $v(t)$. But this is not all. To obtain either the scaling or the log-normal distribution, additional assumptions are required. But these additional assumptions have no intuitive meaning, which makes both conclusions unconvincing. However, the models that lead to the exponential are still slightly better. In fact, we can argue that the apparent normality of the "density" $p(v)$ in the central zone of v's simply means that $-\log p(v)$ may be represented by a parabola in that zone, whereas for large v's it is represented by a straight line. Such a parabolic interpolation needs no limit theorems of probability for its justification; it applies to any regularly concave curve, at least in the first approximation.

In models of the asymptotic Pareto law for incomes, further nonintuitive assumptions are necessary to rationalize $\alpha > 1$.

Various other models of the normal or exponential distributions often occur in statistical thermodynamics. These models – and their translation in terms of economics – are essentially equivalent. In particular, they assume that there is no institutional income structure; all income recipients are treated as entrepreneurs. However, a rewording of the classical theories can make them applicable to many possible institutional structures of wage and salary earners. This was done in Lydall 1959. I had independently rediscovered Lydall's model and had discussed it in the original text of this paper, as submitted on June 19, 1959. The predictions of the L-stable model and of the Lydall model coincide for $1 < \alpha < 2$. This shows that the distribution that corresponds to the least amount of organization could be "frozen" without modifying it, reinterpreted as a wage distribution, and allowed to evolve along conceptually quite different lines. This fact has great relevance to the problem of the value of α since in highly organized industrial societies, α has tended to increase beyond 2.

This article attempts to show that *one need not abandon the analogy of statistical physics to avoid the weaknesses which mar existing theories*. There will be no need for the transformation $V = \log U$, nor for economic counterparts of such conditions as the presence or the absence of a bottom to the tube in which Brownian motion is studied. That is, this will not be an implicit attempt to force income into the structure of statistical thermodynamics but an explicit attempt to generalize the statistical methods of thermodynamics to cover the economic concept of income.

2. L-STABLE RANDOM VARIABLES

2.1 Analysis of the definition of the notion of income; random variables that are invariant under addition, up to scale

One of the principal claims of this paper is that it is impossible to "explain" why the distribution of income is scaling, and does not follow some other distribution, without first wondering why essentially the same distribution continues to be followed by "income," despite changes in the definition of this term.

This invariance is, of course, very important to census analyzers, because it means that large changes in methods of estimating income have an unexpectedly small effect on the distribution.

2.1.1. Analysis of the notion of income. We shall argue that there are several ways of distinguishing different sources of U. Therefore, U may be written in different ways as the sum of elements, such as (A) agricultural, commercial or industrial incomes; (B) incomes in cash or in kind; (C) ordinary taxable income or capital gains; (D) incomes of different members of a single taxpaying unit, and so forth. Label the income categories in a certain decomposition of U as $U_i (1 \le i \le N)$. We assume that every method of reporting or estimating income corresponds to the observation of the *sum* of the U_i corresponding to some subset of indices i. This quite reasonable consideration imposes the restriction that the scale of incomes themselves is more intrinsic than any transformed variables, such as $\log U$.

The same kind of division may also be performed in the direction of time and the year is unlikely to be an intrinsic unit of time.

Consider the incomes in the categories U_i to be statistically independent. This assumption idealizes the actual situation. *A priori*, this abstraction may seem a bad first approximation since when income is divided into two categories U' and U'', such as agricultural and industrial incomes, the observed values u' and u'' are usually very different. However, when the parts are independent and follow the asymptotic scaling distribution, we shall find that one would actually expect them to be very unequal. Hence, although such an inequality cannot be a confirmation of independence, at least it does not contradict it in this case.

2.1.2. Paul Lévy's "stability," meaning invariance under addition, up to scale. Under the assumptions described in Section 2.1.1, the only probability distribution for income that could possibly be observed must be such that, if U' and U'' follow this distribution (up to a scale transformation and up to the choice of origin), then $U' + U''$ must also follow the same distribution. That is, given $a' > 0$, b', $a'' > 0$, and b'', there must exist two constants $a > 0$ and b such that

$$(a'U + b') + (a''U + b'') = aU + b.$$

Such a probability distribution, its density and U itself are said to be L-stable under addition.

The family of all distributions which satisfy this requirement was constructed in Lévy 1925. In addition to the Gaussian distribution, it includes nonGaussian distributions that are asymptotically scaling with some $0 < \alpha < 2$. In other words, the additive property of U and the behavior of $P(u)$ under the asymptotic Pareto distribution (both of which disappear if the scale of U is changed), turn out to be precisely sufficient and necessary for the application of Lévy's theory of L-stable distributions.

Furthermore, a L-stable U satisfies $E(U) < \infty$ when $1 < \alpha < 2$.

2.1.3. "Positive" L-stable distributions. In the study of income, we restrict ourselves to the extreme cases when $p(-u)$ decreases very much faster than $p(u)$ when $u \rightarrow \infty$. Those L-stable variables may be called positive, an abbreviation for "maximally skewed in the positive direction." For convenience, this paper writes "L-stable distributions" without reminding each time that the variable is positive.

It is, obviously, not strictly true that the distribution of income is invariant over the whole range of u, with respect to a change in definition of income. But we argue the following: if the variable U is not Gaussian, and if its distribution is very skew, the most reasonable "first order" assumption about income is that it is a L-stable variable.

2.1.4. The L-stable probability density. Unfortunately, this density cannot be expressed in a closed analytic form, but is determined indirectly by its bilateral Laplace transform

$$G(b) = \int_{-\infty}^{\infty} \exp(-bu)\,|\,dP(u)\,| = \int_{-\infty}^{\infty} \exp(-bu)p(u)du.$$

In the L-stable case, $G(b)$ is defined for $b > 0$, and takes the form

$$G(b) = \exp\{(b\tilde{u})^{\alpha} - Mb\},$$

which depends on three parameters. The exponent α satisfies $1 < \alpha < 2$, $\tilde{u}$ is a positive scale parameter, and M is a location parameter; When $1 < \alpha < 2$, then $M = E(U)$.

Behavior of $P(u)$ for $u \to \infty$. The behavior of $G(b)$ for $b \to 0$ shows that

$$P(u) \sim \frac{u^{-\alpha}(\tilde{u})^{\alpha}}{\Gamma(1-\alpha)},$$

where Γ denotes the Euler gamma-function.

For other values of u, the L-stable distribution is necessarily obtained numerically on the computer. Sample graphs of densities appear in Figure 1. {P.S. 1996: An added comment is found, at the end of this chapter, in *Annotation to Figure 1.*} More detailed graphs are found in M & Zarnfeller 1961.

We see that as long as α is not close to 2, the L-stable density curve becomes very rapidly indistinguishable from a uniformly scaling curve of the same α. For this reason, the origin of the uniform curve must be chosen properly, and one must set $\tilde{u} = \tilde{u}[\Gamma(1-\alpha)]^{-1/\alpha}$. The two curves converge near $u = E(U)$ when α is in the neighborhood of 3/2 and at even

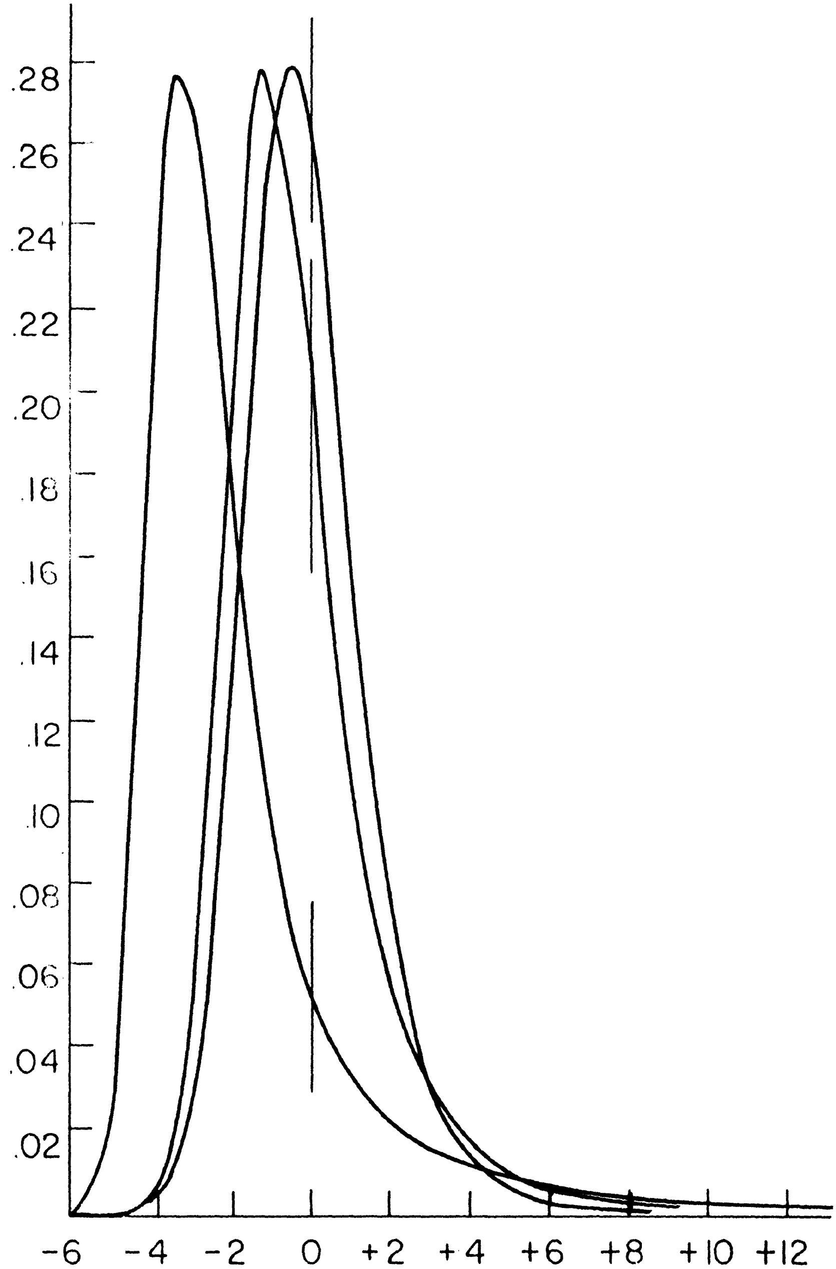


FIGURE E10-1. Plot in natural coordinates of the probability densities of the reduced L-stable random variables for $M = EU = 0, \beta = 1$, and the following values of the exponent: $\alpha = 1.2$ (to the left), 1.5 (center), and 1.8 (to the right).

smaller values of u when α is less than 3/2. That is, the asymptotic behavior of $p(u)$, derived from $G(b)$, is attained very rapidly.

Behavior of $p(u)$ for $u \rightarrow -\infty$. This density is > 0 for all u, as $u \rightarrow -\infty$, it decreases faster than in the Gaussian case. Therefore, large negative values of u may be safely disregarded. Indeed, Appendix I proves that

$$\log[-\log p(u)] \sim \frac{\alpha}{\alpha - 1} \log |u|.$$

The middle range of values of U. In the bell, the graph of the L-stable probability density is skew; this is the behavior one finds in the empirical data and hopes to derive in a theoretical curve.

2.1.5. *Non-positive L-stable distributions.* The Gaussian distribution is of course stable, and it is the only L-stable distribution with a finite variance. Furthermore, all stable variables with a finite mean are either sums (obviously) or differences of L-stable variables scaled by arbitrary positive coefficients. The bilateral generating function is no longer defined, and we must consider the usual characteristic function (ch.f.). For a L-stable variable, the ch.f. is immediately obtained as

$$\varphi(\zeta) = \int_{-\infty}^{\infty} e^{iu\zeta} p(u) du = \exp\left\{ iM\zeta - |\zeta|^{\alpha} (\tilde{u})^{\alpha} \left\{ 1 - \frac{i\zeta}{|\zeta|} \tan \frac{\alpha\pi}{2} \right\} \Big| \cos \frac{\alpha\pi}{2} \Big| \right\}.$$

Hence, denoting by $(1+\beta)/(1-\beta)$ the ratio of the positive and negative components, the general L-stable variable with $E(U) < \infty$ has a ch.f. of the following form, where $\tilde{u} \geq 0$, $|\beta| \leq 1$ and $1 < \alpha < 2$:

$$\varphi(\zeta) = \exp\left\{ iM\zeta - |\zeta|^{\alpha} (\tilde{u})^{\alpha} \left\{ 1 - \frac{i\beta\zeta}{|\zeta|} \tan \frac{\alpha\pi}{2} \right\} \Big| \cos \frac{\alpha\pi}{2} \Big| \right\}.$$

The same formula extended to $0 < \alpha < 1$ gives L-stable variables with $E(U) = \infty$. The limit value $\alpha = 2$ yields the Gaussian variable, and the crossover value $\alpha = 1$ yields the Cauchy variable and related skew variables.

2.2 Role of the L-stable distributions in central limit theorems

The L-stable distributions have another (equivalent) property: they are the only possible limit distributions of weighted sums of identical and independent random variables.

To prove that every L-stable distribution is such a limit, assume that the variables U_i themselves are L-stable. This meaning of "behavior like" expresses that the sum $\sum_{i=1}^{N} U_i$ is a variable of the form $a(N)U + b(N)$. Hence, $\{a(N)\}^{-1}$ $\{(\sum_{i=1}^{N} U_i) - b(N)\}$ has the same distribution function as U. Conversely, assume that a certain normed sum of the variables U_i has a limit. The N-th normed sum, where $N = n + m$, can be written as

$$A(N)\sum_{i=1}^{N} U_i - B(N)$$
$$= \frac{A(N)}{A(n)}\left\{A(n)\sum_{i=1}^{n} U_i - B(n)\right\} + \frac{A(N)}{A(m)}\left\{A(m)\sum_{i=1}^{m} U_i - B(m)\right\} + C(N).$$

By letting both n and m tend to infinity a long proof found in Gnedenko & Kolmogorov 1954 shows that the definition of a L-stable variable must be satisfied by the limit of $A(N)\sum_{i=1}^{N} U_i - B(N) + C(N)$.

We also mention the following related theorem due to Gnedenko. For the convergence of $N^{-1/\alpha}\sum_{i=1}^{N} U_i - B(N)$ to a L-stable limit, it is necessary and sufficient that $0 < \alpha < 2$. For $u > 0$, $F(u) = 1 - P(u) = 1 - \{1 + e'(u)\}$ $(u/u_1)^{-\alpha}$, where $e'(u) \to 0$ as $u \to \infty$. For $u < 0$, $F(u) = \{1 + e''(u)\}$ $(u/u_2)^{-\alpha}$, where $e''(u) \to 0$ as $u \to -\infty$. That is, if U is decomposed into a sum of a *large* number of components U_i, we need not resort to the above argument of observational invariance. Assume that (1) the sum is not Gaussian (which is a conspicuous fact) for $u \gg 1$, (2) the expected value of the sum is finite (which is also a fact) and (3) the probability of $-u$ is much less than that of u. Under these assumptions, no further hypothesis about the distribution of the parts is necessary to conclude that the sum can only be a L-stable variable. We shall develop these points in more detail, and in Section 2.6 we shall show that the above restrictions may be reduced to broad "qualitative" properties, such as the likelihood that two independent variables U' and U'' contribute unequally to the sum $U' + U''$.

Note that the preceding explanation in terms of sums of many contributions does not involve the $\log U$ transformation, which leads to the lognormal distribution and contradicts scaling.

These results may be better than expected because the limit distribution might not hold for U, which is the sum of *only a few* random components. That is, to increase the number of components of U, one must abandon at some point the hypothesis that the components are independent. The larger the number of components, the less independent they become. This difficulty is not unique to the problem of income but is acutely present in all social science applications of probability theory.

The difficulty is already apparent in physics, for example, in arguments meant to explain why a given noise is Gaussian. But in most physical problems there exists a sufficiently large zone between the systems which are so small that they are impossible to subdivide and those systems which are so large that they can no longer be considered homogeneous. No such zone exists in most problems of economics, so that a successful application of a limit theorem may seem too good to be true.

In summary, it cannot be strictly true that the additive components are independent and have the same distribution (up to scale). However, the acceptability of a given distribution is greatly increased when no other distribution is reducible to limit arguments. Hence, suppose that a sum of many components is not Gaussian, is skewed, and such that $E(U) < \infty$; then the most reasonable first assumption concerning the sum is that it follows the L-stable distribution.

More precisely, the L-stable distribution can be given two extreme interpretations. The "minimal" interpretation observes that it is correct asymptotically, is sufficiently easy to handle, and is useful in the first approximation. (After all, the Gaussian itself is frequently a good first approximation to distributions that, actually, are certainly not Gaussian.)

At the other extreme, we may take the L-stable distribution entirely seriously and try to check its ability to predict some properties of income distribution that otherwise would seem independent of the asymptotic Pareto distribution. We believe that such predictions were, in fact, achieved. This provides some supporting evidence for a "maximal" interpretation of the L-stable distribution, which regards its invariance and limit properties as being "explicative."

2.3 L-stable distributions in abstract probability theory

L-stable distributions are well-known in abstract probability theory. Therefore, we might have introduced them without special motivation, other than the fact that the known asymptotic behavior of $P(u)$ and its recently computed behavior for intermediate u make it an attractive interpolation for income data (M 1959p). Unfortunately, the behavior of the L-stable distributions, in many other ways, is quite different from the behavior most statisticians learn from constant handling of the Gaussian distribution. To show their suitability to the present problem, we proceed to study them heuristically from the viewpoint of addition and extremal values; this will be followed by statements of some rigorous results.

The three main drawbacks we have attributed to the classical theories have no counterpart in our argument. Indeed, many variants of our argument lead to the same result; no change of scale of U is necessary, the behavior of $P(u)$ being *exactly* what is needed, and α is "near" 3/2.

2.4 Divergence of the variance of U

A consequence of the limitation $1 < \alpha < 2$ is that the second moment of U is infinite while the first moment is finite. Indeed,

$$E(U) = \int_{-\infty}^{\infty} u dF(u) = -\int_{-\infty}^{\infty} u dP(u) = P(u)du < \infty,$$

$$E(U^2) = \int_{-\infty}^{\infty} u^2 dF(u) = -\int_{-\infty}^{\infty} u^2 dP(u) = 2uP(u)du = \infty.$$

(We assume that the behavior of $P(u)$ for negative u does not lead to a convergence problem for $E(U)$.)

The last result holds for both uniform and asymptotic Pareto variables if $1 < \alpha < 2$, but it fails to hold for the density $p(u) = ku^{-(\alpha+1)}\exp(-bu)$, which is described in Section 1.4. This provides a new and important test of our conjecture that $b = 0$.

The finiteness of $E(U)$ means that if the u_i are samples from a Pareto distribution, the empirical mean $E_N = \Sigma_{i=1}^{N} u_i/N$ tends to $E(U)$ with probability 1 (Kolmogorov's uniform distribution of large numbers). In addition, E_N is a good estimate for $E(U)$, if N is large. Now consider $S_N = (1/N)\Sigma_{i=1}^{N} (u_i - E_N)^2$. If $\alpha > 2$, S_N tends to a finite limit $S(U)$, the finite sum being a good estimate of the limit and $\Sigma_{i=1}^{N} (U_i - E_N)/\sqrt{N}$ tends to a Gaussian variable. This result changes little if one adds an exponential factor with small b. If $\alpha < 2$ and $b = 0$ the limit of S_N is infinity, and S_N grows without limit like $N^{1/\alpha - 1}$. To the contrary, $S(U)$ is finite if $b > 0$.

Therefore, the usefulness of the exponential factor $\exp(-bu)$ may be tested by checking whether or not S_N continues to increase with N in the case of the largest sample available. We could not make the direct test, but an indirect test results from the following observation: the ordering of the different populations by "increasing inequality" presumably should be identical with their ordering by decreasing α. On the other hand, more usual measures of inequality are given by S_N, S_N/E_N or $\sqrt{S_N}/E_N$. These two methods of ordering populations have been compared and found to be entirely contradictory. This result ceases to be absurd if one recalls that the values of N in the different samples we compared range from 10^2 to 10^3. And it seems that even if $S(U)$ was finite, it would not be approached, even with the largest samples. In that case, *irrespective of any theory*, it is preferable to take $b=0$ and $S(U)=\infty$. Furthermore, no function of S_N is adequate to compare degrees of inequality, except perhaps between samples of identical size.

Another test of the usefulness of the approximation $S=\infty$ is provided by the relative contribution to S_N of the largest of the u_i; As predicted from the theory of the L-stable distribution, this relative contribution is very large; it is close to 1/2 for the Wisconsin incomes, Hanna et al 1948. Truncating U to avoid $S=\infty$ distorts the whole problem.

2.5 Heuristic study of the sums of two independent random variables in the exponential, Gaussian and asymptotically scaling cases

Let U' and U'' be two independent random variables with the same probability density $p(u)$, and let $p_2(u)=\int_{-\infty}^{\infty}p(x)p(u-x)dx$ be the density of the sum $U=U'+U''$. We assume that u may vary from 0 to $+\infty$ and compare the behavior for $u \rightarrow \infty$ of $p(u)$ and $p_2(u)$.

To understand the behavior of $p_2(u)$, it is useful to stress the following three forms of the graph of $\log[p(u)]$.

The simplest form is the linear graph $\log[p(u)]=\log C - bu$.

In the next simplest cases – illustrated in Figure 2 – the graph of $\log[p(u)]$ is convex or concave over the whole range of variation of u', hence $-\log p(x)-\log p(u-x)$ has an extremum for $x=u/2$.

Terminology. Having never learned the exact meaning of the terms *convex* and *concave*, I play it safe with *cap convex* for $y=-x^2$ and *cup convex* for $y=x^2$.

2.5.1. *The graph of* log*p*(*u*) *is rectilinear.* In the exponential case,

$$p(u')=\begin{cases} C\exp(-bu') & \text{if } u'\geq 0,\\ 0 & \text{if } u'<0.\end{cases}$$

Then,

$$p_2(u) = \int_0^u C^2 \exp(-bx) \exp[-b(u-x)]dx$$
$$= \int_0^u C^2 \exp(-bu)dx = C^2 u \exp(-bu).$$

Thus, all the values of u' contribute equally to $p_2(u)$.

2.5.2. The graph of **log***p*(*u*) *is cap convex:* $d^2 \log p(u)/du^2 \leq 0$, *for all u.* This convexity implies that $p(u)$ decreases rapidly as $u \to \infty$, and the integrand $p(x)p(u-x)$ has a *maximum* for $x = u/2$. If that maximum is sufficiently strong, the integral $\int_{-\infty}^{\infty} p(x)p(u-x)dx$ largely consists of the contributions of a small interval of values of x, near $u/2$. Hence, *a large value of u is likely to result when the two contributions u' and u'' are almost equal.*

This conclusion may sound obvious, but in fact need not be true, in particular our main point will soon be that it ceases to be true if $-\log p$ is cup-convex.

For the Gaussian density,

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} \text{ and } p_2(u) = \frac{1}{\sigma\sqrt{2}\sqrt{2\pi}} \exp\left\{-\frac{u^2}{4\sigma^2}\right\}.$$

The Gaussian density is therefore preserved under addition, except that the scale factor σ is multiplied by $\sqrt{2}$. The same result can be obtained heuristically by arguing that $p(x)p(u-x)$ remains near its maximum value $\{p(u/2)\}^2$ over some interval of width $D/2$ on each side of $u/2$ and is negligible elsewhere. This yields the approximate estimate

$$p_2(u) \sim p(u/2)p(u/2)D = \frac{D}{\sigma^2} 2\pi \exp\left\{-\frac{u^2}{4}\sigma^2\right\}.$$

In other words, the approximate estimate is correct, if one takes for the width the value $D = \sigma\sqrt{\pi}$ independent of u.

2.5.3. The graph of **log** ***p*(*u*)** *is cup convex:* $d^2 \log p(u)/du^2 \geq 0$, *for all u.* Next, when $p(u)$ decreases slowly, $p(x)p(u-x)$ has a *minimum* for $x = u/2$. Note that, for every probability density, $p(x) \to 0$ as $x \to \infty$ or $x \to -\infty$. The assumed cup convexity of $-\log p(x)$ requires x to be bounded from below. In the simplest example, u' can only be $\geq \tilde{u}$, and $\tilde{u}$ is its most probable value. If so, $p(x)p(u-x)$ will have two maxima, one for $x=\tilde{u}$ and one for $x=u-\tilde{u}$. If they are sufficiently uniform, $p_2(u)$ will mostly come from the contributions of the neighborhoods of these maxima. That is, the "error term"

$$2\int_0^{u/2} p(x)p(u-x)dx - 2\int_0^{u/2} p(x)p(u)dx$$

will reduce to contributions of value of x very different from $x=0$. These contributions being small,

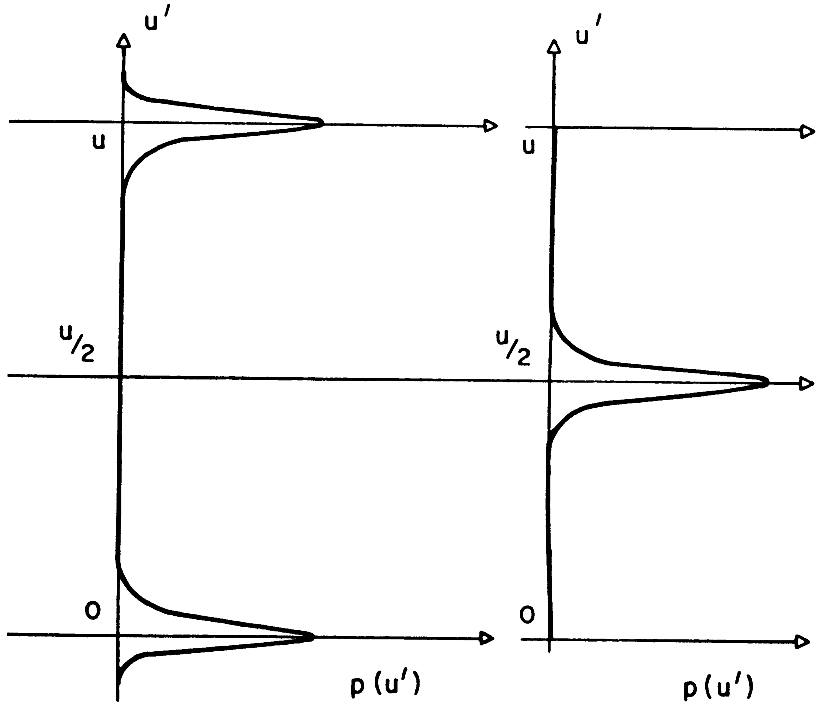


FIGURE E10-2. Conditional probability density of the addend U' when the value u of $U'+U''$ is large, for two distributions of U' and U'': a bell with an asymptolically scaling tail (left) and the Gaussian (right). For the exponential, $p(u')$ is a constant.

$$p_2(u) \sim 2p(u)\int_0^{u/2} p(x)dx.$$

Finally, if u is large, $\int_0^{u/2} p(x)dx \sim \int_0^{\infty} p(x)dx$ so that $p_2(u) \sim 2p(u)$. Hence,

$$\int_u^{\infty} p_2(x)dx = P_2(u) \sim 2P(u).$$

Note that a large value of u is now likely to be the sum of a relatively small value of either u' (or u'') and of a value of u'' or u') that is very close to $u - \tilde{u}$. The two addends are likely to be very unequal. But the problem is entirely symmetric so that $E(u' \mid u) = u/2$. (See also Appendix II.)

In addition, we have proved that, given two variables U' and U'' of the slowly decreasing type, the distribution of the larger of these variables has the same asymptotic behavior as the distributions of their sum. Another derivation of this result starts from the derivation of the distributions of $\max(U', U'')$. Clearly, the probability that u is larger than $\max(U', U')$ is the probability that u is larger than both u' and u''. Hence, $1 - P_m(u) = 1 - \Pr\{\max(U', U'') > u\} = [1 - P(u)]^2$. For large u and small $P(u)$, this becomes $P_m(u) \sim 2P(u)$. That is, for slowly decreasing densities, $P_m(u) \sim P_2(u)$.

A prototype of the slowly decreasing probability density is the uniform scaling variable. In that case, we can write

$$P_2(u) \sim 2P(u) \sim P(2^{-1/\alpha}u).$$

That is, the sum of two independent and identical uniform scaling variables is asymptotic scaling with unchanged α and $\tilde{u}$ multiplied by $2^{1/\alpha}$.

Likewise, any asymptotic scaling distribution will be invariant under addition, up to the value of $\tilde{u}$. The proof requires a simple refinement of the previous argument, to cover the case where $-\log p(x)$ is cup (or cap) concave for large values of x. One can show in this way that asymptotic scaling is preserved under the addition of two (or a few) independent random variables. There is no self-contradiction in the observed fact that this distribution holds for part of the range of incomes as well as for the whole range. That is, the exact definition of the term "income" may not be a matter of great concern. But, conversely, it is unlikely that the observed data on $P(u)$ for large u will be useful in discriminating among several different definitions of "income."

The asymptotic scaling and the Gaussian distributions are the only distributions strictly having the above invariance ("L-stability") property. They will be distinguished by the criterion of "equality" versus "inequality" between u' and u'', when $u = u' + u''$ is known and large. (See Section 2.6.) In Section 2.8 we shall cite another known result concerning L-stable probability distributions.

We may also need to know the behavior of $p_2(u)$ when the addends differ, so that the density $p'(u')$ of U' decreases slowly and the density $p''(u'')$ of U'' decreases rapidly. In that case, a large u is likely to be equal to u' plus some "small fluctuation." In particular, a Gaussian error of observation on an asymptotic scaling variable is negligible for large u.

2.6 Addition and division into two for L-stable variables; criterion of (in)equality of the addends

We have shown that the behavior of the sum of two variables is determined mainly by the convexity of $-\log p(u)$. We shall later show that this criterion is, in general, insufficient to study random variables. However, if we limit ourselves to L-stable random variables, the convexity of $-\log p(u')$ is sufficient to distinguish between the case of the Gaussian distribution and of all other L-stable distributions. That is, these two cases may be distinguished by the criterion that approximate equality of the parts of a Gaussian sum is in contrast with the great inequality between the parts in all other cases, in particular for L-stable distributions.

Thus far this distribution has been used only to derive the distributions of $U' + U''$. Suppose now that the value u of U is given and that we wish to study the distribution of u' or of $u'' = u - u'$.

The Gaussian case. If the *a priori* distribution of U' is Gaussian with mean M and variance σ^2, then the *a priori* distribution of U is Gaussian with mean $2M$ and variance $2\sigma^2$. The conditional distribution of u', given u, is then given by

$$p(u' \mid u) = \frac{\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(u'-M)^2}{2\sigma^2} \right] \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(u-u'-M)^2}{2\sigma^2} \right]}{2^{-1/2} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(u-2M)^2}{4\sigma^2} \right]}$$

$$= 2^{1/2} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(u'-u/2)^2}{\sigma^2} \right].$$

Thus, the conditional probability is Gaussian with expectation $u/2$ and variance $\sigma^2/2$.

A striking feature of the result is that the expectation of U' is $u/2$; otherwise, the distribution of U' *does not depend on u*, as seen in the right-hand side of Figure 2.

The L-stable case for $\alpha < 2$. In income studies, the problem of division enters into questions such as the following. If we know the sum of the agricultural and industrial incomes of an individual, and if the *a priori* distributions of both these quantities are L-stable with the same α, then what is the distribution of the agricultural income. This case is more involved than the Gaussian, because we do not know an explicit analytic form for the distribution of U' or U''; we can, however, do some numerical plotting, as seen in the left-hand side of Figure 2).

If the sum U takes a very large value u, we find that the distribution of U' has two very sharp maxima, near u'_{max} and $u - u'_{max}$. As u decreases, the actual shape of this distribution of u' changes, instead of simply being translated, as in the Gaussian case. When u becomes small, more maxima appear. They then merge, and the distribution of u' has the same overall shape as in the Gaussian case. Finally, as u becomes negative and very large, the distribution of u' continues to have a single maximum.

Hence, bisection provides a very sharp distinction in this respect between the Gaussian and all other L-stable distributions.

Now, given a fairly small N, what is the distribution of $(1/N)$-th of a L-stable variable? In the Gaussian case, this $(1/N)$-th remains Gaussian for all N, with a mean value of u/N and a variance of $(N-1)\sigma^2/N$. In the non Gaussian case, each part of a large u may be small or may be close to u. Most of the N parts will be small, but, with a high probability, the largest part will be close to the whole.

The situation is less intuitive when N becomes very large. However, Lévy has proved that the necessary and sufficient condition for the limit of the sum ΣU_i to be Gaussian is that the value u_i of the largest of the summands is negligible compared to the whole. On the contrary, if the limit is L-stable nonGaussian and such that $E(U)=0$, both the sum and the largest of the summands will increase roughly as $N^{1/\alpha}$. It is shown in Darling 1952 that their ratio tends toward a limit, which is a random number having a distribution dependent on α.

Section 2.2 argued that U is the sum of N components without knowing N. A posteriori, this assumption is quite acceptable, because the largest (or the few largest) of the components contribute to the whole a proportion that is substantial and essentially independent of the number N of components. This eminently desirable feature of the L-stable theory is an important confirmation of its usefulness.

However, if a L-stable income is small, its components are likely to be of the same order of magnitude, like in the Gaussian case. This has an important effect on the problem mentioned near the end of Section 2.1 Assume that one has a Census category in which most income are rather small and that, when U is decomposed into parts, the sizes of the parts tend to be proportional to their *a priori* sizes. One may assimilate this behavior to that of t, a Gaussian distribution used to represent an unskilled worker's income (in which case the decomposition may refer to such things as the lengths of time during which parts of income were earned). But such behavior may also be that of a L-stable variable considered for small values of u. As a result, the apparently fundamental problem of splitting income into two parts so that only one follows the L-stable distribution is bound to have some solutions which are unassailable but impossible to justify positively. Hence, it is questionable whether this problem is really fundamental.

2.7 Addition of many asymptotically scaling variables: reason for the asymptotic invariance of the asymptotic scaling density if $0<\alpha<2$ and its non-invariance in the case $\alpha>2$

The reasoning of Section 2.5, if applied to the sum $W_N=\Sigma_{i=1}^{N}U_i$ of N asymptotic scaling variables, yields

$$P_N(u)=\Pr\{W_N>u\}\sim NP(u)\sim P(uN^{-1/\alpha}).$$

This relationship may be expressed alternatively as follows. Let a be any small probability, let $u(a)$ be the value of U such that $P[u(a)]=a$ and let $w_N(a)$ be the value of W_N such that $P_N[w_n(a)]=a$. The above approximation then becomes

$$w_N(a) \sim N^{1/\alpha} u(a).$$

Either way, the distribution of $W_N N^{-1/\alpha}$ is independent of N for large values of w; in other terms, W_N "diffuses" like $N^{1/\alpha}$. Actually, there are two obvious limitations to the validity of the approximation $P_N \sim NP$.

First limitation of the $N^{1/\alpha}$ rule for diffusion. This limitation applies when $\alpha > 1$, and hence $E(U) < \infty$. The above approximation for $w_n(a)$ is only valid if $E(U) = 0$. This gives the only possible choice of origin of U_i such that, when $N \rightarrow \infty$, the distribution of $W_N N^{-1/\alpha}$ tends to a limit over a range of w_N where $P_N(w_N)$ does not decrease to zero. To show this, write $V = U + c$; then $V_N = U_N + Nc$ and $V_N N^{-1/\alpha} = U_N N^{-1/\alpha} + cN^{1-1/\alpha}$. When $N \rightarrow \infty$, the last term increases without limit so that $V_N N^{-1/\alpha}$ cannot have a nontrivial limit distribution if $U_N / N^{-1/\alpha}$ has one. Furthermore, if $U_N N^{-1/\alpha}$ has a limit distribution, U_N / N has the degenerate limit 0 so that we must assume that $E(U) = 0$.

If, on the contrary, $\alpha < 1$, the above argument fails because $N^{1-1/\alpha}$ tends to zero. Therefore $W_N N^{-1/\alpha}$ could have a limit distribution on a non-decreasing range of values of U, whatever the origin of U. (In any event, $E(U) = \infty$ for $\alpha < 1$, so that the origin could not be $E(U)$).

Second limitation of the $N^{1/\alpha}$ rule for diffusion. This limitation applies when $\alpha > 2$. In that case, the relationship $P_N \sim NP$ can hold at best in a zone of values w_N such that the total probability $P_N(W_N) \rightarrow 0$ as $N \rightarrow \infty$. This limitation is due to the fact that U has a finite variance $D(U)$ Therefore, one can apply the classical central limit theorem to U. That is, now we can assert that

$$\lim_{N \rightarrow \infty} \Pr \left\{ \frac{W_N - NE(U)}{[ND(U)]^{1/2}} < x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-y^2/2) dy.$$

This holds over a range of values of x where the probability $P_N \rightarrow 1$ as $N \rightarrow \infty$. That is, an increasingly large range of values of W_N will eventually enter the Gaussian zone, in which $W_N - NE(U)$ diffuses like $N^{1/2}$, which increases more rapidly than $N^{1/\alpha}$. As the counterpart, *the total probability of the value of x such that $P_N \sim NP$, must tend to zero as $N \rightarrow \infty$.* Note that the $N^{1/2}$ diffusion is not radically changed if U is truncated to be less than some fixed bound. That is, the $N^{1/2}$ diffusion represents the behavior of those values of W_N that are sums of comparably small contributions whereas the $N^{1/\alpha}$ diffusion represents the behavior of sums of a single large term and many very small ones.

Now we can draw conclusions concerning the relationship between the behavior of the largest of N terms U_i and the behavior of their sum. In the asymptotically scaling case with $N=2$, the two problems are identical for all α, but as N increases, they become distinct. The problem of the maximum is the only one which remains simple and continues to lead to an asymptotically scaling variable. On the contrary, the concavity of $-\log p(u')$ is not a sufficiently stringent criterion to discriminate between those cases where $P_N \sim NP$ does or does not apply, over a range of values of u having a fixed probability.

2.8 Infinite divisibility of the L-stable distributions

If U is L-stable, we can write

$$U = A(N)\sum_{i=1}^{N} U_i - B(N) = \sum \{A(N)U_i - B(N)/N\} = \sum V_{Ni},$$

where the U_i are independent values of U and $V_{N,i} = A(N)U_i - B(N)/N$. Hence, for every N, U can be considered as a sum of N independent and identically distributed variables $V_{N,i}$. This property defines the notion of *infinite divisibility* for a random variable. Suppose that U is a L-stable variable with $1 < \alpha < 2$ and $E(U) = 0$. Then the infinite divisibility of U is made obvious by writing

$$\log\varphi(\zeta) = C\int_0^\infty (e^{i\zeta x} - 1 - i\zeta x)\,|d(x^{-\alpha})|.$$

To divide a L-stable variable by N, we need only replace C by C/N. This preserves the form of the function $\log\varphi(\zeta)$, as it should, because $(1/N)$-th of a L-stable variable is itself stable. If ε is sufficiently small.

$$\log\varphi(\zeta) \sim \log\varphi(\zeta, \varepsilon) = C\int_\varepsilon^\infty (e^{i\zeta x} - 1 - i\zeta x)\,|d(x^{-\alpha})|.$$

The contribution of $i\zeta C\int_\varepsilon^\infty x\,|d(x^{-\alpha})|$ to $\log\varphi(\zeta, \varepsilon)$ is to displace U by a nonrandom quantity. The essential term in $\log\varphi(\zeta, \varepsilon)$, namely,

$C\int_{\varepsilon}^{\infty}(e^{i\zeta x}-1)\,|\,d(x-\alpha)\,|$, is a limit of approximations of the form $\Sigma(e^{i\zeta x}-1)\,|\,\Delta(x^{-\alpha})\,|$.

Thus, one can represent *U* as a limit of sums of Poisson variables. To each increment *dx* of the variable *x*, there is a corresponding contribution to *U* equal to *x* multiplied by a Poisson variable of expected value $C\,|\,d(x^{-\alpha})\,|$. This means that a L-stable variable may be considered as a sum of variables, each of which is closely related to the uniform scaling distribution. The uniform scaling distribution may be truncated at any $\varepsilon > 0$ because the term $i\zeta x$ of $\log\varphi(\zeta)$ removes the divergence of the integral $\int_0^1(e^{i\zeta x}-1)\,|\,dx^{-\alpha}\,|$.

Appendix III describes considerations that give an intuitive basis to infinite divisibility.

3. CONCLUSION

I hope that this first application of the L-stable distributions will stimulate interest in a more detailed study of their properties. Most of the usual procedures of statistics must be revised when variance is infinite, and new questions arise. It will be important to learn how to best choose the origin u'' to insure that the L-stable density $p(u)$ and the uniform scaling density $\alpha(u-u'')^{-\alpha-1}(\tilde{u})^{\alpha}$ coincide over as large a range of values of *u* as possible. To validate a good fit, one should compare the L-stable curve with the empirical data in the region of intermediate values of *u*.

A final problem concerns the sign of $\alpha - 2$. We have referred to it several times, but a further discussion can be pursued only within the framework of a theory of L-stable processes. (One indication may already be found in Section 2 of M 1959p.) To settle this problem, it will probably be necessary to introduce some dependence between the additive components of *U*; this must be done carefully, however, to avoid obtaining a wholly indeterminate answer.

APPENDIX I: THE L-STABLE DENSITY FOR $1 < \alpha < 2$, AS $u \rightarrow -\infty$

This behavior is not reported in the literature but is yielded by the following heuristic argument. First, the bilateral generating function (a rarely used expression, but one that is useful in this context) takes the form $G(b) = \exp(b^{\alpha})$. This follows from the commonly used characteristic function by standard theorems on Fourier transforms in the complex plane. The existence of $G(b)$ implies that, as $u \rightarrow -\infty$, $p(u)$ must decrease faster than any expression of the form $\exp(|bu|)$. Expressed in terms of $v=-u$ and $f(v)=-\log p(v)$ the convergence of $G(b)$ implies that, or $v \to \infty$, $f(v)$ increases faster than any linear form of v. Furthermore,

$$G(b)=\int_{-\infty}^{\infty} \exp(bv)p(v)dv=\int_{-\infty}^{\infty} \exp[bv-f(v)]dv=\int_{-\infty}^{\infty} \exp[h(v)]dv.$$

If b is large, the integrand $\exp(h)$ is maximum for $v=w$, where w is the solution of $b=f'(w)$. Near w, we can write

$$h(v)=[bw-f'(w)]-(1/2)(v-w)^2 f''(w)+(1/6)(v-w)^3 f^{(3)}(w)+\ldots.$$

The terms of order 0 and 2 yield the approximation

$$G(b) \sim \frac{\exp[bw-f'(w)]}{\sqrt{2\pi/f''(w)}}.$$

Let us verify whether or not this approximation can reduce to $\exp(b^\alpha)$, with a $f(v)$ of the form Kv^c. With the term $\exp[bw-f'(w)]$, this goal is achieved by taking $c=\alpha(\alpha-1)^{-1}$. Extending our attention to the term involving $f''(w)$ weakens the result somewhat; instead of $\log[-\log p(v)]=\log K+c\log v$, we can only assert that $\log(-\log p)/\log v \to c$ as $v \to \infty$. Finally, consider the terms of orders other than 0 and 2. The contribution to $G(b)$ of the term in $(v-w)^2$ is nonnegligible only as long as $v-w$ is of the order of magnitude of

$$[f''(w)]^{-1/2} \sim w^{1-\alpha[2(\alpha-1)]^{-1}}.$$

In this range, the term in $(v-w)^3$ is of the order of magnitude of $w^{-\alpha/[2(\alpha-1)]}$, and is negligible. Similarly, the terms of higher order do not modify the behavior of $p(v)$. {P.S. 1996: See, at the end of this reprint, the *Annotation of Appendix I.*}

APPENDIX II: REPRESENTATIVE ASYMPTOTICALLY SCALING VARIABLES THAT ILLUSTRATE INVARIANCE UNDER ADDITION

The case $0<\alpha<1$. Consider the discrete variable whose discrete (one-sided) generating function is

$$Z(b) = \sum_{n=0}^{\infty} \exp(-bn)p(n) = 1 - C(1 - e^{-b})^{\alpha}, \text{ where } 0 < C < 1.$$

The corresponding $p(u)$ satisfy $0 < p(n) < 1$ and $\sum_{n=1}^{\infty} p(n) = 1$. For $n \gg 1$,

$$p(n) \sim \frac{Cn^{-(\alpha+1)}}{\Gamma(-\alpha)}.$$

The sum of two variables of this type has the following generating function

$$Z_2(b) = Z^2(b) - 1 - 2C(1 - e^{-b})^{\alpha} + C^2(1 - e^{-b})^{2\alpha}.$$

It follows that

$$P_2(n) \sim \frac{2Cn^{-(\alpha+1)}}{\Gamma(-\alpha)} + \frac{C^2 n^{-(2\alpha+1)}}{\Gamma(-2\alpha)} = 2p(n) + \text{ correction.}$$

For large n, the second part of the right-hand term becomes negligible compared to the first part. If $\alpha = 1/2$, this second part vanishes so that the range of values of n in which it may be neglected increases as $\alpha \rightarrow 1/2$.

The case $1 < \alpha < 2$. To obtain an acceptable generating function, it is now necessary to add a factor in $(1 - e^{-b})$. Consider, for example,

$$Z(b) = 1 - C(1 - e^{-b}) + C'(1 - e^{-b})^{\alpha}$$

where $0 < \alpha C' \leq C \leq C' + 1$ so that $C' \leq (\alpha - 1)^{-1}$. Here

$$p_2(n) \sim \frac{2Cn^{-(\alpha+1)}}{\Gamma(-\alpha)} + \frac{2CC'(\alpha+1)n^{-(\alpha+1)}}{\Gamma(-\alpha)} + \frac{C'^2 n^{-(2\alpha+1)}}{\Gamma(-2\alpha)}.$$

For large n, the second and third terms become negligible for all α. The ratio of the coefficients of the first and second terms depends little on α; but the ratio of the coefficients of the third and first terms is ruled by $\Gamma(-\alpha)/\Gamma(-2\alpha)$, which is zero for $\alpha = 3/2$ but may become large elsewhere. As a result, the third term may be important over a large range of values of n.

The case $\alpha > 2$. Each time α increases past an integer, the sign of $C(1 - e^{-b})$ must be changed, and another polynomial term must be added if $Z(b)$ is to remain a generating function. The number of corrective terms of $p_2(n) - 2p(n)$ increases, as well as the range of values of n in which the corrective terms are appreciable.

Similarly, as more than two terms are added, $p_N(n) - Np(n)$ fails to be negligible over an increasing range of values of n. Let $N \rightarrow \infty$, and observe the weighted sums of the variables U_i.

Conclusion. For $0 < \alpha < 1$, it is sufficient to consider the expression $W_N = N^{-1/\alpha}\Sigma U_i$; its g.f. is $Z^N(N^{-1/\alpha b})$, which tends to $\exp(-Cb^{\alpha})$ when $N \rightarrow \infty$, as it should.

If $1 < \alpha < 2$, one must consider the expression $W^N = N^{-1/\alpha}\Sigma(U_i - M)$, where M is the expectation ζU_i and is easily found to be C. The g.f. of W_N is clearly $\exp(NCb)Z^N(N^{-1/\alpha}b)$. As $N \rightarrow \infty$, it tends to $\exp(Cb^{\alpha})$, as expected. It is easily seen that when $M \neq C$, the g.f. of W_N does not have a nondegenerate limit.

If $\alpha > 2$, no linear renormalization of U_i can eliminate the square term Kb^2 from $\log[Z(b)]$. Hence, the best normalized sum of the U_i is the classical $N^{-1/2}\Sigma(U_i - M)$, which tends to a Gaussian for all $\alpha > 2$.

APPENDIX III: HOLTSMARK'S PROBLEM OF ATTRACTION IN AN INFINITE AND UNIFORM CLOUD OF IDENTICAL STARS

The relation between scaling and L-stable distributions and the need for the convergence factor $i\zeta x|d(x^{-\alpha})|$, both discussed in Section 2.7 look artificial. But they become very intuitive against the background of a physical problem posed and solved in Holtsmark 1919. The original context was in spectroscopy, but Chandrasekhar 1943 describes the problem in a more perspicuous restatement that involves Newtonian attraction.

Postponing convergence problems, consider a very large sphere of radius R, within which N stars of unit mass are distributed at random, uniformly and independently. A final star being located at the center Ω of the sphere. We wish to compute the resultant of the Newtonian attractions exerted on the star at Ω by the N other stars. Units will be chosen such that two stars of unit mass attract each other with the force $r^{-2} = u$. Let $\delta = N(4R^3\pi/3)^{-1}$ be the average density of stars, and let $\tilde{u} = R^{-2}$.

First consider a thin pencil (or infinitesimal cone) covering dS spherical radians, having its apex at Ω and extending in one direction from Ω. This pencil is a sum of cells, each of which is contained between r and $r+dr$ and includes the volume $dV = dSd(r^3) = dS\,|d(u^{-3/2})|$. Knowing that there is one star in the pencil dS, the conditional probability of its being in the cell of volume dV would be given by the uniform scaling distribution

$$\left|\frac{dSd(u^{-3/2})}{dSR^3}\right| = \left|d(u/\tilde{u})^{-3/2}\right|.$$

The characteristic function of this distribution is a fairly involved function $\varphi(\zeta\tilde{u})$. If there were N stars in the pencil, the probability that the attraction on Ω is u would have the ch.f. $\varphi^N(\zeta\tilde{u})$, which becomes increasingly more involved as $N \to \infty$.

The problem is simplified if the density δ is left constant but $R \to \infty$ and $N \to \infty$. Then one can assume that the number of stars in the cell of volume dV is not fixed but given by a Poisson random variable with δ as the expected density: one can easily move from one problem to the other by slightly changing the distribution of stars that are far from Ω and contribute little to u.

In this Poisson approximation, the stars located in the volume dV exert a total force that is a multiple of $u = r^{-2}$, the multiplier being a Poisson variable with expected value $\delta dS\,|d(u^{-3/2})|$. That is, the total force exerted on Ω will be the sum of a number of independent discrete jumps. The expected relative number of jumps, with a value between a and b will be $\delta dS(a^{-3/2} - b^{-3/2})/\delta dS\tilde{u}^{-3/2}$. That is, it will follow the uniform scaling distribution. The ch.f. of the total contribution of the pencil dS will be approximated by the integral

$$\log\varphi_R(\zeta) = \delta dS \int_{R^{-2}}^{\infty} (e^{i\zeta u} - 1)\,|d(u^{-3/2})|.$$

Extending the integral to $u = \infty$ raises no convergence difficulty. But careless extension of the integration to $u = 0$ ($R = \infty$) would lead to divergence: while each of the distant stars contributes little attraction, their number is such that their total expected attraction is infinite. However, the difference between the attraction and its mean is finite. Indeed, the fluctuation in the contribution of far away stars has the characteristic function

$$\varphi(\zeta) = \exp\left\{\delta dS \int_0^{R^{-2}} (e^{i\zeta u} - 1 - i\zeta u) \, | d(u^{-3/2}) | \right\},$$

which converges and tends to zero as $R \to \infty$. For the sake of convenience, the same correction $i\zeta u$ may be used for all values of u since its effect for large u only adds a finite term to U. Hence, the difference between the attraction of the stars in the pencil dS and the mean value of this attraction is a positive L-stable variable with $\alpha = 3/2$. {P.S. 1996: this *removal of infinities* is a form of a more general procedure called renormalization.}

In this context, the meaning of the L-stability is easy to understand. Consider two clouds of "red" and "blue" stars, having the same density and filling the same pencil dS. The difference between the forces exerted on Ω by red or blue stars alone, or by both together, reduces to a scale factor and does not affect the analytic form of their distributions.

A large negative value u can occur only if there is an abnormally small number of stars in the pencil dS. Each individual "missing" star contributes little to U, so their total contribution is also small, unless their number is very large, which is very unlikely.

To the contrary, a very large positive u may result from the presence of a *single* star near Ω, *irrespective of the density of stars elsewhere*. This event is far more likely than the combination of events required for a negative u. It is easy to check that the asymptotic behavior for large u is the same for the total attraction U and the attraction of the nearest star. {P.S. 1996: See at the end of this reprint, the *Annotation to Appendix III*.}

ACKNOWLEDGEMENT. This paper was started while the author was associated with the University of Geneva (under partial support from the Rockefeller Foundation), and later with the University of Lille and the École Polytechnique in Paris.

&&&&&&&&&&&& ANNOTATIONS &&&&&&&&&&&&

Editorial comment. The original denotes sums of random variables by + surrounded by a circle. Figure 1 was reorganized with three large-size superposed curves replacing small curves next to each other. The term *law* was preserved when it refers to an empirical regularity, but replaced by *distribution* when it refers to a probability. Some long sections were broken up into subsections. Published *Errata* were implemented.

Annotation to Figure 1. This figure is redrawn using the original data: instead of three small plots side to side, as in M 1960, we see three large plots printed over one another. Linear coordinates were used because a goal of this paper was to show that the L-stable has a bell somewhat like the lognormal's, hence, is not unreasonable from the viewpoint of interpolating the "Paretian" tail of the actual income data. An unfortunate but inevitable consequence of using a linear plot is that the tails are far off-scale. If the scales were chosen to allow a wide range of the abscissa, the tails of the L-stable densities would become indistinguishable from the horizontal axis, while the bells would draw very close to a vertical half-axis. Such illegible graphs are effectively not worth plotting. In any event, theory yields the tail density $\sim u^{-\alpha-1}$ and the tail probability $P(u) \sim u^{-\alpha}$.

A remarkable feature that is seen on this plot is that the most probable value of u has the same density for all the values that were considered.

Much more extensive data were available, however, and a plot on log-log coordinates became interesting as soon as M 1963b{E14} tackled the symmetric L-stable and raised the issue of how the tails attach to the bell, especially when α is close to 2. The plot for the maximally asymetric case was then prepared, but was only distributed as a figure of M 1963j; it is reproduced in Appendix IV of this chapter. Needless to say, many such log-log graphs were published since 1963.

Annotation to Appendix I: behavior of the density in the short tail of the maximally skew L-stable densities. The result this Appendix obtained heuristically has been confirmed; the reference is Zolotarev 1986.

Annotation to Appendix III: removal of an infinite expectation reinterpreted as a form of "renormalization." In physics, an original "brute force" approach often yields a divergent integral, but a closer look (often, a much more difficult one, technically) yields a finite difference between two infinite terms. The formula Appendix III derives for the L-stable density is an independent example of this very general procedure. It helps establish that the concept of renormalization had several parallel incarnations.

&&&& POST-PUBLICATION APPENDICES &&&&

M 1963i{E10} corrected, and amplified upon, the comments in M 1960i{E10} concerning $\alpha \sim 2$. Those earlier comments were deleted, and the proposed replacements inserted, in edited form.

APPENDIX IV (ADAPTED FROM M 1963i, j) ESTIMATION BIASES THAT SEEM TO YIELD $\alpha > 2$

The L-stable densities have a bell followed by one or two scaling tails. But where do those tails begin, and how smoothly do they merge with the bell? Those questions are immaterial in mathematics, but essential to concrete applications. They are answered by numerical tables, but for three sets of parameter values the L-stable density is known explicitly.

Figure 1 plotted some numerically obtained L-stable densities in natural coordinates. In Figure 3, the corresponding tail probabilities are replotted on log-log coordinates for a larger number of αs. This yields immediately the principal observation of this appendix: near $\alpha \sim 2$ the L-stable density exhibits an inverted *S*-shape. As a result, statistical fitting of the theoretical distribution by a straight line would readily yield biased estimates of α, larger than the true value, and perhaps even larger than 2. This finding will serve in Appendix V to argue that L-stable distributions may in fact also represent actual data that appear to be scaling with $\alpha > 2$, a value incompatible with L-stability.

IV.1 Continuity of the skew L-stable density as function of α

This appendix excludes $\alpha = 1$ with $\beta \neq 0$ and sets $\gamma = 1$ and $\delta = 0$. If so, once again, the characteristic function of the L-stable distribution is

$$\varphi(\zeta) = \exp\left[i\delta\zeta - |\zeta|^{\alpha}\gamma\{1 - (i\beta\zeta/|\zeta|)\tan(\tfrac{1}{2}\alpha\pi)\}|\cos(\tfrac{1}{2}\alpha\pi)|\right].$$

This $\varphi(\zeta)$, the resulting tail probability $\Pr\{U > u\} = P(u)$ and the probability density $-P'(u)$ are continuous functions of α and β. A major complication is that the bell behavior is not known, with three already-mentioned exceptions. But the tail behavior of $-P'(u)$ is known; with only two exceptions, there are two scaling tails with α as exponent. The first exception is the Gaussian limit case $\alpha = 2$, for which β plays no role and both tails are shorter than any scaling. The second exception, $\alpha \neq 2$ and $\beta = \pm 1$, when one tail is very short, is at the center of M 1960i{E10}.

IV.2 True and apparent scaling for $|\beta| = 1$ and α close to $\alpha = 2$

For $\alpha = 2$, the graph of $\log[-P'(u)]$ versus $\log u$ drops exponentially. For $\alpha = 2 - \varepsilon$, its form must be an inverted letter S. Near the inflection, the local slope becomes much larger than $\alpha + 1 = 3 - \varepsilon$, then it falls down and eventually stabilizes at the theoretical $\alpha + 1$. A straight line approximating more than the strict tail is necessarily of slope well above three.

Similar results apply to the tail probability plotted in Figure 3. The largest data corresponds to the region in which the slope varies near its maximum; however, the dispersion of sample values makes it hard or

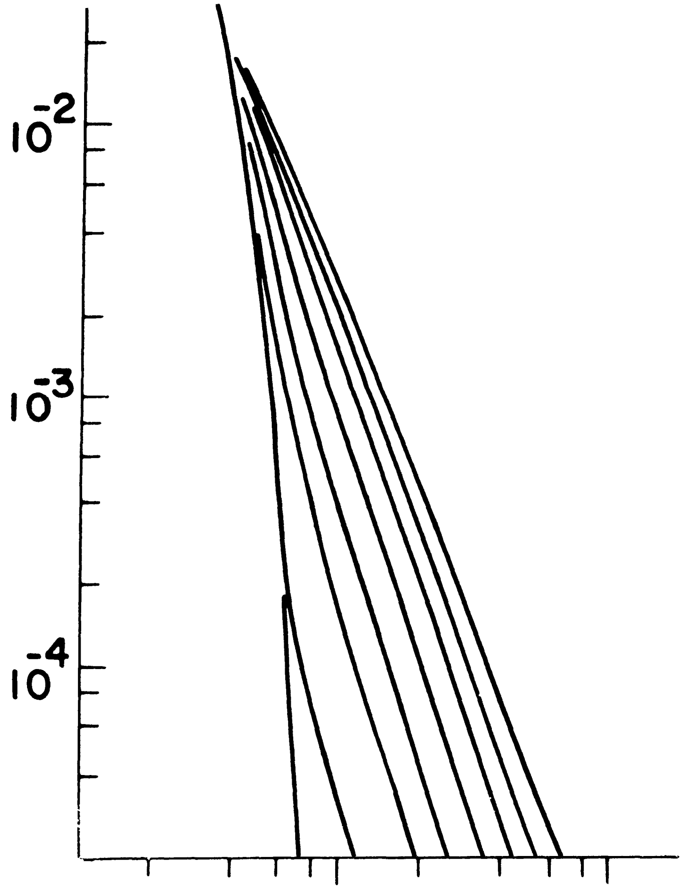


FIGURE E10-3. Densities of certain maximally skew L-stable distributions. These densities are defined as having the normalized Fourier transform $\varphi(z) = \exp\{-|z|^{\alpha}[1 - i|z|z^{-1}\tan(\alpha\pi/2)]\}$. Reading from the lower left to the upper right corner, α takes the values 2, 1.99, 1.95, 1.9, 1.8, 1.7, 1.6 and 1.5.

impossible to observe the curvature of the population function $\log[\Pr\{U > u\}]$. Hence, it will be irresistibly tempting to interpolate the curve of empirical data by a straight line of slope greater than the asymptotic α, often greater than 2. Those results continue to hold as α decreases but becomes more and more attenuated.

From the viewpoint of data analysis, moving the origin of u away from $\tilde{U} = EU = 0$ could broaden the scaling range, that is, insure that the graph of $\log[Pr\{U > u\}]$ versus $\log(u - \tilde{u})$ becomes straight over a longer range of values of u.

IV.3 Ranges of scaling near the Cauchy and the "Cournot"

In two explicitly known cases, namely the Cauchy and the "Cournot" (maximally skew with $\alpha = 1/2$), there is a simple explicit formula for L-stable density, hence also for the discrepancy between it and its scaling approximation. As the parameters are made to differ slightly from those special cases, this discrepancy varies smoothly. It is called a "regular" perturbation.

The Cournot special case $\alpha = 1/2$ and $\beta = \pm 1$. This, the only skew L-stable distribution for which the density has an explicit analytic expression, enters into the study of the returns of Brownian motion $B(t)$ to its initial value $B(0)$. In this case,

$$-P'(u) = (\text{constant})u^{-3/2} \exp(-1/u).$$

In log-log coordinates, and setting $\log u = v$, one obtains

$$-\log[-P'(u)] = (\text{constant}) - (3/2)\log u - 1/u = (\text{constant}) - 3v/2 - e^{-v}.$$

This log-log graph is asymptotic for $v \to \infty$ to a straight line of slope $-3/2 = -\alpha - 1$, and for all v it lies below this asymptote. The α estimated from a straight portion of the empirical curve underestimates the true value $\alpha + 1$.

To broaden the scaling range, one is tempted to choose $\tilde{u}$ so as to increase the straightness of the graph of $\log[Pr\{U > u\}]$ versus $\log(u - \tilde{u})$. Observe that the maximum of $-P'(u)$ is attained for $u = 2/3$, but choosing $\tilde{u} = 2/3$ would make our graph even less straight. For $\alpha = 1/2$, it is necessary to choose a negative value of $\tilde{u}$.

Blind adherents of the lognormal distribution will see a bell recalling an inverted parabola, with shortening to the left and outliers to the right.

The Cauchy case $\alpha = 1$ and $\beta = 0$. For the well-known density given by $-P'(u) = 1/\pi(1+u^2)$, the log-log graph is asymptotic to a straight line of absolute slope $\alpha + 1$, and for all v it lies below this asymptotic.

The skew case with $\alpha = 1$. It is of interest to the distribution of firm sizes, but has no analytic expression. As α decreases from 2 to 1 in formula (*) of Section 1, $EU \rightarrow \infty$ and so do the mode and the median. As α increases from 0 to 1, $EU = \infty$, but the mode and the median $\rightarrow \infty$.

APPENDIX V (ADAPTED FROM M 1963j): OVERALL AND DISAGGREGATED DISTRIBUTIONS OF PERSONAL INCOME

V.1 Practical implications of Appendix IV

When the exponent α is near 1.5, the L-stable density rapidly reaches its asymptotically scaling behavior, and it can be markedly skew. This is why M 1960i{E10} could claim the L-stable with this range of values of α as a promising candidate to represent the distribution of personal incomes in non-industrialized countries originally studied by Pareto. In industrialized countries, however, it is reported that the empirical exponent increased to $\alpha \sim 2$, or even $\alpha > 2$. The largest values of α were reported after World War II. If $\alpha > 2$ were to be confirmed, and/or to become the rule, the L-stable would cease to be a reasonable candidate to represent the distribution of income.

Appendix IV showed that estimation bias causes the observed α of an L-stable density to exceed the true (and asymptotic) α, and even to exceed 2. But the extent of straightness of the log-log graphs with $\alpha \sim 2$ remains puzzling. Besides, the existence of an income tax has changed income-before-taxes in industrialized countries, as well as income-after-taxes, and generally led to more complicated economic structures which examine the distributions of personal income within disaggregated age and professional groups: those distributions exhibit the same inverted *S*-shape and are less skew than the overall distribution. Therefore, the situation near $\alpha \sim 2$ is complicated, but the L-stable distribution remains in contention.

V.2 The lack of skewness of the L-stable densities near α=2

The L-stable distribution for $\alpha \sim 2$ is symmetric, while the data are skew. But skewness is greatest in the overall distribution of income, while most distributions of incomes coming from narrowly defined sources are practically symmetric. When most low or middle-income occupations are taken out, the distribution of the remainder is less skew than the overall distribution. Hence, the study of high-income occupations should not be concerned about skewness. Of course, the above argument does note explain why the apparent slope extrapolated from the tail of the curve also represents part of the middle-income data. This fact may be covered by the argument in M 1962g.

V.3 The distribution of personal income viewed as "mixture" of disaggregated distributions

Several portions of this book express my deep reluctance to account for nonGaussian distributions as being mixtures of simpler ones. This reluctance vanishes, however, when the participating distribution can actually be separated. One such case is illustrated in Figure 4.

M 1963e{E3} shows that the scaling distribution is invariant under mixing: If the U_n are such that $\Pr\{U > u\} \sim C_n u^{-\alpha}$, where α is independent of n, and if one mixes data from those various distributions in the proportions p_n, one obtains a random variable U_W such that $\Pr\{U_W > u\} \sim (\Sigma p_n C_u) u^{-\alpha}$. This results from the fact that the doubly logarithmic graphs of the scaling distribution are straight lines with no "kink" that any mixture could smooth off.

But Appendix IV shows that the L-stable distribution does have two clear-cut kinks, especially if α is near 2. Therefore, even if the distribution of income is L-stable within sufficiently narrowly defined categories, the overall distribution would appear scaling with a high α.

4. Verification of the preceding conjecture on empirical data

When writing M 1963i{E10}, I knew nothing of the actual distribution of income within narrow *high*-income categories. Data that came to my attention in 1963 suggest that the decomposition described in the preceding section constituted a correct prediction and that I was too cautious in handling the consequences of my thoughts. Figure 4 is drawn from graphs communicated privately by H. S. Houthakker, and used by him (for entirely different purposes) in the preparation of Houthakker 1959. The main fact is that the curves whose absolute slope is "on the whole" the largest are also the least straight.

Another inference is that α changes during a person's lifetime, thus throwing doubt on all "diffusion models" which obtain the scaling distribution as the limit state resulting from the action of various transformations, and throwing doubt also on the model in M 1961e{E3}. But M 1962q{E12} probably remains valid.

The curves of Figure 4 are distinguished as follows.

Reading from the lower left to the upper right corner, the bold lines refer to the following levels of education, as measured by the number of years in school: None, 1-4, 5-7, 8, high school 1-3, high school 4, college 1-3, college 4 or plus.

Reading from the lower left to the upper right corner, the bold lines refer to the following age groups: 14-15, 16-17, 18-19, 20-21, 22-24, 25-29, 30-34, 45-54. The dashed line refers to all income-earners with 8 years of schooling and aged 25 or more.

Reading from the lower left to the upper right corner, the bold lines refer to the following age groups: 20-21, 22-24, 25-29, 30-34, 45-54. The dashed line refers to all income-earners with 4 years of college or more and aged 25 or more.

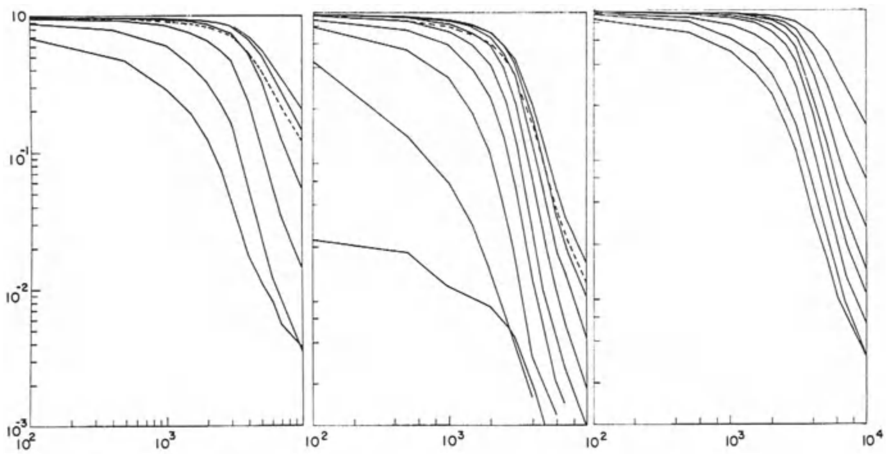


FIGURE E10-4. Distribution of 1949 incomes in the U.S.A. A) all persons from 35 to 44 years of age. B) persons with 8 years of school. C) persons with 4 years of college or more. Horizontally: Income u in dollars. Vertically: Number of persons with an income exceeding u.

Econometrica: 29,1961, 517-543.

E11

L-stability and multiplicative variation of income

✦ **Abstract.** This paper describes a theory of the stationary stochastic variation of income based on a new family of nonGaussian random functions, $U(t)$. This approach is intimately connected with random walks of log $U(t)$, but no use is made of the "principle of proportionate effect." Instead, the model is based upon the fact that there exist limits for *sums* of random functions, in which the effect of chance in time is *multiplicative*. This feature provides a new type of motivation for the widespread, convenient, and frequently fruitful use of the logarithm of income, considered as a "moral wealth."

I believe that these new stochastic processes will play in linear economics, for example in certain problems of aggregation. The reader will find that the results are easily rephrased in terms of diverse economic quantities other than income. As a result, the tools to be introduced may be as important as the immediate results to be achieved. In particular, the distribution and variation of city sizes raises very similar problems. ✦

THIS CHAPTER is devoted to the variation of personal income, rather than its distribution at a point in time. Sections and paragraphs printed in a smaller font can be omitted without disrupting the continuity.

1. INTRODUCTION AND GENERAL CONSIDERATIONS

1.1. The asymptotic scaling distribution

It is well known that the fundamental observation concerning the distribution of income is an *asymptotic form of the* scaling distribution (Pareto

1896). Let $U(t)$ be the function that gives an individual's yearly income, U, as a function of the year during which it was earned, t. The asymptotic scaling distribution then states that there exist two constants, C and α, such that, when t is fixed, one has

$$\Pr\{U(t) > u\} \sim Cu^{-\alpha}.$$

The notation $A(u) \sim B(u)$ means that the ratio $A(u)/B(u) \to 1$ as $u \to \infty$.

M 1960i{E10} further assumed that $1 < \alpha < 2$. The requirement that $\alpha < 2$ eliminated certain cases whose discussion had to be reserved for another occasion. When α is between 1 and 2, one can study the asymptotic scaling distribution without having to refer to the variation of income in time. The key to the model is the decomposition of income *itself* into additive parts.

The "synchronic" character of the models in M 1960i{E10} was, of course, a little unusual. Indeed, in most other theories, the *distribution* of income comes out as the steady-state that corresponds to some "diachronic" model of the *variation* of income in time. In fact, the most usual approach is of the following type: One starts by assuming (sometimes only implicitly or indirectly) that there exists an origin for U such that the next value of income, $U(t+1)$, is equal to $u(t)$ multiplied by some random factor (except, perhaps, when $u(t)$ is small). Selecting $E(U)$ as the origin of U (except when otherwise noted), the above "law of random proportionate effect" amounts to postulating that there exists a constant B, such that the variation of $\log[U(t)+B]$ in time results from the addition of random increments. As a result, the expression $\log[U(t)+B]$ is considered as being more intrinsic than $U(t)$ itself.

Additional assumptions are, however, necessary to insure that a steady-state distribution for $\log(U+B)$ exists, and that it is the exponential distribution, as required by the asymptotic scaling character of U itself. One must also eliminate solutions similar to the lognormal law of Gibrat 1932, according to which $\log(U+B)$ is Gaussian.

It is particularly convenient to derive the existence of a steady state of U from the following set of hypotheses, which are very similar to those used by Champernowne 1953. Suppose that $\log U(t)$ *performs a random walk, or a diffusion, with a downward trend and with a reflecting layer*. That is, make the following assumptions:

(a) $U(t)$ is a *Markovian sequence* in discrete time, so that $U(t+1)$ depends mostly on chance and on the known value of $u(t)$, but not on earlier values of income.

(b) As long as $u(t)$ is large, $\log U(t+1) - \log u(t) = T$ is a random variable *independent* of $u(t)$. If the values of $\log U(t)$ are multiples of some unit, the model is a "random walk," otherwise, a "diffusion" process.

(c) As long as $u(t)$ is large, the expected drift $E[U(t+1) - u(t) \mid u(t)]$ remains *negative*.

(d) When u becomes small, the asymptotic transition probabilities of $\log U(t)$ are modified so as to *prevent all $u(t)$ from eventually vanishing*, as they would if a process following (a), (b) and (c) were left to act for all $u(t)$. This last hypothesis implies a diffuse form of the "reflecting barrier" of the usual theory of random walk or diffusion, as treated, for example, in Feller 1957. We call this a "reflecting layer."

The above four statements will be referred to as the *asymptotic Champernowne conditions*. If they are verified, one can show that, whatever the initial distributions of income, $V(t)$ eventually becomes scaling with $\alpha < 1$, If, on the contrary, (c) and (d) are not verified, one may obtain the normal law for $\log(U + B)$, instead of the exponential.

The above assumptions are so simple in form and so little "unexpected," that it is perfectly reasonable to say that Champernowne has given an "explanation" of scaling. Moreover, some data are provided in Champernowne 1953 as an empirical basis for (b) and (c). But empirical data concerning the transition probabilities of $U(t)$ are negligible in quantities and quality, next to the of data concerning the distribution of $U(t)$ for fixed t. Therefore, the most convincing basis for the second edition of Champernowne is still provided by the general, indirect evidence of the validity of the law of random proportionate effect.

This law is not so well established, however, that it would not be most worthwhile to find other, and preferably very different, theoretical grounds for the widespread, convenient, and frequently successful use of the logarithm of income. The third and fourth assumptions could also be strengthened by new and different evidence. These are the minimum aims we hope to achieve in the present paper. To achieve them, we shall be careful to remain within the natural scale of income itself.

This paper will be limited to stationary – time-invariant – processes. This assumption is a natural a point of departure, but is objectionable. For example, the periods of unquestionable stationarity of income variation have been short in recent times, so that it may be that the unconditioned distribution of $U(t)$ never reached the steady state resulting from the repeated application of any given set of laws of change. I shall not examine this difficulty any further, however, but shall point out that small α are observed for less developed countries, which are those for which the periods of stationarity may be the longer. Hence, nonstationarity may arise in the same cases as the problem raised by the sign of $\alpha - 2$. This does not imply that they are necessarily related, but both must be examined more carefully – sometime later.

1.2. The study of the evolution of income in time. Inapplicability of the usual theory of second order stationary random processes

Working in the scale of income itself immediately presents problems when one tries to describe the stochastic mechanism generating $U(t)$. The applied mathematician is frequently presented with such tasks these days, and his automatic first response is to investigate what can be obtained from the application of the "classical" theory of the second order random processes, due to Wiener, Khinchin, Wold and many others (see Hannan 1960). The first step is to form the *covariance function* $C(s) = E\{U(t)U(t+s)\}$, which must be assumed to be finite for all. The Fourier transform of $C(s)$ is the density of the various spectral components in the harmonic decomposition of $U(t)$ into a sum of sine and cosine components. The properties of $U(t)$ that follow from the second degree expression $C(s)$ are called "second order properties."

Unfortunately, if $1 < \alpha < 2$, *the theory of weak second order processes is not applicable to income.* To see why, it suffices to note that $C(0) = E(U^2) = \infty$ when $U(t)$ is asymptotically scaling with $1 < \alpha < 2$, and $C(s)$ remains infinite for small s because of the strong dependence observed between incomes of the same individual over not too distant years.

When the second order theory ceases to apply, there is nothing of comparable generality to replace it, and we are prevented in this paper from the customary resort to a spectral study of $\log U(t)$. Let us recall, however, that the second-order theory is only rarely used in full generality. In many cases, there are good reasons to replace a stationary process by a Gaussian process with the same covariance $C(s)$. This paper generalizes Gaussian processes to the case where the covariance is infinite.

The first reason given for the use of Gaussian processes is that they are simple analytically and fully determined by $C(s)$. Hence, whenever one is interested only in second order properties (or whenever only $C(s)$ is available), one may be content with a Gaussian approximation, even when the actual process is far from normal. The same is true when simplicity is a primary consideration. All these reasons have no counterpart when $C(s)$ is infinite.

On the other hand, one can also argue that many empirically motivated random processes cannot really be very far from *actually* being Gaussian, because of the role played by the normal distributions in the theory of limits of sums of random variables and because of related properties of "L-stability." It turns out that the Gaussian processes are not alone in having these properties, as we shall now proceed to show.

1.3. Limits of linear aggregates of parts of income, Gaussian and NonGaussian stable limits. Definition of the L-stable random variables and of certain L-stable scaling stochastic processes

1.3.1. Definitions of L-stability. The random variable U is called an L-stable random variable. It is the limit as $N \to \infty$ of a sequence of expressions $U_N = A(N)\sum_{i=1}^{N} V_i - B(N)$, which are linearly weighted sums of independent and identically distributed random variables V_i. The expressions V_i and U may be scalar or vectorial.

Similarly, L-stable random sequences in discrete time are the possible limits, as $N \to \infty$, of sequences of weighted sums of N components $V_i(t)$ having the following properties: For different and any fixed i, the scalars $V_i(t)$ can be interdependent. However, for any fixed t, and different i, the scalars $V_i(t)$ are independent and identically distributed. Similarly, for any fixed set of $t_k(1 \leq k \leq K)$ and different i, the vectors $\{V_i(t_1), ..., V_i(t_k), ..., V_i(t_K)\}$ are independent and identically distributed.

1.3.2. The above definitions involve linear expressions and passages to the limit. Both procedures are familiar in economics. For example, one may argue that the L-stability of $U(t)$ is an "asymptotic" assumption about the "details of the nature of income," and "ought to" be satisfied in a first approximation. But one also assumes usually that the covariance of the random process $U(t)$ is finite. Under these conditions, it is well-known that only the stable processes are Gaussian, while the process of variation of $U(t)$ is conspicuously *nonGaussian.*

This apparent contradiction has been used to invalidate all arguments based on decomposing income *itself* into additive parts. But contradictions disappear if the second moment need not be finite. *Then a wide family of nonGaussian stable elements becomes available to the model builder.*

This family is too broad for our present purposes. One must first restrict it by assuming that $E(U)$ is finite. In addition, large negative values of income cannot occur, except perhaps with a very small probability. (And even if they were completely impossible, one could accept a theoretical probability law that makes them sufficiently unlikely, in the same way as one accepts the Gaussian law as representative of the distribution of many strictly positive quantities.)

One is then left with the family of scalar L-stable distributions (M 1960i = Chapter XX) and with the random functions studied in this paper. The very existence of these random elements shows that to save the usual limit argument, it may not be necessary to apply it to log $U(t)$. Instead, the variation of income (as well as its distribution, see M 1960i) may perhaps be described with the help of processes obtained from a Gaussian $U(t)$ by replacing every normal distribution by a nonGaussian stable distribution having the same number of dimensions.

I shall show in the present paper that this recipe is indeed successful, in the sense that a wide family of scaling L-stable sequences $U(t)$ satisfy Champernowne's conditions, in the first approximation. This can be shown without assuming a priori the law of random proportionate effect. In the second approximation, I shall show that the effect of chance upon a L-stable process of our type may be described for large $u(t)$, by

$$U(t+1) = TU(t) + R,$$

where T is a random variable (except in one degenerate case), and where R is either zero or a L-stable random variable independent of T.

Other conditions of Champernowne remain valid. We think that the above equation in $U(t)$ is a quite natural and acceptable improvement over the usual statement of the principle of random proportionate effect.

1.3.3. The replacement of the normal by a nonGaussian stable distribution cannot be casual, because deep differences exist between the two kinds of limits for weighted sums of identical random variables. The *first difference* refers to the meaning that can be given to the statement that "each of the addends $A(N)V_i - B(N)/N$ provides only a small proportion of the sum U_N." Clearly, since $A(N)$ tends to zero with $1/N$, the same is always true of the probability that $|A(N)V_i - B(N)/N|$ be greater than any positive ε. Therefore, one can always say that, as $N \rightarrow \infty$, the "a priori" or "ex ante" values of the additive components of U_N all become relatively very small. But, of course, this implies nothing concerning the "a posteriori" or "ex post" values of $A(N)v_i - B(N)/N$, because the number of these expressions is equal to N and becomes very large. In fact it is well known that, *in order that the "a posteriori" largest of the addends contribute negligibly to the sum, it is necessary and sufficient that the sum be Gaussian.* Therefore, if the sum is not Gaussian, the largest term is not negligible *a posteriori*. In fact, it is a substantial part of the whole, and is even predominant if the ex post value u_N is sufficiently large.

This raises a very important question: when economists speak of a variable as the sum of very many "small components," do they wish to imply that the components are small only *a priori* or also *a posteriori?* We believe that in economics, as well as in other behavioral sciences, one typically finds that if an event is due to the additive combination of many causes, a few of these may contribute a very substantial part of the whole. Suppose then, that one has to choose between an economic probability distribution derived from the Gaussian, and a distribution associated with another stable law. It is quite possible that the Gaussian law be *a priori* eliminated from consideration because of its prediction concerning the relative *a posteriori* size of the few largest of many contributions to an additive whole. (See M 1963b{E14} and Section 9 of M 1962e{E12}.)

Consider another aspect of the distinction between the Gaussian limit and all others. As is well known, the normal law is obtained whenever the variances of the addends are finite, a property that implies little about the analytic behavior of the probability density $p(u)$, for large values of the argument u.

In comparison, the following necessary and sufficient conditions of convergence to a nonGaussian stable limit will appear to be extremely stringent. In the one-dimensional case, the variables V_i must satisfy a barely weakened form of the asymptotic scaling distribution. Otherwise, one cannot choose a set of two functions $A(N)$ and $B(N)$, so that the weighted sum of the V_i converges to a limit. Similarly (see Rvaceva 1954, Sakovich 1956, Takano 1954, 1955), when $U(t)$ is a nonGaussian stationary sequence, the transition probabilities of the contributing sequences behave according to the law embodied in the expression $V_i(t+1) = TV_i(t) + R$, with T and R as in the last formula of Section 1.3.2. (Actually, one must accept addends that are slightly more general than either of the above cases.)

Consequently, one can express only mild surprise at the fact that L-stable random scalars are themselves asymptotically scaling. M 1960i, or at the fact that, asymptotically, L-stable random processes $U(t)$ behave exactly like their additive components $V_i(t)$.

But the mildness of our surprise should not be exaggerated. After all, the term "scientific explanation" can be used for statements that are intuitive and simple, yet imply the more complicated statements one wishes to account for. In the present case, the assumption that income is not Gaussian but, rather, is a sum of very many independent and identically distributed components is indeed simple and intuitive, and it requires, and hence "explains," both the fact that income itself must be a L-stable variable *and* the fact that its components must have some special properties, analogous to those of the whole.

It may be noted that stochastic models frequently violate the above idea of what is a scientific explanation. Because of the weakness of the condition of validity of many limit theorems of the calculus of probability, the properties of the limits imply very little about the parts.

1.4. Linear aggregation of independent incomes; inversion of the argument of Section 1.3

1.4.1. The exclusive use of addition on the natural scale of incomes is important from the viewpoint of the place which the present theory of income distribution takes within the broader context of economics. Suppose, in particular, that one is entirely satisfied with the usual foundation for the asymptotic Champernowne assumptions. In that case, the introduction of the slight formal complications of the new theory will undoubtedly be postponed; but one is likely to encounter them later.

To show this, let us examine the behavior of the aggregate income of a group of individuals, when each person's income follows the same asymptotic Champernowne process with an infinite covariance (for example, suppose that each person's income follows a L-stable process). If the group is sufficiently large, its aggregate income will then follow a L-stable process, which will depend upon the size of the aggregate only through the values of two functions $A(N)$ and $B(N)$, which occur in the unique scaling that makes $A(N)\sum V_i - B(N)$ converge to a normalized L-stable limit. As a matter of fact, the aggregate will also be "asymptotically" Markovian (see Section 2.3.4). In such a case, it is fully meaningful to speak of an "aggregate distribution of income."

Suppose, on the contrary, that the processes of variation of individual income have an infinite covariance, but are neither Champernowne processes nor the slight generalizations mentioned in Section 1.3. Then no linear scaling will be able to make the aggregated income tend towards a nondegenerate limit (i.e., a limit other than a constant, because constants can always be obtained by making $A(N)$ grow very fast). As a result, the probability laws of aggregates will depend upon their sizes in a critical fashion.

1.4.2. The preceding observation suggests an inversion of the argument proposed in Section 1.3. and M 1960i{E10}. To begin with, if one adds essentially positive, independent, and identically distributed processes, the aggregate can only be a Gaussian or another L-stable variable (or a constant). Hence, L-stable processes are bound to occur in economics, sooner or later, for example, in the case of aggregates of different incomes. Now suppose that individual income itself can also be represented as an aggregate of independent and identical parts. In that case, even though the properties of the parts may permanently remain somewhat conjectural, L-stable processes will also provide a model for the variation of individual income.

Of course, the identity and independence of the parts of income *is not* an intuitively obvious assumption. It is best justified by the desire to deny (at least tentatively) that economics can reach any genuinely "microscopic" level and to assert in particular that individual income is not a microeconomic quantity. One might say that identity and independence correspond to a first approximation of complete "disorder," in which no part is allowed a priori to contribute more than any other part.

For another comment concerning the independence between the aggregated parts, see Section 1.9.

If we also required that the different parts be roughly equal a posteriori, we would restrict ourselves to the Gaussian case. If we did not require that the different parts contribute equally *a priori*, we could generalize our theory to the so-called self-decomposable laws of Paul Lévy (which include some asymptotic Pareto variables with $\alpha > 2$.) But the model would cease to be genuinely macroeconomic, and it would also have other drawbacks. I hope to address these issues at a later date.

The above decomposition procedure has also been applied by the author to several other problems in social science, especially in linguistics and sociology.

1.5. A consequence of L-stability: the invariance argument

The original definition of L-stability given by Paul Lévy differs from the preceding definition via a limit process. It received more emphasis in M 1960i, and we shall now present it as a theorem. The point of departure there was the surprising fact that asymptotic scaling seems to hold for "income" no matter which exact definition is chosen for that concept. That is, it applies to partial sums of isolable parts of income as well as to the total sum. It turns out that L-stable variables and processes satisfy a strong form of the invariance property:

Suppose that U is a random scalar, vector, or process, and let a' and a'' be positive quantities. We wish to be able to write $(a'U' + b') + (a''U'' + b'')$ as $(aU^\circ + b)$, where U', U'', and U° are realizations of U, and a is positive. It can be shown that the necessary and sufficient condition for this relation to hold is that U be a L-stable element.

A priori the invariance argument may also appear to be "explanatory" only in a asymptotic meaning of the word. But actually, very similar procedures are found to be perfectly satisfactory in several branches of physics. The fact that both hypothesis and conclusions are better established in physics than in economics is irrelevant if one is concerned only with the acceptability of the present argument as an "explanation."

The above invariance means the following: given the kind of extreme uncertainty which exists in economics concerning the "things" which are represented by the available quantitative data, "the" properties of the distribution of income could never have been observed if that distribution were not L-stable. From a certain viewpoint, the "Nature" that "chooses" the distribution of income is as cooperative with the statistician as it could ever be (since if Nature went a step further and made income nonrandom, the statistician would become unnecessary).

1.6. Structure of one and two-dimensional L-stable random elements

1.6.1. L-stable scalars. L-stable scalars (M 1960i{E10}) nonGaussian stable variables with a finite mean value, and with a very small probability of taking large negative values. Their probability densities $p(u)$ are deduced from their bilateral generating functions, which take the form:

$$G(\sigma) = \int_{-\infty}^{\infty} \exp(-\sigma u)\, p(u) du = \exp[-M\sigma + (u^*\sigma)^{\alpha}],$$

where $\sigma > 0$, $u^* > 0$ and $1 < \alpha < 2$. It then follows that the characteristic functions are of the form

$$\phi(\zeta) = \int_{-\infty}^{\infty} e^{i\zeta u} p(u) du = \exp\left[iM\zeta - |u^*\zeta|^{\alpha} \left\{ 1 - \frac{i\zeta}{|\zeta|} \tan \frac{\alpha\pi}{2} \right\} |\cos \frac{\alpha\pi}{2}| \right].$$

If $M = E(U) = 0$ and if $u^* = 1$, one has what is called a *"normalized"* or *"reduced" L-stable variable.* It will hereafter be denoted by L_α. Its density was tabulated in M & Zarnfaller 1961.

When the positive σ in the above expression $G(\sigma)$, is replaced by a complex variable β, one obtains a function of a complex variable having a singularity at the origin. $\phi(\zeta)$ is the value of that function along the imaginary axis and it should be determined by following that sheet of the Riemann surface of $G(\beta)$, which makes $G(\sigma)$ real when σ is real and positive.

Three classical properties of the L-stable variables are worth recalling:

(A) Each L-stable density follows the asymptotic scaling distribution:

$$\text{for } u \rightarrow \infty \text{ one has } p_\alpha(u) \sim \frac{1}{\Gamma(-\alpha)} u^{-(\alpha+1)}.$$

(B) Let $aL^\circ_\alpha = a'L'_\alpha + a''L''_\alpha$, where L'_α, L''_α, and L°_α are independent realizations of L_α. Then $G(a'\sigma)G(a''\sigma) = G(a\sigma)$ and $a^\alpha = (a')^\alpha + (a'')^\alpha$.

(C) Keeping the same notation, the mean value of $a'L'_\alpha$, when conditioned by the known value u of aL°_α, is $E(a'L'_\alpha | u) = u(\alpha')^\alpha / (a)^\alpha$. To prove this, suppose first that $(a')^\alpha/(a)^\alpha = m/n$ is rational. Introduce a n L-stable variable $L_{i\alpha}$, such that $aL^\circ_\alpha = \sum_{i=1}^{n}(a/n^{1/\alpha})L_{i\alpha}$, while $a'L'_\alpha = \sum_{i=1}^{m}(a/n^{1/\alpha})L_{i\alpha}$. Although the conditioned variables $\{(a/n^{1/\alpha})L_{i\alpha} \mid u\}$ are not independent, it is known that their expected values can be added. Hence, we can write: $E\{(a/n^{1/\alpha})L_{i\alpha} \mid u\} = u/n$, and $E(a'L'_\alpha \mid u) = mu/n = u(a')^\alpha/(a)^\alpha$. The generalization to an irrational ratio $(a')^\alpha/(a)^\alpha$ is immediate.

1.6.2. Two-dimensional L-stable vectors. Lévy 1937-1954 (Section 63) shows that, if $E(X, Y)$ is infinite and is chosen as the origin, *the most general stable nonGaussian two-vector is a sum of independent L-stable contributions from every direction of the plane.* Further, the L-stable scalar representing the length of the contribution from the angle $(\theta, \theta + d\theta)$ must have zero mean. As to its scale parameter, u^*, it may depend upon θ, and will be denoted by $[dD(\theta)]^{1/\alpha}$, where the function $D(\theta)$ is bounded and nondecreasing (the reasons for introducing the exponent $1/\alpha$ in this notation will soon be made apparent).

Let $\overline{V}(\theta)$ be the unit vector of direction θ. Then the most general stable nonGaussian vector (X, Y) with a finite mean is of the form:

$$(X, Y) = \int_0^{2\pi} L_\alpha(\theta)\overline{V}(\theta)[dD(\theta)]^{1/\alpha}.$$

This expression is a *stochastic integral.* If $D(\theta)$ is reduced to a finite number of jumps, such an integral is simply a sum of random variables. Otherwise, it is constructed with the help of auxiliary random sums, in the same way as the ordinary integral is constructed with the help of ordinary sums. Of course, the definition of a stochastic integral has raised some problems, but we shall not dwell on them in this paper, because we shall concentrate on $D(\theta)$ reduced to jumps.

The counterpart of the relation $(a)^\alpha = (a')^\alpha + (a'')^\alpha$ of Section 1.6.1 is the fact that, if one adds two L-stable vectors with D-functions $D'(\theta)$ and $D''(\theta)$, the sum is a L-stable vector with D-function $D(\theta) = D'(\theta) + D''(\theta)$. This is the first motivation for the exponent $1/\alpha$.

By *definition,* "positive" L-stable vectors will be expressions of the above form, (X, Y), constructed with D-functions that satisfy $0 \le \theta \le \pi/2$, meaning that θ lies in the first quadrant of the plane *including* the positive parts of the coordinate axes.

1.7. Projection properties of L-stable vectors

1.7.1. Factor analysis. The proofs of many properties of L-stable vectors are simplified if the function $D(\theta)$ is reduced to a *finite* number F of jumps. But most properties proved in this case are valid quite generally and we shall henceforth emphasize this particular case.

Let us begin by defining our notation. The positions of the discontinuities of $D(\theta)$ will be denoted by $\theta_f (1 \leq f \leq F)$, and the unit of directions θ_f will be labeled $\overline{V}(\theta_f) = \overline{V}_f$. Write

$$dD(\theta_f) = J_f, \; L_\alpha(\theta_f) = L_{\alpha f}, \; X_f = L_{\alpha f}(\cos \theta_f)(J_f)^{1/\alpha}, \; Y_f = L_{\alpha f}(\sin \theta_f)(J_f)^{1/\alpha}.$$

Moreover, if $0 \leq \theta_f \leq \pi/2$, write

$$(pf)^{1/\alpha} = (J_f)^{1/\alpha}(\cos \theta_f),(\bar{q}_f)^{1/\alpha} = (J_f)^{1/\alpha}(\sin \theta_f).$$

Finally, if $\theta_f \neq \pi/2$ or $3\pi/2$, write $\tan \theta_f = \tau_f$.

Under these conditions, one has the following relations:

$$(X, Y) = \sum_{f=1}^{F} L_{\alpha f} \overline{V}_f (J_f)^{1/\alpha},$$

$$X = \sum_{f=1}^{F} L_{\alpha f}(\cos \theta_f)(J_f)^{1/\alpha} = \sum_{f=1}^{F} X_f, \text{ and}$$

$$Y = \sum_{f=1}^{F} L_{\alpha f}(\sin \theta_f)(J_f)^{1/\alpha} = \sum_{f=1}^{F} Y_f.$$

This last variant of the construction emphasizes that the relation between X and Y is a direct generalization of the classical statistical technique of "factor analysis." The principal change is that, while the classical approach is usually developed for the case where the $L_{\alpha f}$ are Gaussian variables, we shall assume that the $L_{\alpha f}$ are *nonGaussian* and are L-stable random scalars. As we shall see, a surprisingly large number of changes will be brought about by this replacement (see also M 1962e{E12}. It should therefore be stressed that the construction of X and Y was *not* postulated arbitrarily, but was shown to be essentially the only one compatible with L-stability.

Let us also borrow the terminology of factor analysis. $\overline{V_f}L_{\alpha f}$ will be called the f-th *diachronic factor* of the set (X, Y) (The term "diachronic" emphasizes that X and Y will be successive values of a random function.) $L_{\alpha f}$ will be the *length* of the f-th factor. $(J_f)^{1/\alpha}\cos\theta_f$ and $(J_f)^{1/\alpha}\sin\theta_f$ will be the *diachronic loadings* of the f-th factor.

***1.7.2. Stability of the unconditioned projections* X and Y.** In the finite case, the L-stability of X and Y follows immediately from factor analysis, because sums or differences of L-stable scalars are L-stable. In fact, they are the only nonGaussian stable scalars with a finite expectation. If (X, Y) is a "positive" L-stable vector, that is, if $D(\theta)$ varies only in the first quadrant, both projections are themselves L-stable scalars:

X has a scale parameter $(u^*)^\alpha$ equal to

$$\sum_{f=1}^{F}(\cos\theta_f)^\alpha J_f=\sum_{f=1}^{F}p_f=\int_0^{\pi/2}(\cos\theta)^\alpha dD(\theta).$$

Y has a scale parameter equal to

$$\sum_{f=1}^{F}(\sin\theta_f)^\alpha J_f=\sum_{f=1}^{F}q_f=\int_0^{\pi/2}(\sin\theta)^\alpha dD(\theta).$$

Suppose now that $X=U(t)$ and $Y=U(t+1)$, where $U(t)$ is a stationary time series. Further, for the sake of simplicity, assume that the unconditioned $U(t)$ is a *reduced* L-stable scalar. Then we must have

$$\sum_{f=1}^{F}(\cos\theta_f)^\alpha J_f=\sum_{f=1}^{F}(\sin\theta_f)^\alpha J_f=\sum_{f=1}^{F}p_f=\sum_{f=1}^{F}q_f=1$$

which will make it possible to interpret the p_f or q_f as being probabilities. Note that the above relation makes it unnecessary to postulate separately that $D(\theta)$ is bounded.

All the above arguments still hold when $D(\theta)$ is not reduced to F jumps, that is, if the "number of factors" is infinite. To see this, one must grant that the relation $(a)^\alpha=(a')^\alpha+(a'')^\alpha$ also applies to the addition of infinitesimal parts. The scale parameter of X is the ordinary integral $\int(\cos\theta)^\alpha dD(\theta)$; this shows why the notation $dD(\theta)$ was chosen in the first place. Stationarity then requires

$$\int_0^{\pi/2} (\cos\theta)^\alpha dD(\theta) = \int_0^{\pi/2} (\sin\theta)^\alpha dD(\theta) = 1.$$

1.7.3. Degenerate Markovian nonGaussian stable sequences. Once again, let $X = U(t)$ and $Y = U(t+1)$. Explicit formulas for the passage from $U(t)$ to $U(t+1)$ are available only in *three* degenerate cases.

One factor. Invariant $U(t)$. If there is only *one* diachronic factor, it must be located along the main diagonal of the coordinate axes. Then $U(t)$ will always remain identical to its initial L-stable value $U(t_0)$.

Two degenerate factors. Independent $U(t)$. If there are two diachronic factors, situated along the coordinate axes, the successive values of $U(t)$ are independent L-stable scalars.

It should be noted here that, in the Gaussian case, independence of $U(t)$ and $U(t+1)$ is synonymous with the "isotropy" of the vector $(U(t), U(t+1))$. That is, the angle of this vector with the X-axis – or any other axis – is uniformly distributed over $(0, 2\pi)$. Here, on the contrary, isotropy would require a continuous $D(\theta)$ of the form $k\theta$, where k is a constant and θ can vary from 0 to 2π. This is entirely different from a vector with independent L-stable components.

Two degenerate factors. Autoregressive $U(t)$. Suppose again that there are *two* diachronic factors, having the directions θ' and θ'', where $\theta'' = \pi/2$, while θ' is equal neither to 0 nor $\pi/2$. Let $J(\theta') = J'$, $J\theta'' = J''$. Stationarity and normalization require that

$$(\cos\theta')^\alpha J' = (\sin\theta')^\alpha J' + J'' = 1.$$

Hence, $\tan\theta' = \tau' \leq 1$. Furthermore, it is clear that $X' = X$. Hence,

$$U(t+1) = \tau' U(t) + \left\{ \begin{array}{c} \text{a Lévy stable scalar increment that is independent} \\ \text{of } u(t) \text{ and has a scale parameter } (u^*)^\alpha = J'' \end{array} \right\}.$$

This case is *autoregressive: the effect of chance is additive,* since the multiplication by τ' is independent of chance.

In the Gaussian case, the above degenerate cases exhaust all the possibilities: due to the geometric properties of quadratic forms in Euclidean space, every two-dimensional Gaussian vector may be used to construct an autoregressive $U(t)$.

Things are completely different in the nonGaussian cases. Unless $D'(\theta)$ and $D''(\theta)$ coincide almost everywhere, they lead to *different* vectors

(X, Y), and the effective number of factors is *not* by any means bounded by 2, the dimension of (X, Y). Therefore, the system of factors is uniquely determined for (X, Y). But as one would expect, two functions $D'(\theta)$ and $D''(\theta)$ that "differ little," lead to vectors that also "differ little" (for examples, see Section 2.4).

As a result, particular interest attaches to forms of $D(\theta)$ that have no Gaussian analog. We shall see in Section 2 that the nondegenerate case also has direct economic significance.

However, those nondegenerate $D(\theta)$ are, strictly speaking, incompatible with the Markovian hypothesis, as we shall see in Section 1.9.

1.7.4. Projections of a L-stable vector on other axes. These projections are also L-stable. For example, the two-year average of values of a positive L-stable process, $(1/2)(U(t)+U(t+1))$, is a L-stable variable. The increment $U(t+1)-U(t)$ is also L-stable, but is not a L-stable variable.

Similarly, the set of projections of a vector (X, Y) on any two axes yields a L-stable two-dimensional vector. For example, if the vector (X, Y) is L-stable, so is the vector having X and $Y-X$ as orthogonal coordinates.

1.7.5. Interpretation of X and Y as successive values of some stationary stochastic process, preferably Markovian. So far, we have considered only the case where $X=U(t)$ and $Y=U(t+1)$, when $U(t)$ is income. But it might be desirable to take into account the fact that the $U(t+1)$ may depend upon a past more remote than $u(t)$. Such dependence falls within the present general framework if it is supposed that $U(t)$ is deducible from the knowledge of the past values, $Z(t)$, $Z(t-1)$, $Z(t-2)$, etc., of some function $Z(t)$, generated by one of our L-stable stationary processes.

To cover these, and other, possibilities, we shall from now on denote by $Z(t)$ the stochastic process to which the present study is devoted.

1.8. Linear regression properties of positive L-stable random vectors and processes

1.8.1. The construction of Y, where X is known in the finite case. Suppose that there is a finite number of factors, as in 1.7.1. Then, in order to construct Y, one must perform the following three operations:

(A) Divide x at random into F components X_f. The distribution of each X_i is the conditional distribution of a L-stable scalar of index α and scale p_f, when the value of the sum of X_f is known. (If $\theta=\pi/2$ is the direction of a factor, the corresponding p_f must be zero.)

(B) Form the expressions $y_f = \tau_f x_f$. (The indeterminate product $\infty \bullet 0$, which corresponds to $\pi/2$, is always taken to be zero.)

(C) Add the y_f. (If there is a diachronic factor along $\pi/2$, one must also add a variable independent of x, contributed by the random variable $L_\alpha(dD(\pi/2))^{1/\alpha}$.

Consider now a random sequence $Z(t)$. The vector (X, Y) is interpreted successively as $(Z(t-1), Z(t))$ and as $(Z(t), Z(t+1))$. The first interpretation introduces a set of numbers $z_f(t)$, which are the "y_f" involved in the transformation from $z(t-1)$ to $Z(t)$. The second interpretation introduces a second set of numbers $z_f(t)$, which are the "x_f" involved in the transformation from $z(t)$ to $Z(t+1)$.

A particularly illustrative case of the double interpretation of $z_f(t)$ is encountered in moving average processes of the form

$$Z(t) = \sum_{s=0}^{\infty} L(t-s)[K(s)]^{1/\alpha},$$

where the $L(t)$ are independent, reduced, L-stable random variables, and where $K(s)$ satisfies $\sum_{s=0}^{\infty} K(s) < \infty$. Obviously, one can write:

$$Z(t-1) = [K(0)]^{1/\alpha} L(t-1) + [K(1)]^{1/\alpha} L(t-2) + \ldots$$
$$Z(t) = [K(0)]^{1/\alpha} L(t) + [K(1)]^{1/\alpha} L(t-1) + [K(2)]^{1/\alpha} L(t-2) + \ldots.$$

One sees immediately that the joint distribution of $Z(t)$ and $Z(t-1)$ is a L-stable two-dimensional distribution. *Note that the successive decompositions of $Z(t)$ are not independent, but deduced from each other by a kind of translation.*

The function $D(\theta)$, varies by jumps, which may be infinite in number. In every case, it has a discontinuity at $\theta_0 = \pi/2$, of amplitude $dD(\pi/2) = K(0)$. The number and position of the other jumps depend upon the function $K(s)$. If $K(s) > 0$ for all s, there are jumps for every θ_f of the form

$$\theta_f = \tan^{-1}\left\{\left\{\frac{K(f)}{K(f-1)}\right\}^{1/\alpha}\right\},$$

with

$$J(\theta_f) = K(f)(\sin\theta_f)^{-\alpha} = K(f)\{1 + (\cot\theta_f)^2\}^{\alpha/2} = \{[K(f)]^{2/\alpha} + [K(f-1)]^{2/\alpha}\}^{\alpha/2}.$$

The exponential kernel $K(t) = \exp(-\gamma t)$ corresponds to the case where all θ_f are identical (except for θ_0), and leads to the autoregressive Markovian process.

One might try to generalize factor analysis by assuming that the X_f are not independent, even before the condition $\Sigma X_f = x$ is imposed upon them. If one assumes that the X_f have a nontrivial L-stable distribution, however, the generalization can rather easily be shown to be void.

1.8.2. Linearity of the regression of Y on X. Let us recall the notation of Section 1.7.1 and write Σ' for the summation that excludes $\theta = \pi/2$, even if this is the direction of a diachronic factor. We then obtain (independent of the sign of x, i.e., of $x - E(X)$),

$$E(Y\,|\,x) = \sum{}'(\tan\theta_i)E(X_i\,|\,x) = x\sum{}'(\tan\theta_i)p_i = x\sum{}'\tau_i p_i = xE(T).$$

(The notation $E(T)$ for $\Sigma'\tau_i p_i$ will be justified shortly.) Moreover, $E(t) = \Sigma'\tau_i p_i$ can be written as $\Sigma'\tau_i p_i/\Sigma q_i = \Sigma'\tau_i p_i/[\Sigma'\tau_i^\alpha p_i + q(\pi/2)]$. Writing τ as $H(\tau^\alpha)$, we see that the function H is increasing and its convexity is turned up, so that

$$E(T) \le \frac{\sum' H(\tau_i^\alpha)p_i}{\sum' \tau_i p_i} \le H\left\{\frac{\sum' \tau_i^\alpha p_i}{\sum' \tau_i^\alpha p_i}\right\}$$

by convexity. The last term is equal to $H(1) = 1$. In short, we have shown that *$E(Y\,|\,x)$ is the product of x by a positive scalar at most equal to 1.*

The same result also applies if $D(\theta)$ is not finite. But, aside from special cases, it fails (except in special cases) if $D(\theta)$ could vary in the second to fourth quadrants of the plane.

1.9. On approximately Markovian L-stable scaling sequences

At this stage, we must make more specific assumptions concerning our process $Z(t)$. It would have been convenient if $Z(t)$ could be Markovian as well as L-stable. Unfortunately, this is impossible, except in the cases where the vector $[Z(t), Z(t+1)]$ takes one of the forms of Section 1.7.3. That is, if $[Z(t), Z(t+1)]$ is a nondegenerate L-stable vector and $Z(t)$ is Markovian, vectors such as $[Z(t), Z(t+1), Z(t+2)]$ are not L-stable.

More generally, if $Z(t)$ is Markovian with a memory equal to M, it cannot in general be the aggregate of independent processes: each of the aggregated processes may influence the values which the other processes will take $M+1$ steps later.

Hence, two desirable, but somewhat far-fetched approximations to the facts are not, strictly speaking, mutually compatible. From our present viewpoint, however, this is not a problem. One can show that there exist L-stable scaling processes such that, if the value of $Z(t-1)$ is large, $Z(t)$ depends only upon chance and upon $z(t-1)$, and not explicitly upon previous values of Z. For these processes, the first Champernowne conditions is satisfied only in the range of values of $Z(t-1)$ that interest us here. In this range, it also happens that the behavior of $Z(t)$ is the same as that of a strictly Markovian sequence for which the joint distribution of successive $Z(t)$ is given by some two-dimensional L-stable vector (X, Y). The latter process is simpler to describe, and we shall now proceed to study its asymptotic properties.

The reader should be warned that, *in Section 2, the term "stable Markovian sequence" will be used in the somewhat loose sense* which we have just motivated.

2. APPROXIMATIONS TO CERTAIN STABLE MARKOVIAN RANDOM SEQUENCES

2.1. The effect of the form of $D(\theta)$ upon the behavior of $Z(t)$

Most of the properties of the scalars and vectors studied in Section 1 are also true of the Gaussian prototypes of all stable distributions (except, in some expressions, for the replacement of 2 by α). Those properties did not involve any approximation, and they depended little upon the shape of the function $D(\theta)$ (except when we assumed that $D(\theta)$ varied in the first quadrant only). But, as we have noted, important differences appear as soon as one goes beyond this section, because the law of $Z(t)$, given $z(t)$, cannot always be reduced to the universal form that corresponds to an autoregressive behavior. $Z(t)$ may behave in many ways, depending on the shape of $D(\theta)$. Things simplify drastically, however, if $D(\theta)$ is "near" in form to either the function corresponding to time-invariant $Z(t)$ or to that of the autoregressive $Z(t)$.

In the first case, we shall show that, as long as $z(t)$ is large, a stationary Markovian $Z(t)$ will behave in such a fashion that $\log Z(t)$ performs a random walk or a diffusion. The last two of the asymptotic Champernowne conditions are also satisfied, as a side-product of the linear regression proved in Section 1.8.2. Thus, we shall derive *Champernowne's conditions* and thereby achieve one of the principal aims of this paper.

If, on the contrary, $D(\theta)$ is "near" the function that corresponds to the autoregressive $Z(t)$, the behavior of $Z(t)$ will be "near" the autoregressive behavior.

In sum, the effect of chance upon the evolution of $Z(t)$ may be either *multiplicative* or *additive*. It can also be *a mixture of the two*. Everything depends upon how much $D(\theta)$ varies near $\theta = \pi/2$. For the purposes of this article, it is unnecessary to describe formally the respective domains of validity of the two extreme approximations.

2.2. Approximations to the behavior of a L-stable variable U', when the value of the sum of U' and of U'' is known and is large

The construction of Section 1.8.1 requires the probability distribution of X_f to be conditioned by the known value x. This reduces to the problem, already raised at the end of Section 1.6.1, of describing the distribution of the variable $U' = \alpha' L'_\alpha$, when one knows the value u of the sum $aL_\alpha = a'L'_\alpha + a''L''_\alpha$. The density of the distribution of U' obviously takes the form

$$p(u' \mid u) = \frac{(1/a')p_\alpha(u'/a')(1/a'')p_\alpha[(u-u') \mid a'']}{(1/a)p_\alpha(u/a)}.$$

Unfortunately, this expression is, in general, quite complicated.

Suppose, however, that u is positive and large. Then we find that $p(u' \mid u)$ consists of two "bells," with some small additional probability between the bells. This probability distribution plays an important role in M 1962e{E12}.

The first "bell" is located near $u'=0$ and its shape is deduced from a reduced L-stable density by the following transformations: (a) Very large values of u' (values greater than $u/2$) are truncated. (b) The scale of abscissas, which is dominated by the term $(1/\alpha')p_\alpha(u'/a')$ of $p(u'\,|\,u)$, is multiplied by a'. (c) The scale of ordinates is dominated by the term

$$\frac{(1/a'')p_\alpha[(u-u')/a'']}{(1/\alpha)p_\alpha(u/a)}$$

of $p(u'\,|\,u)$; this term remains close to

$$\frac{(1/a'')p_\alpha(u/a'')}{(1/a)p_\alpha(u/a)},$$

which in turn can be approximated by $(a'')^\alpha/(a)^\alpha$ (since, when u is large, $p_\alpha(u)$ is asymptotically scaling). Hence, defining the event A' by the fact that $a'L'_\alpha$ falls within the bell located near $u'=0$, we see that the probability of A' is itself close to $p''=(a'')^\alpha/(a)^\alpha$. Furthermore, if U' is conditioned by both u and A', its mean value is close to zero (and tends to zero as $u \rightarrow \infty$) and its most probable value is close to $-m(\alpha)a'$, where $m(\alpha)$ is the distance between the mean and the mode of a reduced L-stable variable L_α.

The properties of the second bell, located near $u'=u$, and of the corresponding event A'', are obtained by permuting a' with a'' and u' with $u''=u-u'$.

As a result, *a conditioned L-stable variable can be reduced to a combination of two truncated L-stable variables, corresponding to the realization of either one of two mutually exclusive events, A' and A'':*

When A' has the probability p' then u'' is equal to a truncated unconditioned L-stable variable $a''L''_\alpha$ and $u'=u-u''$.

When A'' has the probability p'' then u' is equal to a truncated unconditioned L-stable variable $a'L'_\alpha$ and $u''=u-u'$.

Clearly, when u decreases, the two bells eventually merge. But even before this happens, the truncation of the smaller of the two contributions to u first affects the mean of the bell, then its median, and finally its mode.

It is interesting to study the lower bound of the range of u where the above approximation is applicable. If α is close to 2, u must be very large indeed for the approximation to hold. But if α is close to 1, the mode remains unaffected even when u becomes negative and gets fairly close (on the positive side) to the mode of the unconditioned $\alpha' L'_\alpha$. Consider the interesting case when α is close to 3/2. If p' and p'' are both close to 1/2, it seems that u may be as small as 4 or 5 times $m(\alpha)$, without displacing the modes very much from the untruncated case. If p' is much smaller than p'', u may become even smaller.

Large negative values of u. In this case, the conditioned $p(u' \mid u)$ is very close to a Gaussian distribution of variance equal to

$$c(\alpha)p'p'' \mid u \mid^{-(2-\alpha)/(\alpha-1)},$$

where $c(\alpha)$ is independent of u. Except in the limit case of $\alpha = 2$, this variance depends upon u and tends to zero with $1/u$.

2.3. Markovian stationary stable sequences associated with random walks of the logarithm of $Z(t)$

2.3.1. *A nondegenerate two-factor sequence.* Let the joint distribution of $Z(t)$ and of $Z(t+1)$ involve two factors, having the directions θ' and θ'', such that $0 < \theta' < \theta'' < \pi/2$. Let $dD(\theta') = J'$, $dD(\theta'') = J''$, $J'(\cos\theta')^\alpha = p'$, and $J''(\cos\theta'')^\alpha = p''$. Let L'_α and L''_α be two independent reduced L-stable scalars. Let $X' = L'_\alpha(\cos\theta')(J')^{1/\alpha}$, and similarly for X'', Y', Y''. Finally, let $\tau' = \tan\theta'$, and $\tau'' = \tan\theta''$. Then we can write:

$$\begin{aligned} X &= X' + X'' = L'_\alpha(J')^{1/\alpha}\cos\theta' + L''_\alpha(J'')^{1/\alpha}\cos\theta', \\ Y &= Y' + Y'' = L'_\alpha(J')^{1/\alpha}\sin\theta' + L''_\alpha(J'')^{1/\alpha}\sin\theta' \\ &= X'\tau' + X''\tau'' = X\tau'' - X'(\tau'' - \tau') = X\tau' + X''(\tau'' - \tau'). \end{aligned}$$

Assuming that x (or y) is large, the results of Section 2.2 can be used to greatly simplify the decomposition of X into X' and X'' (or of Y into Y' and Y'').

However, even if x is large and overwhelmingly due to x', it does not necessarily follow that y is also large and is due to the part y' that corresponds to x'. For this to be the case, we *suppose that both τ' and τ'' are close to 1, which requires the directions of the two diachronic factors to be close to the main diagonal.* In that case, if x is large, it is overwhelmingly likely that the contribution of one of the factors is much larger than that of the other, so that one of the following mutually exclusive events will be realized:

Event A', of probability p'. Here x is predominantly due to X', and X'' is approximately equal to a truncated L-stable variable $(p'')^{1/\alpha}L_{t\alpha}$. Furthermore,

$$Y = x\tau' + (\tau'' - \tau')(p'')^{1/\alpha}L_{t\alpha}.$$

In particular, as $x \to \infty$, one has:

$$E(Y \mid x, A') \to x\tau'; \quad \text{Mode } (Y \mid x, A') \to x\tau' + m(\alpha)(p'')^{1/\alpha}(\tau'' - \tau'),$$

the second convergence being far faster than the first.

Event A'', of probability p''. Here, x is predominantly due to X'', and X' is approximately equal to a truncated L-stable variable $(p')^{1/\alpha}L_{t\alpha}$. Furthermore,

$$Y = x\tau'' - (\tau'' - \tau')(p')^{1/\alpha}L_{t\alpha}.$$

In particular, as $x \to \infty$, one has:

$$E(Y \mid x, A'') \to x\tau''; \quad \text{Mode } (Y \mid x, A'') \to x\tau'' + m(\alpha)(p')^{1/\alpha}(\tau'' - \tau'),$$

the second convergence being far faster than the first.

Consider also the variable $\log Y$, as conditioned by $\log x$ and by either A' or A''. For large y, the curvature of the function $\log y$ is small (and tends to 0 with $1/y$). Hence,

$$E(\log Y \mid \log x) \to \log x + \log T,$$

where the scalar T equals either τ' (with probability p') or τ'' (with probability p'').

2.3.2. ***Concerning Bernoulli's concept of "moral wealth."*** Consider a process whose successive values are ruled by the joint law of the vector (X, Y) of the last section. We have just proved the following result:

Let $Z(t)$ be a stationary Markovian stable sequence with two diachronic factors close to the main diagonal. Then, as long as $z(t)$ remains large, $Z(t+1)$ will, "on the average," be equal to either $z(t)\tau'$ or to $z(t)\tau''$, with respective probabilities p' and p''. Log $Z(t+1)$ *will, "on the average," be equal to either* $\log z(t) + \log \tau'$ *or to* $\log z(t) + \log \tau''$. *Hence,* $\log Z(t)$ *will, "on the average," perform a random walk.*

Combined with the linear regression of $Z(t+1)$, given $z(t)$, the above result yields the asymptotic conditions of Champernowne and leads to a proof of our fundamental result, as stated in the simplest possible case.

But our statement concerning random walks is not completely satisfactory. *Mathematically,* while the effects of truncation by $z(t)$ decrease with $1/z(t)$, the "average" is the slowest, among three typical values of a truncated variable, to converge to its "true" value, which is zero. *Empirically,* the concept of a "moral wealth" is not very well represented by $\log[Z(t)E(Z)]$. It would be better to express the moral wealth as $\log[Z(t)+B]$, where B is a positive constant (for Bernoulli), and lies between $E(Z)$ and the "nominal" origin of z.

It happens that expressions of precisely this form are encountered in the study of the "most likely" behavior of $Z(t)$. Indeed, writing $B' = m(\alpha)[(p')^{1/\alpha} + (p'')^{1/\alpha}]$ and $B'' = m(\alpha)[(q')^{1/\alpha} + (q'')^{1/\alpha}]$, we can make the following statement:

The sequence $Z(t)$ *will "most likely" be such that* $z(t) + B'$ *is followed by* $Z(t+1) + B'' = T[z(t) + B']$, *where the random variable* T *takes the values* τ' *and* τ'' *with the respective probabilities* p' *and* p''.

Suppose, moreover, that $B' = B''$ (as in Figure 1, in which the vectors of (X, Y) are symmetric with respect to the main diagonal). Then one can say that $\log[Z(t)+B]$ "most likely" performs a random walk. Unfortunately, it is clear that, in general, we have $B' \neq B''$. But the expression $x^{1/\alpha} + (1-x)^{1/\alpha}$, which is involved in both B' and B'', stays close to its maximum, $2^{1-1/\alpha}$, except perhaps when x is very close to 0 or 1. Hence, B' and B'' are likely to be almost equal if the loadings p' and p'' (respectively, q' and q'') are not too different from each other. In that case, the diachronic factors are not only close to the main diagonal but close to being symmetrical with respect to that line. Therefore:

Let $Z(t)$ *be one of our stationary Markovian stable sequences with two diachronic factors close to the main diagonal and close to being symmetric with respect to that line. Then, as long as* $z(t)$ *remains large, the behavior of* $\log[Z(t)+B]$ *is approximated by a random walk.*

Of course, we cannot *prove* that the above expression $Z(t)+B$ can "really" be considered as representing income as counted from some "subsistence income," $-B$, and that $[Z(t)+B]$ can be considered as a "moral wealth." It would be very pleasing, indeed, if this were so. But perhaps our more general equation, of the form $Z(t+1) = Tz(t) + TB' - B''$, is an improvement over "simple" multiplicative behavior.

In the above argument, it is noteworthy that asymptotic scaling still holds for values of $z(t)$ for which the "bells" of the conditioned distribution of X', knowing x, are not well separated. These values of z can become far smaller than those for which the random walk approximation

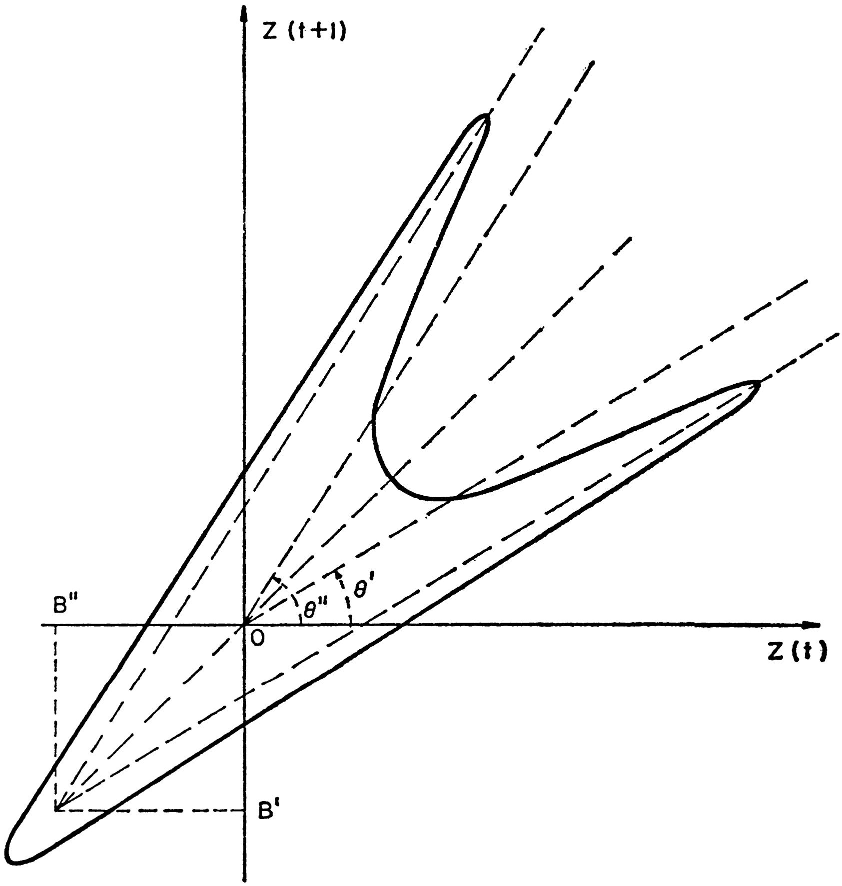


FIGURE E11-1. An example of the form taken by the lines of constant probability density of the vector $[Z(t), Z(t+1)]$, in the case where two factors are symmetric with respect to the main diagonal.

is valid. There is nothing inexplicable in this situation, however, because asymptotic scaling is far better established than the random walk behavior of log $Z(t)$.

2.3.3. ***An F-factor sequence.*** Suppose that there are F diachronic factors for (X, Y). The notation of Section 1.7.1, with $0 < \theta_f < \pi/2$ yields

$$X = \sum_{f=1}^{F} L_{\alpha f}(J_f)^{1/\alpha}(\cos \theta_f); \qquad Y = \sum_{f=1}^{F} L_{\alpha f}(J_f)^{1/\alpha}(\sin \theta_f).$$

Once again, we wish to be certain that when x and y are large, it is because both are predominantly due to the projections of a single loaded factor $(J_\alpha)^{1/\alpha} L_{\alpha f} \overline{V}_f$. This will be the case if x is very large and if there is no diachronic factor close to either axis. If this is the case, it will be overwhelmingly likely that the large size of x will be due mostly to the realization of one of the following F mutually exclusive events. (The indexed f satisfies $1 \leq f \leq F$.)

Event A_f (of probability p_f). The contribution x_f equals $x - \Sigma_{i \neq f} x'_i$, where the X'_i are (approximately) independent of x and equal to truncated L-stable variables, $L_{t\alpha i}(p_i)^{1/\alpha}$. Hence, ordering the τ_f to form an increasing sequence, we have

$$\begin{aligned}(Y \mid x, A_f) &= x\tau_f + \sum_{i \neq f} L_{t\alpha i}(p_i)^{1/\alpha}(\tau_i - \tau_f) \\ &= x\tau_k + \sum_{i > f} L_{t\alpha i}(p_i)^{1/\alpha}(\tau_i - \tau_f) - \sum_{i < f} L_{t\alpha i}(p_i)^{1/\alpha}(\tau_f - \tau_i).\end{aligned}$$

The truncations are neglected, then (the approximation becomes increasingly good as x increases), each of the sums $\Sigma_{i>f}$, $\Sigma_{i<f}$ becomes a L-stable variable, so that

$$(Y \mid x, A_f) = x\tau_f + L'_{\alpha f}\left\{\sum_{i>f} p_i(\tau_i - \tau_f)^\alpha\right\}^{1/\alpha} - L''_{\alpha f}\left\{\sum_{i<f} p_i(\tau_f - \tau_i)^\alpha\right\}^{1/\alpha}.$$

In particular, we have

$$E(Y \mid x, A_f) = x\tau_f; \qquad \text{Mode } (Y \mid x, A_f) = x\tau_f + m_f(\alpha).$$

The expressions $m_f(\alpha)$ refer to *differences* of L-stable variables, so that we do not know of any explicit formula for them. We may, however, state the following:

Let $Z(t)$ *be a stationary Markovian stable sequence with* F *diachronic factors, all close to the main diagonal. Then as long as* $z(t)$ *remains large,* $Z(t+1)$ *will, "on the average," be equal to one of the quantities* $z(t)\tau_k$, *with respective probabilities* p_k. *Thus,* $\log Z(t+1)$ *will, "on the average," be equal to one of the quantities* $\log z(t) + \log \tau_k$. *Hence,* $\log Z(t)$ *will, "on the average," perform a random walk.*

Now consider the "most likely" behavior. It is unfortunately impossible to tell whether the straight lines, $y = x\tau_f(\alpha) + m_f(\alpha)$, have a common point and, *a fortiori*, whether they intersect near the main diagonal. But this may be achieved by imposing sufficiently stringent conditions upon the vector (X, Y). Otherwise, the quantity B that enters in Bernoulli's "moral wealth," $\log[Z(t)+B]$, would not be defined precisely, only as a rough order of magnitude.

As x becomes smaller, the random walk steps are increasingly "smeared;" the breakdown of the random walk happens increasingly faster as α gets close to 2 and as F increases.

Addition of two L-stable processes. Let the functions $D'(\theta)$ and $D''(\theta)$ be different, but let them both lead to random walks of $\log Z(t)$. In this case, the function $D(\theta) = D'(\theta) + D''(\theta)$ also reduces to jumps near $\theta = \pi/4$, hence also leads to a random walk. This closure property is not valid for autoregressive processes: if the θ' of the addends are different, the sum is not autoregressive.

2.3.4. *Sums of Champernowne sequences: a heuristic approach.* (See Section 2.5 of M 1960i{E10}.) Let $W_i(t)$ be independent and identically distributed Champernowne sequences, with α restricted only to being greater than 1.

Consider then the unconditioned sum $Z_N(t) = A(N)\sum_{i=1}^{N} W_i(t) - B(N)$. If $z_2(t)$ is large, this fact is overwhelmingly likely to be due to only one of the asymptotic Pareto addends, $W_1(t)$ and $W_2(t)$. Hence, $Z_2(t)$ will behave in time like either of the two terms of which it is made. This means that Champernowne's properties are preserved in the addition of two addends. The same continues to hold for the addition of three or a few more parts, irrespective of the sign of $\alpha - 2$.

But now let N become very large. Then, if $\alpha > 2$, the sum $Z_N(t)$ remains valid in a region of high values of $z_N(t)$, the total probability of which tends towards zero. If, on the contrary, $\alpha < 2$, the sum tends towards a L-stable process and Champernowne's properties remain valid within a region of values of z_N, the probability of which tends towards a finite limit.

Consider then the directions such that $\log(\tan \theta_f)$ is equal to one of the steps which $\log W_i(t)$ may perform when it is large. These will also be the directions of the factors of the limit of the vector $(Z_N(t), Z_N(t+1))$. Moreover, it is easily ascertained that *the factors of the vector* ($Z_N(t)$, $Z_N(t+1)$, $Z_N(t+2)$) *have such a structure that* $Z_N(t)$ *is a "Markovian" stable sequence.*

2.4. Stable sequences which are not associated with random walks of the logarithm of $Z(t)$

2.4.1. The effect of a diachronic factor close to the Y-axis. Let there be two diachronic factors. If $\theta'' = \pi/2$, one has the autoregressive case of Section 1.7.3.

Suppose now that $\theta'' = \pi/2 - \varepsilon$, while θ' is close to $\pi/4$. Then, the division of x into x' and x'' can still be approximated by the device introduced in Section 2.2. But if x'' is small and x' large, it may very well happen that $y'' = x''(\tan \theta'')$ is of the same order of magnitude as $y' = x'(\tan \theta')$. Therefore, the random walk breaks down, even if one starts from an instant of time for which $z(t)$ happens to be large. The variation of $Z(t)$ must instead be represented by the following mixed approximation:

With a small probability p'', x will be predominantly due to x''. The reason why p'' is small, is because J'' is less than 1 and $(\cos \theta'')^{\alpha} \sim \varepsilon^{\alpha}$. Therefore, the event A'' happens very rarely. But when it does happen, x is multiplied by the enormous coefficient $\tan \theta'' \sim 1/\varepsilon$.

With a probability p', close to 1, x will be mostly contributed by x'. In that case, x'' will be approximately a L-stable variable with scale factor $J''\varepsilon^{\alpha}$, truncated at a value of, say, $x/2$. Then, y'' will be a L-stable variable with scale factor $J''\varepsilon^{\alpha}$, truncated at $x/2$. Hence, y'' will be a L-stable variable with scale factor $J''\varepsilon^{\alpha}(\tan \theta'')^{\alpha} \sim J''$, truncated at the meaninglessly large number x/ε. As a result, the behavior of $Z(t)$ will be quite indistinguishable from an autoregressive sequence obtained by making $\varepsilon = 0$, without changing J''.

As $\varepsilon \to 0$, the probability p'' of the rare event also tends to 0, as does the expected value of the multiplier, $\tan \theta''$ (this expected value is $\tan \theta'' \sim J''\varepsilon^{\alpha-1}/p''$). It is, of course, satisfactory to find that the behavior of $Z(t)$ varies continuously with $D(\theta)$.

2.4.2. The effect of a diachronic factor close to the X-axis. The limit case, where $\theta' = 0$ and θ'' is close to $\pi/4$, is the inverse of an autoregressive sequence. Here, the random walk predicts that $z(t)$ will be followed by 0 with probability p'. But, of course, the effect of the other component, x'', will prevent this from happening; instead, $Z(t+1)$ will be a *reduced* L-stable variable, independent of $z(t)$. That is, if $z(t)$ starts from some large value, it will increase further, exponentially, through successive multiplications by $\tan \theta'' > 1$, until the first occurrence of the event A'.

When A' happens, the process starts again from scratch, and it is very unlikely to start with a large value of $z(t)$.

Suppose now that θ' is very small. Then, if A' occurs, $Z(t+1)$ will be equal to a new L-stable variable, independent of the past and having a scale factor slightly less than 1, plus a contribution from the past in the form of $z(t)\varepsilon$. Things again change continuously as ε tends to zero.

We think that the above indications suffice to show what kind of behavior can be expected of L-stable processes not associated with random walks.

3. CONCLUSION

This paper emphasizes the case in which the behavior of $Z(t)$ performs a random walk or a diffusion. In such a process, $Z(t)$ cannot be multiplied by any very large or very small factor. But we also showed that a more general expression for $Z(t)$ is such that $Z(t+1) = TZ(t) + R$, with T and R independent of each other and of the past. The random term R allows for occasional large discontinuities of $Z(t)$ and seems to provide a useful improvement over the principle of random proportionate effect.

One must not be too dogmatic about the exact degree of relevance of our theory to the study of income. Since it starts with extreme idealizations, the theory may be interpreted in many ways. The strongest interpretation would assert that one or more of the motivating properties of our processes has explanatory value (there may be disagreement about which is the best explanation) and that the theory as a whole has predictive value. The weakest interpretation would stress only two facts: (a) success in dispelling the apparent contradiction between the aggregation of parts and the nonGaussian character of the whole; and (b) starting from hypotheses that are reasonably close to facts, derivation of results that are close (or very close) to the facts.

In any event, it is clear that I strongly believe that the theory of stable nonGaussian random elements is essential to the urgent task of extending the stock-in-trade of stochastic models.

Dedication and acknowledgement

Dédié à mon Maître, Monsieur Paul Lévy, Professeur honoraire à l'Ecole Polytechnique.

This work was started in 1957 while the author was associated with Columbia and Cornell Universities, under contract Nonr 266(04), task NR(047,005). The work was continued at the University of Lille, France, and at the Ecole Polytechnique in Paris. Professor Lawrence R. Klein, Mr. Richard E. Levitan and a referee have made several useful suggestions concerning the organization and style.

The original text notes that the model reported in this paper was first stated in a memorandum I wrote in 1957, while at the University of Geneva.

&&&&&&&&&&&& ANNOTATIONS &&&&&&&&&&&&

The scientific and mathematical value of Champernowne 1953. This paper gives extensive play to a reference I soon came to view as overblown. Its basic result was familiar: the exponential distribution of log U can be generated by a random walk with a downward drift and a reflecting boundary. Building on this base, one can design an infinity of mechanisms that seem more general, because they demand a cumbersome notation, but in fact bring nothing significant.

Champernowne 1953 is an example of such idle make-work and mathematical overkill with no practical or mathematical interest. Champernowne taught economics at Oxford, then Cambridge, but found a permanent niche in the literature for the number C whose decimal development starts with 0 and continues with all the integers in sequence without leaving any gap. Thus, C begins as follows:

$$C = 0.123456789101112131415161718192021 \ldots$$

This number is non-random, but passes all tests for randomness. Norbert Wiener gave an earlier example, but the memorable form is C.

The Quarterly Journal of Economics: 76, 1962, 57-85

E12

Scaling distributions and income maximization

• *Chapter Foreword.* Judged by the illustrations, this chapter is an exercise in something close to linear programming. The formulas show that this exercise is carried out in a random context ruled by scaling distributions, including the Pareto law for the distribution of personal income. To help both concepts become "broken in" and better understood, I investigated this and other standard topics afresh with the Gaussian replaced by the scaling distribution. For many issues, major "qualitative" changes follow. As is often the case, the root reason lies in sharp contrast in the convexity of probability isolines, between the circles of the form "$x^2 + y^2 =$ constant," which characterize independent Gaussian coordinates, and the hyperbolas of the form "$xy =$ constant," which characterize independent scaling coordinates. The resulting changes concern an issue central to this book and especially in Chapter E5: evenness of distribution associated with the Gaussian, versus concentration associated with the scaling.

This paper examines many special cases in turn, which tends to be tedious, but the style is not devoid of dry whimsy and revealing near-autobiographical considerations. As the reader must know, a key feature of my scientific life is that rarely, if ever, does it involve an established field, one that requires well-defined skills provided by a recognized and trained combination of nature and nurture. To the contrary, most of my work relies on an idiosyncratic combination of seeing and reckoning. Not a few brave souls attempt to work in this way, but very few succeed. The reader will find that this chapter tackles that statistical observation: it does not really attempt to "explain" it, only shows that it does not fit in a Gaussian universe, but fits comfortably in a universe in which randomness is scaling. •

I. THE EMPIRICAL DISTRIBUTION OF PERSONAL INCOME; SCOPE OF THE ASYMPTOTIC LAW OF PARETO

The distribution of income follows the asymptotic law of Pareto. That is, there exist two constants α and $\tilde{u}$, such that, for sufficiently large values of u, the relative number of income earners with income $> u$, takes the scaling form $P(u) = 1 - F(u) = \Pr\{U > u\} \sim (u/\tilde{u})^{-\alpha}$. The exponent α is the asymptotic slope of the curve giving $\log P(u)$ as function of $\log u$. Pareto 1896 reports that α lies between 1 and 2; perhaps its value may exceed 2 in some recent cases (other than the salaries of which we speak below.) In any event, little significance attaches to the value $\alpha = 1.5$, which Pareto emphasized.

In the nonasymptotic part, the shape of $F(u)$ is irregular and sometimes even multimodal, but can be represented by a mixture of simpler distributions. As a matter of fact, the distributions relative to sufficiently narrowly defined "occupations" are unimodal curves, for which skewness and average income tend to increase simultaneously. It is reported that this result was known to Moore 1911, but I did not find this reference. Similar results can be found in Westergaard & Nyboelle 1928. The most accessible reference is Miller 1955. Within narrow occupations the distributions of income typically behave as follows.

At the lower end of the scale, certain types of pay, such as wages for unskilled laborers, tend to have symmetric distributions, which (for all that we know) may even be Gaussian. Here, the inequality of incomes can be attributed primarily to the variability of the number of days worked for a fixed daily wage.

Farther up the scale, several different categories of income have a markedly skewed distribution, but a very small probability of attaining high values. Such income categories have been successfully fitted by the lognormal distribution, which is rationalized by considering the logarithm of income as the sum of very many components.

Finally, there are several income categories that follow the asymptotic law of Pareto, *but with different values of* α. For example, the alpha of salaries may by found between 4 and 5. But the set of those incomes that remain after all simple professional categories have been excluded has an α that is traditionally contained between 1 and 2. This exponent dominates the distribution of very high incomes, considered irrespectively of occupation.

II. SCOPE OF THE PRESENT PAPER

One could explain the data of Section 1 in two stages. At first, one would treat different income categories as being conditioned by independent and possibly different economic mechanisms. Later, the obviously strong interaction between the various distributions would be taken into account by a separate model meant to explain the observed distribution of the GNP between various categories.

The aim of the present paper is bolder; to suggest one possible reason why several values of the α exponent, and several non-scaling categories, may well coexist in a single, strongly interacting, economic community.

It will be assumed that each individual must choose *one* of N possible "occupations," P_n, with $1 \leq n \leq N$ (Roy 1951). Two occupations that make the same offers to every income-earner may be considered identical. Further, we assume that each individual knows the incomes, U_n, which he could derive from various occupations P_n. Finally, we suppose that each individual chooses the occupation that offers him the most. If several P_n make the same maximal offer, he chooses one at random. Some income-earners hold several part-time jobs, and a "professional" income is often supplemented from investments. We could have stated some simple facts concerning this possibility, but we shall abstain.

We assume that, within the total population, the distributions of the various U_n are random. We also assume a certain interdependence between the various offers. The conclusion is that *if the overall income distribution is scaling with exponent α, the offers accepted from each "occupation" taken separately will also be scaling, but with an exponent of the form $w(n)\alpha$.* The quantity $w(n)$ is an *integer* called the *"weight"* of P_n; it is found to be equal the *number of factors that must be large simultaneously, in order that the offer U_n be accepted and turn out to be large.* This concept of weight is central to this paper. If the weight is large, the law of Pareto becomes practically useless and should be replaced by the lognormal distribution.

The word "offer" used to denote the U_n is a terminological device, not meant to imply that all occupations are organized in bureaucratic hierarchies and that all incomes are salaries. The same approach holds if an income-earner wanders around until nobody offers more to move again. Some persons' best choice may be too obvious to be performed consciously. In any event, there is no doubt that other mechanisms can generate the same mixtures of scaling distributions.

The hypothesis of income maximization presupposes more order in the economic organization than is strictly necessary to derive the desired results. For example, different people could maximize their income among different numbers, N, of possibilities, and the observed distribution of income would be a mixture of those which correspond to different values of N. For example, N may be random.

III. LINEAR FACTOR ANALYSIS OF THE RENTAL PRICE OF AN INDISSOLUBLE BUNDLE OF ABILITIES

Let us now analyze the interdependence between the offers U_n and the economic significance of the existence of several *different* offers for the use of the services of the *same* man.

It is natural to assume that all offers are conditioned by the same broad set of "attributes," but that different professions "weigh" the attributes quite differently. "Positive" attributes include "intelligence," "character," "mathematical ability," "inherited fortune," and there are "negative attributes." However, inherited fortune is the only item for which there is an objective measure. The others raise so many questions (sociological and economic, and also, perhaps, psychological) that it seems better to leave them unspecified, even though it is inevitable and useful for the reader to identify these attributes mentally with some widely used psychological concepts.

There is a widely held belief that the distributions of psychological factors are Gaussian. However, the distributions of uncorrected psychological scores are usually very skew and have long tails in one direction. (See Thurstone 1947.) There is also much evidence, due to Lotka, H.T. Davis and others, that the distribution of certain psychological characteristics may be scaling; but we do not want to enter into this discussion.

This paper's results could be derived under several different types of dependence between the offers and the abilities. Linear dependence is time-honored in economics. U_n $(1 \leq n \leq N)$ being the rental price which the occupation P_n is ready to pay, suppose U_n can be written as a nonhomogeneous *linear* form of F independent factors $V_f (1 \leq f \leq F)$. Each factor is randomly distributed in the population, and "measures" one or several "abilities." The letter F should not be confused with the function $F(u)$. Then one can write:

$$U_n = \sum_{f=1}^{F} a_{nf}V_f + a_{n0} + a_nE_n.$$

The E_n are "error terms," which we mention for the sake of completeness, but shall neglect. A different very reasonable correction was introduced by Machlup 1960, who notes that the occupation of inventor is, for some, the best paying current option, while others would never consider doing anything else, hence would invent even for a very small reward. Similar considerations obviously apply more generally. The factors V_f may be common to several offers; their units will be chosen later. The nonrandom a_{nf} are called "factor loadings;" if the factors are considered as different commodities, the loadings play the role of *prices*. (The fact that the same commodity may have different prices with respect to different buyers, P_n, is a result of the impossibility of renting the different factors to different employers; we intend to discuss this question elsewhere.) The nonhomogeneous term a_{n0}, which can well be negative, expresses the value of any "body" to the occupation P_n.

The above technique is like the factor analysis used in psychometry. However, the replacement of the Gaussian factors by scaling quantities (see Section V) changes the theory beyond resemblance. In particular, the usual difficulties of factor analysis disappear (to be replaced, presumably, by fresh ones); we hope to take up this topic elsewhere.

Note that "family incomes" provide an intermediate example between cases where the factors are indissolubly linked or separable.

IV. REGIONS OF ACCEPTANCE OF DIFFERENT OFFERS

The prospective employee has to maximizes the nonlinear function U_n. Let the random variable W_n designate the incomes *accepted* for the occupation P_n, and let W designate all accepted incomes.

Clearly, the region of acceptance R_n of the offer U_n will be the intersection of the $N-1$ half-hyperspaces defined by the equations

$$U_n - U_i \geq 0 \qquad (i \neq n).$$

The domain of values of v_f in which the accepted income is $W < u$ is familiar from linear programming: it is the convex hull of the N planes $U_n = u$ and of the F planes $V_f = 0$. If, for every u, the plane $U_n = u$ is exterior to that convex hull, the occupation P_n may "economically unessential," and must rely on noneconomic inducements to recruit unless the error terms E_n provide it with the personnel which it needs.

It may be noted that the above regions of acceptance are independent of the probability distribution of the factors. This is unrealistic, insofar as the factor loadings, considered as prices, must eventually be determined by supply and demand, which in turn will sharply depend upon the available combinations of factors. But before synthesizing the whole theory, we must analyze its various parts. (See Section 9 for some semblance synthesis.)

Now consider a few examples.

The offers are homogeneous forms of the V_f. The accepted P_n will only depend on the relative values of the factors, that is, on what we may call the "factor profile" of the individual.

A consequence is that the regions of acceptance are polygonal cones, all having their apex at the origin of the factor space. Unboundedly large incomes will be encountered in every profession.

For the case $F=2$, the typical shape of the regions R_n and of the surfaces $W=u$, is presented in Figure 1. Figure 2 shows what can happen when $F=3$. The weights $w(n)$ marked on these figures were described in Section 2 and will be derived in Section 7. The dashed lines of Figure 1 will be explained in Section 5.

In Figures 2a to 2e, if one factor is present in very large quantity, it fully determines the chosen occupation, so that the values of other factors are irrelevant. But in Figure 2f the occupation is not determined by the value of the high factor V_1 alone; it is also greatly influenced by V_2 and V_3, which are present in small quantities. Small changes in these other factors can change the optimal value of n.

As to the surfaces $W=u$, they continue to be obtained from the surface of unit income by multiplying all coordinates by u.

Linear and nonhomogeneous offers. The regions of acceptance R_n need no longer be open cones. Some may be closed, leading to bounded incomes. The corresponding occupations will clearly not be very sensitive to the values of the factors. Fanning-out regions R_n, to the contrary, correspond to occupations that are quite sensitive to the factors. An example is given in Figure 3.

If the offers U_n, as functions of the V_f, were nonlinear but homogeneous of the first order, the regions of acceptance would be cones with their apex at the origin. But they would not by polygonal and, if the origin is excluded, R_n need not be simply connected.

V. THE ASSUMPTION THAT THE FACTORS ARE SCALING

The fundamental and least obvious "input" of the theory will be the assumption that each factor is asymptotically scaling. Hence, there exist some constants $\tilde{u}_f$ and $\alpha > 1$ such that, when u increases to infinity,

$$1 - F_f^0(u) = \Pr\{V_f > u\} \sim (u/\tilde{u}_f)^{-\alpha}.$$

Assume that offers can be factor-analyzed linearly, that an occupation is chosen by income maximization and that the expressions $1 - F_f^0(u\tilde{u}_f)$ have the same asymptotic behavior for all f. Then the following can be proven (but not here): in order that the overall income distribution be scaling, the same must be true of every factor taken separately.

The scaling characters of every factor will be one of the "axioms." *Strict scaling* will considerably simplify the mathematical arguments,

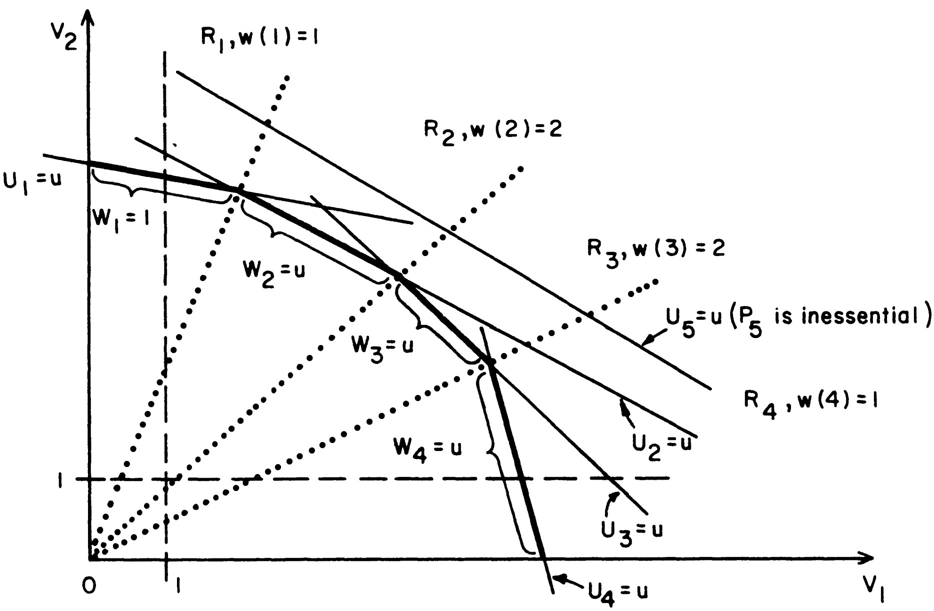


FIGURE E12-1. Regions of acceptance in the case of two homogenous factors. Dots: boundaries between different R_ns. Thin lines: $U_n = u$. Thick line: $W = u$.

although it is unquestionably not a very good representation of the facts. This distribution states that

$$1 - F(u) = (u/\tilde{u})^{-\alpha} \text{ for } u > \tilde{u} \text{ and } F(u) = 0 \text{ for } u < \tilde{u}.$$

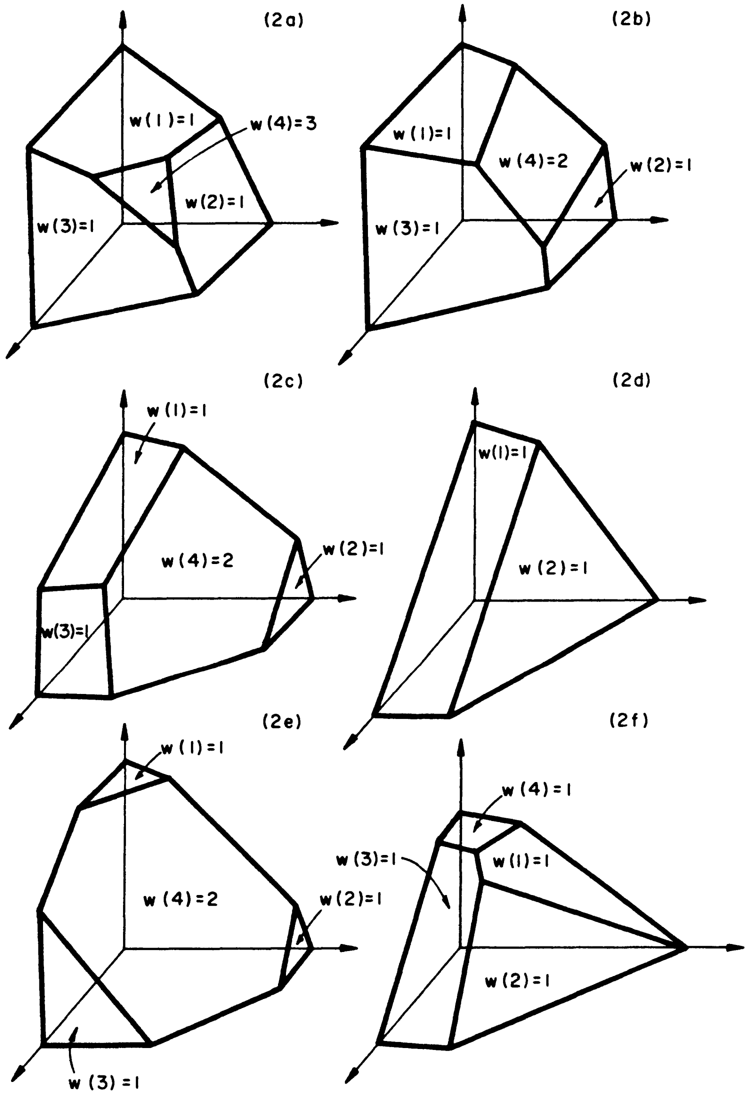


FIGURE E12-2. Six miscellaneous examples of the shape of the iso-surface $W = u$, in the case of three homogenous factors.

Another special example is the L-stable distribution. Its density $p_\alpha(u)$ is not expressed in closed analytic form, only through the bilateral generating function

$$G(s) = \int_{-\infty}^{\infty} \exp(-su)p_\alpha(u)du = \exp[-Ms + (s\tilde{s})^\alpha],$$

where $M = EU$, and $1 < \alpha < 2$. This is a nuisance, but the L-stable distribution has two major advantages:

(a) Its scaling tail continues, for medium and small values of the variable, by a "bell" that is skew but not excessively so.

(b) The Appendix to this section shows that this distribution reduces the "axiom" that the V_f are scaling, to considerations on sums of random variables. This reduction is intimately connected with the models of income distribution and variation in M 1960i{E10} and M 1961e{E11}, and with other applications I hope to publish in due time.

Surfaces of constant probability. Examine strict scaling first. Since it is considered only for the sake of mathematical simplicity, and since the parameters a_{nf} take care of the scale of factors, we can suppose that the $\tilde{u}$ of all the factors are equal to one. The joint density function of the F factors then takes the form

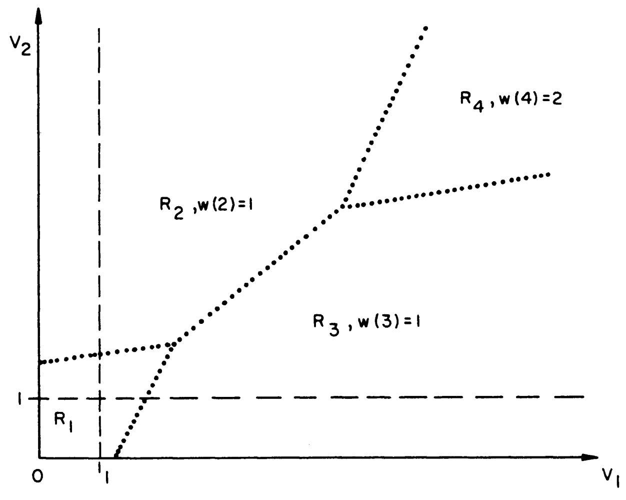


FIGURE E12-3. Regions of acceptance in the case of two inhomogeneous factors.

$$\alpha^F(\Pi v_f)^{-(\alpha+1)}\Pi dv_f.$$

The probability iso-surfaces are the hyperboloids defined by the equations

$$\Pi v_f = \text{constant},$$

truncated to the region in which all the $v_f \geq 1$.

Consider the surface $\sum_{f=1}^F v_f a_{nf} = u_n$. If the restrictions $v_f \geq 1$ are neglected, the point having the coordinates u_n / Fa_{nf} is the "least probable" one on that surface. It happens to be precisely at this point that the amounts paid for the different factors are equal to each other and to u_n / F. Similarly, every surface with a convexity inverse to that of the above hyperboloids also includes a unique least probable point. For example, there is a unique least probable point on the surface along which the best offer is equal to u.

If one or more of these coordinates turn out to be smaller than 1, some "corner" difficulties are encountered; we will not enter into them here.

In any event, the points for which $v_f < 1$ for at least one f are never encountered; they are even less probable than (u_n / Fa_{nf}). But we need not worry about this either.

Examine now the case of factors for which the scaling "tail" is preceded by some kind of "bell." The probability iso-surfaces then takes the form that Figure 4 exemplifies for the case $F = 2$. If u_n is sufficiently large, the surface $\sum_{f=1}^F v_f a_{nf} = u_n$ includes a "least probable" point.

Finally, consider the region in which at least one of the factors is large. The probability density will mostly concentrate in "finger-shaped" regions "parallel" to the F axes; there, one factor is large while others are close to their most probable values. The remaining probability will mostly concentrate in "web-like" regions "parallel" to one of the planes joining two coordinate axes; there, two of the factors are large, while the others are close to their most probable values. There is extremely little probability in the region away from every coordinate axis.

The concentration of the probability near the coordinate axes and planes means that relatively few bundles of abilities contain large quantities of several factors, simultaneously. That is, the problem of pricing such bundles, mentioned in Section 4, will not be so frequently encountered as one might have feared. As a corollary, these prices may be less well determined than they would be in a more active market; this may be a rather important point.

VI. CONCLUSION: OFFERS ACCEPTED FROM DIFFERENT OCCUPATIONS WILL BE SCALING, WITH EXPONENTS GIVEN BY INTEGRAL MULTIPLE OF THAT OF THE FACTORS

For all factor loadings, as we know, the variables W_n (namely the *offers accepted* from the professions P_n) will either be bounded, or satisfy

$$\Pr\{W_n > u\} \sim (u/u_n)^{-w(n)\alpha}.$$

Let us show that $w(n)$ equals the smallest number of different factors that must be large simultaneously, if W_n is to be large. To prove this

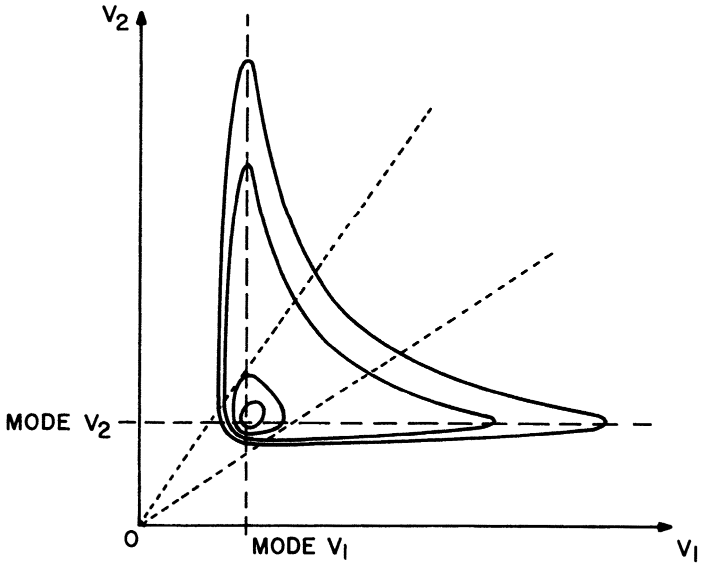


FIGURE E12-4. Example of probability iso-lines in the case of two factors, when the probability density curve of each factor has scaling tails and a "bell."

In the L-stable case, this graph continues without bound for negative values of v_1 and v_2 because for $\alpha > 1$ the L-stable density is not restricted to positive values of v. However, large negative values of v are so unlikely (less probable than corresponding large negative exceedances of the Gaussian law), that they are not worth worrying about.

(P.S. 1996. In the 1962 original, this figure was drawn incorrectly.)

result, recall that the regions of acceptance R_n are cones (Section IV), and that the joint probability density of the factors is concentrated in some "fingers," some "webs," etc. (Section V). It follows that the acceptance of the offer from occupation P_n may be due to either of the following circumstances.

The region R_n may contain a probability "finger." In that case, the large size of U_n will usually be due to *one* of their factors being very large. Other factors are either close to the most probable value, or are so lightly loaded (i.e., so little valued), that they do not influence U_n markedly. In this case, *the weight is* $w(n) = 1$.

The region R_n may contain no finger, but a half-line in a web in which two factors are large. In that case, those who choose such a profession will mostly find that most of their income comes from two factors, i.e., $w(n) = 2$.

More generally, consider the linear manifolds of the factor space spanned by the positive parts of one or more of the coordinate axes. Some of these subspaces touch the portion of the hyperspace $U_n = u$ which is contained within the domain of acceptance R_n of profession P_n. Then $w(n)$ is equal to the dimension of the subspace of lowest dimensionality that is touched by this portion of $U_n = u$.

At least one occupation will require only one factor and have a weight $w(n) = 1$; but every P_n may have a weight equal to 1. In general, whichever the sign of $N - F$, all the weights may equal 1 (see Figures 2d and 2f).

Now examine more closely the case $F = 2$. If there are one or two essential occupations, both have the weight 1. If there are $N > 2$ essential occupations with unbounded incomes, two of them have the weight 1. For them, the region of acceptance R_n (which is asymptotically bounded by two straight lines) contains the infinite part of the line $v_1 = 1$ or the infinite part of the line $v_2 = 1$. The other $N - 2$ occupations have weight 2. This distinction is visible at a glance on Figures 1 and 3.

Every occupation on Figure 1 makes offers based upon the values of both factors. But consider the occupations of weight 1; the expression $a_{n1}v_1/u_u$, which is the relative rent of the first factor, varies either from 0 to some maximum r_n, or from some minimum s_n to 1. For the occupations of weight 2, this ratio varies from some non-zero minimum s_n to some maximum r_n less than 1. In this sense, one may say that $w(n)$ represents the number of factors that influence the distributions of W_n in an essential fashion; i.e, the number of factors that can never be negligible with respect to the value of W_n.

The above result means that *the larger the number of factors* that must all rate highly in order for the offer U_n to be preferred to all others, *the higher the value of $w(n)$ and the fewer the number of people* deriving a very high income from the occupation P_n.

In other words, if all the offers are linear forms in terms of scaling factors (or homogeneous forms of the first order), the occupations that require a mixture of two or more different "abilities" will attract very much fewer exceptionally talented people than the occupations that require only one ability. Most of the highly paid people will be completely specialized in a way that is socially advantageous at the time.

If encouraged to train certain abilities further, in order to increase their income, those highly paid people are likely to train their best ability and become even more specialized. This tendency would be a reasonable point of departure for an alternative explanation of the distribution of the factors, using a "learning" process. For example, the process recalled in the Appendix to Section 6, could be interpreted in learning terms. In this context, the choice of a one-factor profession, or of one requiring w factors, could well be conditioned by factors other than maximization of current – or expected future – income.

Consider from this viewpoint an originally "interdisciplinary" enterprise, that is, one requiring a certain combination of various talents, to which a sufficient social advantage is initially attributed to insure that it starts. Such an enterprise will attract people of high ability in large numbers only if it turns out to continue to require some single factor, which until then was either unrewarded, unsuspected, or mixed with other factors.

VII. EVALUATION OF THE ASYMPTOTIC WEIGHT OF AN OCCUPATION

The derivation of the value of $w(n)$ is immediate when $F=2$ and $w(n)=2$. Suppose, to begin with, that the boundaries of R_n are two halflines, issuing from the origin and strictly contained between the halfline $v_1=0$, $v_2=1$ and the half-line $v_2=0$, $v_1=1$. Suppose also that both factors are strictly scaling with the same α. Then

$$\Pr\{U=W_n>u\}=\Pr\{U_n>u \text{ and } U_i<U_n \text{ for } i\neq n\}$$
$$=\iint \alpha^2 x^{-(\alpha+1)} dx\, dy.$$

This double integral is taken over a domain delimited by two straight lines passing through 0 and a cross line. The boundary may also include two portions of the lines $v_1 = 1$ or $v_2 = 1$ but this need will not arise when u is large enough, as will be assumed. Then one can choose some reference income $\tilde{w}$ and write $x = \tilde{x}(u/\tilde{w})$ and $y = \tilde{y}(u/\tilde{w})$, so that

$$\Pr \{U = W_n > u\} = (u/\tilde{w})^{-2\alpha} \int\int \alpha^2 \tilde{x}^{-(\alpha+1)} \tilde{y}^{-(\alpha+1)} d\tilde{x} d\tilde{y}.$$

The integral now extends over the domain in which $U = W_n > \tilde{w}$, hence is equal to $\Pr \{U = W_n > \tilde{w}\}$. As a result,

$$\Pr \{W_n > u\} = \frac{u^{-2\alpha}}{\tilde{w}^{-2\alpha}} \Pr \{W_n > \tilde{w}\} = Cu^{-2\alpha}.$$

This shows that $w(n) = 2$. The asymptotic scaling behavior is achieved as soon as the restrictions $v_1 > 1$ and $v_2 > 1$ cease to influence the shape of the domain of integration of $\Pr \{W_n > u\}$.

Similarly, if $F > 2$, an occupation has a weight equal to F if R_n is a cone of apex 0, defined by the following condition: if V is a unit vector of a direction contained in R_n, all F projections of V on the coordinate axes are nonvanishing (i.e., if R_n contains no half-line spanned by less than F coordinate vectors). This is so because

$$\Pr \{W_n > u\} = \int_{F \text{ times}} x^F (v_1 v_2 v_3 \ldots)^{-(\alpha+1)} dv_1 dv_2 dv_3 \ldots dv_F,$$

Here, the integral extends over a domain such that one can again change the integration variables to show that

$$\Pr \{W_n > u\} = \text{ a constant } u^{-F\alpha}.$$

Therefore, the weight is equal to F, as announced.

Now, returning to the case $F = 2$, suppose that R_n is bounded by two parallel lines having a direction different from that of either coordinate axis. It is easily shown that

$$\Pr \{W_n > u\} \sim \tilde{C} u^{-(2\alpha+1)}.$$

Actually, parallel boundaries have an infinitesimal "likelihood" of occurrence. But it is useful to consider them as an auxiliary step for the case where the "lateral" boundaries of R_n no longer converge to the origin. It is then possible to decompose R_n into a centered cone, plus or minus one or two parallel strips. Hence,

$$\Pr\{W_n > u\} \sim Cu^{-2\alpha} + C^* u^{-2\alpha-1}.$$

The second term becomes negligible if u is sufficiently large. Again, $w(n) = 2$.

Now, consider in more detail the ratio $a_{n1}v_1/u_n$, among individuals known to receive the income $u_n = u$ from P_n. Suppose that the cone R_n is centered at 0, and compare the two points for which $y'/x' = \tan\theta'$ and $y''/x'' = \tan\theta''$. One has $x' = uC'\cos\theta'$, $y' = uC'\sin\theta'$, $x'' = uC''\cos\theta''$, $y'' = uC''\sin\theta''$, where C' and C'' depend upon the factor loadings. That is, the ratio of the densities at the two points is

$$\frac{\alpha^2 x'^{-(\alpha+1)} y'^{-(\alpha+1)}}{\alpha^2 x''^{-(\alpha+1)} y''^{-(\alpha+1)}} = \frac{(C')^{-2(\alpha+1)}(\cos\theta')^{-(\alpha+1)}(\sin\theta')^{-(\alpha+1)}}{(C'')^{-2(\alpha+1)}(\cos\theta'')^{-(\alpha+1)}(\sin\theta'')^{-(\alpha+1)}}.$$

It is independent of u. The same result will hold asymptotically if the cone R_n is not centered at 0: the distribution of the ratio of factors, among individuals with a high u obtained from P_n, is independent of u if the occupation has the weight 2.

Note also that L_n, defined as the least probable point on the line $U_n = u$, has the following property. If L_n is not within R_n, those who choose P_n are mostly located near the edge of R_n, that is, farthest from L_n.

Being mostly from this specific area, those persons show a certain homogeneity. But if L_n is within R_n, the most probable combinations of factors are located near either edge of R_n, so that *two* particularly likely kinds of factor profiles are present.

Now, consider one of the two occupations of weight 1. For a profession that makes offers that depend upon one factor

$$\Pr\{u < W_n < u + du\} = \iint \alpha^2 x^{-(\alpha+1)} y^{-(\alpha+1)} dx\, dy,$$

extended over a strip bounded by $x = su$, $x = s(u + du)$, $y = 1$ and $y = ru$, where s and r are constants depending upon the factor loadings. Hence,

$$\Pr \{u < W_n < u + du\} = \alpha s^{-\alpha} u^{-(\alpha+1)} du[1 - (ru)^{-\alpha}].$$

For large u, the term $(ru)^{-\alpha}$ becomes negligible, so that $w(n) = 1$.

If an occupation of weight 1 makes offers that depend upon both factors, the calculation is very slightly more involved, but it is easy to show that $\Pr \{W_n > u\}$ is bounded by two expressions of the form $c'u^{-\alpha}$ and $c''u^{-\alpha}$, which again shows that the weight is equal to 1.

Consider now the ratio $a_{n1}v_1/u_n$ for individuals choosing an occupation of weight 1, such as P_4 of Figure 1. The quotient of the proportions of individuals for whom the ratios of the factors are $y'/x' = \tan\theta'$ or $y''/x'' = \tan\theta''$, is still independent of u. However, as u increases, y/x can take increasingly smaller values, corresponding to increasingly specialized individuals. As a result, the distribution of $a_{n1}v_1/u_n$ concentrates increasingly near the point of complete specialization. That is, even if the $F-1$ factors other than the most important one ceased to be rewarded, most of the highly paid people in an occupation of weight 1 would hardly notice the difference.

If $F > 2$ and $w(n) < F$, the proofs are even more cumbersome, but quite straightforward; we shall not describe them in detail.

VIII. BEHAVIOR OF THE DENSITIES BEFORE THE ASYMPTOTIC SCALING RANGE; LOGNORMAL APPROXIMATION

When there are two strongly scaling factors, each occupation has a minimum income which is attained by some people for whom one or both factors take their minimum value 1. For simplicity's sake, let us concentrate on Figure 5a, in which P_1 and P_2 have the weight 1 and P_3 has the weight 2, and the minimal incomes are related by $u''_1 > u''_2 > u''_3$. In this case, no income less than u''_3 is *ever* accepted.

Now let $u''_3 < u < u''_2$. By inspection of the figure, all such incomes come from P_3, and they are in no way influenced even by the existence of the other two occupations. Since U_3 is a weighted sum of two strongly scaling variables, it is asymptotically scaling. Its asymptotic behavior may begin to apply before the influence of P_1 and P_2 begins to be felt.

Suppose now that $u''_2 < u < u''_1$. Then, the accepted income will come from either P_3 or P_2. There are two occupations and two factors and both accepted income distributions W_3 and W_2 will have weight 1.

Finally, if $u > u''_1$, one may accept any of the three offers. This consigns the occupation P_3 to a zone that does not contain either coordinate axis, so that its weight reaches its asymptotic value 2.

In a more realistic case, the scaling "tail" is preceded by some kind of a "bell." If so, the sharp corners characteristic of strict scaling disappear and the above transitions occur slowly and gradually, as shown in Figure 5b and 5c.

The argument extends without difficulty to the nonasymptotic behavior of the probability density of offers accepted from an occupation P_n of weight greater than 2. Before it settles down to the behavior corresponding to its asymptotic weight, the density traverses stages in which certain pure-factor professions have not yet become tempting to anyone, so that P_n behaves "as if" its weight were 1, then 2, etc. The higher the final $w(n)$, the higher the number of possible intermediate regions (the weight can skip some integers, but it cannot decrease). The transitions between successive regions are not actually abrupt, except in the unreal-

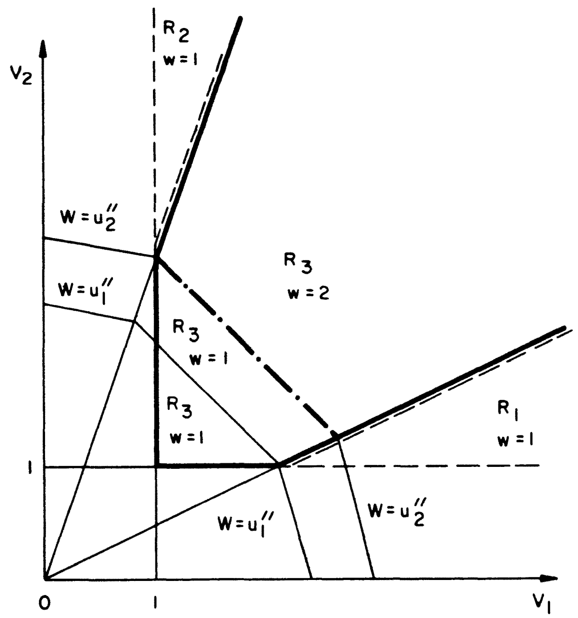


FIGURE E12-5. Example of successive appearance of new occupations, as u increases.

istic case of strongly scaling factors. That is, all that we can assert in general is that

$$S(u) = \frac{-d \log [dF(u)/du]}{d(\log u)}$$

increases with u, and asymptotically tends to some limit of the form $w(n)\alpha$.

Such a behavior affects the very meaning of asymptotics. Suppose, for example, that the α of the factors is about 1.5. Then, an occupation with weight 3 will have an α close to 5, which is about the highest value of α that one has a realistic chance of ever observing. For higher weights, the probability of encountering high values of u will be too small to allow a reliable estimation of $w(n)$. In a way, weights values greater than 3 may as well be considered infinite.

As a result, within a category of weight greater than 3, the asymptotic value of $S(u)$ (as defined in the latest displayed formula) ceases to describe the distribution of income. It is important to know how fast $S(u)$ increases in the usual range of $\log u$. For example, useful approximations will be given by polynomial expansions of $\log [dF(u)/du]$. The approximation $-\log [dF(u)/du] = c' + c''(\log u - c''')^2$ would make such income categories appear to be lognormally distributed. Or, if $S(u)$ increases even faster,

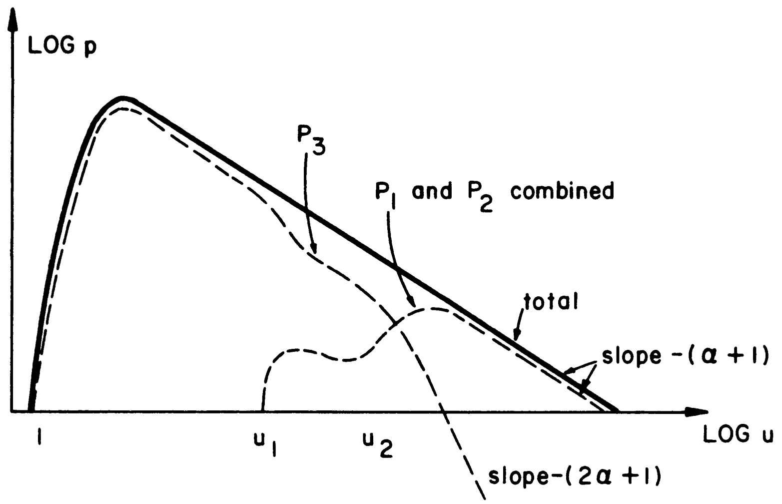


FIGURE E12-6. Example of probability density curves for different occupations in bilogarithmic coordinates.

$F(u)$ may appear to be Gaussian. In both cases, the normal character of the distribution would simply be linked to a polynomial representation of $-\log$ (density), and not to any probability limit theorem.

This terminates the detailed justification of the results which we set out to prove.

A critical question remains. The preceding justification of approximate lognormality involves no limit theorem of probability. How does it relate to the usual explanation, according to which the logarithm of income is Gaussian, because it is the sum of very many components? Our answer shall be addressed to the reader acquainted with L-stable random processes, and shall be very brief.

First, when studying the convergence weighted sums of random variables of the form $A(N)\Sigma X_i - B(N)$, it is customary to only consider cases where "each contribution to the sum is negligibly small in relative value." But this statement is ambiguous. An ex-ante condition is $\Pr\{|A(N)X_i - B(N)/N| > \varepsilon\}$, where ε is fixed, decreases to zero with $1/N$. But it is quite possible that, ex-post, the largest addend remain nonnegligible. As a matter of fact, from the condition that each addend be negligible both ex-post *and* ex-ante, it would follow that the limit is Gaussian. When the limit is a L-stable random variable, the largest addend is

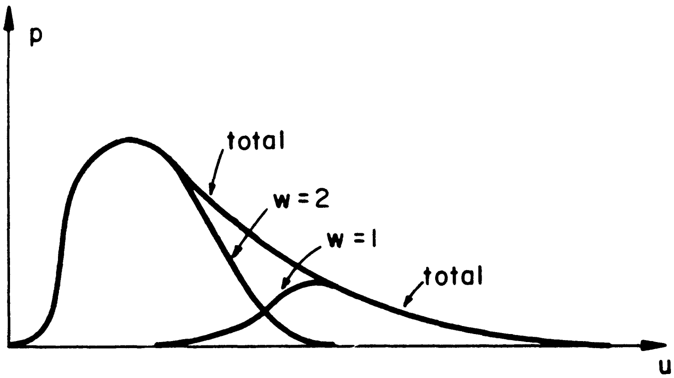


FIGURE E12-7. Example of probability density curves, as plotted in ordinary coordinates.

ex-post; not negligible, in fact, if the value of the limit is large, then the largest addend usually contributes an overwhelming percentage of the sum.

Now consider an occupation P_n for which the weight of accepted incomes is quite large, which requires F to be large. The incomes *offered* by P_n were sums of the scaling contributions of many factors, so that they were themselves L-stable variables (assuming that $1 < \alpha < 2$). Hence, for most prospective employees, an appreciable portion of U_n will be contributed by the very few most influential factors. However, the set of people for whom this holds is precisely the same as the set of people who will *not* choose P_n as their occupation. As a result, the offers *accepted* from P_n will be such that *no* small group of factors contribute to an appreciable proportion of W_n, just as in the case for the contributions to a Gaussian sum.

This leads to an important point. We believe that scaling behavior with $1 < \alpha < 2$ is quite usual for linearly aggregated economic quantities, and for linear functions of such aggregates. Hence, the actual contributions to these quantities will often be of very unequal sizes. However, this rule is broken when nonlinear operations are applied. The model of aggregation combined with maximization is an example of such nonlinear relationship between the inputs (V_f) and the outputs (W_n).

IX. ELASTICITIES OF THE DISTRIBUTION OF PEOPLE AMONG VARIOUS OCCUPATIONS

Let us investigate the effect of changes in the "loadings" of the various factors upon the numbers of people choosing various occupations.

First consider an occupation having a weight equal to 1, and a region of acceptance that does not present the special feature of Figure 2f. The effect of widening its fan of acceptance decomposes into two parts.

On one hand, it will add some people who rank highly in the factor V_f that predominates for that profession. However, since the majority of high-ranking people are extremely specialized, the increase of their numbers will be less than proportional to the increase of the width of the fan of acceptance: the elasticity relative to high rank will be *smaller than one.* The additional outlay necessary to increase the number of people in such an occupation will mostly go towards increasing the incomes of people who choose that occupation anyway. (A different and independent treatment of the same topic may be found in Machlup 1960.)

On the other hand, widening the fan of acceptance also adds people having a value of V_f lower than the smallest value encountered before the raise of salary levels. When the U_n are all homogeneous linear forms, and the factors are strongly scaling, the elasticity relative to low-ranking people is infinite. In any case, it is very high. Needless to say, an addition of "inferior candidates" is commonly observed when one wishes to increase the number of people in a very specialized activity, by increasing its rewards throughout the range of u.

(A simple device for avoiding such effects altogether consists of dividing a profession into several parts, each having a different reward system, even though their members are not professionally segregated. Note also that if the region R_n of a profession of weight 1 presents the special features of Figure 2f, the problem of elasticity is more involved. However, this section does not propose to list all possibilities, only to indicate some likely behaviors.)

Consider now an occupation of weight 2. We saw that those who rank highest in that profession are very likely to be those near one or both edges of the region of acceptance R_n. Hence, if R_n is widened at such an edge, the elasticity of the number of high ranking people choosing P_n will be *greater than* 1. If the fan of acceptance is widened at an edge of lowest density, the elasticity will be less than 1; But the average of the elasticities relative to both edges will always be greater than 1. The behavior of low-ranking people will again be different from that of high-ranking people, but will not change the general nature of the result. Similar considerations apply a fortiori to occupations with weights greater than 2.

But this is not the whole story. If one occupation changes its rewards so as to increase or decrease the number of its employees, what happens to other P_n's? For example, if a single-factor occupation disappears entirely as a result of a small change in its U_n there will be a tremendous over-supply of its former employees in the high ranges. Such a change in the rewards of a single-factor occupation will inevitably influence the rewards of the nearest two- or three-factor P_n's. That is, if the value of a factor decreases if taken alone, it will also decrease when taken in association with other factors. The elasticity limited to direct effects of change of price will be rather meaningless. (See Figure 6.)

X. REGIONAL DIFFERENCES IN THE VALUE OF α

The specialization associated with low α may well be impossible except in certain geographical regions of a country. Suppose then that geographical mobility is as complete as the professional mobility which we assume. In that case, low weights may be absent in some regions, and the overall exponent, equal to the α of the occupation of lowest weight, would vary from region to region. If this effect were confirmed, this paper may be worth translating to terms of regional economics.

APPENDICES

This paper's predictions seem in surprisingly good agreement with the data described in Section 1. Besides, the original assumptions are simpler than the conclusions which we reach. Hence, while the body of the paper asks the reader to accept those assumptions without much discussion, we feel that we achieve an "explanation" of the observed gross data.

All the same, it is good to investigate the sensitivity of this explanation with respect to changes of axioms and to see whether this explanation

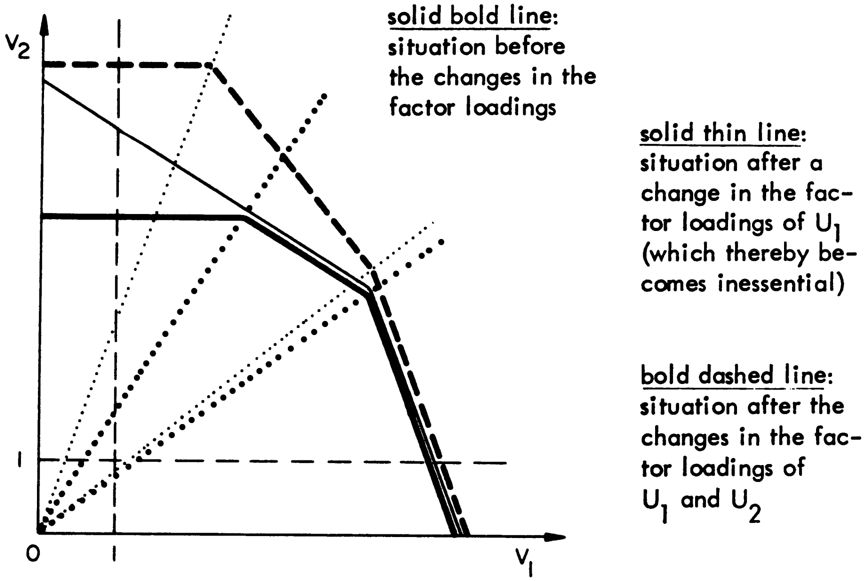


FIGURE E12-8. Example of the change in the form of the curve $W = u$, as factor loadings are modified.

could not be reduced to still more "elementary" terms. This is the purpose of two appendices.

APPENDIX TO SECTION V. ON THE SCALING CHARACTER OF THE FACTORS AND THE USE OF LINEAR FACTOR ANALYSIS

The "L-stable" approach. As discussed in M 1960i{E10}, the L-stable distributions can be obtained, like the Gaussian, as limits of weighted sums of independent and identically distributed random variables. Conversely, the L-stable distributions are the only nonGaussian limits with a finite expectation and rare large negative values.

This means that additive decomposition of incomes into parts does not contradict their being obviously nonGaussian. It is, therefore, extremely tempting to argue that the factors must be L-stable because they can be written as the additive aggregates of many different influences. Granting that the factors are surely not Gaussian, they could only be L-stable, with α that need not be the same for different values of f.

However, we would very much hesitate to take this viewpoint. After all, L-stable variables avoid the addition of parts of income in a nonlinear scale such as that of the logarithm of income. It would be paradoxical to end up by adding contributions in the conjectural scales of the "factors."

Fortunately, the addition of conjectural parts can be avoided with the help of an N-dimensional Euclidean space having the offers U_n as coordinates. It is useful to define $J_0 = (a_n)^\alpha$, $J_f = (\Sigma_n (a_{nf})^2)^{\alpha/2}$ and $c_{nf} = a_{nf}(J_f)^{-1/\alpha}$, so that $\Sigma_n (c_{nf})^2 = 1$ for all f, and the c_{nf} can be considered as directing cosines of some direction D_f in N-space. Then,

$$U_n = \sum_f c_{nf}(J_f)^{1/\alpha} V_f + (J_0)^{1/\alpha} E_n + a_{n0} ,$$

The vector of coordinates U_n is the sum of vectors placed along the directions D_f such that their lengths are scaling scalars having the scales J_f. As shown in Lévy 1937 (see also M 1961e{E11}), if the lengths of the above vectors are L-stable scalars, one obtains the finite version of the most general L-stable vector with infinite variance and finite mean. That is, suppose that the vector having the N offers for coordinates, is the sum (in the scale of dollars and cents) of very many contributions. Then, the relationship between the offers must be given by our linear L-stable factor analysis, or at most, by a variant of this analysis in which the number of factors is infinite.

In conclusion, not only can we justify the L-stable character of the factors, but also the linear character of our factor analysis.

The Champernowne approach. Let the logarithm of each factor perform a random walk, with downward trend and a reflecting boundary layer. The distributions of the factors eventually become – and stay – scaling.

This is a straightforward adaptation of an argument which has long been used to justify the Pareto law of income, and which was put into a good form in Champernowne 1953. In the simplest case, $\log(V_f)$ can only increase by some quantity v'', decrease by v'', or remain constant. If $\log V_f > \tilde{v}$, the probabilities of the above three events are assumed to be independent of V_f and to be equal, respectively, to p, $q > p$, and $1 - p - q$. For smaller $\log V_f$, this simple random walk is so modified, that V_f is prevented from becoming infinitely small. Under these conditions, one finds that V_f is asymptotically scaling with $\alpha = (1/v'') \log(q/p)$.

This argument can be used to further explain the nature of the weight $w(n)$.

Suppose first that P_n has weight 1, that is, the ratio $a_{n1}v_1/u_n$ can vary from zero to some maximum r_n. If the factor V_1 is multiplied by $\exp(\pm v'')$, the total income is multiplied by a term contained between 1 and $1 + r_n[\exp(\pm v'') - 1]$; that is, it can change only a little. On the contrary, if the factor V_2 changes, the total u_n changes in approximately the same ratio as V_2. The behavior of the whole is, of course, not a random walk. But, in a not-so-rough approximation, one may still argue as if it were one. One could even assume that the logarithm of income varies following the same values of v'', p and q as the logarithms of the factors. Hence, income from P_n will have the same α as the factors.

Examine now an occupation of weight 2, and suppose that p and q are sufficiently small to allow us to neglect the probability of seeing both factors increase or decrease simultaneously in the same time interval. Suppose also that $a_{n1}v_1/u_n$ is contained between 1/3 and 2/3. Then, there is a total probability $p + p = 2p$ that u_n be multiplied by a term contained between

$$\frac{2}{3} + \frac{1}{3}\exp(v'') \sim \exp\left(\frac{v''}{3}\right) \text{ and } \frac{1}{3} + \frac{2}{3}\exp(v'') \sim \left(\frac{2v''}{3}\right);$$

Also, there is a probability $2q$ that u_n be multiplied by a term contained between

$$\sim \exp\left(-\frac{v''}{3}\right) \text{ and } \sim\left(-\frac{2v''}{3}\right).$$

Again, W_n no longer follows the simple random walk, but the orders of magnitude of the up and down steps are $v''/2$ and $-v''/2$, with respective probabilities $2p$ and $2q$. This gives "alpha" $=(2/v'')\log(2p/2q)=2\alpha$, and the weight is 2, as expected.

Because of the approximation of the behavior of u_n by a simple random walk, the above heuristic argument is not a complete proof. Besides, it was rather obvious that, if the components of income are of comparable sizes and behave independently of each other) their sum is less sensitive to random changes of the parts than if the whole is mostly due to one contribution. But it was not obvious that the choice of a profession, if it involves the maximization of a linearly factored-out income, can be used to separate the bundles of factors in which one factor predominates, from those in which both factors contribute comparably.

APPENDIX TO SECTION VII. UNEQUAL FACTOR EXPONENTS AND NONSCALING FACTORS

If the factors are strongly scaling with different values of α_f the surfaces of equal probability have the equation

$$\Pi(v_f)^{(\alpha_f+1)} = \textit{constant}.$$

Their convexity is the same as when the α's are equal. However, the concept of weight disappears in this case. Instead, the incomes accepted from each P_n will be asymptotically scaling, and their "exponent" will be the *sum* $\alpha = \alpha_{f_1} + \alpha_{f_2} + \alpha_{f_3} + \ldots$ relative to those factors V_{f_i} that must be large simultaneously, if U_n is to be large. For example, if all the α_f are between 1.33 and 2, the occupations requiring two factors will have exponents in the range 4 to 6; the higher numbers of factors become meaningless. Hence, it is still possible to say that, the larger the number of factors that must be large simultaneously, the larger the scaling exponent, and conversely.

Although the weight is no longer defined, the overall properties of the largest income remain as in the body of this paper.

Very different results follow when one changes the convexity of the surfaces of constant probability. In the Gaussian case, the densities of the factors are of the form

$$\frac{1}{\sigma_f\sqrt{2\pi}}\exp\left\{-\frac{(v_f-\tilde{v}_f)^2}{2\sigma_f^2}\right\},$$

and the surfaces of constant probability are the ellipsoids

$$\sum_{f=1}^{F}\frac{(v_f-\tilde{v}_f)^2}{\sigma_f^2}=\text{constant}.$$

In the region of high positive exceedances, the convexity of these surfaces is the same as that of a sphere of center 0. On each surface $\Sigma v_f a_{nf}=u_n$, there is a point of highest probability. There are no fingers or webs along which the probability is concentrated. {P.S. 1996. A digressive paragraph of the original is omitted here, because it was found to be in error.}

Granting this convexity, examine the proportion of various factors most likely to be encountered in the high income range. In the scaling case, such proportions were most often determined by going *as far as possible* from the least likely point on a surface, that is, as far as possible from the unique point of contract between the surface $U=u$ and a surface of constant probability. These two surfaces have the same convexity and nothing whatsoever can be said concerning the number and the position of points of contact; that is, concerning the degree of specialization usually encountered among highly paid people. In conclusion – even if the normal and lognormal laws were not in contradiction with the overall distribution of income – maximization of linearly factored-out incomes would get us nowhere.

Exponentially distributed factors. Factors with densities of the form $\exp(-v_f)$ are far more interesting. Again, scale parameters are, in effect, inserted through the a_{nf}. The surfaces of constant probability are the hyperplanes $\Sigma v_f=$ constant. On a surface $\Sigma v_f a_{nf}=u_n$, the points of largest or smallest probability are in general concentrated along the edges, that is, at least one coordinate is f. However, there is definitely a point of smallest probability on every surface having the same convexity as spheres of center 0, for example, on every surface along which the best offer is equal to any u.

Moreover, suppose that, for every fixed n, the F factor loadings a_{nf} are all different, the largest being $\max a_{nf}$. It follows that the probability distribution of the offer U_n behaves for large u like $C_n \exp(-u/\max_f a_{nf})$, irrespective of the values of the factor loading other than the largest, on the condition that they remain strictly smaller. If there are G factor loadings equal to the largest, the behavior of the probability is $C_n u^{G-1} \exp(-u/\max_f a_{nf})$. For small G and large u, this decrease does not differ very much from the exponential.

Similarly, the accepted offer's density will be $\propto C'_n \exp(-\beta_n u)$. The important thing is the form of β_n. For the occupations which are influenced mostly by a single factor, $\beta_n = 1/(\max_f a_{nf})$, and the asymptotic behaviors of U_n and W_n are identical; one can say that the weight of the accepted W_n, *relative to the offer* U_n, is equal to unity.

But for occupations influenced by more than one factor, β_n is much more complicated; the "relative" weight can become arbitrarily close, but never equal, to 1. The reason for this difference with the scaling case can be explained by considering a representative space having for coordinates the expressions $\exp(V_f)$. These expressions are now scaling, so that the lines of constant probability have the characteristic form with fingers and webs. But the regions of acceptance of the different occupations are not bounded by straight lines, rather by fractional parabolas having equations of the form $\exp v_1 = (\exp v_2)^\gamma$, with some constant γ that can become very close to 0 or to 1. As a result, the "fan" constituted by the region of acceptance R_n can increase slower or faster than the straight edge fans encountered in linear factor analysis of income.

The above argument matters because, instead of factor-analyzing the incomes U_n themselves. we could factor-analyze the expressions $\log(U_n)$ as linear forms with respect to exponentially distributed factors. In this case, the offers made and accepted would still be weakly scaling, but their α exponents would have no simple relation with each other – such as that implied in the concept of weight – and could vary in continuous fashion.

This variation of α provides a distinction between the factor analysis of incomes and of their logarithms. Unfortunately, it is difficult to conceive of an experimental test between the two methods. However, the exponential approach has two undesirable features: the linear decomposition of log (income) – instead of income – and the fact that α may depend upon the exact values of the loading a_{nf} instead of more fundamentally structural quantities, such as the number of factors contributing to u in an essential way.

This example shows particularly clearly that linear factor analysis of income yields different results if performed in its own scale U, in the scale of log U, or in the scale of some other monotone increasing function of U. It also indicates that a finger – and web-like behavior for the surfaces of constant probability, demand that the densities of the factors decrease *far slower* than it is for any of the usual probability density functions. Their behavior need not be scaling, but log (probability density) must be cup-convex.

Dedication and Acknowledgement

Dedié à Monsieur Maurice Fréchet, Membre de l'Académie des Sciences, Professeur Honoraire à la Sorbonne.

Initiated in 1958, while I was at the University of Lille, France, this work was set aside until Professor Hendrik S. Houthhakker, of Harvard, encouraged me to resume it and pointed out Roy 1951 and Tinbergen 1956. Those very stimulating papers take a different approach and yield different predictions.

Useful comments upon an earlier version of this paper were made by my IBM colleagues Ralph E. Gomory and Richard E. Levitan.

E13

Industrial concentration and scaling

✦ **Abstract.** Since the theme of concentration is mentioned in the title of this book, and this book boasts several alternative entrances, it is appropriate that yet another introductory chapter, even a very brief one, should take up the phenomenon of *industrial concentration*, and the meaning – or absence of meaning – of the notion of average firm size in an industry. ✦

INDUSTRIAL CONCENTRATION CAN BE SIGNIFICANT, nearly constant through time, and not overwhelmingly dependent of the number of firms included in the industry. Unfortunately, this very important phenomenon is elusive and difficult to study.

For example, suppose that the number of firms and the degree of concentration are large, and consider the fraction that implements the notion of average size of a firm. The numerator greatly depends on the sizes of the largest firms. The denominator mostly depends on the number of the very small ones and depends on what is called a "firm," therefore is ill-determined in many ways. A qualitative study of the ratio demands an analytical formula that accounts for firms of every size, but no such formula is generally accepted. In its absence, the question, "What is the average firm size?," cannot be given a sensible answer. By contrast, Chapter E7 argues that the questions that can be handled sensibly include the evaluation of the degree of concentration. Altogether, the identification of sensible questions is a wide-open prime topic for fractal analysis.

Unfortunately, my knowledge of this topic is thirty years old and even then consisted in a fresh analysis of data collected to fit the needs of others, not mine. Therefore, this chapter and the descriptions of my newest models of price variation did mostly hope to encourage others to a more careful experimental study.

An alert reader will find that most of the fractal program of research on industrial concentration is also implicitly presented between the lines of other chapters. But the topic is so important that it is good to restate basic points in a manner that is focussed, informal, and not dependent on a detailed acquaintance with the rest of this book. This chapter does *not* attempt to *explain* either the phenomenon of concentration, or the underlying size distribution. Its only goal is to demonstrate the indissoluble link between parallel "qualitative" features of physics, economics and randomness.

Scaling and L-stable versus lognormal distribution (Chapter E9) . For the sake of completeness, it is good to sketch this old controversy once again.

Statisticians who favor the lognormal insist that the bulk of firms is distributed along a skew "bell," that the firms in the tail are too few to matter, and that in the bell log U is fitted adequately by a Gaussian probability density.

The uniform scaling distribution, which is not in contention, satisfies

$$\Pr\{U > u\} = (u/\tilde{u})^{-\alpha} \text{ for } u > \tilde{u} \text{ and } \Pr\{U > u\} = 1 \text{ for } u < \tilde{u}.$$

The presence of a most probable minimum value $\min U = \tilde{u}$ is unrealistic. So is the absence of a bell.

As to the asymptotic scaling distribution, it only asserts that

$$\Pr\{U > u\} \propto u^{-\alpha} \text{ for large } u;$$

it says nothing about a minimum value and/or a bell. Those who favor this alternative are restricted in the questions they can ask. Those who favor this alternative are restricted in the question they can ask.

Thus, the lognormal fit is unquestionably superior for a high proportion of the total number of firms. By contrast, the scaling fit claims to be superior to a high proportion of the total size of an industry. Note that, from their viewpoint of concentration, the numerous firms in the distribution's bell count only through their average; the only firms that count individually are those few in the distribution's tail. From their viewpoint, the nice fit of the lognormal's bell is immaterial.

In addition, the general shape of the bell is, qualitatively, about the same for the lognormal density and the asymptotically scaling density called L-stable, which is discussed throughout this book. Given the inadequacy of the data relative to small firms, there is little point in invoking detailed statistical tests to distinguish between different equations for the bell. The assumption that the L-stable applies to firm sizes must mostly be tested in the range of large firms, where the L-stable and the scaling coincide. What counts in that range is that the lognormal's tail is thoroughly unacceptable.

To appreciate the next point, one needs to be aware of the distinction fully described in Chapter E5 between *mild, slow* and *wild* "states" of randomness. Among random variables, a first criticism distinguishes preGaussian randomness, as opposed to wild randomness by the applicability of the law of large numbers and the central limit theorem with a Gaussian limit and the Fickian $\sqrt{N}$ weighting factor. The scaling distribution with a small α is the prototype of wild randomness; the central limit theorem it satisfies is neither Gaussian nor Fickian. A second criterion defines mild randomness, whose prototype is the Gaussian as opposed to long-tailed randomness. Finally, the lognormal distribution is the prototype of slow randomness, which is the intermediate state between mild and wild.

This chapter argues that concentration cannot be accommodated in any straightforward way by any preGaussian size distribution, that is, by any distribution having finite variance, as is, for example, the case for the lognormal. In the case of preGaussian distributions, concentration can only be a "transient" phenomenon that is compatible only with small, not large values of N. By contrast, concentration that persists for large N is an intrinsic characteristic of size distributions with infinite expectation; in practice, this means size distributions that are at least roughly scaling. An exponent $\alpha < 1$ allows for concentration ratios independent of N. That is, as shown in Chapter E7, such distributions describe an industry in which the size of the largest firm is non-negligible compared to the sum of the sizes of all firms, the second largest is non-negligible compared to all firms except the first, etc.. In general, the relative share of the r largest firms is a function mainly of the α exponent of the scaling distribution, and there is no difficulty in fitting that parameter so that the sum of the sizes of the four largest firms together constitute any reasonable percentage of total industry size. When the scaling exponent satisfies $1 < \alpha < 2$, the same is true if the variables U_N are replaced by $U_N - EU_N$. This is a delicate change, because EU, while finite, is not well-determined in practice, as we saw. In practice, the market share of the largest firms is significant if the number of firms is small; as the number of firms increases, concentration does tend to zero, but very slowly.

To repeat, this chapter approaches the ancient conflict of scaling versus lognormal by arguing that the observed concentration ratios afford persuasive evidence *against* the lognormal distributions, and *in favor* of scaling with $\alpha < 1$, which implies an infinite first moment, or at most α not much above 1.

Chapter E5 also introduces a helpful metaphor that associates the mild, slow and wild states of randomness, respectively, with the solid, liquid and gas states of matter. Everyone agrees that the distribution of firm sizes is far from Gaussian, hence the proper metaphor is not a "soft," but a "hard" object. This leaves open the question of whether it is a real solid or a glass. Physics describes glasses as being extremely viscous liquids that mimic solids and happen to be particularly difficult to study. In the same way, the lognormal distribution averages mildly in the very long run, but in the short run it mimics the wild randomness of the scaling in very treacherous fashion. One can call it a sheep in wolf's clothing.

A metaphor will never solve a scientific conflict, but may make it lose some of its bite. The question should be viewed as having changed: it is no longer which of two claimants is closest to a deep truth, but which claimant has the more useful basis for predictions concerning meaningful questions, in particular, the degree of concentration. The fractal viewpoint based on the scaling distribution makes specific predictions described in Chapter E7 and E9. The nature of those predictions may not be welcome, and they may be either falsified or confirmed by new data, but they are available, simple (though not given by elementary formulas), and easy to describe and test. To the contrary, lognormal fitting yields predictions that are analytically unmanageable, hence of no practical use.

A consequence is that the possible failure of the predictions based on the scaling distribution would *not* mean that the distribution of firm sizes proves after all not to be "hard," it would not even mark the triumph of the lognormal. But, it would mark a significant postponement of the hope of achieving a simple rational approach to industrial concentration. A similar situation is described in M 1963e{E3}, which argues that the alternative to scaling is not one non-scaling distribution or another, but a kind of lawlessness.

Finiteness of the population moments is a central issue here. Scaling and L-stable distributions with an exponent satisfying $1 < \alpha < 2$ have *finite* expectation. They occur throughout this book in the context of financial data and the distribution of personal income. Firm sizes force us to move one more step away from mild randomness, and to deal primarily with scaling and/or L-stable distributions with an exponent satisfying < 1, hence an *infinite* expectation.

A few words may help those who are not familiar with the themes of this book. Distributions with an infinite expectation were recognized by the mathematicians. To his contrary, the few examples glimpsed in the course of scientific work were viewed as anomalous, not as proper and non-pathological. To paraphrase a famous line that is quoted in M 1982F{*FGN*}, p. 38, it used to be the case that nearly all statisticians "turned away from this lamentable plague in fear and horror." Some continue to do so! In sharp contrast, a major lesson of fractal geometry is that many aspects of reality are best viewed as exhibiting mathematical behaviors one had learned to view as pathological. This lesson applies in this and many other contexts.

An important property of most well-known frequency distributions was already alluded to in this chapter, but is worth amplifying a bit here. Take N sample values U_n drawn independently from the same distribution, and let $N \rightarrow \infty$. The average of the U_n converges to a constant (this is the law of large numbers), and it is possible to weight the sum $\sum_{n=1}^{N} U_n$ in such a way that, as N increases, the distribution of the weighted sample sum tends to the normal or Gaussian distribution (this is central limit theorem.) Random variables exhibiting this property are said to be "attracted to the Gaussian," and I prefer the shorter term, "preGaussian."

Elementary textbooks prove those theorems for special cases, and stress the fact that attraction holds under much wider conditions. But there are exceptions. For this chapter's purposes, the most significant indirect criterion for being pre-Gaussian is the following. For every ΔU, the weighted sum of the quantities $U_m - \Delta U$ tends towards a Gaussian limit if, and only if, the largest *ex-post* contribution is asymptotically negligible compared to the sum. Here is a consequence expressed in the terms of this chapter: if the size distribution of firms were a lognormal, or any other distribution attracted by the Gaussian, the degree of concentration would tend to zero as the number of firms increases.

The acknowledged weakness of the preceding argument is that it concerns an asymptotic long-run, while "in the long-run we shall all be dead." Another maxim one must not forget is that "it is better to be approximately right than certifiably wrong." It is too much to expect asymptotics to yield a correct *quantitative* representation of the middle-run, but it is legitimate to expect it to yield a degree of *qualitative* understanding.

Now, what about the census data on the total number of firms in an industry and its relation to the degree of concentration? Those data suggest that concentration ratios vary little and in non-systematic fashion with changes in the number of firms in the industry. To argue that a preGaussian distribution is appropriate as a description of firm sizes, artificial modifications are required, or the argument becomes complicated transients.

For example, suppose we abandon the idea that firms are drawn at random and independently from a statistical population, and assume, instead, that the giants collude to divide the market among all firms. If so, *any* degree of concentration could be justified under lognormal or any other hypothesis. It may be difficult to accept the view that conspiracy is so widespread and effective. By contrast, it is not necessary to introduce such external determining forces if the statistical theory is founded upon a distribution with infinite first moment (practically a synonym for a scaling distribution with $\alpha < 1$.)

To sum up, change in industry size with unchanged concentration ratio is a consequence of a scaling distribution of firm sizes. Size distributions with finite population moments up to at least the second order predict a concentration that decreases with industry size. The unvarying concentration ratios are important, because they necessarily relate to oligopolistic market phenomena.

PART IV: THE 1963 MODEL OF PRICE CHANGE

My efforts to improve on Bachelier's Brownian *model started with markets on which the dominant factor is the highly nonGaussian nature of the distribution's tails. In IBM Report NC-87,* "The Variation of certain speculative prices, *published on March 26, 1962, the title pointedly meant* not necessarily all. *For external publication, the hefty NC-87 was split. Its core became M 1963b{E14}, which is the centerpiece of this part. My interests having changed, what was left of NC-87 appeared years later, as M 1967b{E15}; this awkward composite adds some data on cotton and continues with wheat, railroad securities and interest rates, but also includes answers to criticism of M 1963b{E14} and other material that made the text closer to being self-contained. A final small piece from NC-87 is published for the first time as Pre-publication Appendix I to Chapter E14.*

&&&&&&&&&&&&&&&&&&&&&&&&&

The Journal of Business, 36, 1963, 394-419 & 45, 1972, 542-3
Econometrica, 31, 1963, 757-758. Current Contents, 14, 1982, 20.

E14

The variation of certain speculative prices

✦ **Pre-publication abstract (M 1962a).** The classic model of the temporal variation of speculative prices (Bachelier 1900) assumes that successive changes of a price $Z(t)$ are independent Gaussian random variables. But, even if $Z(t)$ is replaced by $\log Z(t)$, this model is contradicted by facts in four ways, at least:

(1) Large price changes are *much more frequent* than predicted by the Gaussian; this reflects the "excessively peaked" ("leptokurtic") character of price relatives, which has been well-established since at least 1915.

(2) Large practically instantaneous price changes *occur often*, contrary to prediction, and it seems that they must be explained by causal rather than stochastic models.

(3) Successive price changes *do not* "look" independent, but rather exhibit a large number of recognizable patterns, which are, of course, the basis of the technical analysis of stocks.

(4) Price records *do not* look stationary, and statistical expressions such as the sample variance take very different values at different times; this nonstationarity seems to put a precise statistical model of price change out of the question.

I shall show that there is a simple way to solve difficulties (1), (2) and (4), and – to some extent – difficulty (3). This will imply that it is *not* necessary to give up the stationary stochastic models. Suppose indeed that the price relatives are so extremely leptokurtic (1), as to lead to infinite values for the population variance, and for other population moments beyond the first. This could – and indeed does – explain the erratic behavior of the sample moments (4), and the sample paths generated by such models would indeed by expected to include large discontinuities (2). Additionally, some features of the dependence between successive changes (3) could be taken into account by injecting a comparatively limited weakening asymptotic? of the hypothesis of independence; that is, "patterns" that have such a small probability in a Gaussian function that their occurrence by chance is practically impossible, now acquire a credibly large probability of occurring by chance.

As known in the case of the Cauchy distribution, having an infinite variance does not prevent a distribution from being quite proper, but it does make it quite peculiar. For example, the classical central limit theorem is inapplicable, and the largest of M addends is not negligibly small but rather provides an appreciable proportion of their sum. Fortunately, these peculiar consequences actually happen to describe certain well-known features of the behavior of prices.

The basic distribution with an infinite variance is scaling with an exponent between 1 and 2. My theory of prices is based upon distributions with two scaling tails, as well as upon L-stable distributions. The latter are akin to the scaling law, and appear in the first significant generalization of the classical central limit theorem. My theory is related to my earlier work on the distribution of personal income. ✦

I. INTRODUCTION

Louis Bachelier is a name mentioned in relation to diffusion processes in physics. Until very recently, however, few people realized that his path-breaking contribution, Bachelier 1900, was a by-product of the construction of a random-walk model for security and commodity markets. Let $Z(t)$ be the price of a stock, or of a unit of a commodity, at the end of time period t. Then, Bachelier's simplest and most important model assumes that successive differences of the form $Z(t+T)-Z(t)$ are independent Gaussian random variables with zero mean and with variance proportional to the differencing interval T.

That simplest model implicitly assumes that the variance of the differences $Z(t+T)-Z(t)$ is independent of the level of $Z(t)$. There is reason to expect, however, that the standard deviation of $\Delta Z(t)$ will be proportional to the price level, which is why many authors suggest that the original assumption of independent increments of $Z(t)$ be replaced by the assumption of independent and Gaussian increments of $\log_e Z(t)$.

Despite the fundamental importance of Bachelier's process, which has come to be called "Brownian motion," it is now obvious that it does not account for the abundant data accumulated since 1900 by empirical economists. Simply stated, *the empirical distributions of price changes are usually too "peaked" to be viewed as samples from Gaussian populations.* To the best of my knowledge, the first to note this fact was Mitchell 1915. But unquestionable proof was only given by Olivier 1926 and Mills 1927. Other evidence, regarding either $Z(t)$ or $\log Z(t)$, can be found in Larson 1960, Osborne 1959 and Alexander 1961.

That is, the histograms of price changes are indeed unimodal and their central "bells" are reminiscent of the "Gaussian ogive." But there are typically so many "outliers" that ogives fitted to the mean square of price changes are much lower and flatter than the distribution of the data themselves (see, Fig. 1). The tails of the distributions of price changes are in fact so extraordinarily long that the sample second moments typically vary in an erratic fashion. For example, the second moment reproduced in Figure 2 does not seem to tend to any limit even though the sample size is enormous by economic standards.

It is my opinion that these facts warrant a radically new approach to the problem of price variation in speculative markets. The purpose of this paper will be to present and test a new model that incorporates this belief. (A closely related approach has also proved successful in other contexts; see M 1963e{E3}. But I believe that each of the applications should stand on their own feet and I have minimized the number of cross references.

The model I propose begins like the Bachelier process as applied to $\log_e Z(t)$ instead of $Z(t)$. The major change is that I replace the Gaussian distribution throughout by "L-stable," probability laws which were first described in Lévy 1925. In a somewhat complex way, the Gaussian is a limiting case of this new family, so the new model is actually a generalization of that of Bachelier.

Since the L-stable probability laws are relatively unknown, I shall begin with a discussion of some of the more important mathematical properties of these laws. Following this, the results of empirical tests of the L-stable model will be examined. The remaining sections of the paper will

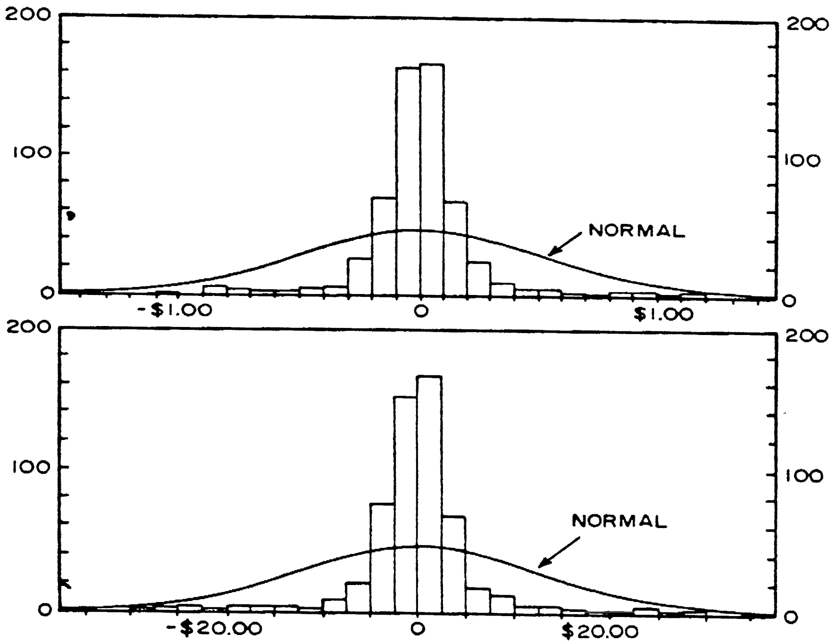


FIGURE E14-1. Two histograms illustrating departure from normality of the fifth and tenth difference of monthly wool prices, 1890-1937. In each case, the continuous bell-shaped curve represents the Gaussian "interpolate" from -3σ to 3σ based upon the sample variance. Source: Tintner 1940.

then be devoted to a discussion of some of the more sophisticated mathematical and descriptive properties of the L-stable model. I shall, in particular, examine its bearing on the very possibility of implementing the stop-loss rules of speculation.

II. MATHEMATICAL TOOLS: L-STABLE DISTRIBUTIONS

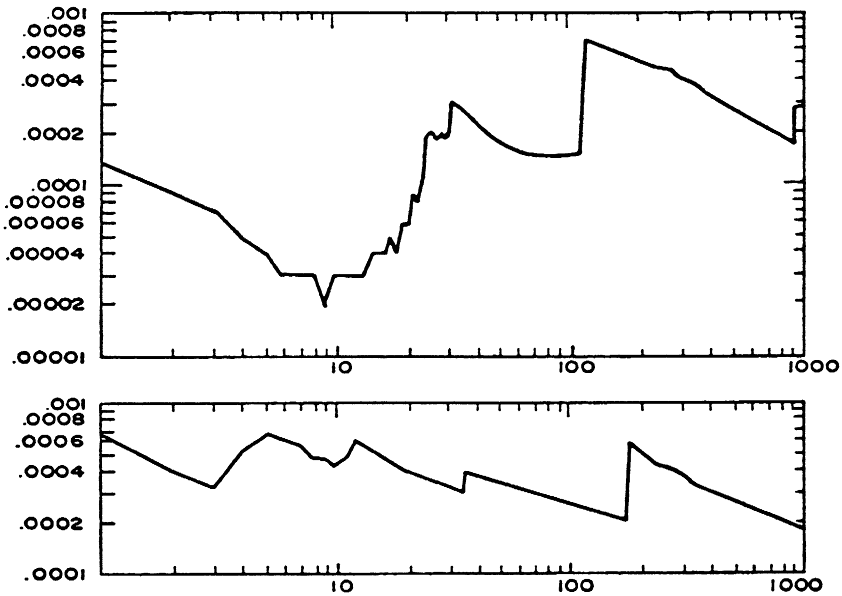


FIGURE E14-2. Both graphs represent the sequential variation of the sample second moment of cotton price changes. The horizontal scale represents time in days, with two different origins T_0. On the upper graph, T_0 was September 21, 1900; on the lower graph, T_0 was August 1, 1900. The vertical lines represent the value of the function

$$(T - T_0)^{-1} \sum_{t=T_0}^{t=T} [L(t, 1)]^2,$$

where $L(t, 1) = \log_e Z(t+1) - \log_e Z(t)$ and $Z(t)$ is the closing spot price of cotton on day t. I am grateful to the United States Department of Agriculture for making these data available.

II.A. "L-stability" of the Gaussian distribution and generalization of the concept of L-stability

One of the principal attractions of the modified Bachelier process is that the logarithmic relative

$$L(t, T) = \log_e Z(t + T) - \log_e Z(t),$$

is a Gaussian random variable for *every* value of T; the only thing that changes with T is the standard deviation of $L(t, T)$. This feature is the consequence of the following fact:

Let G' and G'' be two independent Gaussian random variables, of zero means and of mean squares equal to σ'^2 and σ''^2, respectively. Then the sum $G' + G''$ is also a Gaussian variable of mean square equal to $\sigma'^2 + \sigma''^2$. In particular, the "reduced" Gaussian variable, with zero mean and unit square, is a solution to

$$s'U + s''U = sU, \qquad \text{(S)}$$

where s is a function of s' and s'' given by the auxiliary relation

$$s^2 = s'^2 + s''^2. \qquad (\mathrm{A}_2)$$

It should be stressed that, from the viewpoint of the equation (S) and relation A_2, the quantities s', s'', and s are simply scale factors that "happen" to be closely related to the root-mean-square in the Gaussian case.

The property (S) expresses a kind of L-stability or invariance under addition, which is so fundamental in probability theory that it came to be referred to simply as *L-stability*. The Gaussian is the only solution of equation (S) for which the second moment is finite – or for which the relation A_2 is satisfied. When the variance is allowed to be infinite, however, (S) possesses many other solutions. This was shown constructively by Cauchy, who considered the random variable U for which

$$\Pr\{U > u\} = \Pr\{U < -u\} = 1/2 - (1/\pi)\tan^{-1} u,$$

so that its density is of the form

$$d \Pr\{U < u\} = \frac{1}{\pi(1 + u^2)}.$$

For this law, integral moments of all orders are infinite, and the auxiliary relation takes the form

$$(A_1) \qquad s = s' + s'',$$

where the scale factors s', s'', and s are not defined by any moment.

The general solution of equation (S) was discovered by Lévy 1925. (The most accessible source on these problems is, however, Gnedenko & Kolmogorov 1954.) The logarithm of its characteristic function takes the form

$$(L) \qquad \log \int_{-\infty}^{\infty} \exp(iuz) d \Pr\{U < u\} = i\delta z - \gamma |z|^{\alpha} \left\{ 1 + \frac{i\beta z}{|z|} \tan \frac{\alpha\pi}{2} \right\}.$$

It is clear that the Gaussian law and the law of Cauchy are stable and that they correspond to the cases ($\alpha = 2$; β arbitrary) and ($\alpha = 1$; $\beta = 0$), respectively.

Equation (L) determines a family of distribution and density functions $\Pr\{U < u\}$ and $d \Pr\{U < u\}$ that depend continuously upon four parameters. These four parameters also happen to play the roles the Pearson classification associates with the first four moments of U.

First of all, the α is an index of "peakedness" that varies in $]0, 2]$, that is, from 0 (excluded) to 2 (included). This α will turn out to be intimately related to the scaling exponent. The β is an index of "skewness" that can vary from -1 to $+1$, except that, if $\alpha = 1$, β must vanish. If $\beta = 0$, the stable densities are symmetric.

One can say that α and β together determine the "type" of a stable random variable. Such a variable can be called "reduced" if $\gamma = 1$ and $\delta = 0$. It is easy to see that, if U is reduced, sU is a stable variable with the same α, β and δ, and γ equal to s^{α}. This means that the third parameter, γ, is a scale factor raised to the power of α. Suppose now that U' and U'' are two independent stable variables, reduced and having the same values for α and β. It is well-known that the characteristic function of $s'U' + s''U''$ is the product of those of $s'U'$ and of $s''U''$. Therefore, the equation (S) is readily seen to be accompanied by the auxiliary relation

$$(A) \qquad s^{\alpha} = s'^{\alpha} + s''^{\alpha}.$$

More generally, suppose that U' and U'' are stable, have the same values of α, β and of $\delta = 0$, but have different values of γ (respectively, γ' and γ''), the sum $U' + U''$ is stable and has the parameters α, β, $\gamma = \gamma' + \gamma''$ and $\delta = 0$. Now recall the familiar property of the Gaussian distribution, that when two Gaussian variables are added, one must add their " variances." The variance is a mean-square and is the square of a scale factor. The role of a scale factor is now played by γ, and that of a variance by a scale factor raised to the power α.

The final parameter is δ; strictly speaking, equation (S) requires that $\delta = 0$, but we have added the term $i\delta z$ to (PL) in order to introduce a location parameter. If $1 < \alpha \leq 2$, so that $E(U)$ is finite, one has $\delta = E(U)$. If $\beta = 0$, so that the stable variable has a symmetric density function, δ is the median or modal value of U. But when $0 < \alpha < 1$, with $\beta \neq 0$, δ has no obvious interpretation.

II.B. Addition of more than two stable random variables

Let the independent variables U_n satisfy the condition (PL) with values of α, β, γ, and δ equal for all n. The logarithm of the characteristic function of

$$S_N = U_1 + U_2 + \ldots U_n + \ldots U_N$$

is N times the logarithm of the characteristic function of U_n, and equals

$$i\,\delta N z - N\gamma |z|^{\alpha}[1 + i\beta(z/|z|)\tan(\alpha\pi/2)].$$

Thus S_N is stable with the same α and β as U_n, and with parameters δ and γ multiplied by N. It readily follows that

$$U_n - \delta \text{ and } N^{-1/\alpha}\sum_{n=1}^{N} U_n - \delta$$

have identical characteristic functions and thus are identically distributed random variables. (This is, or course, a most familiar fact in the Gaussian case, $\alpha = 2$.)

The generalization of the classical "$T^{1/2}$ law." In the Gaussian model of Bachelier, in which daily increments of $Z(t)$ are Gaussian with the standard deviation $\sigma(1)$, the standard deviation of $\Delta Z(t)$, where Δ is taken over T days, is equal to $\sigma(T) = T^{1/2}\sigma(1)$.

The corresponding prediction of my model is as follows: Consider any scale factor such as the intersextile range, that is, the difference between the quantity U^{+} which is exceeded by one-sixth of the data, and the quantity U^{-} which is larger than one-sixth of the data. It is easily found that the expected range satisfies

$$E[U^{+}(T) - U^{-}(T)] = T^{1/\alpha}E[U^{+}(1) - U^{-}(1)].$$

We should also expect that the deviations from these expectations exceed those observed in the Gaussian case.

Differences between successive means of $Z(t)$. In all cases, the average of $Z(t)$, taken over the time span $t^0 + 1$ to $t^0 + N$, can be written as:

$$(1/N)[Z(t^0 + 1) + Z(t^0 + 2) + ...Z(t^0 + N)]$$
$$= (1/N)\{N\, Z(t^0 + 1) + (N - 1)[Z(t^0 + 2) - Z(t^0 + 1)] + ...$$
$$+ (N - n)[Z(t^0 + n + 1) - Z(t^0 + n)] + ...[Z(t^0 + N) - Z(t^0 + N - 1)]\}.$$

To the contrary, let the average over the time span $t^0 - N + 1$ to t^0 be written as

$$(1/N)\{N\, Z(t^0) + (N - 1)[Z(t^0) - Z(t^0 - 1)]...$$
$$+ (N - n)[Z(t^0 - n + 1) - Z(t^0 - n)]...$$
$$+ [Z(t^0 - N + 2) - Z(t^0 - N + 1)]\}.$$

Thus, if the expression $Z(t + 1) - Z(t)$ is a stable variable $U(t)$ with $\delta = 0$, the difference between successive means of values of Z is given by

$$U(t^0) + [(N - 1)/N][U(t^0 + 1) + U(t^0 - 1)]$$
$$+ ...[(N - n)/N][U(t^0 + n) + U(t^0 - n)]$$
$$+ ...[U(t^0 + N - 1)...U(t^0 - N + 1)].$$

This is clearly a stable variable, with the same α and β as the original U, and with a scale parameter equal to

$$\gamma^0(N) = [1 + 2(N-1)^\alpha N^{-\alpha} + \ldots 2(N-n)^\alpha N^{-\alpha} + \ldots + 2]\gamma(U).$$

As $N \to \infty$, one has

$$\frac{\gamma^0(N)}{\gamma(U)} \to \frac{2N}{(\alpha+1)},$$

whereas a genuine monthly change of $Z(t)$ has a parameter $\gamma(N) = N\gamma(U)$. Thus, the effect of averaging is to multiply γ by the expression $2/(\alpha+1)$, which is smaller than 1 if $\alpha > 1$.

III.C. L-stable distributions and scaling

Except for the Gaussian limit case, the densities of the stable random variables follow a generalization of the asymptotic behavior of the Cauchy law. It is clear, for example, that as $u \to \infty$, the Cauchy density behaves as follows:

$$u \Pr\{U > u\} = u \Pr\{U < -u\} \to 1/\pi.$$

More generally, Lévy has shown that the tails of *all* nonGaussian stable laws follow an asymptotic form of scaling. There exist two constants, $C' = \sigma'^\alpha$ and $C'' = \sigma''^\alpha$, linked by $\beta = (C' - C'')/(C' + C'')$, such that,

$$\text{when } u \to \infty,\ u^\alpha \Pr\{U > u\} \to C' = \sigma'^\alpha \text{ and } u^\alpha \Pr\{U < -u\} \to C'' = \sigma''^\alpha.$$

Hence, *both* tails are scaling if $|\beta| \neq 1$, a solid reason for replacing the term "stable nonGaussian" by the less negative one of "*L-stable.*" The two numbers σ' and σ'' share the role of the standard deviation of a Gaussian variable. They will be denoted as the "standard positive deviation" and the "standard negative deviation," respectively.

Now consider the two extreme cases: when $\beta = 1$, hence $C'' = 0$, and when $\beta = -1$, hence $C' = 0$). In those cases, one of the tails (negative and positive, respectively) decreases faster than the scaling distribution of index α. In fact, one can prove (Skorohod 1954-1961) that the short tail withers away even faster that the Gaussian density so that the extreme cases of stable laws are, for all practical purposes, J-shaped. They play an important role in my theory of the distributions of personal income and of city sizes. A number of further properties of L-stable laws may therefore be found in my publications devoted to these topics. See M 1960i{E10}, 1963p{E11} and 1962g{E12}.

II.D. The L-stable variables as the only possible limits of weighted sums of independent, identically distributed addends

The L-stability of the Gaussian law can be considered to be only a matter of convenience, and it often thought that the following property is more important.

Let the U_n be independent, identically distributed random variables, with a finite $\sigma^2 = E[U_n - E(U)]^2$. Then the classical central limit theorem asserts that

$$\lim_{N \to \infty} N^{-1/2}\sigma^{-1}\sum_{n=1}^{N}[U_n - E(U)]$$

is a reduced Gaussian variable.

This result is, of course, the basis of the explanation of the presumed occurrence of the Gaussian law in many practical applications relative to sums of a variety of random effects. But the essential thing in all these aggregative arguments is not that $\Sigma[U_n - E(U)]$ is weighted by any special factor, such as $N^{-1/2}$, but rather that the following is true:

There exist two functions, $A(N)$ and $B(N)$, such that, as $N \to \infty$, the weighted sum

(L) $$A(N)\sum_{n=1}^{N} U_n - B(N),$$

has a limit that is finite and is not reduced to a nonrandom constant.

If the variance of U_n is not finite, however, condition (L) may remain satisfied while the limit ceases to be Gaussian. For example, if U_n is stable nonGaussian, the linearly weighted sum

$$N^{-1/\alpha}\sum(U_n - \delta)$$

was seen to be *identical in law* to U_n, so that the "limit" of that expression is already attained for $N = 1$ and a stable nonGaussian law. Let us now suppose that U_n is asymptotically scaling with $0 < \alpha < 2$, but not stable. Then the limit exists, and it follows the L-stable law having the same value of α. As in the L-stability argument, the function $A(N)$ can be chosen equal to $N^{-1/\alpha}$. These results are crucial but I had better not attempt to rederive them here. The full mathematical argument is available in the literature. I have constructed various heuristic arguments to buttress it. But experience shows that an argument intended to be illuminating often comes across as basing far-reaching conclusions on loose thoughts. Let me therefore just quote the facts:

The Doeblin-Gnedenko conditions. The problem of the existence of a limit for $A(N)\Sigma U_n - B(N)$ can be solved by introducing the following generalization of asymptotic scaling (Gnedenko & Kolmogorov 1954). Introduce the notations

$$\Pr\{U > u\} = Q'(u)u^{-\alpha}; \ \Pr\{U < -u\} = Q''(u)u^{-\alpha}.$$

The term *Doeblin-Gnedenko condition* will denote the following statements: (a) when $u \rightarrow \infty$, $Q'(u)/Q''(u)$ tends to a limit C'/C''; (*b*) there exists a value of $\alpha > 0$ such that for every $k > 0$, and for $u \rightarrow \infty$, one has

$$\frac{Q'(u) + Q''(u)}{Q'(ku) + Q''(ku)} \rightarrow 1.$$

These conditions generalize the scaling distribution, for which $Q'(u)$ and $Q''(u)$ themselves tend to limits as $u \rightarrow \infty$. With their help, and unless $\alpha = 1$, the problem of the existence of weighting factors $A(N)$ and $B(N)$ is solved by the following theorem:

If the U_n are independent, identically distributed random variables, there may exist no functions $A(N)$ and $B(N)$ such that $A(N)\Sigma U_n - B(N)$ tends to a proper limit. But, if such functions $A(N)$ and $B(N)$ exist, one knows that the limit is one of the solutions of the L-stability equation (S). *More precisely, the limit is Gaussian if, and only if, the U_n has finite variance; the limit is nonGaussian if, and only if, the Doeblin-Gnedenko conditions are satisfied for some $0 < \alpha < 2$. Then, $\beta = (C' - C'')/(C' + C'')$ and $A(N)$ is determined by the requirement that*

$$N \Pr\{U > uA^{-1}(N)\} \rightarrow C'u^{-\alpha}.$$

(For all values of α, the Doeblin-Gnedenko condition (b) also plays a central role in the study of the distribution of the random variable $\max U_n$.)

As an application of the above definition and theorem, let us examine the product of two independent, identically distributed scaling (but not stable) variables U' and U''. First of all, for $u > 0$, one can write

$$\Pr\{U'U'' > u\} = \Pr\{U' > 0;\ U'' > 0;\ \text{and} \log U' + \log U'' > \log u\}$$
$$+ \Pr\{U' < 0;\ U'' < 0;\ \text{and}\ \log |U'| + \log |U''| > \log u\}.$$

But it follows from the scaling distribution that

$$\Pr\{U > e^z\} \sim C' \exp(-\alpha z) \text{ and } \Pr\{U < -e^z\} \sim C'' \exp(-\alpha z),$$

where U is either U' or U''. Hence, the two terms P' and P'' that add up to $\Pr\{U'U'' > u\}$ satisfy

$$P' C'^2 \alpha z \exp(-\alpha z) \text{ and } P'' C''^2 \alpha z \exp(-\alpha z).$$

Therefore,

$$\Pr\{U'U'' > u\} \sim \alpha (C'^2 + C''^2)(\log_e u) u^{-\alpha}.$$

Similarly,

$$\Pr\{U'U'' < -u\} \sim \alpha 2C'C''(\log_e u) u^{-\alpha}.$$

It is obvious that the Doeblin-Gnedenko conditions are satisfied for the functions $Q'(u) \sim (C'^2 + C''^2)\alpha \log_e u$ and $Q''(u) \sim 2C'C''\alpha \log_e u$. Hence the weighted expression

$$(N \log N)^{-1/\alpha} \sum_{n=1}^{N} U'_n U''_n$$

converges toward a L-stable limit with the exponent α and the skewness

$$\beta = \frac{C'^2 + C''^2 - 2C'C''}{C'^2 + C''^2 + 2C'C''} = \left[\frac{C' - C''}{C' + C''} \right]^2 \geq 0.$$

In particular, the positive tail should always be bigger than the negative tail.

II.E. Shape of the L-stable distributions outside the asymptotic range

There are closed expressions for three cases of L-stable densities: Gauss ($\alpha = 2, \beta = 0$), Cauchy ($\alpha = 1, \beta - 0$), and third, ($\alpha = 1/2; \beta = 1$). In every other case, we only know the following: (a) the densities are always unimodal; (b) the densities depend continuously upon the parameters; (c) if $\beta > 0$, the positive tail is the fatter – hence, if the mean is finite (i.e., if $1 < \alpha < 2$), it is greater than the median.

To go further, I had to resort to numerical calculations. Let us, however, begin by interpolative arguments.

The symmetric cases, $\beta = 0$. For $\alpha = 1$, one has the Cauchy density $[\pi(1 + u^2)]^{-1}$. It is always *smaller* than the scaling density $1/\pi u^2$ toward which it converges as $u \rightarrow \infty$. Therefore, $\Pr\{U > u\} < 1/\pi\, u$, and it follows that for $\alpha = 1$, the doubly logarithmic graph of $\log_e[\Pr\{U > u\}]$ is entirely on the left side of its straight asymptote. By continuity, the same shape must appear when α is only a little higher or a little lower than 1.

For $\alpha = 2$, the doubly logarithmic graph of the Gaussian $\log_e \Pr\{(U > u)\}$ drops very quickly to negligible values. Hence, again by continuity, the graph must also begin with decreasing rapidly when α is just below 2. But, since its ultimate slope is close to 2, it must have a point of inflection corresponding to a maximum slope greater than 2, and it must begin by "overshooting" its straight asymptote.

Interpolating between 1 and 2, we see that there exists a smallest value of α, call it $\tilde{\alpha}$, for which the doubly logarithmic graph begins by overshooting its asymptote. In the neighborhood of $\tilde{\alpha}$, the asymptotic α can be measured as a slope even if the sample is small. If $\alpha < \tilde{\alpha}$, the asymptotic slope will be underestimated by the slope of small samples; for $\alpha > \tilde{\alpha}$ it will be overestimated. The numerical evaluation of the densities yields a value of $\tilde{\alpha}$ in the neighborhood of 1.5. A graphical presentation of the results of this section is given in Figure 3.

The skew cases. If the positive tail is fatter than the negative one, it may well happen that the doubly logarithmic graph of the positive tail begins by overshooting its asymptote, while the doubly logarithmic graph of the negative tail does not. Hence, there are two critical values of α^0, one for each tail. If the skewness is slight, if α lies between the critical values, and if the sample size is not large enough, then the graphs of the two tails will have slightly different overall apparent slopes.

II.F. Joint distribution of independent L-stable variables

Let $p_1(u_1)$ and $p_2(u_2)$ be the densities of U_1 and of U_2. If both u_1 and u_2 are large, the joint probability density is given by

$$p^0(u_1, u_2) = \alpha C'_1 u_1^{-(\alpha+1)} \alpha C'_2 u_2^{-(\alpha+1)} = \alpha^2 C'_1 C'_2 (u_1 u_2)^{-(\alpha+1)}.$$

The lines of equal probability belong to hyperbolas $u_1 u_2 = \text{constant}$. They link together as in Figure 4, into fattened signs +. Near their maxima, $\log_e p_1(u_1)$ and $\log_e p_2(u_2)$ are approximated by $\alpha_1 - (u_1/b_1)^2$ and $\alpha_2 - (u_2/b_2)^2$. Hence, the probability isolines are of the form

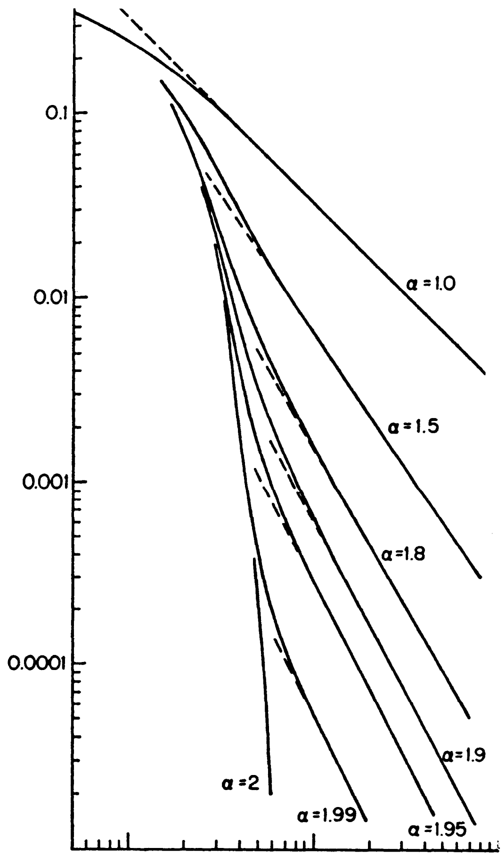


FIGURE E14-3. The various lines are doubly logarithmic plots of the symmetric L-stable probability distributions with $\delta = 0$, $\gamma = 1$, $\beta = 0$ and α as marked. Horizontally: $\log_e u$; vertically: $\log_e \Pr\{U > u\} = \log_e \Pr\{U < -u\}$. Sources: unpublished tables based upon numerical computations performed at the author's request by the IBM T. J. Watson Research Center.

$$(u_1/b_1)^2 + (u_2/b_2)^2 = \text{constant}.$$

The transition between the ellipses and the "plus signs" is, of course, continuous.

II.G. Distribution of U_1 when U_1 and U_2 are independent L-stable variables and $U_1 + U_2 = U$ is known

This conditional distribution can be obtained as the intersection between the surface that represents the joint density $p_0(u_1, u_2)$ and the plane $u_1 + u_2 = u$. Thus, the conditional distribution is unimodal for small u. For large u, it has two sharply distinct maxima located near $u_1 = 0$ and near $u_2 = 0$.

More precisely, the conditional density of U_1 is given by $p_1(u_1)p_2(u - u_1)/q(u)$, where $q(u)$ is the density of $U = U_1 + U_2$. Let u be positive and very large; if u_1 is small, one can use the scaling approximations for $p_2(u_2)$ and $q(u)$, obtaining

$$\frac{p_1(u_1)p_2(u - u_1)}{q(u)} \sim \frac{C'_1}{C'_1 + C'_2} p_1(u_1).$$

If u_2 is small, one similarly obtains

$$\frac{p_1(u_1)p_2(u - u_1)}{q(u)} \sim \frac{C'_2}{(C'_1 + C'_2)} p_2(u - u_1).$$

In other words, the conditional density $p_1(u_1)p_2(u - u_1)/q(u)$ looks as if two unconditioned distributions, scaled down in the ratios $C'_1/(C'_1 + C'_2)$ and $C'_2/(C'_1 + C'_2)$, had been placed near $u_1 = 0$ and $u_1 = u$. If u is negative, but very large in absolute value, a similar result holds with C''_1 and C''_2 replacing C'_1 and C'_2.

For example, for $\alpha = 2 - \varepsilon$ and $C'_1 = C'_2$, the conditional distribution is made up of two almost Gaussian bells, scaled down to one-half of their height. But, as α tends toward 2, these two bells become smaller and a third bell appears near $u_1 = u/2$. Ultimately, the two side bells vanish, leaving a single central bell. This limit corresponds to the fact that when the sum $U_1 + U_2$ is known, the conditional distribution of a Gaussian U_1 is itself Gaussian.

III. EMPIRICAL TESTS OF THE L-STABLE LAWS: COTTON PRICES

This section has two different aims. From the viewpoint of statistical economics, its purpose is to motivate and develop a model of the variation of speculative prices based on the L-stable laws discussed in the previous section. From the viewpoint of statistics considered as the theory of data analysis, it shows how I use the theorems concerning the sums $\sum U_n$ to build a new test of the scaling distribution. Before moving on to the main points of the section, however, let us examine two alternative ways of handling the large price changes which occur in the data with frequencies not accounted for by the normal distribution.

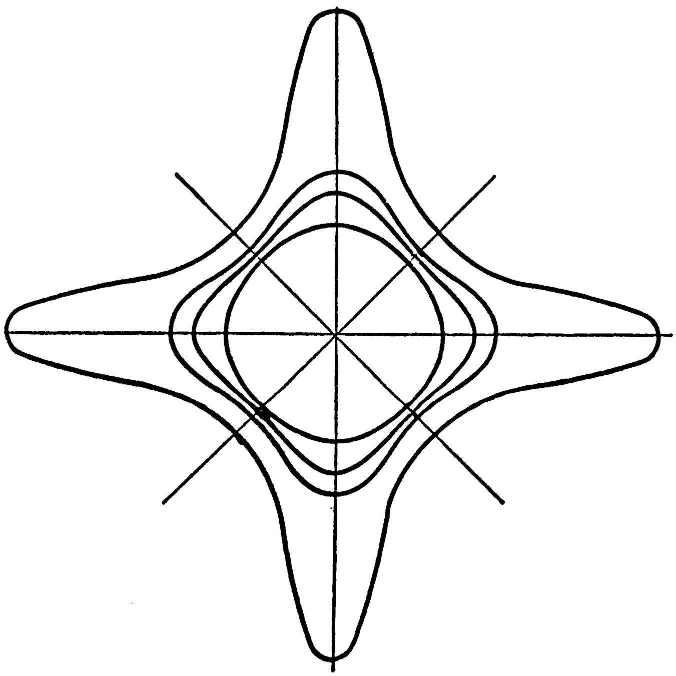


FIGURE E14-4. Joint distribution of successive price relatives $L(t, 1)$ and $L(t+1, 1)$.

If $L(t, 1)$ and $L(t+1, 1)$ are independent, their values should be plotted along the horizontal and vertical coordinates axes.

If $L(t, 1)$ and $L(t+1, 1)$ are linked by the model in Section VII, their values should be plotted along the bisectors, or else the figure should be rotated by 45°, before $L(t, 1)$ and $L(t+1, 1)$ are plotted along the coordinate axes.

III.A. Explanation of large price changes as due to causal or random "contaminators"

One very common approach is to note that, a posteriori, large price changes are usually traceable to well-determined "causes," and should be eliminated before one attempts a stochastic model of the remainder. Such preliminary censorship obviously brings any distribution closer to the Gaussian. This is, for example, what happens when the study is limited to "quiet periods" of price change. Typically, however, no discontinuity is observed between the "outliers" and the rest of the distribution. In such cases, the notion of outlier is indeterminate and arbitrary. above censorship is therefore usually indeterminate.

Another popular and classical procedure assumes that observations are generated by a mixture of two normal distributions, one of which has a small weight but a large variance and is considered as a random "contaminator." In order to explain the sample behavior of the moments, it unfortunately becomes necessary to introduce a larger number of contaminators, and the simplicity of the model is destroyed.

III.B. Introduction of the scaling distribution to represent price changes

I propose to explain the erratic behavior of sample moments by assuming that the corresponding population moments are infinite. This is an approach that I used successfully in a number of other applications and which I explained and demonstrated in detail elsewhere.

In practice, the hypothesis that moment are infinite beyond some threshold value is hard to distinguish from the scaling distribution. Assume that the increment, for example,

$$L(t, 1) = \log_e Z(t+1) - \log_e Z(t)$$

is a random variable with infinite population moments beyond the first. This implies that $\int p(u)\, u^2 du$ diverges but $\int p(u)\, u du$ converges (the integrals being taken all the way to infinity). It is of course natural, at least in the first stage of heuristic motivating argument, to assume that $p(u)$ is somehow "well-behaved" for large u. If so, our two requirements mean that, as $u \to \infty$, $p(u)u^3$ tends to infinity and $p(u)u^2$ tends to zero.

In other words: $p(u)$ must somehow decrease faster than u^{-2} and slower than u^{-3}. The simplest analytical expressions of this type are asymptotically scaling. *This observation provided the first motivation of the* *present study*. It is surprising that I could find no record of earlier application of the scaling distribution to two-tailed phenomena.

My further motivation was more theoretical. Granted that the facts impose a revision of Bachelier's process, it would be simple indeed if one could at least preserve the following convenient feature of the Gaussian model. Let the increments,

$$L(t, T) = \log_e Z(t + T) - \log_e Z(t),$$

over days, weeks, months, and years. In the Gaussian case, they would have different scale parameters, but the same distribution. This distribution would also rule the fixed-base relatives. This naturally leads directly to the probabilists' concept of L-stability examined in Section II.

In other words, the facts concerning moments, together with a desire for a simple representation, led me to examine the logarithmic price relatives (for unsmoothed and unprocessed time series relative to very active speculative markets), and check whether or not they are L-stable. Cotton provided a good example, and the present paper will be limited to the examination of that case.

Additional studies. My theory also applies to many other commodities (such as wheat and other edible grains), to many securities (such as those of the railroads in their nineteenth-century heyday), and to interest rates such as those of call or time money. These examples were mentioned in my IBM Research Note NC-87 (dated March 26, 1962). Later papers {P.S. 1996: see M 1967j{E15}} shall discuss these examples, describe some properties of cotton prices that my model fails to predict correctly and deal with cases when few "outliers" are observed. It is natural in these cases to favor Bachelier's Gaussian model – a limiting case in my theory as well as its prototype.

III.C. Graphical method applied to cotton price changes

Let us first describe Figure 5. The horizontal scale u of lines 1*a*, 1*b*, and 1*c* is marked only on lower edge, and the horizontal scale u of lines 2*a*, 2*b*, and 2*c* is marked along the upper edge.

The vertical scale gives the following relative frequencies:

$$\text{(A)} \quad \begin{cases} (1a) & \text{Fr}\,\{\log_e Z(t + \text{ one day}) - \log_e Z(t) > u\}, \\ (2a) & \text{Fr}\,\{\log_e Z(t + \text{ one day}) - \log_e Z(t) < -u\}, \end{cases}$$

both for the daily closing prices of cotton in New York, 1900-1905. (Source: the United Stated Department of Agriculture.)

(B)
$$\begin{cases} (1b) & \text{Fr}\,\{\log_e Z(t+\text{ one day}) - \log_e Z(t) > u\}, \\ (2b) & \text{Fr}\,\{\log_e Z(t+\text{ one day}) - \log_e Z(t) < -u\}, \end{cases}$$

both for an index of daily closing prices of cotton in the United States, 1944-58. (Source: private communication from Hendrick S. Houthakker.)

(C)
$$\begin{cases} (1c) & \text{Fr}\,\{\log_e Z(t+\text{ one month}) - \log_e Z(t) > u\}, \\ (2c) & \text{Fr}\,\{\log_e Z(t+\text{ one month}) - \log_e Z(t) < -u\}, \end{cases}$$

both for the closing prices of cotton on the 15th of each month in New York, 1880-1940. (Source: private communication from the United States Department of Agriculture.)

The theoretical $\log \Pr\{U > u\}$, relative to $\delta = 0$, $\alpha = 1.7$, and $\beta = 0$, is plotted as a solid curve on the same graph for comparison.

If it were true that the various cotton prices are L-stable with $\delta = 0$, $\alpha = 1.7$ and $\beta = 0$, the various graphs should be horizontal translates of each other. To ascertain that, on cursory examination, the data are in close conformity with the predictions of my model, the reader is advised to proceed as follows: copy on a transparency the horizontal axis and the theoretical distribution and to move both horizontally until the theoretical curve is superimposed on one or another of the empirical graphs. The only discrepancy is observed for line *2b*; it is slight and would imply an even greater departure from normality.

A closer examination reveals that the positive tails contain systematically fewer data than the negative tails, suggesting that β actually takes a small negative value. This is confirmed by the fact that the negative tails, but not the positive, begin by slightly "overshooting" their asymptote, creating the expected bulge.

III.D. Application of the graphical method to the study of changes in the distribution across time

Let us now look more closely at the labels of the various series examined in the previous section. Two of the graphs refer to daily changes of cotton prices, near 1900 and 1950, respectively. It is clear that these graphs do not coincide, but are horizontal translates of each other. This implies that between 1900 and 1950, the generating process has changed only to the extent that its scale γ has become much smaller.

Our next test will concern monthly price changes over a longer time span. It would be best to examine the actual changes between, say, the middle of one month and the middle of the next. A longer sample is available, however, when one takes the reported monthly *averages* of the price of cotton; the graphs of Figure 6 were obtained in this way.

If cotton prices were indeed generated by a stationary stochastic process, our graphs should be straight, parallel, and uniformly spaced. However, each of the 15-year subsamples contains only 200-odd months, so that the separate graphs cannot be expected to be as straight as those relative to our usual samples of 1,000-odd items. The graphs of Figure 6 are, indeed, not quite as neat as those relating to longer periods; but, in the absence of accurate statistical tests, they seem adequately straight and uniformly spaced, except for the period of 1880-96.

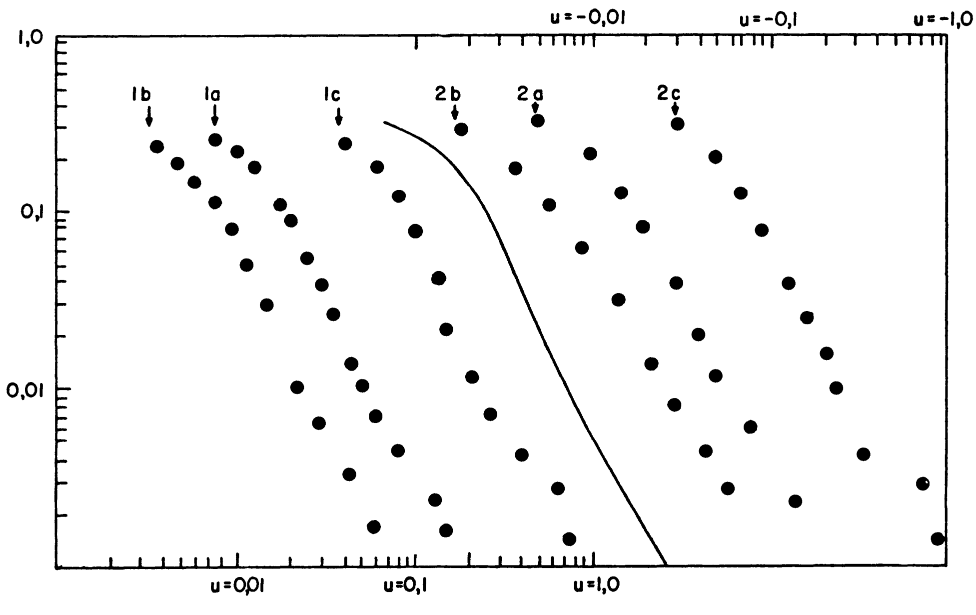


FIGURE E14-5. Composite of doubly logarithmic graphs of positive and negative tails for three kinds of cotton price relatives, together with a plot of the cumulated density function of a stable distribution.

I conjecture therefore, that, since 1816, the process generating cotton prices has changed only in its scale, with the possible exception of the periods of the Civil War and of controlled or supported prices. Long series of monthly price changes should therefore be represented by *mixtures* of L-stable laws; such mixtures remain scaling. See M 1963e{E3}.

III.E. Application of the graphical method to study effects of averaging

It is, of course, possible to derive mathematically the expected distribution of the changes between successive monthly means of the highest and lowest quotation; but the result is so cumbersome as to be useless. I have, however, ascertained that the empirical distribution of these changes does not differ significantly from the distribution of the changes between the monthly means, obtained by averaging all the daily closing quotations over a month. One may, therefore, speak of a single average price for each month.

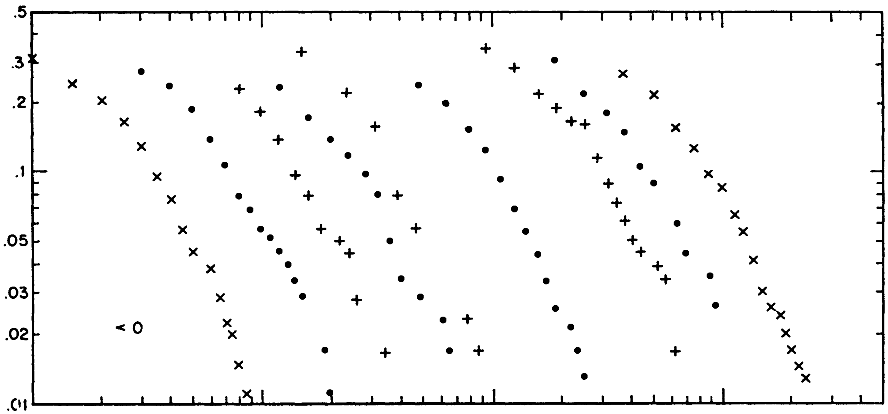


FIGURE E14-6. A rough test of stationarity for the process of change of cotton prices between 1816 and 1940. The horizontal axis displays negative changes between successive monthly *averages*. (Source: *Statistical Bulletin No. 99 of the Agricultural Economics Bureau, United States Department of Agriculture.*) To avoid interference between the various graphs, the horizontal scale of the kth graph from the left was multiplied by 2^{k-1}.) The vertical axis displays relative frequencies Fr $(U < -u)$ corresponding respectively to the following periods (from left to right): 1816-60, 1816-32, 1832-47, 1847-61, 1880-96, 1896-1916, 1916-31, 1880-1940.

Moving on to Figure 7, we compare the distribution of the averages with that of actual monthly values. We see that, overall, they only differ by a horizontal translation to the left, as predicted in Section IIC. Actually, in order to apply the argument of that section, it would be necessary to rephrase it by replacing $Z(t)$ by $\log_e Z(t)$ throughout. However, the geometric and arithmetic averages of daily $Z(t)$ do not differ much in the case of medium-sized overall monthly changes of $Z(t)$.

But the largest changes between successive averages are smaller than predicted. This seems to suggest that the dependence between successive daily changes has less effect upon actual monthly changes than upon the regularity with which these changes are performed. {P.S. 1996: see Appendix I of this chapter.}

III.F. A new presentation of the evidence

I will now show that the evidence concerning daily changes of cotton price strengthens the evidence concerning monthly changes, and conversely.

The basic assumption of my argument is that successive daily changes of log (price) are independent. (This argument will thus have to be revised when the assumption is improved upon.) Moreover, the population second moment of $L(t)$ seems to be infinite, and the monthly or yearly price changes are patently nonGaussian. Hence, the problem of whether any limit theorem whatsoever applies to $\log_e Z(t+T) - \log_e Z(t)$ can also be answered *in theory* by examining whether the daily changes satisfy the Pareto-Doeblin-Gnedenko conditions. *In practice,* however, it is impossible to attain an infinitely large differencing interval T, or to ever verify any condition relative to an infinitely large value of the random variable u. Therefore, one must consider that a month or a year is infinitely long, and that the largest observed daily changes of $\log_e Z(t)$ are infinitely large. Under these circumstances, one can make the following inferences.

Inference from aggregation. The cotton price data concerning daily changes of $\log_e Z(t)$ appear to follow the weaker asymptotic? condition of Pareto-Doeblin-Gnedenko. Hence, from the property of L-stability, and according to Section IID, one should expect to find that, as T increases,

$$T^{-1/\alpha}\{\log_e Z(t+T) - \log_e Z(t) - T\,E[L(t,1)]\}$$

tends towards a L-stable variable with zero mean.

Inference from disaggregation. Data seem to indicate that price changes over weeks and months follow the same law, except for a change of scale. This law must therefore be one of the possible nonGaussian limits, that is, it must be L-stable. As a result, the inverse part of the theorem of Section IID shows that the daily changes of $\log Z(t)$ must satisfy the Doeblin-Gnedenko conditions. (The inverse D-G condition greatly embarrassed me in my work on the distribution of income. It is pleasant to see that it can be put to use in the theory of prices.) A few of the difficulties involved in making the above two inferences will now be discussed.

Disaggregation. The D-G conditions are less demanding than asymptotic scaling because they require that limits exist for $Q'(u)/Q''(u)$ and for $[Q'(u)+Q''(u)]/[Q'(ku)+Q''(ku)]$, but not for $Q'(u)$ and $Q''(u)$ taken separately. Suppose, however, that $Q'(u)$ and $Q''(u)$ still vary a great deal in the useful range of large daily variations of prices. In this case, $A(N)\Sigma U_n - B(N)$ will not approach its own limit until *extremely* large

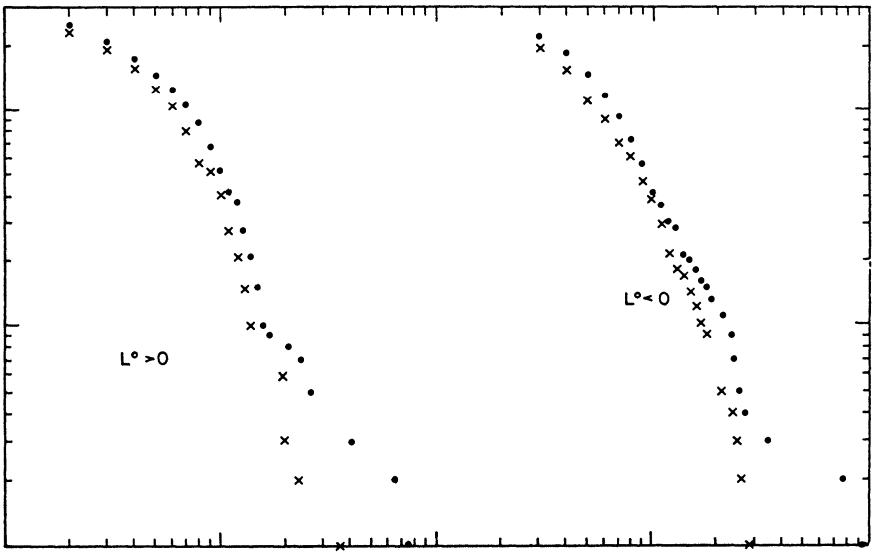


FIGURE E14-7. These graphs illustrate the effect of averaging. Dots reproduce the same data as the lines 1*c* and 2*c* of Figure 5. The × 's reproduce distribution of $\log_e Z^0(t+1) - \log_e Z^0(t)$, where $Z^0(t)$ is the average spot price of cotton in New York during the month *t*, as reported in the *Statistical Bulletin* No. 99 of the Agricultural Economics Bureau, United States Department of Agriculture.

values of N are reached. Therefore, if one believes that the limit is rapidly attained, the functions $Q'(u)$ and $Q''(u)$ of daily changes must vary very little in the tails of the usual samples. In other words, it is necessary, after all, that daily price changes be asymptotically scaling.

Aggregation. Here, the difficulties are of a different order. From the mathematical viewpoint, the L-stable law should become increasingly accurate as T increases. Practically, however, there is no sense in even considering values of T as long as a century, because one cannot hope to get samples sufficiently long to have adequately inhabited tails. The year is an acceptable span for certain grains, but here the data present other problems. The long available yearly series do not consist of prices actually quoted on some market on a fixed day of each year, but are averages. These averages are based on small numbers of quotations, and are obtained by ill-known methods that are bound to have varied in time. From the viewpoint of economics, two much more fundamental difficulties arise for very large T. First of all, the model of independent daily L's eliminates from consideration every "trend," except perhaps the exponential growth or decay due to a nonvanishing δ. Many trends that are negligible on the daily basis would, however, be expected to be predominant on the monthly or yearly basis. For example, the effect of weather upon yearly changes of agricultural prices might be very different from the simple addition of speculative daily price movements.

The second difficulty lies in the "linear" character of the aggregation of successive L's used in my model. Since I use natural logarithms, a small $\log_e Z(t+T) - \log_e Z(t)$ will be indistinguishable from the relative price change $[Z(t+T) - Z(t)]/Z(t)$. The addition of small L's is therefore related to the so-called "principle of random proportionate effect." It also means that the stochastic mechanism of prices readjusts itself immediately to any level that $Z(t)$ may have attained. This assumption is quite usual, but very strong. In particular, I shall show that if one finds that $\log Z(t + \text{ one week}) - \log Z(t)$ is very large, it is very likely that it differs little from the change relative to the single day of most rapid price variation (see Section VE); naturally, this conclusion only holds for independent L's. As a result, the greatest of N successive daily price changes will be so large that one may question both the use of $\log_e Z(t)$ and the independence of the L's.

There are other reasons (see Section IVB) to expect to find that a simple addition of speculative daily price changes predicts values too high for the price changes over periods such as whole months.

Given all these potential difficulties, I was frankly astonished by the quality of the prediction of my model concerning the distribution of the changes of cotton prices between the fifteenth of one month and the fifteenth of the next. The negative tail has the expected bulge, and even the most extreme changes of price can be extrapolated from the rest of the curve. Even the artificial excision of the Great Depression and similar periods would not affect the results very greatly.

It was therefore interesting to check whether the ratios between the scale coefficients, $C'(T)/C'(1)$ and $C''(T)/C''(1)$, were both equal to T, as predicted by my theory whenever the ratios of standard deviations $\sigma'(T)/\sigma'(s)$ and $\sigma''(T)/\sigma''(s)$ follow the $T^{1/\alpha}$ generalization of the "$T^{1/2}$ Law," which was referred to in Section IIB. If the ratios of the C parameters are different from T, their values may serve as a measure of the degree of dependence between successive $L(t, 1)$.

The above ratios were absurdly large in my original comparison between the daily changes near 1950 of the cotton prices collected by H. Houthakker, and the monthly changes between 1880 and 1940 of the prices given by the USDA. This suggested that the price varied less around 1950, when it was supported, than it had in earlier periods. Therefore, I also plotted the daily changes for the period near 1900, which was chosen haphazardly, but not actually at random. The new values of $C'(T)/C'(1)$ and $C''(T)/C''(1)$ became quite reasonable: they were equal to each other and to 18. In 1900, there were seven trading days per week, but they subsequently decreased to five. Besides, one cannot be too dogmatic about estimating $C'(T)/C'(1)$. Therefore, the behavior of this ratio indicated that the "apparent" number of trading days per month was somewhat smaller than the actual number.

{P.S. 1996. Actually, I had badly misread the data: cotton was *not* traded on Sundays in 1900, and correcting this error improved the fit of the M 1963 model; see Appendix IV to this Chapter.}

IV. WHY ONE SHOULD EXPECT TO FIND NONSENSE MOMENTS AND NONSENSE PERIODICITIES IN ECONOMIC TIME SERIES

IV.A. Behavior of second moments and failure of the least-squares method of forecasting

It is amusing to note that the first known nonGaussian stable law, namely, the Cauchy distribution, was introduced in the course of a study of the method of least squares. A surprisingly lively argument followed the reading of Cauchy 1853. In this argument, Bienaymé 1853 stressed that a method based upon the minimization of the sum of squares of sample deviations cannot reasonably be used if the expected value of this sum is known to be infinite. The same argument applies fully to the problem of least-squares smoothing of economic time series, when the "noise" follows a L-stable law other than that of Cauchy.

Similarly, consider the problem of least-squares forecasting, that is, of the minimization of the expected value of the square of the error of extrapolation. In the L-stable case, this expected value will be infinite for every forecast, so that the method is, at best, extremely questionable.

One can perhaps apply a method of "least ζ-power" of the forecasting error, where $\zeta < \alpha$, but such an approach would not have the formal simplicity of least squares manipulations. The most hopeful case is that of $\zeta = 1$, which corresponds to the minimization of the sum of absolute values of the errors of forecasting.

IV.B. Behavior of the sample kurtosis and its failure as a measure of the "peakedness" or "long-tailedness" of a distribution

Pearson proposed to measure the peakedness or long-tailedness of a distribution by the following quantity, call "kurtosis"

$$\text{kurtosis} = -3 + \frac{\text{fourth population moment}}{\text{square of the second population moment}}.$$

In the L-stable case with $0 < \alpha < 2$, the numerator and the denominator both have an infinite expected value. One can, however, show that the sample kurtosis + 3 behaves proportionately to the following "typical" value

$$\frac{(\frac{1}{N}\text{ (the most probable value of } \sum L^4)}{\left\{\frac{1}{N}\text{ (the most probable value of } \sum L^2)\right\}^2}$$
$$= \frac{(\text{a constant})N^{-1+4/\alpha}}{\{(\text{a constant})\ N^{-1+2/\alpha}\}^2} = (\text{a constant})N.$$

It follows that the kurtosis is expected to increase without bound as $N \to \infty$. For small N, things are less simple, but presumably quite similar.

In this light, examine Cootner 1962. This paper developed the tempting hypothesis that prices vary at random as long as they do not wander outside a "penumbra", defined as an interval that well-informed speculators view as reasonable. But random fluctuations triggered by ill-informed speculators will eventually let the price go too high or too low. When this happens, the operation of well-informed speculators will induce this price to come back within the "penumbra." If this view of the world were correct, one would conclude that the price changes over periods of, say, fourteen weeks would be smaller than expected if the contributing weekly changes were independent.

This theory is very attractive a priori, but could not be generally true because, in the case of cotton, it is not supported by the facts. As for Cootner's own justification, it is based upon the observation that the price changes of certain securities over periods of fourteen weeks have a much smaller kurtosis than one-week changes. Unfortunately, his sample contains 250-odd weekly changes and only 18 fourteen-week periods. Hence, on the basis of general evidence concerning speculative prices, I would have expected, a priori, to find a smaller kurtosis for the longer time increment. Also, Cootner's evidence is not a proof of his theory; other methods must be used in order to attack the still very open problem of the possible dependence between successive price changes.

IV.C. Method of spectral analysis of random time series

These days, applied mathematicians are frequently presented with the task of describing the stochastic mechanism capable of generating a given time series $u(t)$, known or presumed to be random. The first response to such a problem is usually to investigate what is obtained by applying a theory of the "second-order random process." That is, assuming that $E(U) = 0$, one forms the sample covariance

$$r(\tau) = \frac{1}{N-\tau} \sum_{t=T^0+1}^{t=T^0+N-\tau} u(t)u(t+\tau),$$

which is used, somewhat indirectly, to evaluate the population covariance

$$R(\tau) = E[U(t)U(t+\tau)].$$

Of course, $R(\tau)$ is always assumed to be finite for all τ. The Fourier transform of $R(\tau)$ is the "spectral density" of the process $U(t)$, and rules the "harmonic decomposition" of $U(t)$ into a sum of sine and cosine terms.

Broadly speaking, this method has been very successful, though many small-sample problems remain unsolved. Its applications to economics have, however, been questionable even in the large-sample case. Within the context of my theory, there is, unfortunately, nothing surprising in this finding. Indeed,

$$2E[U(t)U(t+\tau)] = E[U(t)+U(t+\tau)]^2 - E[U(t)]^2 - E[U(t+\tau)]^2.$$

For time series covered by my model, the three variances on the right hand side are all infinite, so that spectral analysis loses its theoretical motivation. This is a fascinating problem, but I must postpone a more detailed examination of it.

V. SAMPLE FUNCTIONS GENERATED BY L-STABLE PROCESSES; SMALL-SAMPLE ESTIMATION OF THE MEAN "DRIFT"

The curves generated by L-stable processes present an even larger number of interesting formations than the curves generated by Bachelier's Brownian motion. If the price increase over a long period of time happens a posteriori to have been exceptionally large, one should expect, in a L-stable market, to find that most of this change occurred during only a few periods of especially high activity. That is, one will find in most cases that the majority of the contributing daily changes are distributed on a fairly symmetric curve, while a few especially high values fall way outside this curve. If the total increase is of the usual size, to the contrary, the daily changes will show no "outliers."

In this section these results will be used to solve one small-sample statistical problem, that of the estimation of the mean drift δ, when the other parameters are known. We shall see that there is no "sufficient statistic" for this problem, and that the maximum likelihood equation does not necessarily have a single root. This has severe consequences from the viewpoint of the very definition of the concept of "trend."

V.A. Some properties of sample paths of Brownian motion

The sample paths of Brownian motion very much "look like" the empirical curves of time variation of prices or of price indexes. This was noted by Bachelier and (independently of him and of each other) by several modern writers (see especially Working 1934, Kendall 1953, Osborne 1959 and Alexander 1964), At closer inspection, however, one sees very clearly the effect of the abnormal number of large positive and large negative changes of $\log_e Z(t)$. At still closer inspection, one finds that the differences concern some of the economically most interesting features of the generalized central-limit theorem of the calculus of probability. It is therefore necessary to discuss this question in detail, beginning with a review of some classical properties of Gaussian random variables.

Conditional distribution of a Gaussian addend $L(+\tau, 1)$, knowing the sum $L(t, T) = L(t, 1) + \ldots + L(t + T - 1, 1)$. Let the probability density of $L(t, T)$ be

$$\frac{1}{\sqrt{2\pi\sigma^2 T}} \exp\left\{ -\frac{u - \delta T^2}{2T\sigma^2} \right\}.$$

It is then easy to see that, if one knows the value of u of $L(t, T)$, the density of any of the quantities $L(t+\tau, 1)$ is given by

$$\frac{1}{2\pi\sigma^2(T-1)/T} \exp\left\{ -\frac{(u' - u/T)^2}{2\sigma^2(T-1)/T} \right\}.$$

This means that each of the contributing $L(t+\tau, 1)$ equals u/T plus a Gaussian error term. For large T, that term has the same variance as the unconditioned $L(t, 1)$ – one can in fact prove that the value of u has little influence upon the size of the largest of those "noise terms." One can therefore say that, whatever its value, u is roughly uniformly distributed over the T time intervals, each contributing negligibly to the whole.

Sufficiency of u for the estimation of the mean drift δ from the $L(t+\tau, 1)$. In particular, δ has vanished from the distribution of any $L(t+\tau, 1)$ conditioned by the value of u. In the vocabulary of mathematical statistics u is a "sufficient statistic" for the estimation of δ from the values of all the $L(t+\tau, 1)$. That is, whichever method of estimation a statistician may favor, his estimate of δ must be a function of u alone. The knowledge of intermediate values of $\log_e Z(t+\tau)$ is of no help. Most methods recommend estimating δ from u/T and extrapolating the future linearly from the two known points, $\log_e Z(t)$ and $\log_e Z(t+T)$. Since the causes of any price movement can be traced backwards only if the movement is of sufficient size, all that one can explain in the Gaussian case is the mean drift interpreted as a trend. Bachelier's model, which assumes a zero mean for the price changes, can only represent the movement of prices once the broad causal parts or trends have been removed.

V.B. One value from a process of independent L-stable increments

Returning to the L-stable case, suppose that the values of γ, of β (or of C' and C'') and of α are known. The remaining parameter is the mean drift δ; one must estimate δ starting from the known $L(t, T) = \log_e Z(t+T) - \log_e Z(t)$.

The unbiased estimate of δ is $L(t, T)/T$, while the estimate matches the observed $L(t, T)$ to its a priori *most probable* value. The "bias" of the maximum likelihood is therefore given by an expression of the form $\gamma^{1/\alpha} f(\beta)$, where the function $f(\beta)$ must be determined from the numerical table of the L-stable densities. Since β is mostly manifested in the relative sizes of the tails, its evaluation requires very large samples, and the quality of predictions will depend greatly upon the quality of one's knowledge of the past.

It is, of course, not at all clear that anybody would wish the extrapolation to be unbiased with respect to the mean of the change of the *logarithm* of the price. Moreover, the bias of the maximum likelihood estimate comes principally from an underestimate of the size of changes that are so large as to be catastrophic. The forecaster may very well wish to treat such changes separately, and to take into account his private opinions about many things that are not included in the independent-increment model.

V.C. Two values from a L-stable process

Suppose now that T is even and that one knows $L(t, T/2)$ and $L(t+T/2, T/2)$, and thus also their sum $L(t, T)$. Section IIG has shown that when the value $u = L(t, T)$ is given, the conditional distribution of $L(t, T/2)$ depends very sharply upon u. This means that the total change u is not a sufficient statistic for the estimation of δ; in other words, the estimates of δ will be changed by the knowledge of $L(t, T/2)$ and $L(t+T/2, T/2)$.

Consider, for example, the most likely value δ. If $L(t, T/2)$ and $L(t+T/2, T/2)$ are of the same order of magnitude, this estimate will remain close to $L(t, T)/T$, as in the Gaussian case. But suppose that *the actually observed* values of $L(t, T/2)$ and $L(t+T/2, T/2)$ are very unequal, thus implying that at least one of these quantities is very different from their common mean and median. Such an event is most likely to occur when δ is close to the observed value of either $L(t+T/2, T/2)/(T/2)$ or $L(t, T/2)/(t/2)$.

As a result, the maximum likelihood equation for δ has two roots, one near $2L(t, T/2)/T$ and the other near $2L(t+T/2, T/2)/T$. That is, the maximum-likelihood procedure says that one of the available items of information should be neglected, since any weighted mean of the two recommended extrapolations is worse than either. But nothing says which item should be neglected.

It is clear that few economists will accept such advice. Some will stress that the most likely value of δ is actually nothing but the most probable value in the case of the uniform distribution of a priori probabilities of δ. But it seldom happens that a priori probabilities are uniformly distributed. It is also true, of course, that they are usually very poorly determined. In the present problem, however, the economist will not need to determine these a priori probabilities with any precision: it will be sufficient to choose the most likely *for him* of the two maximum-likelihood estimates.

An alternative approach (to be presented later in this paper) will argue that successive increments of $\log_e Z(t)$ are not really independent, so that the estimation of δ depends upon the order of the values of $L(t, T/2)$ and $L(t+T/2, T/2)$, as well as upon their sizes. This may help eliminate the indeterminacy of estimation.

A third alternative consists in abandoning the hypothesis that δ is the same for both changes $L(t, T/2)$ and $L(t+T/2, T/2)$. For example, if these changes are very unequal, one can fit the data better by assuming that the trend δ is not linear but parabolic. In a first approximation, extrapolation would then consist in choosing among the two maximum-likelihood estimates the one which is chronologically the latest. This is an example of a variety of configurations which would have been so unlikely in the Gaussian case that they would have been considered nonrandom, and would have been of help in extrapolation. In the L-stable case, however, their probability may be substantial.

V.D. Three values from the L-stable process

The number of possibilities increases rapidly with the sample size. Assume now that T is a multiple of 3, and consider $L(t, T/3)$, $L(t+T/3, T/3)$, and $L(t+2T/3, T/3)$. If these three quantities are of comparable size, the knowledge of $\log Z(t+T/3)$ and $\log Z(t+2T/3)$ will again bring little change to the estimate based upon $L(t, T)$.

But suppose that one datum is very large and the others are of much smaller and comparable sizes. Then, the likelihood will have two local maximums, well separated, but of sufficiently equal sizes as to make it impossible to dismiss the smaller one. The absolute maximum yields the estimate $\delta = (3/2T)$ (sum of the two small data); the smaller local maximum yields the estimate $\delta = (3/T)$ (the large datum).

Suppose, finally, that the three data are of very unequal sizes. Then the maximum likelihood equation has *three* roots.

This indeterminacy of maximum likelihood can again be lifted by one of the three methods of Section VC. For example, if only the middle datum is large, the methods of nonlinear extrapolation will suggest a logistic growth. If the data increase or decrease – when taken chronologically – a parabolic trend should be tried. Again, the probability of these configurations arising from chance under my model will be much greater than in the Gaussian case.

V.E. A large number of values from a L-stable process

Let us now jump to the case of a very large amount of data. In order to investigate the predictions of my L-stable model, we must first reexamine the meaning to be attached to the statement that, in order that a sum of random variables follow a central limit of probability, it is necessary that each of the addends be negligible relative to the sum.

It is quite true, of course, that one can speak of limit laws only if the value of the sum is not *dominated* by any single addend known in advance. That is, to study the limit of $A(N)\Sigma U_n - B(N)$, one must assume that, for every n, Pr $|A(N)U_n - B(N)/N| \geq \varepsilon$ tends to zero with $1/N$.

As each addend decreases with $1/N$, their number increases, however, and the condition of the preceding paragraph does not by itself insure that the largest of the $|A(N)U_n - B(N)/N|$ is negligible in comparison with the sum. As a matter of fact, the last condition is true only if the limit of the sum is Gaussian. In the scaling case, on the contrary, the ratios

$$\frac{\max |A(N)U_n - B(N)/N|}{A(N)\sum U_n - B(N)} \quad \text{and} \quad \frac{\text{plex pssum of } k \text{ largest } |A(N)U_n - B(N)/N|}{A(N)\sum U_n - B(BN)}$$

tend to nonvanishing limits as N increases (Darling 1952 and Arov & Bobrov 1960). In particular, it can be proven that, when the sum $A(N)\Sigma U_n - B(N)$ happens to be large, the above ratios will be close to *one*.

Returning to a process with independent L-stable $L(t)$, we may say the following: If, knowing α, β, γ, and δ, one observes that $L(t, T =$ one month) is *not* large, the contribution of the day of largest price change is likely to be nonnegligible in relative value, but it will remain small in absolute value. For large but finite N, this will not differ too much from the Gaussian prediction that even the largest addend is negligible.

Suppose, however, that $L(t, T =$ one month) is *very* large. The scaling theory then predicts that the sum of few largest daily changes will be very close to the total $L(t, T)$. If one plots the frequencies of various values of $L(t, 1)$, conditioned by a known and very large value for $L(t, T)$, one should expect to find that the law of $L(t + \tau, 1)$ contains a few widely "outlying" values. However, if the outlying values are taken out, the conditioned distribution of $L(t + \tau, 1)$ should depend little upon the value of the conditioned $L(t, T)$. I believe this last prediction to be well satisfied by prices.

Implications concerning estimation. Suppose now that δ is unknown and that one has a large sample of $L(t + \tau, 1)$'s. The estimation procedure then consists of plotting the empirical histogram and translating it horizontally until its fit to the theoretical density curve has been optimized. One knows in advance that the best value will be very little influenced by the largest outliers. Hence, "rejection of the outliers" is fully justified in the present case, at least in its basic idea.

V.F. Conclusions concerning estimation

The observations made in the preceding sections seem to confirm some economists' feeling that prediction is feasible only if the sample size is both very large and stationary, or if the sample size is small but the sample values are of comparable sizes. One can also make predictions from a sample size of one, but here the availability of a unique estimator is due only to ignorance.

V.G. Causality and randomness in L-stable processes

We mentioned in Section VA that, in order to be "causally explainable," an economic change must be large enough to allow the economist to trace back the sequence of its causes. As a result, the only causal part of a Gaussian random function is the mean drift δ. The same is true of L-stable random functions when their changes happen to be roughly uniformly distributed.

But it is not true in the cases where $\log_e Z(t)$ varies greatly between the times t and $t+T$, changing mostly during a few of the contributing days. Then, the largest changes are sufficiently clear-cut, and are sufficiently separated from "noise," to be explained causally, just as well as the mean drift.

In other words, a careful observer of a L-stable random function will be able to extract causal parts from it. But if the total change of $\log_e Z(t)$ is neither very large nor very small, there will be a large degree of arbitrariness in this distinction between causal and random. Hence, it would not be possible to determine whether the predicted proportions of the two kinds of effects are empirically correct.

In sum, the distinction between the causal and the random areas is sharp in the Gaussian case and very diffuse in the L-stable case. This seems to me to be a strong recommendation in favor of the L-stable process as a model of speculative markets. Of course, I have not the slightest idea why the large price movements should be representable in this way by a simple extrapolation of movements of ordinary size. I have come to believe, however, that it is very desirable that both "trend" and "noise" be aspects of the same deeper "truth." At this point, we can adequately describe it but cannot provide an explanation. I am certainly not antagonistic to the goal of achieving a decomposition of economic "noise" into parts similar to the trend, and to link various series to each other. But, until we come close to this goal, we should be pleased to be able to represent some trends as similar to "noise."

V.H. Causality and randomness in aggregation "in parallel"

Borrowing a term from elementary electrical circuit theory, the addition of successive daily changes of a price may be denoted by the term "aggregation in series," the term "aggregation in parallel" applying to the operation

$$L(t, T) = \sum_{i=1}^{I} L(i, t, T), = \sum_{i=1}^{I} \sum_{\tau=0}^{T-1} L(i, t+\tau, 1),$$

where i refers to "events" that occur simultaneously during a given time interval such as T or 1.

In the Gaussian case, one should, of course, expect any occurrence of a large value for $L(t, T)$ to be traceable to a rare conjunction of large changes in all or most of the $L(i, t, T)$. In the L-stable case, one should, on the contrary, expect large changes $L(t, T)$ to be traceable to one or a small number, of the contributing $L(i, t, T)$. It seems obvious that the L-stable prediction is closer to the facts.

If we add up the two types of aggregation in a L-stable world, we see that a large $L(t, T)$ is likely to be traceable to the fact that $L(i, t+\tau, 1)$ happens to be very large for one or a few sets of values of i and of τ. These contributions would stand out sharply and be causally explainable. But after a while, they should rejoin the "noise" made up of the other factors. The next rapid change of $\log_e Z(t)$ should be due to other "causes." If a contribution is "trend-making," in the above sense, during a large number of time-increments, one will naturally doubt that it falls under the same theory as the fluctuations.

VI. PRICE VARIATIONS IN CONTINUOUS TIME AND THE THEORY OF SPECULATION

The main point of this section is to examine certain systems of speculation, which appear advantageous, and to show that, in fact, they cannot be followed in the case of price series generated by a L-stable process.

VI.A. Infinite divisibility of L-stable variables

In theory, it is possible to interpolate $L(t, 1)$ indefinitely. That is, for every N, one can consider that a L-stable increment

$$L(t, 1) = \log_e Z(t+1) - \log_e Z(t)$$

is the sum of N independent, identically distributed random variables. The only difference between those variables and $L(t, 1)$ is that the constants γ, C' and C'' are N times smaller in the parts than in the whole.

In fact, it is possible to interpolate the process of independent L-stable increments to continuous time, assuming that $L(t, dt)$ is a L-stable variable with a scale coefficient $\gamma(dt) = dt\,\gamma(1)$. This interpolated process is a very important "zeroth" order approximation to the actual price changes. That is, its predictions are without doubt modified by the mechanisms of the market, but they are very illuminating nonetheless.

VI.B. Path functions of a L-stable process in continuous time

Mathematical models of physical or of social sciences almost universally assume that all functions can safely be considered to be continuous and to have as many derivatives as one may wish. Contrary to this expectation, the functions generated by Bachelier have no derivatives, even though they are indeed continuous. In full mathematical rigor, "there is a probability equal to 1 that they are continuous but nondifferentiable almost everywhere, but price quotations are always rounded to simple fractions of the unit of currency. If only for this reason, we need not worry about mathematical rigor here.

In the scaling case things are quite different. If my process is interpolated to continuous t, the paths which it generates become discontinuous in every interval of time, however small (in full rigor, they become "almost surely almost everywhere discontinuous"). That is, most of their variation occurs through noninfinitesimal "jumps." Moreover, the number of jumps larger than u and located within a time increment T is given by the law $C'T|d(u^{-\alpha})|$.

Let us examine a few aspects of this discontinuity. Again, very small jumps of $\log_e Z(t)$ could not be perceived, since price quotations are always expressed in simple fractions. More interesting is the fact that there is a nonnegligible probability of witnessing a price jump so large that supply and demand cease to be matched. In other words, the L-stable model can be considered as predicting the occurrence of phenomena likely to force the market to close. In a Gaussian model, such large changes are so extremely unlikely that the occasional closure of the markets must be explained by nonstochastic considerations.

The most interesting fact is, however, the large probability predicted for medium-sized jumps by the L-stable model. Clearly, if those medium-sized movements were oscillatory, they could be eliminated by market mechanisms such as the activities of the specialists. But if the movement is all in one direction, market specialists could at best transform a discontinuity into a change that is rapid but progressive. On the other hand, very few transactions would then be expected at the intermediate smoothing prices. As a result, even if the price Z_0 is quoted transiently, it may be impossible to act rapidly enough to satisfy more than a minute fraction of orders to "sell at Z_0." In other words, a large number of intermediate prices are quoted even if $Z(t)$ performs a large jump in a short time; but they are likely to be so fleeting, and to apply to so few transactions, that they are irrelevant from the viewpoint of actually enforcing a "stop loss order" of any kind. In less extreme cases – as, for example, when borrowings are over-subscribed – the market may have to resort to special rules of allocation.

These remarks are the crux of my criticism of certain systematic trading methods: they would perhaps be very advantageous *if only they could* be followed systematically; but, in fact, they *cannot* be followed. I shall be content here with a discussion of one example of this kind of reasoning.

VI.C. The fairness of Alexander's "filter" game

Alexander 1964 has suggested the following rule of speculation: "If the market goes up 5%, go long and stay long until it moves down 5%, at which time sell and go short until it again goes up 5%."

This procedure is motivated by the fact that, according to Alexander's interpretation, data would suggest that "in speculative markets, price changes appear to follow a random walk over time; but ... if the market has moved up x%, it is likely to move up more than x% further before it moves down x%." He calls this phenomenon the "persistence of moves." Since there is no possible persistence of moves in any "random walk" with zero mean, we see that if Alexander's interpretation of facts were confirmed, it would force us to seek immediately a model better than the random walk.

In order to follow this rule, one must, of course, watch a price series continuously in time and buy and sell whenever its variation attains the prescribed value. In other words, this rule can be strictly followed if and only if the process $Z(t)$ generates continuous path functions, as for example in the original Gaussian process of Bachelier.

Alexander's procedure cannot be followed, however, in the case of my own first-approximation model of price change in which there is a probability equal to one that the first move *not smaller* than 5% is *greater* than 5% and *not equal* to 5%. It is therefore mandatory to modify the filter method: one can at best recommend buying or selling when moves of 5% are *first exceeded*. One can prove that the L-stable theory predicts that this is game also fair. Therefore, evidence – as interpreted by Alexander – would again suggest that one must go beyond the simple model of independent increments of price.

But Alexander's inference was actually based upon the discontinuous series constituted by the closing prices on successive days. He assumed that the intermediate prices could be interpolated by some continuous function of continuous time – the actual form of which need not be specified. That is, whenever there was a difference of *over* 5% between the closing price on day F' and day F'', Alexander implicitly assumed that there was at least one instance between these moments when the price had gone up *exactly* 5 per cent. He recommends buying at this instant, and he computes the empirical returns to the speculator as if he were able to follow this procedure.

For price series generated by my process, however, the price actually paid for a stock will almost always be *greater* than that corresponding to a 5% rise; hence the speculator will almost always have paid *more* than assumed in Alexander's evaluation of the returns. Similarly, the price received will almost always be *less* than suggested by Alexander. Hence, at best, Alexander overestimates the yield corresponding to his method of speculation and, at worst, the very impression that the yield is positive may be a delusion due to overoptimistic evaluation of what happens during the few most rapid price changes.

One can, of course, imagine contracts guaranteeing that the broker will charge (or credit) his client the actual price quotation nearest by excess (or default) to a price agreed upon, irrespective of whether the broker was able to perform the transaction at the price agreed upon. Such a system would make Alexander's procedure advantageous to the speculator, but the money he would be making, on the average, would come from his broker and not from the market, and brokerage fees would have to be such as to make the game at best fair in the long run.

VII. A MORE REFINED MODEL OF PRICE VARIATION, TAKING ACCOUNT OF SERIAL DEPENDENCE

Broadly speaking, the predictions of my main model seem to me to be reasonable. At closer inspection, however, one notes that large price changes are not isolated between periods of slow change; they rather tend to be the result of several fluctuations, some of which "overshoot" the final changes. Similarly, the movements of prices in periods of tranquility seem to be smoother than predicted by my process. In other words, large changes tend to be followed by large changes – of either sign – and small changes tend to be followed by small changes, so that the isolines of low probability of $[L(t, 1), L(t-1, 1)]$ are **X**-shaped. In the case of daily cotton prices, Hendrik S. Houthakker stressed this fact in several conferences and in private conversation.

Such an **X**-shape is easily obtained by a 90° rotation from the "+ shape" which was observed when $L(t, 1)$ and $L(t-1, 1)$ are statistically independent and symmetric (Figure 4). This rotation introduces the two expressions:

$$S(t) = (1/2)[L(t, 1) + L(t-1, 1)] = (1/2)[\log_e Z(t+1) - \log_e Z(t-1)]$$

and

$$D(t) = (1/2)[L(t, 1) - L(t-1, 1)]$$
$$= (1/2)[\log_e Z(t+1) - 2\log_e Z(t) + \log_e Z(t-1)].$$

It follows that in order to obtain **X**-shaped empirical isolines, it would be sufficient to assume that the first and second finite differences of $\log_e Z(t)$ are two L-stable random variables, independent of each other, and naturally of $\log_e Z(t)$ (Figure 4). Such a process is invariant by time inversion.

It is interesting to note that the distribution of $L(t, 1)$, conditioned by the known $L(t-1, 1)$, is asymptotically scaling with an exponent equal to $2\alpha + 1$. A derivation is given at the end of this section. For the cases, we are interested in, $\alpha > 1.5$, hence $2\alpha + 1 > 4$. It follows that the conditioned $L(t, 1)$ has a finite kurtosis; no L-stable law can be associated with it.

Let us then consider a Markovian process with the transition probability I have just introduced. If the initial $L(T^0, 1)$ is small, the first values of $L(t, 1)$ will be weakly asymptotic scaling with a high exponent $2\alpha + 1$, so that $\log_e Z(t)$ will begin by fluctuating much less rapidly than in the case of independent $L(t, 1)$. Eventually, however, a large $L(t^0, 1)$ will appear. Thereafter, $L(t, 1)$ will fluctuate for some time between values of the orders of magnitude of $L(t^0, 1)$ and $-L(t^0, 1)$. This will last long enough to compensate fully for the deficiency of large values during the period of slow variation. In other words, the occasional sharp changes of $L(t, 1)$ predicted by the model of independent $L(t, 1)$ are replaced by oscillatory periods, and the periods without sharp change are shown less fluctuating then when the $L(t, 1)$ are independent.

We see that, if α is to estimated correctly, periods of rapid changes of prices must be considered with the other periods. One *cannot* argue that they are "causally" explainable and ought to be eliminated before the "noise" is examined more closely. If one succeeded in eliminating all large changes in this way, one would indeed have a Gaussian-like remainder. But this remainder would be devoid of any significance.

Derivation of the value $2\alpha + 1$ *for the exponent.* Consider

$$\Pr\{L(t, 1) > u, \text{ when } w < L(t-1, 1) < w + dw\}.$$

This is the product by $(1/dw)$ of the integral of the probability density of $[L(t-1, 1)L(t, 1)]$, over a strip that differs infinitesimally from the zone defined by

$$S(t) > (u+w)/2; w + S(t) < D(t) < w + S(t) + dw.$$

Hence, if u is large as compared to w, the conditional probability in question is equal to the integral

$$\int_{(u+w)/2}^{\infty} C'\alpha s^{-(\alpha+1)} C'\alpha (s+w)^{-(\alpha+1)} ds \sim (2\alpha+1)^{-1}(C')^2\alpha^2 2^{-(2\alpha+1)} u^{-(2\alpha+1)}.$$

&&&& POST-PUBLICATION APPENDICES &&&&

These four appendices from different sources serve different purposes.

APPENDIX I (1996): THE EFFECTS OF AVERAGING

The M 1963 model of price variation asserts that price changes between equally spaced *closing* times are L-stable random variables. As shown momentarily, the model also predicts that changes between monthly average prices are L-stable.

To the contrary, Figure 7 suggests that the tails are *shorter* than predicted and the text notes that this is a token of interdependence between successive price changes.

The incorrect prediction. If $L(0) = 0$ and $L(t)$ has independent L-stable increments, consider the increment between the "future" average from 0 to t and the value at t. Integration by parts yields

$$\frac{1}{t}\int_0^t L(s)ds - L(t) = -\frac{1}{t}\int_0^t s dL(s).$$

The r.h.s. is a L-stable random variable for which (scale)$^\alpha$ equals

$$t^{-\alpha}\int_0^t s^\alpha ds = (\alpha + 1)^{-1} t.$$

The "past increment" is independent of the "future increment," and follows the same distribution. So does the difference between the two

APPENDIX II (MOSTLY A QUOTE FROM FAMA & BLUME 1966): THE EXEMPLARY FALL OF ALEXANDER'S FILTER METHOD

Section VI C of M 1963b criticizes a rule of speculation suggested in Alexander 1961, but does not provide a revised analysis of Alexander's data. However, Alexander's filters did not survive this blow. The story was told by Fama and Blume 1966 in the following terms:

"Alexander's filter technique is a mechanical trading rule which attempts to apply more sophisticated criteria to identify movements in stock prices. An $x\%$ filter is defined as follows: If the daily closing price of a particular security moves up at least x per cent, buy and hold the security until its price moves down at least $x\%$ from a subsequent high, at which time simultaneously sell and go short. The short position is maintained until daily closing prices rises at least $x\%$ above a subsequent low at which time one covers and buys. Moves less than $x\%$ in either direction are ignored.

"Alexander formulated the filter technique to test the belief, widely held among market professionals, that prices adjust gradually to new information.

"The professional analysts operate in the belief that there exist certain trend generating facts, knowable today, that will guide a speculator to profit if only he can read them correctly. These facts are assumed to generate trends rather than instantaneous jumps because most of those trading in speculative markets have imperfect knowledge of these facts, and the future trend of price will result from a gradual spread of awareness of these facts throughout the market [Alexander 1961, p.7].

"For the filter technique, this means that for some values of x we would find that 'if the stock market has moved up $x\%$ it is likely to move up more than x per cent further before it moves down by $x\%$ ' [Alexander 1961, p.26].

"In his Table 7, Alexander 1961 reported tests of the filter technique for filters ranging in size from 5 to 50 per cent. The tests covered different time periods from 1897 to 1959 and involved closing "prices" for two indexes, the Dow-Jones Industrials from 1897 to 1929 and Standard and Poor's Industrials from 1929 to 1959. In general, filters of all different sizes and for all the different time periods yielded substantial profits – indeed profits significantly greater than those of the simple buy-and-hold policy. This led Alexander to conclude that the independence assumption of the random-walk model was not upheld by his data.

"M 1963b [Section VI.C] pointed out, however, that Alexander's computations incorporated biases which led to serious overstatement of the profitability of the filters. In each transaction Alexander assumed that this hypothetical trader could always buy at a prices exactly equal to the low plus x per cent and sell at the high minus x per cent. In fact, because of the frequency of large price jumps, the purchase price will often be somewhat higher than the low plus x per cent, while the sale price will often be below the high minus x per cent. The point is of central theoretical importance for the L-stable hypothesis.

"In his later paper [Alexander 1964, Table 1] Alexander reworked his earlier results to take account of this source of bias. In the corrected tests the profitability of the filter technique was drastically reduced.

"However, though his later work takes account of discontinuities in the price series, Alexander's results are still very difficult to interpret. The difficulties arise because it is impossible to adjust the commonly used price indexes for the effects of dividends. This will later be shown to introduce serious biases into filter results."

Fama & Blume 1966 applied Alexander's technique to series of daily closing prices for each individual security of the Dow-Jones Industrial Average. They concluded that the filter method *does not* work.

Thus, the filters are buried for good, but many "believers" never received this message.

APPENDIX III (1996): ESTIMATION BIAS AND OTHER REASONS FOR $\alpha > 2$

Chapter E10, reproducing M 1960i{E10}, is followed by Post-Publication Appendix IV, adapted from M 1963i{E10}. The body of the present chapter, M 1963b, was written near-simultaneously with that appendix, and very similar comments can be made here. That is, for α close to 2, the diagrams in Figure 3 are inverse S-shaped, therefore, easily mistaken for straight lines with a slope that is $> \alpha$, and even > 2.

A broader structure is presented in Chapters E1 and E6, within which α has no upper bound. Therefore, the remark in the preceding paragraph *must not* be misconstrued. Estimation bias is only one of several reasons why an empirical log log plot of price changes may have a slope that contradicts the restriction [1, 2] that is characteristic of L-stability with $EU < \infty$.

APPENDIX IV (M 1972b): CORRECTION OF AN ERROR IN VCSP

• ***Section foreward.*** The correction of an error in *VCSP* = M 1963b improved in the fit between the data and the M 1963 model, eliminating some pesky descrepancies that *VCSP* had pointed out as deserving a fresh look. •

Infinite variance and of non-Gaussian L-stable distribution of price differentials were introduced for the first time in M 1963b. The prime material on which both hypotheses were based came in part from H.S. Houthakker and in part from the United States Department of Agriculture; it concerned daily spot prices of cotton.

Since then, the usefulness of those hypotheses was confirmed by the study of many other records, both in my work and in that of others. But it has now come to my attention that part of my early evidence suffered from a serious error. In the data sheets received from the USDA, an important footnote had been trimmed off, and as a result they were misread. Numbers which I had interpreted as Sunday closing prices were actually weekly price averages. They were inserted in the blanks conveniently present in the data sheets. My admiring joke about hard-working American cotton dealers of 1900-1905 was backfiring; no one corrected me in public, but I shudder at some comments that must have been made in private about my credibility. The error affected part of Figure 5 of M 1963b: the curves 1*a* and 2*a* relative to that period were incorrect.

After several sleepless nights, this error was corrected, and the analysis was revised. I am happy to report that my conclusion was upheld, in fact, much simplified, and the fit between the theory and the data improved considerably. M 1963b{E14} noted numerous peculiarities that had led me to consider my hypotheses as no more than rough first approximations. For example, the simplest random-walk model implied that a monthly price change is the sum of independent daily price changes. In fact, as I was careful to note, such was the case only if one assumed that a month included an "apparent number of trading days ... smaller than the actual number." The theory also implied that, whenever a monthly price change is large, it is usually about equal to the largest contributing daily price change. In fact, instances when large monthly changes resulted from, say, three large daily changes (one up and two down, or conversely) were more numerous in the data than predicted. Both findings suggested that a strong negative dependence exists between successive price changes. Also, prices seemed to have been more volatile around 1900 than around 1950. After the data have been corrected, these peculiarities have disappeared. In particular, the corrected curves 1*a* and 2*a* are nearly indistinguishable from the corresponding curves 1*b* and 2*b* relative to the Houthhakker data concerning the period 1950-58.

APPENDIX V (M 1982c): A "CITATION CLASSIC"

• ***Section foreward.*** In 1982, the *Citation Index of the Institute of Scientific Information* determined that M 1963b had become a *Citation Classic. Current Contents/Social and Behavioral Sciences* invited me to comment, "emphasizing the human side of the research – how the project was initiated, any obstacles encountered, and why the work was highly cited." •

✦ **Abstract.** Changes of commodity and security prices are fitted excellently by the L-stable probability distributions. Their parameter α is the intrinsic measure of price volatility. The model also accounts for the amplitudes of major events in economic history. An unprecedented feature is that price changes have an infinite population variance. ✦

Early in 1961, coming to Harvard to give a seminar on my work on personal income distributions, I stepped into the office of my host, H. S. Houthakker. On his blackboard, I noticed a diagram that was nearly identical to one I was about to draw, but concerned a topic of which I knew nothing: the variability of the price of cotton. My host had given up his attempt to model this phenomenon and challenged me to take over.

In a few weeks, I had introduced a radically new approach. It preserved the random walk hypothesis that the market is like a lottery or a casino, with prices going up or down as if determined by the throw of dice. It also preserved the efficient marked hypothesis that the market's collective wisdom take account of all available information, hence, the price tomorrow and on any day thereafter will *on the average* equal today's price. The third basis of the usual model is that price changes follow the Gaussian distribution. All these hypotheses, due to Louis Bachelier 1900, were first taken seriously in 1960. The resulting theory, claiming that price (or its logarithm) follows a Brownian motion, would be mathematically convenient, but it fits the data badly.

Most importantly, the records of throws of a die appear unchanged statistically. In comparison, the records of competitive price changes "look *nonstationary*"; they involve countless configurations that seem too striking to be attributable to mere chance. A related observation: the histograms of price changes are very far from the Galton ogive; they are long-tailed to an astonishing degree, due to large excursions whose size is obviously of the highest interest.

My model replaces the customary Gaussian hypothesis with a more general one, while allowing the population variance of the price changes to be infinite. The model is time invariant, but it creates endless configurations, and accounts for all the data, including both the seemingly nonstationary features, and the seemingly nonrandom large excursions.

A visiting professorship of economics at Harvard, 1962-1963, was triggered by IBM Research Note NC-87 (M 1962i), which tackled the prices of cotton and diverse commodities and securities. Also, M 1963b was immediately reprinted in Cootner 1964, along with discussions by E. F. Fama, who was my student at the time, and by the editor. This publication must have affected my election to Fellowship in the Econometric Society. However, after a few further forays in economics, my interest was drawn irresistibly toward the very different task of creating a new fractal geometry of nature. Having learned to live with the unprecedented infinite variance syndrome had trained me to identify telltale signs of divergence in the most diverse contexts, and to account for them suitably.

By its style, my work on prices remains unique in economics; while all the other models borrow from the final formulas of physics, I lean on its basic mental tool (invariance principles) and deduce totally new formulas appropriate to the fact that prices are not subjected to inertia, hence need not be continuous. My work is also unique in its power: the huge bodies of data that it fits involve constant jumps and swings, but I manage to fit everything without postulating that the rules themselves shift and change. Thus, my models are acknowledged as having opened a path toward a new and more realistic economics. Nevertheless, the progress of this new economics is slow, due to inherent mathematical difficulties and to my failure to push its development. All too often (though with notable exceptions, e.g. in Samuelson 1967), a publication devoted to the totally discredited Gaussian model quotes my work largely to show the author's awareness of his own work's limitations, and possibly to assuage his conscience.

&&&&&&&&&&&& ANNOTATIONS &&&&&&&&&&&&

Comment on a generalized L-stability. To avoid unnecessary complication, this chapter does not write down the general definition of L-stability, which adds a constant to the right-hand side of (S). The effect is to introduce additional L-stable laws with $\alpha = 1$; they are of possible relevance in the study of business concentration (Chapter E13).

How this paper came to be written. This paper's publication story is unique in my record, since it led a serious economics journal to follow a practice that is reputed to characterize *The Reader's Digest*. The story begins with M 1962i, IBM Research Note NC-87, dated March 26, 1962. I followed that text when teaching in the Harvard Department of Economics in 1962-63 and it was discussed at the 1962 Annual Meeting of the Econometric Society (see Chapter E17).

Later in 1963, after I had moved from the Harvard Economics Department to the Harvard Applied Physics program, Paul Cootner called from MIT. His book of reprints on *The Random Character of Stock Market Prices* was nearing completion, and he was sorry I was not a contributor. He was very familiar with my Report NC-87 and also knew that I was planning to expand this report into a short book. He wanted me to contribute to his book a short text based on NC-87, but did not want to include unpublished pieces, only actual reprints. Could my work be rushed into print somewhere, anywhere, so he could then reprint it? His publisher would accept page proofs.

I called every economics journal editor in turn. Some asked me to spell my name, others inquired about my field of work, a few knew me, but bemoaned the length of their backlog and the slowness of their referees. Finally, I struck gold with Merton Miller, an editor of the *Journal of Business* of the University of Chicago. He asked for a few hours to check a few things, then called back with a deal: NC87 was already well-known in Chicago, therefore no refereeing was needed; if I could manage to mail a rough version of the paper within a week, he would stop an issue about to go to press, and add my paper to it. The journal would even provide editorial assistance, and there would be no bill for "excess" corrections in proof. This deal could not be turned down, and the paper and its reprint became widely known on the research side of the financial community. At one time, a reprint combining M 1963b and Fama 1963 was given as premium to new subscribers of the *Journal of Business*.

Belated acknowledgement. Only after the paper had appeared did Merton Miller tell me that the editor Miller selected was E. F. Fama, who was no longer my student but on the University of Chicago faculty. Had this information been available in advance, I would have acknowledged Fama's help in my paper. I thanked him verbally, but this was not enough. To thank him in writing, late is better than never.

The Journal of Business 40, 1967, 393-413

E15

The variation of the prices of cotton, wheat, and railroad stocks, and of some financial rates

• *Chapter foreword.* M 1963b{E14} argues that the description of price variation requires probability models less special than the widely used Brownian, because the price relatives of *certain* prices series have a variance so large that it may in practice be assumed infinite. This theme is developed further in the present chapter, which covers the following topics.

1. Restatement of the M 1963{E14} model of price variation, and additional data on cotton.
2. The variation of wheat price in Chicago, 1883-1936.
3. The variation of some railroad stock prices 1857-1936.
4. The variation of some interest and exchange rates.
5. Token contribution to the statistical estimation of the exponent α.

Much of the empirical evidence in this paper was part of IBM Research Note NC-87 (March 1962), from which M 1963b{E14} is also excerpted •

THIS CHAPTER CONTINUES Chapter E14, to be referred to as VCSP.

{P.S. 1996: In view of the current focus on serial dependence, special interest attaches to Section 4. Indeed, changes in interest and exchange rates cannot possibly be independent, hence cannot follow the M 1963 model. The sole question tested in this chapter is whether or not the marginal distribution is L-stable, irrespective of serial dependence. The variation of those records in time may be best studied by the methods that Chapter E11 uses for personal income.}

1. RESTATEMENT OF THE M 1963 MODEL OF PRICE VARIATION, AND ADDITIONAL DATA ON COTTON

One goal of this restatement is to answer certain reservations concerning my L-stable model of price variation. Trusting that those reservations will be withdrawn and not wishing to fan controversy, I shall name neither the friendly nor the unfriendly commentators.

1.1 Bachelier's theory of speculation

Consider a time series of prices, $Z(t)$, and designate by $L(t, T)$ its logarithmic relative

$$L(t, T) = \log_e Z(t, T) - \log_e Z(t).$$

The basic model of price variation, a modification of one proposed in 1900 in Louis Bachelier's theory of speculation, assumes that successive increments $L(t, T)$ are (a) random, (b) statistically independent, (c) identically distributed, and (d) Gaussian with zero mean. The process is called a "stationary Gaussian random walk" or "Brownian motion."

Although this model continues to be extremely important, its assumptions are working approximations that must not be made into dogmas. In fact, Bachelier 1914 made no mention of earlier claims of the empirical evidence in favor of Brownian motion. (To my shame, I missed this discussion when I first glanced through Bachelie 1914 and privately criticized him for blind reliance on the Gaussian. Luckily, my criticism was not committed to print.)

Bachelier noted that his original model contradicts the evidence in at least two ways: Firstly, the sample variance of $L(t, T)$ varies in time. He attributed this to variability of the population variance, interpreting the sample histograms as being relative to mixtures of distinct populations, and observed that the tails of the histogram could be expected to be fatter than in the Gaussian case. Second, Bachelier noted that no reasonable mixture of Gaussian distributions could account for the sizes of the very largest price changes, and he treated them as "contaminators" or "outliers." Thus, he pioneered not only in discovering the Gaussian random-walk model, but also in noting its major weakness.

However, new advances in theory of speculation continues to be best expressed as improvements upon the Brownian model. VCSP shows that an appropriate generalization of hypothesis (d) suffices to "save" (a), (b), and (c) in many cases, and in other cases greatly postpones the moment when the misfit between the data and the theory is such that the latter must be amended.. I shall comment upon Bachelier's four hypotheses, then come to the argument of VCSP.

1.2 Randomness

To say that a price change is random is not to claim that it is irrational, only that it was unpredictable before that fact *and* is describable by the powerful mathematical theory of probability. The two alternatives to randomness are "predictable behavior" and "haphazard behavior," where the latter term is taken to mean "unpredictable and not subject to probability theory." By treating the largest price changes as "outliers," Bachelier implicitly resorted to this concept of "haphazard." This might have been unavoidable, but the power of probability theory has since increased and should be used to the fullest.

1.3 Independence

The assumption of statistical independence of successive $L(t, T)$ is undoubtedly a simplification of reality. It was surprising to see VCSP criticized for expressing blind belief in independence. For examples of reservations on this account, see its Section VII as well as the final paragraph of its Sections III E, III F, and IV B. In defense of independence, I can offer only one observation; very surprisingly, models making this assumption account for many features of price behavior.

Incidentally, independence implies that no investor can use his knowledge of past data to increase his expected profit. But the converse is *not* true. There exist processes in which the expected profit vanishes, but dependence is extremely long range. In such cases, knowledge of the past may be profitable to those investors whose utility function differs from the market's. An example is the "martingale" model of M 1966b{E19}, which is developed and generalized in M 1969e. The latter paper also touches on various aspects of the spectral analysis of economic time series, another active topic whose relations with my work have aroused interest. For example, when a time series is non-Gaussian, its spectral whiteness, that is, the absence of correlation, is compatible with great departures from the random-walk hypothesis.

1.4 Stationarity

One implication of stationarity is that sample moments vary little from sample to sample, as long as the sample is sufficiently long. In reality, price moments often fail to behave in this manner. The notorious fact is understated in the literature, since "negative" results are seldom published – one exception being Mills 1927.

Figure 1 adds to VCSP, and displays the enormous variability in time of the sample second moment of cotton prices in the period of 1900-1905. The points refer to successive fifty-day sample means of $[L(t, 1)]^2$, where the $L(t, 1)$ are the daily price relatives. In Brownian motion, these sample means would have stabilized near the population mean. Since no stabilization is in fact observed, we see conclusively that the price of cotton did not follow a Gaussian stationary random walk.

The usual accounts for this variability claim that the mechanism for price variation changes in time. We shall, loosely speaking, distinguish systematic, random, and haphazard changes of mechanism.

The temptation to refer to systematic changes is especially strong. Indeed, to explain the variability of the statistical parameters of price variation would constitute a worthwhile first step toward an ultimate explanation of price variation itself. An example of systematic change is given by the yearly seasonal effects, which are strong in the case of agricultural commodities. However, Figure 1 goes beyond such effects: not all ends of season are accompanied by large price changes, and not all large price changes occur at any prescribed time in the growing season.

The most controversial systematic changes are due to deliberate changes in the policies of the Federal Government or of the Exchanges. For an example of unquestionable long-term change of this type, take cotton prices (Section III D of VCSP.) All measures of scale of $L(t, T)$ (such as the interquartile interval) did vary between 1816 and 1958. Indeed, lines 1*a* and 2*a* of Figure 5 of VCSP, which are relative to the 1900's, clearly differ from lines 1*b* and 2*b*, which are relative to the 1950's. This clearcut decrease in price variability must, at least in part, be a consequence of the deep changes in economic policy that occurred in the early half of this century. However, precisely because it is so easy to read in the facts a proof of the success or failure of changes in economic policy, the temptation to resort to systematic nonstationarity must be carefully controlled.

{P.S. 1996. The extremely cautious phrasing of the last sentences was rewarded: the large apparent decrease in volatility from 1904 to 1952 proved in fact to be caused by misleadingly presented data; see M 1972b, i.e., Appendix III of Chapter E14}.

In order to model what they perceived to be a "randomly changing" random process, many authors have invoked a random-walk process in which the sizes and probabilities of the steps are chosen by some other process. If this second "master process" is stationary, $Z(t)$ itself is not a random walk but remains a stationary random process.

The final possibility is that the variability of the price mechanism is haphazard, that is, not capable of being treated by probability theory. This belief is, of course, firmly entrenched among nonmathematical economists. But to construct a statistical model one expects to change before it has had time to unfold, can hardly be viewed as a sensible approach.

Moreover, and more importantly, early resort to haphazard variation *need not* be necessary, as is demonstrated by the smoothness and regularity of the graph of Figure 2, which is the histogram of the data of Figure 1.

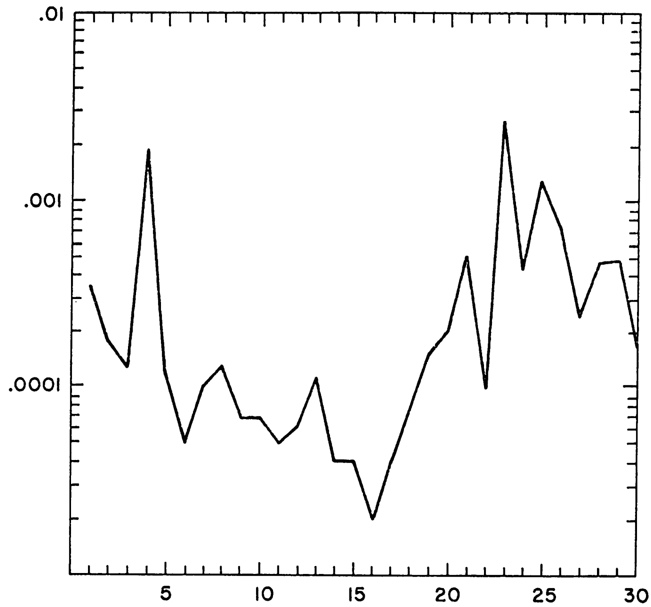


FIGURE E15-1. Sample second moment of the daily change of log $Z(t)$, where $Z(t)$ is the spot price of cotton. The period 1900-1905 was divided into thirty successive fifty-day samples, and the abscissa designates the number of the sample in chronological order. Logarithmic ordinate. A line joins the sample points to improve legibility.

1.5 Gaussian hypothesis

Bachelier's assumption, that the marginal distribution of $L(t, T)$ is Gaussian with vanishing expectation, might be convenient, but virtually every student of the distribution of prices has commented on their leptokurtic (i.e., very long-tailed) character. For an old but eminent practitioner's opinion, see Mills 1927; for several recent theorists' opinions see Cootner 1964. It was mentioned that Bachelier himself regarded $L(t, T)$ as a contaminated mixture of Gaussian variables; see M and Taylor 1967{E22, Sections 1 and 2}.

1.6 Infinite population variance and the L-stable distributions

Still other approaches were suggested to take into account the failure of Brownian motion to fit data on price variation. In all these approaches, each new fact necessitates an addition to the explanation. Since a new set of parameters is thereby added, I don't doubt that reasonable "curve-fitting" is achievable in many cases.

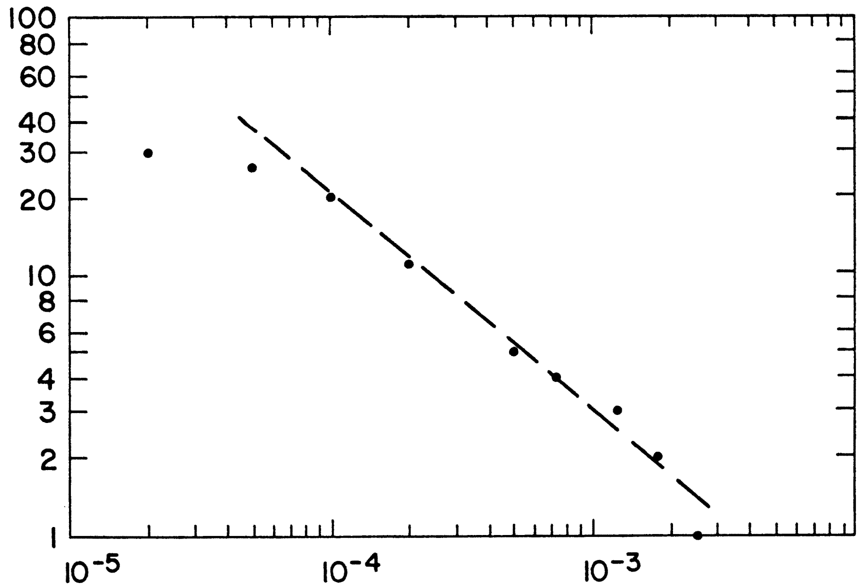


FIGURE E15-2. Cumulated absolute-frequency distribution for the data of Figure 1. Abscissa: log of the sample second moment. Ordinate: log of the absolute number of instances where the sample moment marked as abscissa has been exceeded. The L-stable model predicts a straight line of slope $\alpha/2 \sim 1.7/2$, which is plotted as a dashed line.

However, this form of "symptomatic medicine" (a separate drug for each complaint) could not be the last word! The beauty of science – and the key to its effectiveness – is that it sometimes evolves central assumptions, from which many independently made observations can be shown to follow. These observations are thus organized, and predictions can be made. The ambition of VCSP was to suggest such a central assumption, the infinite-variance hypothesis, and to show that it accounts for substantial features of price series (of various degrees of volatility) without nonstationarity, without mixture, without master processes, without contamination, but with a choice of increasingly accurate assumptions about the interdependence of successive price changes.

When selecting a family of distributions to implement the infinite-variance hypothesis, one must be led by mathematical convenience (e.g.,

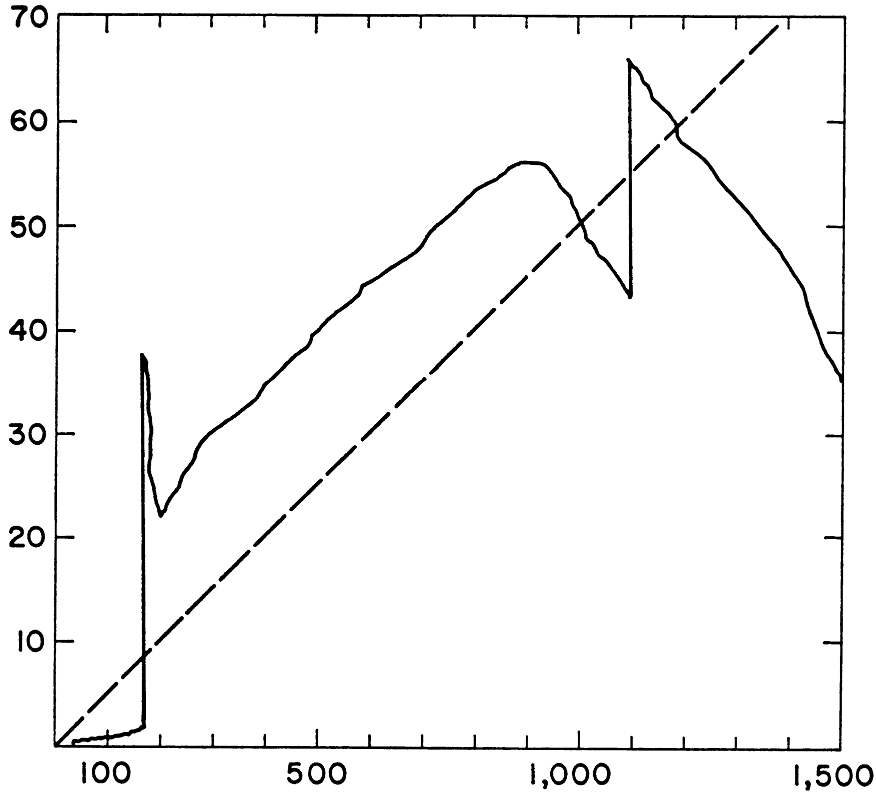


FIGURE E15-3. Sequential variation of the sample kurtosis of the daily changes of log $Z(t)$, where $Z(t)$ is the spot price of cotton, 1900-1905. The abscissa is the sample size. Linear coordinates.

the existence of a ready-made mathematical theory) and by simplicity. For a probability distribution, one important criterion of simplicity is the variety of its properties of "invariance." For example, it would be most desirable to have the same distribution (up to some – hopefully linear – weighting) apply to daily, monthly, etc., price changes. Another measure of simplicity is the role that a family of distributions plays in central limit theorems of the calculus of probability.

In accordance with this logic, VCSP proposed to represent the marginal distribution of $L(t, T)$ by an appropriate member of a family of probability distributions called "L-stable." The L-stable laws measure volatility by a single parameter α ranging between 2 and 0. The simplest members of the family are the symmetric probability densities defined by

$$\rho_\alpha(u) = (1/\pi)\int_0^\infty \exp(-\gamma s^\alpha)\cos(su)ds.$$

Their limit case $\alpha = 2$ is the Gaussian, but my theory also allows non-Gaussian or "L-stable" cases $\alpha < 2$. Suitable α's turn out to represent satisfactorily the data on volatile prices (see VCSP and Section 2 and 3 below).

In assessing the realm of applicability of the M 1963 model, , one should always understand it as including its classical limit. It is therefore impossible to "disprove" VCSP by identifying out price series for which the Gaussian hypothesis may be tenable.

Now to discuss the fact that L-stable variables with $\alpha < 2$ have an infinite population variance; mathematicians sometimes say that they have "no variance." Firstly, one must reassure those who expressed the fear that the sole reason for my finding $E(L^2) = \infty$ was that I inadvertently took the logarithm of zero! Serious concern was expressed at the implication of this feature for statistics, and surprise was expressed at the paradoxically discontinuous change that seems to occurs when α becomes exactly 2.

This impression of paradox is unfounded. The population variance itself cannot be measured, and every measurable characteristic of a L-stable distribution behaves continuously near $\alpha = 2$ as will be seen later in an example. Consequently, there is no "black and white" contrast between the scaling case $\alpha < 2$ and the Gaussian case $\alpha = 2$, but a continuous shading of gray as $\alpha \rightarrow 2$. The finding that the population second moment is discontinuous at $\alpha = 2$ "only" shows that this moment is not well suited to a study of price variation.

In particular, the applicability of second-order statistical methods is questionable. This word could *not* mean "totally inapplicable," because the statistical methods based upon variances suffer no sudden and catastrophic breakdown as α ceases to equal 2. Therefore, to be unduly concerned with a few specks of "gray" in a price series whose α is near 2, may be as inadvisable as to treat very gray series as white. Moreover, statistics would be unduly restricted if its tools were to be used only where they have been fully justified. (As a matter of fact, the quality of statistical method is partly assessed by its "robustness," i.e., the quality of its performance when used without justification.) However, one should look for other methods. For example, as predicted, least-squares forecasting (as applied to past data) would often have led to very poor inferences; least-sums-of-absolute-deviations forecasting, on the other hand, is always at least as good and usually much superior, and its development should be pressed.

1.7 The behavior of the variance L-stable samples and cotton

Define $V(\alpha, N)$ as the variance of a sample of N independent random variables $U_1, \ldots U_n, \ldots U_N$, whose common distribution is L-stable of exponent α. To obtain a balanced view of the practical properties of such variables, one must *not* focus upon mathematical expectations and/or infinite sample sizes. Instead, one should consider quantiles and samples of large but finite size. Let us therefore select a "finite horizon" by choosing a value of N and a quantile threshold q such that events whose probability is below q will be considered "unlikely." Save for extreme cases contributing to a "tail," of probability q, the values of $V(\alpha, N)$ will be less than some function $V(\alpha, N, q)$. This function's behavior tells us much of what we need to know about the sample variance.

As mentioned earlier, when N is finite and $q > 0$, the function $V(\alpha, N, q)$ varies smoothly with α. For example, over a wide range of values of N, the derivative of $V(\alpha, N, q)$ at $\alpha = 2$ is very close to zero, hence $V(\alpha, N, q)$ changes very little from $\alpha = 2.00$ to $\alpha = 1.99$. This insensitivity is due to the fact that $\alpha = 2.00$ and $\alpha = 1.99$ differ only in the sizes that they predict for some outliers; but those outliers belong to those cases whose effects were excluded by the definition of $V(\alpha, N, q)$. Increasing N or decreasing q, decreases the range of exponents in which α is approximable by 2.

A reader who really objects to infinite variance, and is only concerned with meaningful finite-sample problems, may "truncate" U so as to attribute to its variance a very large finite value depending upon α, N, and q.

The resulting theory may have the asset of familiarity, but the specification of the value of the truncated variance will be *useless* because it will tell nothing about the "transient" behavior of $V(\alpha, N, q)$ when N is finite and small. Thus, even when one knows the variance to be finite but very large (as in the case of certain of my more detailed models of price variation; see M 1969e), the study of the behavior of $V(\alpha, N, q)$ is much simplified if one approximates the distribution with finite but very large variance by a distribution with infinite variance. Similarly, it is well known that photography is simplest when the object is infinitely far from the camera. Therefore, the photographer can set the distance at infinity if the actual distance is finite but exceeds some finite threshold dependent on the quality of the lens and its aperture.

1.8 The behavior of the kurtosis for L-stable samples and cotton prices

Pearson's kurtosis measures the peakedness of a distribution by

$$\frac{E(U^4)}{\left[E(U^2)\right]^2} - 3.$$

The discussion of this quantity in VCSP was called obscure, therefore additional detail may be useful. If U is L-stable with $\alpha < 2$, the kurtosis is undetermined, because $E(U^4) = \infty$ and $E(U^2) = \infty$, One can show, however, that, as $N \rightarrow \infty$, the random variable

$$\sum_{n=1}^{N} U_n^4 \left\{ \sum_{n=1}^{N} U_n^2 \right\}^{-2}$$

tends toward a limit that is finite and different from zero. Therefore, the "expected sample kurtosis," defined as

$$E\left\{ N \sum_{n=1}^{N} U_n^4 \left\{ \sum_{n=1}^{N} U_n^2 \right\}^{-2} - 3 \right\},$$

is asymptotically proportional to N.

The kurtosis of $L(t, 1)$ as plotted on Figure 3 for the case of cotton, 1900-1905, indeed increases steeply with N. While exact comparison is impossible because the theoretical distribution was not tabulated, this kurtosis does fluctuate around a line expressing proportionality to sample size. (For samples of less than fifty, the kurtosis is negative, but too small to be read off the figure.)

1.9 Three approximations to a L-stable distribution: implications for statistics and for the description of the behavior of prices

It is important that there should exist a single theory of prices that subsumes various degrees of volatility. My theory is, unfortunately, hard to handle analytically or numerically, but simple approximations are available in different ranges of values of α Thus, given a practical problem with a finite time horizon N, it is best to replace the continuous range of degrees of "grayness" by the following trichotomy (where the boundaries between the categories are dependent upon the problem in question).

The Gaussian $\alpha = 2$ is the best known and simplest. There is no need to worry about a long-tailed distribution, hence one stands a reasonable chance of rapid progress in the study of dependence. For example, one can use spectral methods and other covariance-oriented approaches. Very close to $\alpha = 2$, Gaussian techniques cannot lead one too far astray.

In the zone far away from $\alpha = 2$, another kind of simplicity reigns. Substantial tails of the L-stable distribution are approximations by the scaling distributions with the same α-exponent ruling both tails. A prime example was provided by the cotton prices studied in VCSP. Sections 2 and 3 examine some other volatile price series: wheat, the prices of some nineteenth-century rail securities and some rates of exchange of interest.

The zone of transition between the almost Gaussian and the clearly scaling cases is by far the most complicated of the three zones. It also provides a test of the meaningfulness and generality of the M 1963 model. If it holds, the histogram of price changes is expected to plot on biogarithmic paper as one of a specific family of inverse-S-shaped curves. (Lévy's α-exponent, therefore, is not to be confused with slope of a straight bilogarithmic plot, M 1963e{E3, Appendix III}.) If the M 1963 model fails, the transition between the almost Gaussian and the highly scaling cases would be performed in some other way.

We shall examine in this light the variation of wheat prices and find that it falls into the "light gray" zone of low but positive values for $2 - \alpha$ and medium volatility. Section 2.1 examines wheat data; it is similar in purpose to Fama's 1964 Chicago thesis, Fama 1965, which was the first further test of the ideas of VCSP. To minimize "volatility" and maximize the contrast with my original data, Fama chose thirty stocks of large and diversified contemporary corporations and found their L-stable "grayness" to be unquestionable, although less marked than that of cotton.

2. THE VARIATION OF WHEAT PRICES IN CHICAGO, 1883-1936

Spot prices of cotton refer to standardized qualities of this commodity, but wheat cash prices refer to variable grades of grain. At any given time (say, at closing time), one can at best speak of a *span* of cash prices, and the closing spans corresponding to successive days very often overlap. As a result, Working 1934 chose the week as the shortest period for which one can reasonably express "the" change of wheat price by a single number rather than by an interval. Further interpolation being impossible, the long record from 1883 to 1936 yields a smaller sample than one might have hoped – though a very long one by the standards of economics.

Kendall 1953 suggested that Working's wheat price relatives follow a Gaussian distribution. After all, a casual visual inspection of the histograms of these relatives, as plotted on *natural* coordinates, shows them to be nicely "bell shaped." However, natural coordinates notoriously underestimate the "tails." To the contrary, as seen in Figure 4, probability-paper plots of wheat price relatives are definitely **S**-shaped, though less so than for cotton. As the Gaussian corresponds to $\alpha = 2$ and M 1963b{E14} reports the value $\alpha = 1.7$ for cotton, it is natural to investigate whether wheat is L-stable with an α somewhere between 1.7 and 2.

2.1 The evidence of doubly logarithmic graphs

Figure 3 of VCSP shows that the L-stable distribution predicts that every doubly logarithmic plot of a histogram of wheat price changes should have a characteristic **S**-shape. It would end with a "scaling" straight line of slope near 2, but would start with a region where the local slope increases with u and even begins by markedly exceeding its asymptotic value, M 1963p.

The above conjecture is verified, as seen in Figures 5 and 6. Moreover, by comparing the data relative to successive subsamples of the period 1883-1936, no evidence was found that the *law* of price variation had changed in kind, despite the erratic behavior of the outliers.

When this test was started, one only knew that wheat lies between the highly erratic cotton series and the minimally erratic Gaussian limit.

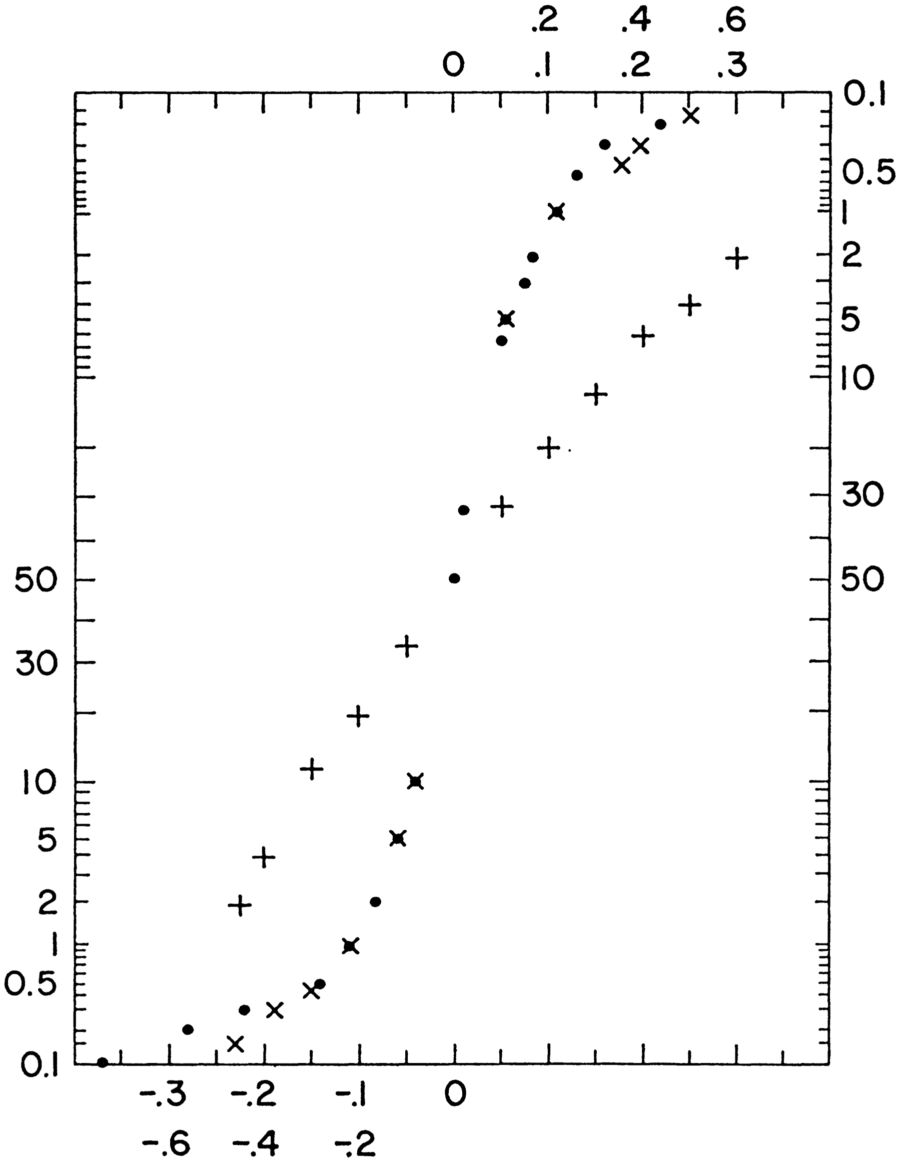


FIGURE E15-4. Probability-paper plots of the distribution of changes of log $Z(t)$, where $Z(t)$ is the spot price of wheat in Chicago, 1883-1934, as reported by Working 1934. The scale − 0.3, − 0.2, − 0.1, 0, 0.1, 0.2, 0.3 applies to the weekly changes, marked by dots, and to the yearly changes, marked by crosses. The other scale applies to changes over lunar months, marked by ×.

Therefore, Figures 5 and 6 are evidence that the L-stable model *predicted* how the price histogram of wheat price changes "should" behave.

To establish the "goodness of fit" of such an S-shaped graph requires a larger sample than in the case of the straight graphs characteristic of cotton. But the available samples are actually smaller. Thus the doubly logarithmic evidence is *unavoidably* less clear-cut than for cotton.

2.2 The evidence of sequential variance

When price series is approximately stationary, one can test whether $\alpha = 2$ or $\alpha < 2$ by examining the behavior of the sequential sample second

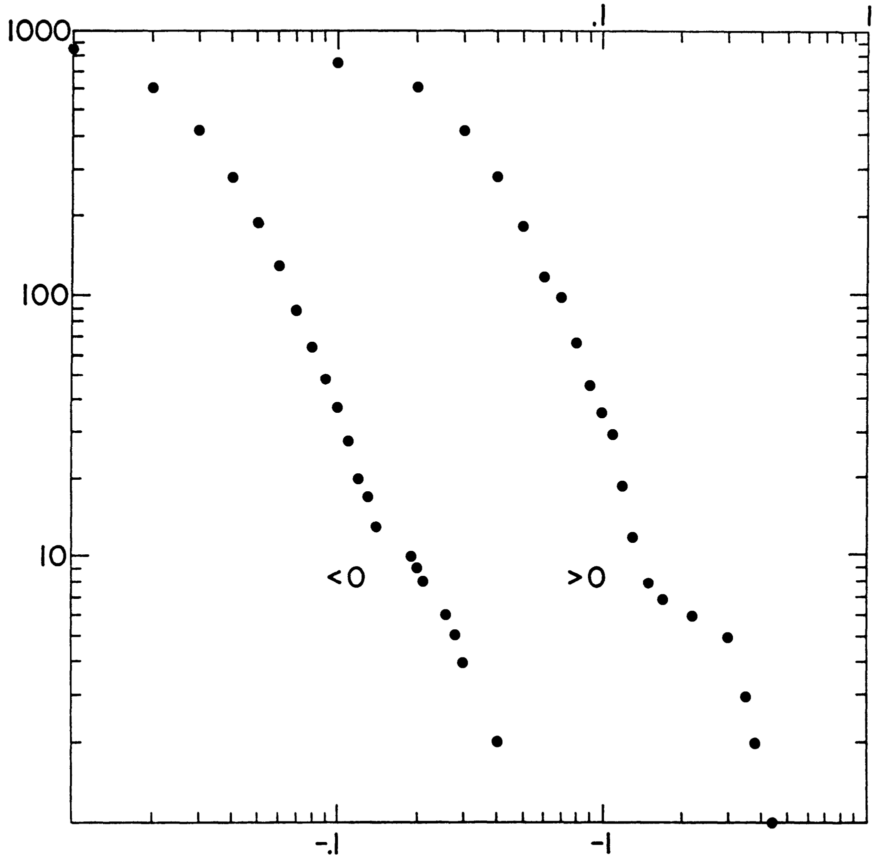


FIGURE E15-5. Weekly changes of log $Z(t)$, where $Z(t)$ is the price of wheat as reported by Working 1934. Ordinate: log of the absolute frequency with which $L \geq u$, respectively $L \leq -u$. Abscissas: the lower scale refers to negative changes, the upper scale to positive changes.

moment. If $\alpha < 2$, the median of the distribution of the sample variance increases as $N^{-1+2/\alpha}$ for "large" N. If $\alpha = 2$, it tends to a limit. *More importantly,* divide the sample variance by its median value; this ratio's variation becomes increasingly erratic as α moves away from 2. Thus, while the cotton second moment increases very erratically, but the wheat second moment should increase more slowly and more regularly. Figure 7 shows that such is indeed the case.

2.3 Direct test of L-stability

The term "L-stable" arose from the following fact: when N such random variables U_n are independent and identically distributed, one has

$$\Pr\left\{N^{-1/\alpha}\sum_{n=1}^{N} U_n \geq u\right\} = \Pr\{U_n \geq u\}.$$

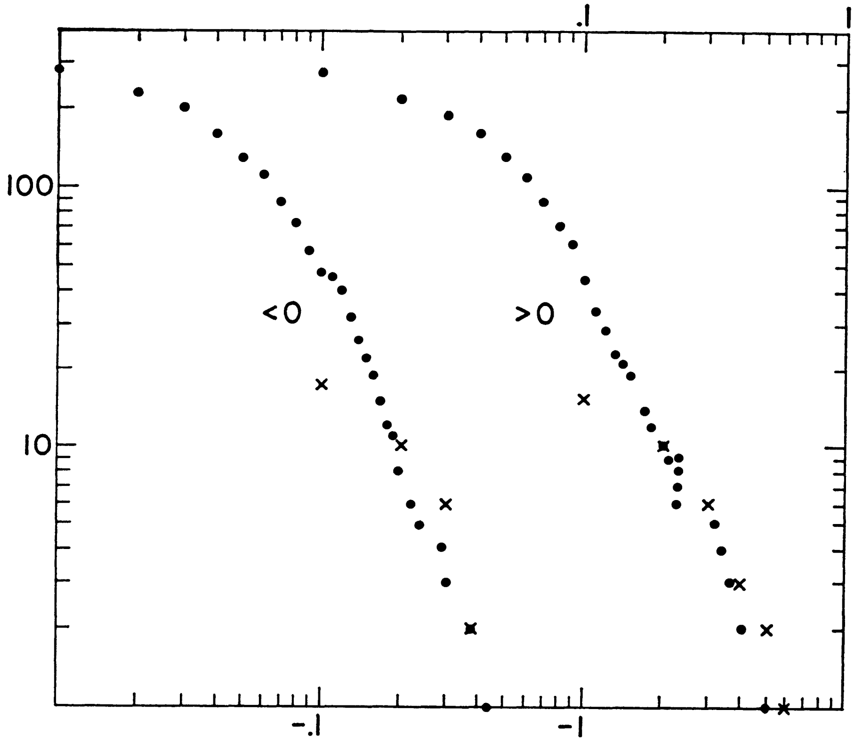


FIGURE E15-6. Changes of log $Z(t)$ over lunar months and years, where $Z(t)$ is the price of wheat as reported in Working 1935. Abscissas and ordinates as in Figure 5.

I settled on $N = 4$. When the random variables U_n are the weekly price changes, $\sum_{n=1}^{4} U_n$ is the price change over a "lunar month" of four weeks. Since α is expected to be near 2, the factor $4^{-1/\alpha}$ will be near 1/2.

One can see in Figure 4 that weekly price changes do indeed have an S-shaped distribution *indistinguishable* from that of one-half of monthly changes. (The bulk of the graph, corresponding to the central bell containing 80% of the cases, was not plotted for the sake of legibility.) Sampling fluctuations are apparent only at the extreme tails and do not appear systematic. Applied in Fama 1965 to common-stock price changes, this method also came in favor of L-stability.

The combination of Figures 5 and 6 provides another test of L-stability. They were plotted with absolute, not relative, frequency as the ordinate, and the L-stable theory predicts that such curves should be

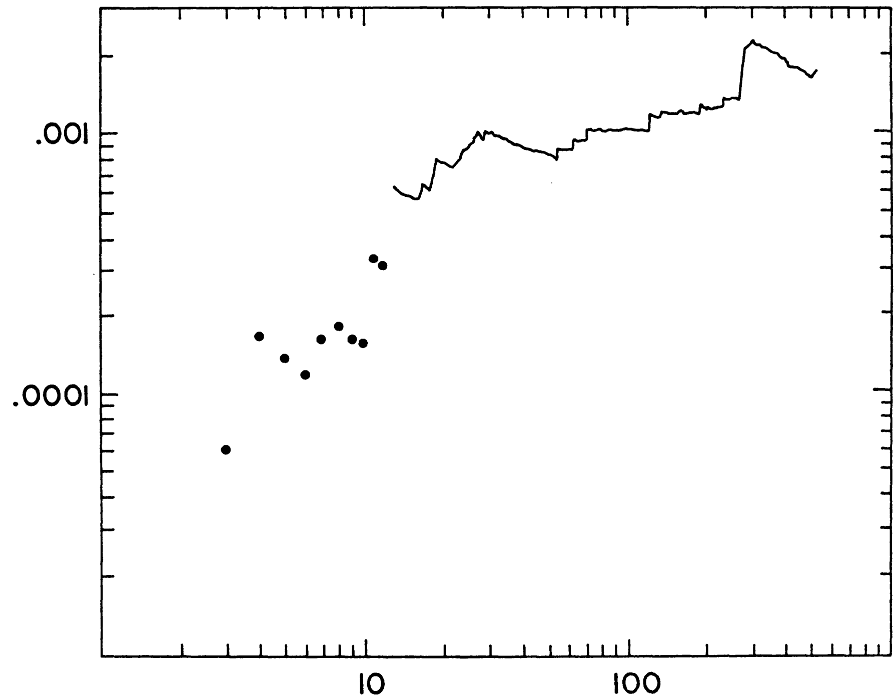


FIGURE E15-7. Sequential variation of the second moment of the weekly changes of log $Z(t)$, where $Z(t)$ is the price of wheat as reported in Working 1934. One thousand weeks beginning in 1896. Bilogarithmic coordinates. For small samples, the sample second moments are plotted separately; for lay samples, they are replaced (for the sake of legibility) by a freely drawn continuous line.

superposable in their tails, except, of course, for sampling fluctuations. (In outline, the reason for this prediction is that, in a L-stable universe, a large monthly price change is of the same order of magnitude as the largest among the four weekly changes adding to this monthly change.) Clearly, wheat data pass this second test also.

It should be stressed that, while the two tests use the same data, they are conceptually *distinct*. Figure 4 compares one-half of a monthly change to the weekly change of the same frequency; Figures 5 and 6 compare monthly and weekly changes of the same size. Stability is thus doubly striking.

2.4 The evidence of yearly price changes

Working 1934 also published a table of average January prices of wheat, and Figure 4 also included the corresponding changes of log $Z(t)$.

Assuming that successive weekly price changes are independent, the evidence of the yearly changes again favors L-stability. It is astonishing that the hypothesis of independence of weekly changes can be consistently carried so far, showing no discernible discontinuity between long-term adjustments to follow supply and demand, which would be the subject matter of economics, and the short-term fluctuations that some economists discuss as "mere effects of speculation."

3. THE VARIATION OF RAILROAD STOCKS PRICES, 1857-1936

For nineteenth-century speculators, railroad stocks were preeminent among corporation securities, and played a role comparable to that of the basic commodities. Unfortunately, Macaulay 1932 reports them incompletely: for each major stock, it gives the mean of the highest and lowest quotation during the months of January; for each month, it gives a weighted index of the high and low of every stock.

I began by examining the second series, even though it is averaged too many times for comfort. If one considers that there "should" have been no difference in kind between various nineteenth-century speculations, one would expect railroad stock changes to be L-stable, and averaging would bring an increase in the slope of the corresponding doubly logarithmic graphs, similar to what has been observed in the case of cotton price averages (Section III E of VCSP). Indeed, Figure 8, relative to the variation of the monthly averages, yields precisely what one expects for averages of L-stable processes with an exponent very close to that of cotton.

Yearly data, to the contrary, are little affected by averaging. Figure 9 should be regarded as made of two parts: the first five graphs concern companies with below average merger activity, the others to companies with above average merger activity.

The first five graphs, in my opinion, proved striking confirmations of the tools and concepts I developed in the study of cotton. Basically, one sees that the fluctuations of the price of these stocks were all L-stable, with the same α-exponent characteristic of the industry and clearly below the critical value 2. (Moreover, they all had practically the same value of the positive and negative "standard deviations," σ' and σ'', defined in VCSP.)

For the companies with an unusual amount of merger activity, the evidence is similar but more erratic.

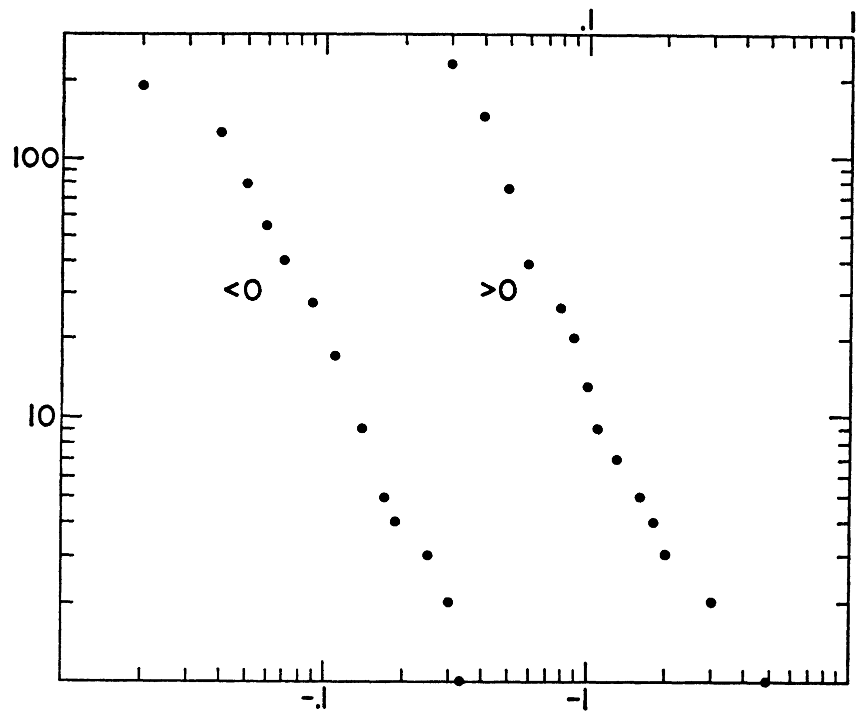


FIGURE E15-8. Monthly changes of log $Z(t)$, where $Z(t)$ is the index of rail stock prices, as reported in Macaulay 1932. Abscissas and ordinates as in Figure 5.

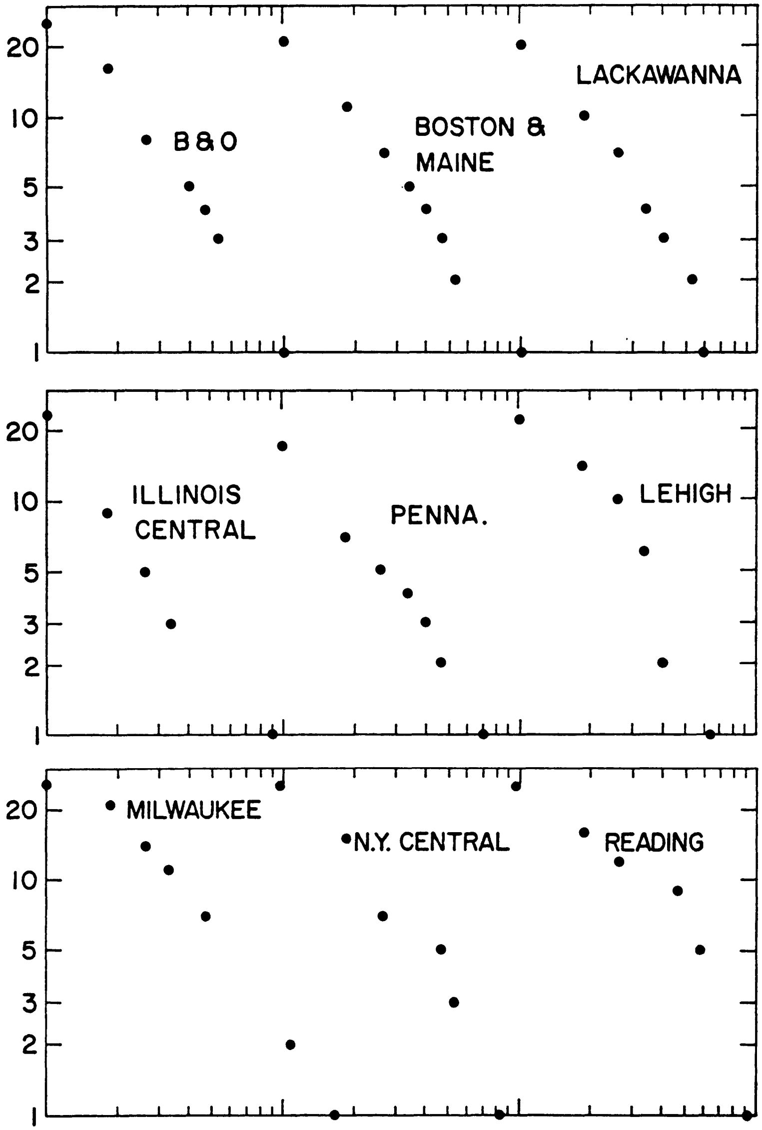


FIGURE E15-9. Yearly changes of log $Z(t)$, where $Z(t)$ is the index based on the prices of nine selected rail stocks in January, as reported in Macaulay 1932. Ordinates: absolute frequencies. Abscissas are not marked to avoid confusion: for each graph, they vary from 1 to 10.

4. THE VARIATION OF SOME INTEREST AND EXCHANGE RATES

Various rates of money – and especially the rate of call money in its heyday – reflect the overall state of the speculative market. One would therefore expect analogies between the behaviors of speculative prices and of speculative rates. But one cannot expect them to be ruled by identical process. For example, one cannot assume (even as rough approximations) that successive changes of a money rate are statistically independent: Such rates would eventually blow up to infinity, or they would vanish. Neither behavior is realistic. As a result, the distribution of $Z(t)$ itself is meaningless when Z is a commodity price, but it is meaningful when it is a money rate. Additionally, a 5% rate of money is a price near equal to 1.05, and for small values of Z, one has $\log(1+Z) \sim Z$. Therefore, in the case of money rates, one should study $Z(t+T) - Z(t)$ rather than $\log Z(t+T) - \log Z(t)$.

4.1 The rate of interest on call money

In Figure 10, the abscissa is based on the data in Macaulay 1932, concerning the excess of rate of call money over its "typical" value, 6%. I have not even attempted to plot the distribution of the other tail of the difference "rate minus 6%," since that expression is by definition very short-tailed, being bounded by 6 per cent, while the positive value of "rate minus 6%" can go sky high (and occasionally did.)

The several lines of Figure 10 correspond, respectively, to the total period 1857-1936 and to three subperiods. They show that call money rates are single-tailed scaling, with an exponent markedly smaller than 2. Scale factors (such as the upper quartile) have changed – a form of non-stationarity – but the exponent α seems to have preserved a constant value, lying within the range in which the scaling distribution is known to be invariant under mixing of data from populations having the same α and different γ; see M 1963e{E3}.

4.2 Other interesting money rates

Examine next the distribution of the classic data collected by Erastus B. Bigelow (Figure 11, dashed line) relative to "street rates of first class paper in Boston" (and New York) at the *end* of each month from January 1836 to December 1860. (Bigelow also reports some rates applicable at the beginnings or middles of the same months, but I disregarded them to avoid the difficulties due to averaging.) The dots on Figure 11 again represent the difference between Bigelow's rates and the typical 6%; their behavior is what we would expect if essentially the same scaling law applied to these rates and those of call money.

Finally, examine a short sample of rates, reported by Davis 1960, on the basis of records of New England textile mills. These rates remained much closer to 6% than those of Bigelow. They are plotted in such a way that the crosses of Figure 11 represent ten times their excess over 6%. The sample is too short for comfort, but, until further notice, it suggests that the two series have differed mostly by their scales.

4.3 The dollar-sterling exchange in the nineteenth century

The exchange premium or discount in effect on a currency exchange seems to reflect directly the difference between the various "forces" that condition the variations of the values of the two currencies taken separately.

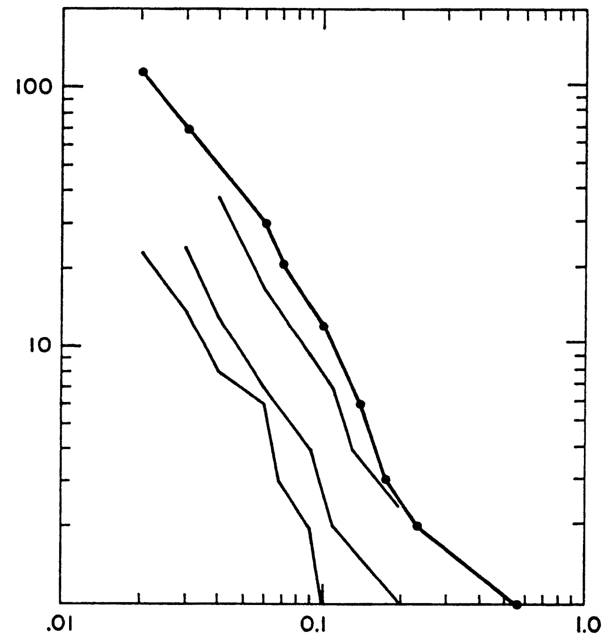


FIGURE E15-10. The distribution of the excess of over 6% of the monthly average of call money rates as reported in Macaulay 1932. Ordinates: absolute frequencies. *Bold line*: total sample 1857-1936. *Thin lines*, read from left to right: subsamples 1877-97, 1898-1936, 1857-76. Note that the second subsample is twice as long as the other two. Thus, the general shape of the curves has not changed except for the scale, and the scale has steadily decreased in time.

This differential quantity has an advantage over the changes of rates: one can consider it without resorting to any kind of economic theory, not even the minimal assumption that price changes are more important than price levels. We have therefore plotted the values of the premium or discount between dollar and sterling between 1803 and 1895, as reported by Davis and Hughes 1960 (Figure 12). This series is based upon operations which involved credit as well as exchange. In order to eliminate the credit component, the authors used various series of money rates. We also plotted the series based upon Bigelow's rates. Note that all the graphs of Figure 11 conform strikingly with the expectations generalized from the known behavior of cotton prices.

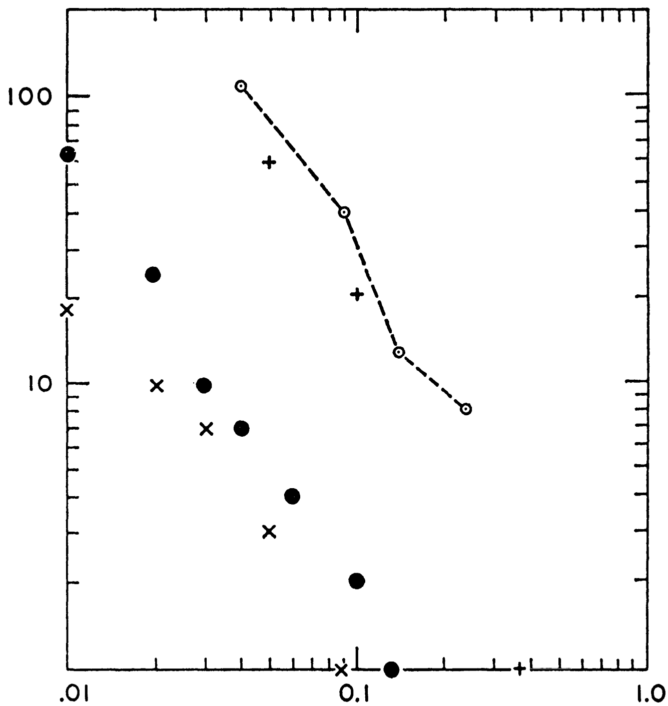


FIGURE E15-11. Four miscellaneous distributions of interest and exchange rates. Reading from the left: two different series of dollar-sterling premium rates. *Crosses*: ten times the excess over 6% of Davis's textile interest rate data. *Dashed line*: excess over 6% of Bigelow's money rates. The four series, all very short, were chosen haphazardly. The point of the figure is the remarkable similarity between the various curves.

5. MAXIMUM LIKELIHOOD ESTIMATION OF α NEAR α= 2

A handicap for the theory of VCSP is that no closed analytic form is known for the L-stable distributions, nor is a closed form ever likely to be discovered. Luckily, the cases where the exponent α is near 1.7 can be dealt with on the basis of an approximating hyperbolic distribution. Now let α be very near 2. To estimate $2-\alpha$ or to test $\alpha=2$ against $\alpha<2$ is extremely important, because $\alpha=2$ corresponds to the Gaussian law and differs "qualitatively" from other values of α. To estimate such an α is very difficult, however, and the estimate will intrinsically be highly dependent upon the number and the "erratic" sizes of the few most "outlying" values of u_n. I hope to show in the present section that simplifying approximations are fortunately available for certain purposes. The main idea is to represent a L-stable density as a sum of two easily manageable expressions, one of which concerns the central "bell," while the other concerns the tails.

5.1 A square central "bell" with scaling tails

The following probability density can be defined for $3/2<\alpha<2$:

$$p(u)=\alpha-3/2 \text{ if } |u|\leq 1 \qquad \text{(adding up to } 2\alpha-3\text{);}$$
$$p(u)=(2-\alpha)\alpha|u|^{-(\alpha+1)} \text{ if } |u|>1 \qquad \text{(adding up to } 4-2\alpha\text{).}$$

When α is near 2, $p(u)$ is a rough first approximation to a L-stable density that lends itself to maximum-likelihood estimation.

Order $u_1, \ldots u_n, \ldots, u_N$, a sample of values of U, by decreasing absolute size, and denote by M the values such that $|u_n|\geq 1$. The likelihood function is

$$\prod p(u_n)=(\alpha-\frac{3}{2})^{N-M}[(2-\alpha)\alpha]^M\left\{\prod_{n=1}^{M}|u_n|\right\}^{-(\alpha+1)}.$$

The logarithm of the likelihood is

$$L(\alpha)=(N-M)\log(\alpha-\frac{3}{2})+M\log[(2-\alpha)\alpha]-(\alpha+1)\sum_{n=1}^{M}\log|u_n|.$$

This $L(\alpha)$ is a continuous function of α. If $M = 0$, it is monotone increasing and attains its maximum for $\alpha = 2$. This is a reasonable answer, since $|U| < 1$ for $\alpha = 2$.

To the contrary, if $M > 0$, $L(\alpha)$ tends to $-\infty$ as $\alpha \to 3/2$ or $\alpha \to 2$. Therefore, $L(\alpha)$ has at least one maximum, and its most likely value $\hat{\alpha}$ is the root of the third-degree algebraic equation

$$\frac{N-M}{\alpha-3/2} - \frac{M}{2-\alpha} + \frac{M}{\alpha} - \sum_{n=1}^{M} \log |u_n| = 0.$$

Thus, $\hat{\alpha}$ only depends on M/N and $M^{-1}\Sigma_{n=1}^{M} \log |u_n| = V$.

For the latter $\log |U|$, conditioned by $\log |U| > 1$, satisfies

$$\Pr\{ \log |U| > u | \log |U| > 0\} = \exp(-\alpha u).$$

Its expected value is $1/\alpha$. Therefore, as $M \to \infty$ and $N \to \infty$,

the terms $\frac{1}{\alpha} - \frac{1}{M} \sum_{n=1}^{M} \log |u_n|$ will tend to 0,

and can be neglected *in the first approximation.* When both M and N are large, the equation in $\hat{\alpha}$ simplifies to the first degree and

$$\hat{\alpha} = 2 - M/2N.$$

This value depends on the u_n through the ratio of these numbers in the two categories $|U| < 1$ and $|U| > 1$, i.e., the relative number of the "outliers" defined by $|U| > 1$.

For example, if M/N is very small, $\hat{\alpha}$ is very close to 2. As N/M barely exceeds 1, $\hat{\alpha}$ nears 3/2. However, this is range in which $p(u)$ is a very poor approximation to a L-stable probability density.

It may be observed that, knowing N, M/N is symptotically Gaussian, and so is $\hat{\alpha}$ for all values of α.

In a second approximation, valid for α near 2, one will use $\alpha = 2$ to compute

$$W = \frac{1}{\alpha} = \frac{1}{M} \sum \log |u_n|.$$

The equation for $\hat{\alpha}$ is now of the second order. One root is very large and irrelevant; the other root is such that $\alpha - (2 - M/2N)$ is proportional to W.

5.2 Scope of estimation based upon counts of outliers

The method of Section 5.1, namely, estimation of $\hat{\alpha}$ from M/N, applies without change under a variety of seemingly generalized conditions:

1. Suppose that the tails are asymmetric, that is,

$$p(u) = (2 - \alpha)\alpha p' u^{-(\alpha+1)} \qquad \text{if } u > 1$$
$$p(u) = (2 - \alpha)\alpha p'' |u|^{-(\alpha+1)} \qquad \text{if } u < -1,$$

where $p' + p'' = 1$. In estimating α, one will naturally concentrate upon the random variable $|U|$, which is the same as in Section 5.1

2. The conditional density of U, given that $|U| < 1$, may be non-uniform as long as it is independent of α. Suppose, for example, that for $|u| < 1, p(u) = (\alpha - 3/2)\, D \exp(-u^2/2\sigma^2)$, where $1/D(\sigma)$ is defined as equal to $\int_{-1}^{1} \exp(-s^2/2\sigma^2)ds$. The likelihood of α then equals

$$[D(\sigma)2^{-1}(\alpha - 3/2)]^{N-M} \exp\left(- \sum_{n=M+1}^{N} \frac{u_n^2}{2\sigma^2} \right) [(2 - \alpha)\alpha]^M \left(\prod_{n=1}^{M} |u_n| \right)^{-(\alpha+1)} .$$

As function of the U_n, the maximum likelihood is as in Section 5.1.

Acknowledgement. This text incorporates several changes suggested by Professor Eugene F. Fama.

The Journal of Business: 36, 1963, 420-429.

E16

Mandelbrot on price variation (a guest contribution by E. F. Fama)

THERE HAS BEEN A TRADITION AMONG ECONOMISTS which holds that prices in speculative markets, such as grain and securities markets, behave very much like random walks. References include Bachelier 1900, Kendall 1953, Osborne 1959, Roberts 1959, Cootner 1962, and Moore 1962. The random walk theory is based on two assumptions: (1) price changes are independent random variables, and (2) the changes conform to some probability distribution. This paper will be concerned with the nature of the distribution of price changes rather than with the assumption of independence. Attention will be focused on an important new hypothesis concerning the form of the distribution which has recently been advanced by Benoit Mandelbrot. We shall see later that if Mandelbrot's hypothesis is upheld, it will radically revise our thinking concerning both the nature of speculative markets and the proper statistical tools to be used when dealing with speculative prices.

I. INTRODUCTION

Prior to the work of Mandelbrot, the usual assumption, which we shall henceforth call the Gaussian hypothesis, was that the distribution of price changes in a speculative series is approximately Gaussian that is, normal. In the best-known *theoretical* expositions of the Gaussian hypothesis, Bachelier 1900 and Osborne 1959 use arguments based on the central limit theorem to support the assumption of normality. If the price changes from transaction to transaction are independent, identically distributed random variables with finite variance, and if transactions are fairly uniformly spaced through time, the central limit theorem leads us to believe that price changes across differencing intervals such as a day, a week, or a month will be normally distributed since they are simple sums of the changes from transaction to transaction. *Empirical* evidence in support of the Gaussian hypothesis has been offered by Kendall 1953 and Moore 1962. Kendall found that weekly price changes for Chicago wheat and British common stocks were "approximately" normally distributed, and Moore reported similar results for the weekly changes in log price of a sample of stocks from the New York Stock Exchange.

Mandelbrot contends, however, that this past research has overemphasized agreements between the empirical distribution of price changes and the normal distribution, and has neglected certain departures from normality which are consistently observed. In particular, in most empirical work, Kendall's and Moore's included, it has been found that the extreme tails of empirical distributions are higher (i.e., contain more of the total probability) than those of the normal distribution. Mandelbrot feels that these departures from normality are sufficient to warrant a radically new approach to the theory of random walks in speculative prices. This new approach, which henceforth shall be called the L-stable hypothesis, makes two basic assertions: (1) the variances of the empirical distributions behave as if they were infinite, and (2) the empirical distributions conform best to the non-Gaussian members of a family of limiting distributions which Mandelbrot has called L-stable. To date, Mandelbrot's most comprehensive work in this area is M 1963b{E14}.

The infinite variance of the L-stable model has extreme implications. From a purely statistical standpoint, if the population variance of the distribution of first differences is infinite, the sample variance is probably a meaningless measure of dispersion. Moreover, if the variance is infinite, other statistical tools (e.g., least-squares regression), which are based on the assumption of finite variance will, at best, be considerably weakened and may in fact give very misleading answers. Because past research on speculative prices has usually been based on statistical tools which assume the existence of a finite variance, the value of much of this work may be doubtful if Mandelbrot's hypothesis is upheld by the data.

In the remainder of this paper we shall examine further the theoretical and empirical content of Mandelbrot's L-stable hypothesis. The first step will be to examine some of the important statistical properties of the L-stable distributions. The statistical properties will then be used to illustrate different types of conditions that could give rise to a L-stable market. After this, the implications of the hypothesis for the theoretical and empirical work of the economist will be discussed. Finally, the state of the evidence concerning the empirical validity of the hypothesis will be examined.

II. THE L-STABLE DISTRIBUTIONS

The derivation of most of the important properties of L-stable distributions is due to Lévy 1925. A rigorous and compact mathematical treatment of the statistical theory can be found in Gnedenko & Kolmogorov 1954. A more comprehensive mathematical treatment can be found in M 1963b{E14}. A descriptive treatment of the statistical theory is found in Fama 1963a.

II.A. The parameters of L-stable distributions

The characteristic function for the L-stable family of distribution satisfies

$$\log f(t) = \log \int_{-\infty}^{\infty} \exp(i\mu t) dP(U < u)$$
$$= i\delta t - \gamma |t|^{\alpha}[1 + i\beta(t/|t|) \tan(\alpha\pi/2)].$$

The characteristic function tells us that L-stable distributions have four parameters: α, β, δ, and γ. The location parameter is δ, and if α is greater than 1, δ is equal to the expectation of mean of the distribution. The scale parameter is γ, while the parameter β is an index of skewness which can take any value in the interval $-1 \leq \beta \leq 1$. When $\beta = 0$ the distribution is symmetric. When $\beta > 0$ (and $1 < \alpha < 2$), the distribution is skewed right (i.e., has a long tail to the right), and the degree of right skewness increases in the interval $0 < \beta \leq 1$ as β approaches 1. Similarly, when $\beta < 0$ (and $1 < \alpha < 2$) the distribution is skewed left, with the degree of left skewness increasing in the interval $-1 \leq \beta < 0$ as β approaches -1.

Of the four parameters of a L-stable distribution, the characteristic exponent α is the most important for the purpose of comparing "the goodness of fit" of the Gaussian and L-stable hypotheses. The character exponent α determines the height of, or total probability contained in, the extreme tails of the distribution, and can take any value in the interval $0 < \alpha \leq 2$. When $\alpha = 2$, the relevant L-stable distribution is the normal distribution. The logarithm of the characteristic function of a normal distribution is $\log f(t) = i\mu t - \sigma^2 t^2/2$. This is the logarithm of the characteristic function of a L-stable distribution with parameters $\alpha = 2$, $\delta = \mu$, and

$\gamma = \sigma^2/2$. When α is in the interval $0 < \alpha < 2$, the extreme tails of the L-stable distributions are higher than those of the normal distribution, with the total probability in the extreme tails increasing as α moves away from 2 and toward 0. The most important consequence of this is that the variance exists (i.e., is finite) only in the extreme case $\alpha = 2$. The mean, however, exists as long as $\alpha > 1$. For a proof of these statements, see Gnedenko & Kolmogorov 1954.

Mandelbrot's L-stable hypothesis states that for distributions of price changes in speculative series, α is in the interval $1 < \alpha < 2$, so that the distributions have means but their variances are infinite. The Gaussian hypothesis, on the other hand, states that α is exactly equal to 2.

It is important to distinguish between the L-stable *distributions* and the L-stable *hypothesis*. Under *both* the L-stable and the Gaussian hypotheses it is assumed that the underlying distribution is L-stable. The conflict between the two hypotheses involves the value of the characteristic exponent α. The Gaussian hypothesis says that $\alpha = 2$, while the L-stable hypothesis says that $\alpha < 2$.

II.B. Estimation of α: the asymptotic scaling range

Since the conflict between the L-stable and Gaussian hypotheses hinges, essentially on the value of the characteristic exponent α, a choice between the hypotheses can be made, in theory, solely by estimating the true value of this parameter. Unfortunately, this is not a simple task. Explicit expressions for the densities of L-stable distributions are known for only three cases: the Gaussian ($\alpha = 2, \beta = 0$), the Cauchy ($\alpha = 1, \beta = 0$), and the coin-tossing case ($\alpha = 1/2$, $\beta = 1$, $\delta = 0$, and $\gamma = 1$). Without density functions, it is very difficult to develop and prove propositions concerning the sampling behavior of any estimators of α that may be used.

Of course, these problems of estimation are not limited to the characteristic exponent α. The absence of explicit expressions for the density functions makes it very difficult to analyze the sampling behavior of estimators of all the parameters of L-stable distributions. The statistical intractability of these distributions is, at this point, probably the most important shortcoming of the L-stable hypothesis.

The problem of estimation is not completely unsolvable, however. Although it is impossible to say anything about the sample error of any given estimator, of α, one can attempt to bracket the true value by using many different estimators. This is essentially the approach that I followed in my dissertation. Three different techniques were used to estimate values of α for the daily first differences of log price for each individual stock of the Dow-Jones Industrial Average. Two of the estimation procedures, one based on certain properties of fractile ranges of L-stable variables, and the other derived from the behavior of the sample variance, were introduced for the first time in the dissertation. An examination of these techniques would take us more deeply into the statistical theory of L-stable distributions than is warranted by the present paper. The third technique, double log graphing, is widely known, however, and will now be discussed in detail.

Lévy has shown that the tails of L-stable distributions for values of α less than 2 follow an asymptotic form of scaling. Consider the distributions following the strong form of this law,

$$\Pr\{U > u\} = (u/V_1)^{-\alpha} \quad u > 0, \tag{1}$$

and

$$\Pr\{U < u\} = (|u|/V_2)^{-\alpha} \quad u < 0, \tag{2}$$

where U is the random variable and the constants V_1 and V_2 are defined by

$$\beta = \frac{V_1^\alpha - V_2^\alpha}{V_1^\alpha + V_2^\alpha}.$$

In this case, of course, β, is the parameter for skewness discussed previously. The asymptotic form of scaling is

$$\Pr\{U > u\} \to (u/V_1)^{-\alpha} \text{ as } u \to \infty \tag{3}$$

and

$$\Pr\{U < u\} \to (|u|/V_2)^{-\alpha} \text{ as } u \to -\infty. \tag{4}$$

Taking logarithms of both sides of expressions (3) and (4), we have,

$$\log \Pr\{U > u\} \to -\alpha(\log u - \log V_1), \quad u > 0 \tag{5}$$

and

$$\log \Pr \{U < u\} \rightarrow -\alpha(\log |u| - \log V_2), \quad u < 0. \tag{6}$$

Expressions (5) and (6) imply that, if $\Pr \{U > u\}$ and $\Pr \{U < u\}$ are plotted against $|u|$ on double log paper, the two lines should become asymptotically straight and have slope that approaches $-\alpha$ as $|u|$ approaches infinity. Double log graphing, then, is one technique for estimating α.

Unfortunately, the simplicity of the double log graphing technique is, in some cases, more apparent than real. In particular, the technique is asymptotic when the characteristic exponent is close to 2. For a discussion, see M 1963p and also Fama 1963a, Chap. IV.

II.C. Other properties of L-stable distributions

The three most important properties of L-stable distributions are (1) the asymptotically scaling nature of the extreme tail areas, (2) L-stability or invariance under addition, and (3) the fact that these distributions are the only possible limiting distributions for sums of independent, identically distributed, random variables. Asymptotic scaling was discussed in the previous Section. We shall now consider in detail the property of L-stability and the conditions under which sums of random variables follow L-stable limiting distributions.

1. *Stability or invariance under addition.* By definition, a L-stable distribution is any distribution that is invariant under addition. That is, the distribution of sums of independent, identically distributed, L-stable variables is itself L-stable and has the same form as the distribution of the individual summands. "Has the same form" is, of course, an imprecise verbal expression of a precise mathematical property. A rigorous definition of L-stability is given by the logarithm of the characteristic function of sums of independent, identically distributed, L-stable variables. This function is

$$n \log f(t) = i(n\delta)t - (n\gamma)\,|t|^{\alpha}\left\{1 + i\beta \frac{t}{|t|}\left(\tan \frac{\alpha\pi}{2}\right)\right\},$$

where n is the number of variables in the sum and $\log f(t)$ is the logarithm of the characteristic function of the individual summands. The above expression is exactly the same as the expression for $\log f(t)$, except that the parameters δ (location) and γ (scale) are multiplied by n. That is, the distribution of the sums is, except for origin and scale, exactly the same as the distribution of the individual summands. More simply, L-stability means that the values of the parameters α and β remain constant under addition.

The above discussion assumes that the individual, L-stable variables in the sum are independent and identically distributed. That is, the distribution of each individual summand has the same values of the four parameters α, β, δ, and γ. It will now be shown that L-stability still holds when the values of the location and scale parameters, δ and γ, are not the same for each individual variable in the sum. The logarithm of the characteristic function of the sums of n such variables, each with different location and scale parameters, δ_j and γ_j, is

$$\sum_{j=1}^{n} \log f_j(t) = i\left(\sum_{j=1}^{n} \delta_j\right)t - \left(\sum_{j=1}^{n} \gamma_j\right)|t|\left(1 + i\beta\frac{t}{|t|}\left(\tan\frac{\alpha\pi}{2}\right)\right).$$

This is the characteristic function of a L-stable distribution with parameters α and β, and with location and the sums of the scale parameters equal, respectively, to the sums of the location and the sums of the scale parameters of the distributions of the individual summands. That is, the sum of the L-stable variables, where each variable has the same values of α and β but different location and scale parameters, is also L-stable with the same values of α and β.

The property of L-stability (in the sense of invariance under addition) is responsible for much of the appeal of L-stable distributions as descriptions of empirical distributions of price changes. The price change in a speculative series for any time interval can be regarded as the sum of the changes from transaction to transaction during the interval. If the changes between transactions are independent, identically distributed, L-stable variables, daily, weekly, and monthly changes will follow L-stable distributions of exactly the same form, except for origin and scale. For example, if the distribution of daily changes is normal with mean μ and variance σ^2, the distributions of weekly (or five-day) changes will also be normal with mean 5μ and variance $5\sigma^2$. It would be very convenient if the form of the distribution of price changes were independent of the differencing interval for which the changes were computed.

2. *Limiting distributions.* It can be shown that L-stability (again, in the sense of invariance under addition) leads to a most important corollary property of L-stable distributions: they are the only possible limiting distributions for sums of independent, identically distributed random variables (Gnedenko & Kolmogorov 1954 pp. 162-63). It is well known that if

such variables have finite variance, the limiting distribution for the sum will be the normal distribution. If the basic variables have infinite variance, however, and if their sums follow a limiting distribution, the limiting distribution must be L-stable with $0 < \alpha < 2$.

It has been proven independently by Gnedenko and Doeblin that in order for the limiting distribution of sums to be L-stable with characteristic exponent α $(0 < \alpha < 2)$, it is necessary and sufficient that (Gnedenko & Kolmogorov 1954 pp. 175-80)

$$\frac{F(-u)}{1-F(u)} \to \frac{C_1}{C_2} \text{ as } u \to \infty, \tag{7}$$

and that for every constant $k > 0$,

$$\frac{1-F(u)+F(-u)}{1-F(ku)+F(-ku)} \to k^{\alpha} \text{ as } u \to \infty, \tag{8}$$

where F is the cumulative distribution function of the random variable U and C_1 and C_2 are constants. Expressions (7) and (8) will henceforth be called the conditions of Doeblin and Gnedenko.

It is clear that any variable that is asymptotically scaling (regardless of whether it is also L-stable) will satisfy these conditions. For example, consider a variable U that is asymptotically scaling but not L-stable. Then as $u \to \infty$

$$\frac{F(-u)}{1-F(u)} \to \left\{ \frac{(|-u|/V_2)}{(u/V_1)} \right\}^{-\alpha} = \frac{V_2{}^{\alpha}}{V_1{}^{\alpha}},$$

and

$$\frac{1-F(u)+F(-u)}{1-F(ku)+F(-ku)} \to \frac{(u/V_1)^{-\alpha}+(|-u|/V_2)^{-\alpha}}{(ku/V_1)^{-\alpha}+(|-ku|/V_2)^{-\alpha}} = k^{\alpha},$$

and the conditions of Doeblin and Gnedenko are satisfied.

To the best of my knowledge, nonstable, asymptotically scaling variables are the only known variables of infinite variance that satisfy conditions (7) and (8). Thus, they are the only known nonstable variables whose sums approach L-stable limiting distributions with characteristic exponents less than two.

III. THE ORIGIN OF A L-STABLE MARKET: SOME POSSIBILITIES

The price changes in a speculative series can be regarded as a result of the influx of new information into the market, and of the reevaluation of existing information. At any point in time, there will be many items of information available. Thus, price changes between transactions will reflect the effects of many different bits of information. The previous Section suggests several ways in which these effects may combine to produce L-stable distributions for daily, weekly, and monthly price changes.

In the simplest case,the price changes implied by individual bits of information may themselves follow L-stable distributions with constant values for the parameters α and β, but possibly different values for the location and scale parameters, δ and γ. If the effects of individual bits of information combine in a simple, additive fashion, then by the property of L-stability the price changes from transaction to transaction will also be L-stable with the same values of the parameters α and β. Since the price changes for intervals such as a day, week, or month are the simple sums of the changes from transaction to transaction, the changes for these intervals will also be L-stable with the same values of the parameters α and β.

Now suppose that the price changes implied by individual items of information are asymptotically scaling, but not L-stable. This means that the necessary and sufficient conditions of Doeblin and Gnedenko will be satisfied. Thus, if the effects of individual bits of information combine in a simple, additive fashion, and if there are very many bits of information involved in a transaction, the distributions of price changes between transactions will be L-stable. It may happen, however, that there are not enough bits of information involved in individual transactions to insure that the limiting L-stable distribution is closely achieved by the distribution of changes from transaction to transaction. In this case, as long as there are many transactions per day, week, or month, the distributions of price changes for these differencing intervals will be L-stable with the same values of the parameters α and β.

Mandelbrot has shown that these results can be generalized even further. (See M 1963e{E5}.) As long as the effects of individual bits of information are asymptotically scaling, various types of complicated combinations of these effects will also be asymptotically scaling. For example, although there are many bits of information in the market at any given time, the price change for individual transactions may depend solely on what the transactors regard as the largest or most important piece of information. Mandelbrot has shown that, if the effects of individual items of information are asymptotically scaling with exponent α, the distribution of the largest effect will also be asymptotically scaling with the same exponent α. Thus the distribution of changes between transactions will be asymptotically scaling, and the conditions of Doeblin and Gnedenko will be satisfied. If there are very many transactions in a day, week, or month, the distributions of price changes for these differencing intervals will be L-stable with the same value of the characteristic exponent α.

In sum, so long as the effect of individual bits of information combine in a way which makes the price changes from transaction to transaction *asymptotically scaling* with exponent α, then, according to the conditions of Doeblin and Gnedenko, the price changes for longer differencing intervals will be L-stable with the same value of α. According to our best knowledge at this time, however, it is necessary that the distribution of the price changes implied by the individuals bits of information be at least *asymptotically scaling* (but not necessarily L-stable) if the distributions of changes for longer time periods are to have *L-stable* limits.

IV. IMPORTANCE OF THE L-STABLE HYPOTHESIS

The L-stable hypothesis has many important implications. First of all, if we retrace the reasoning of the previous Section, we see that the hypothesis implies that there are a larger number of abrupt changes in the economic variables that determine equilibrium prices in speculative markets than would be the case under a Gaussian hypothesis. If the distributions of daily, weekly, and monthly price changes in a speculative series are *L-stable* with $0 < \alpha < 2$, the distribution of changes between transactions must, at the very least, be asymptotically scaling. Changes between transactions are themselves the result of the combination of the effects of many different bits of information. New information, in turn, should ultimately reflect changes in the underlying economic conditions that determine equilibrium prices in speculative markets. Thus, following this line of reasoning, the underlying economic conditions must themselves have an asymptotically scaling character, and are therefore subject to a larger number of abrupt changes than would be the case if distributions of price changes in speculative markets conformed to the Gaussian hypothesis.

The fact that there are a large number of abrupt changes in a L-stable market means, of course, that such a market is inherently more risky for the speculator or investor than a Gaussian market. The variability of a given expected yield is higher in a L-stable market than it would be in a Gaussian market, and the probability of large losses is greater.

Moreover, in a L-stable market speculators cannot usually protect themselves from large losses by means of such devices as "stop-loss" orders. In a Gaussian market, if the price change across a long period of time is very large, the chances are that the total change will be the result of a large number of very small changes. In a market that is L-stable with $\alpha < 2$, however, a large price change across a long interval will more than likely be the result of a few very large changes that took place during smaller subintervals. (For a proof see Darling 1952 or Arov & Bobrov 1960.) This means that if the price level is going to fall very much, the total decline will probably be accomplished very rapidly, so that it may be impossible to carry out many "stop-loss" orders at intermediate prices. (See M 1963b{E14}.)

The inherent riskiness of a L-stable market may account for certain types of investment behavior which are difficult to explain under the hypothesis of a Gaussian market. For example, it may *partially* explain why many people avoid speculative markets altogether, even though at times the expected gains from entering these markets may be quite large. It may also partially explain why some people who are active in these markets hold a larger proportion of their assets in less speculative, liquid reserves than would seem to be necessary under a Gaussian hypothesis.

Finally, the L-stable hypothesis has important implications for data analysis. As mentioned earlier, when $\alpha < 2$, the variance of the underlying L-stable distribution is infinite, so that the sample variance is an inappropriate measure of variability. Moreover, other statistical concepts, such as least-squares regression, which are based on the assumptions of finite variance are also either inappropriate or considerably weakened.

The absence of a finite variance does *not* mean, however, that we are helpless in describing the variability of L-stable variables. As long as the characteristic exponent α is greater than 1, estimators which involve only first powers of the L-stable variable have finite expectation. This means that concepts of variability, such as fractile ranges and the absolute mean deviations, which do involve only first powers, have finite expectation and thus are more appropriate measures of variability for these distributions than the variance.

A fractile range shows the range of values of the random variable that fall within given fractiles of its distributions. For example, the interquartile range shows the range of values of the random variable that fall within the 0.25 and 0.75 fractiles of the distribution. N being the total sample size, the absolute mean deviation is defined as

$$|D| = \sum_{i=1}^{N} \frac{|X_i - \tilde{X}|}{N},$$

V. THE STATE OF THE EVIDENCE

The L-stable hypothesis has far-reaching implications. The nature of the hypothesis is such, however, that its acceptability must ultimately depend on its empirical content rather than on its intuitive appeal. The empirical evidence up to this point *has* tended to support the hypothesis, but the number of series tested has not been large enough to warrant the conclusion that further tests are unnecessary.

For commodity markets, the single most impressive piece of evidence is a direct test of the infinite variance hypothesis for the case of cotton prices. Mandelbrot computed the sample second moments of the daily first differences of the logs of cotton prices for increasing samples of from 1 to 1,300 observations. He found that as the sample size is increased, the sample moment does not settle down to any limiting value but rather continues to vary in absolutely erratic fashion, precisely as would be expected under the L-stable hypothesis. (See M 1963b{E14}.)

Mandelbrot's other tests in defense of the L-stable hypothesis are based primarily on the double log graphing procedure mentioned earlier. If the distribution of the random variable U is L-stable with $\alpha < 2$, the graphs of $\log \Pr\{U < u\}$, u negative, and $\log \Pr\{U > u\}$, u positive, against $\log |u|$ should be curves that become asymptotically straight with slope $-\alpha$. The graphs for the same cotton price data seemed to support the hypothesis that α is less than 2. The empirical value of α for cotton prices appears to be about 1.7.

Finally, in my dissertation (Fama 1963-1965), the L-stable hypothesis has been tested for the daily first differences of log price of each of the thirty stocks in the Dow-Jones Industrial Average. Simple frequency distributions and normal probability graphs were used to examine the tails of the empirical distributions for each stock. In *every* case the empirical distributions were long-tailed, that is, they contained many more observations in their extreme tail areas than would be expected under a hypothesis of normality. In addition to these tests, three different procedures were used to estimate values of the characteristic exponent α for each of the the thirty stocks. The estimates produced empirical values of α *consistently* less than 2. The conclusion of the dissertation is that for the important case of stock prices, the L-stable hypothesis is more consistent with the data than the Gaussian hypothesis.

VI. CONCLUSION

In summary, the L-stable hypothesis has been directly tested only on a limited number of different types of speculative price series. But it should be emphasized that every direct test on unprocessed and unsmoothed price data has found the type of behavior that would be predicted by the hypothesis. Before the hypothesis can be accepted as a general model for speculative prices, however, the basis of testing must be broadened to include other speculative series.

Moreover, the acceptability of the L-stable hypothesis will be improved not only by further empirical documentation of its applicability but also by making the distributions themselves more tractable from a statistical point of view. At the moment, very little is known about the sampling behavior of procedures for estimating the parameters of these distributions. Unfortunately, as mentioned earlier, rigorous, analytical sampling theory will be difficult to develop as long as explicit expressions for the density functions are not known. However, pending the discovery of such expressions, Monte Carlo techniques could be used to learn some of the properties of various procedures for estimating the parameters.

Mandelbrot's L-stable hypothesis has focused attention on a long-neglected but important class of statistical distributions. It has been demonstrated that among speculative series, the first differences of the logarithms of stock and cotton prices seem to confirm to these distributions. The next step must be both to test the L-stable hypothesis on a broader range of speculative series and to develop more adequate statistical tools for dealing with L-stable distributions.

&&&&&&&&&&&& ANNOTATIONS &&&&&&&&&&&&

This exploritory paper greatly helped M 1963b{E14} become understood in the economics community. At the time, I was supervising Fama's Ph.D. from far away, using the telephone, letters, and visits. After his thesis, subtitled "A test of Mandelbrot's Paretian hypothesis" (Fama 1965), Fama stayed in Chicago to pursue the study of my model of price variation. Within a few years, he produced many high quality students, beginning with R. Roll, M. Blume and M. Jensen.

It must be revealed, however, that I never agreed with Fama that prices follow a "random walk." This contention did not seem to be empirically established, see M 1963b{E14}. I saw no reason a priori to expect it to be true, therefore drew Fama's attention to martingales and to Bachelier's efficient market hypothesis.

Fama & Blume 1966 later wrote that "Although independence of successive price changes implies that the history of prices series cannot be used to increase expected gains, the reverse proposition does not hold. It is possible to construct models where successive price changes are dependent, yet the dependence is not of a form which can be used to increase expected profits. In fact, M 1966b{E19} and Samuelson 1965 show that, under fairly general conditions, in a market that fully discounts all available information, prices will follow a martingale which may or may not have the independence property of a pure random walk. In particular, the martingale property implies only that the *expected values* of future prices will be independent of the *values* of past prices; the distributions of future prices, however, *may* very well depend on the values of past prices. In a martingale, though price changes may be dependent, the dependence cannot be used by the trader to increase expected profits. ...In most cases the degree of dependence shown by a martingale will be so small that for practical purposes it will not do great violence to the independence assumption of the random-walk model." I never believed that.

Editorial note. The original title was "Mandelbrot and the stable Paretian hypothesis." I hesitated before changing the title of someone else's paper, but went ahead because of a dislike for the word "hypothesis."

Comments by P. H. Cootner, E. Parzen, and W. S. Morris (1960s) and responses

WHILE TEACHING ECONOMICS AT HARVARD, I spent part of the 1962 Christmas vacation in Pittsburgh PA, where the Econometric Society held its Annual Meeting. M 1962i, which was to provide the substance of M 1963b{E14} and 1967b{E15}, was honored by being made the sole topic of a session. Instead of three talks, each followed by a brief discussion, this session included my talk followed by several discussions. The two that were written down were thoughtful, but demanded a response. Sections 1 and 2 reproduce some telling points of those comments, in italics and between quote marks, and followed by my responses. Two other discussants, Lawrence Fisher and George Hadley of the University of Chicago, left no record. The source of the quite separate contribution by W. S. Morris will be mentioned in Section 3.

1. COMMENTS BY PAUL H. COOTNER (1962-1964), PRINTED IN ITALICS AND FOLLOWED BY THE EDITOR'S RESPONSES

• Cootner (1930-1978), best known for his work on speculation, hedging, commodity prices and the operation of the futures markets, was at the Sloan School of Industrial Management of M.I.T. and later moved to Stanford as C.O.G. Miller Professor of Finance. An influential book he edited, Cootner 1964, reprints M 1963b{E14} and Fama 1963{E16}, but also adds extensive comments of his own, "to make available his skeptical point of view." Those comments began with elegant praise quoted in the Preface, and continued with meticulous criticism not always based on meticulous reading. Beale 1966 (p. 219) wrote that "while I found his remarks useful, I suspect Cootner, the editor, has compromised unduly with Cootner, the protagonist, in this instance." Indeed, I was not asked to respond in print, and detailed responses communicated privately were disregarded.

Most notably, my insistence on dealing with "certain" prices was careful and always implied "not necessarily all," and I never viewed random walk (or even the martingale property) as anything more than a first approximation. Indeed, Section VII of M 1963b{E14} clearly marked my early contact with dependence. I always was sensitive to these two objections, because they keep being brought up to this day.

However, it was proper and necessary that the expert should scrutinize an intensive beginner's "manifesto" with more than ordinary severity. Besides Cootner was helpful in stating in the open what others say in private to this day, and some flaws that he noted were genuine and natural in a first paper: they concerned issues I also recognized, but could not handle until much later – as described in Chapter E1. A belated and muted response appeared in Section 1 of M 1967b{E15}, but had no impact, and Cootner doubtlessly blunted the effects of my eloquence and of that of Fama. As suggested in the Preface, many of those whom my work had shaken and challenged in 1964 were motivated by the inquisitorial tone to wait and see. To conclude this introduction, I regret that Cootner cannot further comment, but believe that even at this late date a purpose is served by responding in writing. •

The same passionate devotion that makes [M 1963b{E14}] such a vivid exposition of ideas lends it a messianic tone which makes me just a bit uncomfortable. I was continually reminded, as I read it, of the needling question a teacher of mine liked to ask – What evidence would you accept in contradiction of your hypothesis? While there is a wealth of evidence presented in the lengthy paper, much of it is disturbingly casual. Much of the evidence is graphical and involves slopes of lines which are not given precise numerical values in the paper. Nor is there any concern that the values of α which fit the tails will also fit the rest of the distribution.

The evidence presented in M 1963b{E14}, and later in M 1967b{E15} was incontrovertible and easily "falsifiable." My tone was casual for two reasons: to be read and provoke discussion, it was necessary to avoid pedantry. More importantly, I was acutely aware, though of course in lesser technical detail than today, that there is a deep contrast between mild and wild randomness, and that the latter called for more than a careful use of standard statistics.

The statisticians' knocking down of graphical evidence was always excessive, and is becoming much less strident in recent years. Computer graphics vindicate my long-held belief that, *in a first analysis,* graphical evidence is unbeatable, both in its power to analyze, and in its power to communicate the results of analysis.

Concerning the "precise numerical" α, Figure 5 of M 1963b{E14} explained in Section II C, incorporates a continuous curve described as the tail probability density of the L-stable distribution for $\alpha = 1.7$. Rough visual estimation was the best anyone could hope for in 1963. Besides, $\alpha \sim 1.7$ was confirmed in diverse contexts (examples being found in Walter 1994) and may perhaps turn out to be of wide occurrence, and even "universal."

In any event, "exact fit" is rarely an issue in science, it is never an issue in first-order models, and I recall distinctly that the mathematician Willy Feller was scathing in his criticism of those who claimed that *any scientific* data could be *exactly* Gaussian. Some data are close enough to Gaussian, and no one is bothered by the need for corrective terms to account for local discrepancies. Similarly, market price data are surely *not* exactly fitted by the L-stable distribution. The only issues worth tackling are the overall shape of the distribution and precise fit in its tails, which are of greatest interest and happen to be the easiest to study. There, the only question is whether or not my L-stable model is a good *first* approximation.

At one of the few places where numbers are mentioned, some [ratios of the scale factors of the distribution of monthly price changes to that of daily price changes] are dismissed as "absurdly large," and "corrected," and in the new data, a number, 18, is accepted as reasonably close to another number which should be somewhere between 21 and 31.

This is a crucial issue. The end of M 1963b{E14} states specifically that successive price changes *fail* to be independent. Unfortunately, when variance is infinite, one cannot use correlations to measure the degree of dependence. I tried out a promising alternative, by defining the "effective" number N_{eff} of trading days in a month as the ratio of scale factors γ^{α} for months and for days. To accept the observed ratio N_{eff}/N_{true} as "reasonably close" to 1, as I did, was simply to conclude that it seemed legitimate to postpone the study of dependence. A ratio $N_{eff}/N_{true} > 1$ would imply "persistence," i.e., a positive dependence between price changes; a ratio $N_{eff}/N_{true} < 1$ would imply "antipersistence," i.e., negative dependence. Not until much later (as seen in Chapter E1) did multifractals extend the scope of scaling to cover this form of dependence.

Multifractals introduce a second descriptive exponent η, and lead to the relation $N_{\text{eff}} \sim (N_{\text{true}})^{\eta}$. Values of $\eta < 1$ (resp., $\eta > 1$) would characterize negative (resp., positive) dependence.

As to the words "absurdly large," they are easy to explain. As read off my Figure 5, the distribution of monthly changes for the period 1880-1940 had a scale factor 10 times larger than the distribution of daily changes for the period around 1950. To obtain a scale change of 10 with α between 1.6 and 1.7 would require between $10^{1.6} = 40$ and $10^{1.7} = 50$ days. This huge value is indeed "absurdly large," yet it was expected in 1963, that the two periods do *not* overlap. M 1972b{E14}, reproduced as appendix I of M 1963b{E14} describes the honest but gross misreading of data sent by the U.S. Department of Agriculture that led me to think that, around 1900, cotton was traded 7 days a week. (It is distressing that someone with Cootner's expertise in commodities did not identify this clear mistake on my part.) In 1963, *everyone* took it for granted that price variability had gone down between 1900 and 1950. It seemed eminently reasonable, therefore, to compare the months for the 1880-1940 period with the days within a period of a few years haphazardly chosen somewhere in the middle of the 1880-1940 span. After the United Stated Department of Agriculture sent better data, the value of 18 days that infuriated Cootner turned out to be an underestimate.

Mandelbrot asserts that the data on cotton spot prices are supported by research into wheat spot prices as well, [but the slope is] 'too close to 2 for comfort.'

The assertion concerning wheat is documented in M 1967b{E15}. Instead of "too close to 2 for comfort," I should have written "too close to 2 for easy estimation of α from the graph and for use as a prime example of a radical theory."

A very interesting, but questionable, proposition about stock price index numbers, that large changes in them are generally traceable to one or a small number of components, is said to be "obviously" correct without any empirical evidence at all.

Touché. This form of concentration remains to this day a nice topic for study. Other forms of concentration are beyond argument, but this particular form is affected by the fact that index numbers may be dominated by market factors that affect every component of the index.

Most of Mandelbrot's data deal with commodity spot prices. These are not, and I repeat, not *the kind of data that we would expect to display Brownian motion in the first place.*

M 1963b{E14} *does not* argue about Brownian motion, but seeks a good description of certain prices' variation.

Anyhow, theory should never blind one to the evidence.

If successive stock price changes were L-stable and independent, *the observed range between the highest and lowest observed* prices *would increase at a rate of* $T^{1/\alpha}$ *instead of at a rate* $T^{1/2}$, *where T is the period of observation. Since the evidence of Stieger and Cootner suggests that the range increases at a rate slower that* $T^{1/2}$, *if stock prices are distributed in L-stable fashion, they must be even more dependent than in the Gaussian case, and thus be an even poorer approximation to a true random walk.*

This is an excellent remark, and it is astonishing that it was not taken up strongly enough to draw everybody's attention. The exponent of T combines tail and dependence effects, and Cootner's observation fits perfectly in the program of investigation I had already mapped out in 1963 (see the comment by W. S. Morris in Section 3) and carried out on and off over the years.

2. COMMENTS BY EMANUEL PARZEN (1962), PRINTED IN ITALICS, AND FOLLOWED BY THE EDITOR'S RESPONSES

• Parzen, best known for a textbook on probability theory and for work on spectra and other methods of analyzing time series, was Professor of Statistics at Stanford, and later moved to SUNY Buffalo, and on to Texas A&M. •

Mandelbrot makes three important contributions.

Its probabilistic *contribution is to make us even more aware than we may have been that the "frequently encountered" probability distributions (Parzen, 1960, pp. 218-221) must be enlarged. ... Along these lines, it is of interest to quote the view of Gnedenko & Kolmogorov 1954 (p. 11). 'There is no doubt that the arsenal of those limit theorems which should be included in future practical handbooks must be considerably expanded in comparison with classical standards. Of course, it is necessary to make some choice. For example, "normal" convergence to the non-normal L-stable laws undoubtedly must already be considered in any comprehensive text in, say, the field of statistical physics'*

The Gnedenko-Kolmogorov line on applications to "statistical physics" caught the attention of many readers and my advance response to Parzen is found in M 1963e{E3}. But it deserves being repeated. In 1958, I was about to publish M 1958p and Kolmogorov was visiting Paris, so I asked him to elaborate. He answered that he knew of no actual application. His surviving students do not, either. This quote may merely demonstrate that the most famous pure mathematician in the USSR of 1949 sometimes needed to invoke hypothetical applications to be left alone.

As further clue, Parikh 1991 describes a visit to Moscow in 1935 by Oskar Zariski. We read that "one evening, Pontryagin and Kolmogorov were putting forward the Marxist view that only applied mathematics had any importance. Zariski [who called himself a Marxist at that time] broke in .... 'Don't you find it difficult to write about topology?' Skillfully side stepping the issue, Kolmogorov answered: 'You must take the term *application* in a wide sense, you know. Not everything must be applied immediately. Almost every view of mathematics is useful for the development of technology, but that doesn't mean that every time you do mathematics you must work on a machine.'"

Its empirical *contribution consists in pointing out economic phenomena which may satisfy L-stable laws, and in developing various heuristic principles (such as the principles of aggregation and disaggregation) for testing ...*

After thirty years during which the idea of scaling grew to conquer several branches of science, it is interesting to see that there was a time when the principle of scaling in economics could be described as "heuristic."

Its statistical *contribution consists of pointing out the need for analytical work on various problems of statistical inference, of which I shall describe two.*

First, *I would like to see a more theoretical treatment of Mandelbrot's doubly logarithmic graphs, and in particular of the [estimation of] the value of α.*

I second this wish wholeheartedly. There is already a good beginning.

Secondly, *I cannot agree that such statistical techniques as the method of moments and spectral analysis have been shown to be without value for economic statistics. Statisticians realize that ... to cast a real problem into one of the well analyzed theoretical frameworks, one may have to manipulate the observed data by a transformation, such as ... square root or logarithm* [or $x - Cx^2$]. *Such transformations [can] yield random variables with finite higher moments.*

Chapter E5 argues that in finance the most important data are the so-called "outliers." Therefore, "stabilization" through a nonlinear change of variable is *self-defeating.* Such transforms bury under the rug all the special features relative to high values of x. The widely condemned graphical methods, when done carefully, are more sensitive to the details of the evidence.

5. COMMENTS BY WILLIAM S. MORRIS (1964): CAN ECONOMICS BE SCIENTIFIC?

"[Mandelbrot's work] is likely to produce profound changes in economics and related social sciences. The tight interplay between theory and objective data found in the ... physical sciences is, I am sure we must all admit, conspicuously absent in economics. This, I believe, is due mainly to the fact that our methods of inductive reasoning... work nicely when the distributions encountered conform to ... the framework of our present statistical techniques, but break down when we try to force these statistical techniques to function in areas where it is impossible for them to do so .... I should like to suggest a manner in which a wealth of important, worthwhile information about economic processes might be gained ....

"1. That we restrict ourselves to raw source data which have not been doctored or averaged in any way. Such data manipulations [usually hide] the extreme values.

"2. That we abandon the ... preconceived notions about how data ought to behave .... The ubiquitous assumption that economic decisions are affected by the difference between the actual value of an economic variable and some theoretically normative or equilibrium value seems to conflict not only with the behavior of economic data but with the degree of human stimulus-response adaptability indicated in numerous psychometric studies.

"3. That we find the simplest stochastic process that can be fit reasonably well. [Chance mechanisms give] us a somewhat greater a priori idea of permissible tolerances than one would have when fitting an algebraic hypothesis.

"A forbidding amount of basic research will have to be done ... [and one needs] compact computer subroutines ....

"While there may be quite a number of special problems for which ... [suffice independent events we need] a less restrictive class of stochastic processes ... support prices, options, pro-ration, liability limitations and the like soon lead to critical factors so inconsistent with the simple random walk that one would expect this approach to break down.

"In the field of stochastic processes, one seems to pass so sharply from the random walk into terrifyingly limitless possibilities ... We shall probably not be able to restrict ourselves to stationary Markov processes, ...

"The criterion by which people of our culture normally judge the success or failure of a scientific discipline is that it provides techniques that work. We reject the ponderous compendium of loosely organized bits of wisdom and search for a unifying principle or technique which somehow brings the entire discipline within the understanding of or, better yet, the control of, the scientist. If his theory works we accept it, however violently it may contradict the unmistakable testimony of our five senses and our common sense as well. The economist who shares this predilection has, I believe, good reasons to hope that in the L-stability preserving Markov process he will find the unifying principle that will put an end to the frustrations and humiliations that have heretofore attended our attempts to invade the mysteries of economic time series."

About William S. Morris and computers on Wall Street. This text confirms that, as already mentioned in the Preface, the author of this text and I were close. He articulated our common concerns very early and very clearly. Born in Canada in 1916, Morris graduated from Princeton with an AB with honors in mathematics. This is why we communicated so easily, and why he sat in the Visiting Committee of the Princeton Mathematics Department when Fine Hall was moving. He was an actuary, joined the Army, and worked for the First Boston Corporation. After he went in business for himself in 1959, his saga was marked by headlines in the First Business page of *The New York Times.* August 17, 1961: *Surprise Bid Wins California Bonds; 100 Million Involved.* August 20, 1961: *Small Concern led by Mathematician in Wall Street Coup.* September 14, 1961: *Second Large California Issue Won by "Insurgent" Bond House. Wall Street Lauds Bid. August Sale Had Drawn Resentment in Trade.* November 21, 1961: *Bond House Installs Own Computer.* May 9. 1963: *'Maverick' Investment Concern Wins 122 Million Bond Issue. Surprise Bid.* May 12, 1963: *Two Large Bond Issues Falter Despite Big Supplies of Money.* May 12, 1963: *Bonds: Morris Sells 62 Million of Washington Power's Dormant Issue at a Discount. Balance Still Unsold.* September 6, 1963: *Man Who Outbid Big Syndicates For Bonds Quits the Business.* January 17, 1964: *Morris Bidding Again, But Loses.*

E18

Computation of the L-stable distributions

✦ **Abstract.** The first tables, M & Zarnfaller 1959, were quite difficult to compute, and the resulting figures were drawn by hand. Today, several dependable programs are available. This brief and casual chapter is neither a systematic discussion nor a full bibliography, only a collection of hopefully useful odds and ends that are close at hand. ✦

WHILE THIS BOOK WAS BEING PLANNED, a stage that lasted many years, it seemed indispensable to include a chapter devoted to numerical evaluation of the L-stable densities. Early on, this topic was difficult and technical, and the M 1963 model that created the need to know was the only scaling model. Today, the computational difficulties have been tamed and new scaling models "compete" with L-stability. Those reasons, and a suddenly increased pressure of time, led me to reduce this chapter to modest bits of history and references.

1. Books on the L-stable distributions

The parametrization of the L-stable distributions. A "comedy of errors," to which I made an unfortunate contribution, is described in Hall 1981. The notation has not reached consensus, for example, Zolotarev gives five different sets of parameters.

Gnedenko & Kolmogorov 1954. This remains the most widely available reference. A comment on it, by E. Parzen, is reproduced in Chapter E17, with a response.

Zolotarev 1983-1986. This wide-ranging and excellent mathematical survey appeared in Russian as Zolotarev 1983. It was surprising and interesting that the introduction of this book from the Soviet era should observe that "In our times there has been a sharp increase of interest in stable laws, due to their appearance in certain socio-economic models".

Section I.3 elaborates, with reference to my work. And the Introduction ends with these words: "It is still comparatively rare to encounter multidimensional stable laws in application (other than Holtsmark's work). However, there is reason to expect that this situation will change in the near future (due first and foremost to applications in economics.)

Most unfortunately, the English translation, Zolotarev 1986, is clumsy and marred by serious misprints in the formulas. Prudence suggests that even the reader who finds Russian to be completely incomprehensible should consider checking the translated formulas against the original ones. The American Mathematical Society told me (a while ago!) that they would welcome a *List of Errata.* If someone comes forth and volunteers to put one together, I would be delighted to provide a xerox of the long out-of-print Russian original.

Recent books. Books on L-stable processes also contain information on the distributions. They include Samorodnitsky & Taqqu 1994, Janicki & Weron 1994, and the forthcoming Adler, Feldman & Taqqu 1997.

2. Selected printed sources of tables and graphics

The oldest, M & Zarnnfaller 1959 (see Section 3), contained both tables and graphics for this maximally skew case with $1 < \alpha < 2$). A few were published in M 1960i{E10} and, in a different format, in M 1960j{Appendix to E10}. As to the symmetric case, the tables corresponding to the Figure 3 of M 1963b{E14} were never published.

The next set of tables and graphics was by Holt & Crow 1973.

The recent books listed in Section 1 are alternative sources.

3. Computer access to current on line programs and worked out tables

When the need arises, my contacts are the following individuals, who have also made many contributions to the theory of L-stability and its applications in finance (for example, by contributing to. Adler et al. 1997.)

J. H. McCulloch (mcculloch.2@osu.edu) Department of Economics, Ohio State University, Columbus, OH 43210-1172.

D. B. Panton (panton@uta.edu) Department of Finance and Real Estate, University of Texas, Arlington, TX 76019-1895.

• The tables due to the preceding two authors are described in McCulloch & Panton 1997 and are available in computer readable form by anonymous FTP at the following two sites:

ecolan.sbs.ohio-state.edu/pub/skewstable

ftp.uta.edu/pub/projects/skewstable.

Web page: http://www.econ.ohio.state.edu/jhm/ios.html.

J. P. Nolan (jpnolan@american.edu) Department of Mathematics and Statistics, American University, Washington, D.C. 20016-8050.

- The Nolan program can be found on his web page, which is

http://www.cas.american.edu/njpnolan.

4. A look far back in time: the preparation of M & Zarnfaller 1959, and scientific computing before the age of FORTRAN

The computer entered my life shortly after I joined IBM in 1958. M 1958p, reporting on a lecture given in 1957, was about to appear. That paper was meant to be a manifesto to advocate the use of L-stable distributions (under the name of "Lévy laws") in the social sciences. The good thing is that my arguments were to pass the test of time. The bad thing is that M 1958p had absolutely no effect. It did not help that it was written in remarkably clumsy French, was not proofread (!), and appeared in a newsletter mailed to psychologists.

But this failure had deeper reasons. In science (contrary perhaps to politics), a manifesto is rarely effective until it is already buttressed by achievement. This is why M 1982F{*FGN*} was to combine a manifesto with a casebook of clear successes. In 1958 and even today, an immediate and intrinsic obstacle to the acceptance of L-stable distributions was that the most useful ones cannot be expressed analytically. I felt that, aside from possible use in science, those distributions were known very incompletely by mathematicians, and that closer numerical and visual acquaintance would be a boon to both the mathematical and the scientific uses.

From a distance, IBM Yorktown seemed the dream location to prepare the required computations, which is one reason why I came as a summer visitor in 1958. (No one dreamed at this time that this visit would extend into thiry-five years of employment.)

Before I left Paris, a French expert told me of a marvelous new invention named FORTRAN, which was expected to revolutionize computer programming. As a matter of fact, this expert asked a special favor: to try and provide him with access to FORTRAN.

However, FORTRAN was not yet generally used at the IBM Research Center, and my first contact with computers was disappointingly old-fashioned. The programmer assigned to my application, Frederick J. Zarnfaller, was young in years, but already a veteran of programming in machine language. Therefore, the computations that led to our joint work were, by the new standards that were already emerging, extraordinarily clumsy and interminable.

After the fact, however, it is a reason for some pride that my roots as a user of computers should reach before the creation of even the most basic tools. As described elsewhere, the same is true of my roots in computer graphics.

The tables that appeared as M & Zarnfaller 1959 (an IBM Report) provided the basis for M 1960i{E10}. The full tables were not detailed, hence not bulky. Naively, I thought they were worth publishing and would interest the probabilists and practical and theoretical statisticians. But I was wrong: no journal expressed the slightest interest in them. At that point in time, interest in the theory of L-stable distributions was at a low point, witness the fact that most of my references, for example Lévy 1925, were already ancient. The only useful textbook on the subject was Gnedenko & Kolmogorov 1954. The standard textbook of probability was then Loève 1955; it was satisfied with devoting a few rushed pages to a topic that altogether lacked generality.

PART V: BEYOND THE M 1963 MODEL

Most pages of my contribution to price variation concerned series that are dominated by the long-tailedness of the distribution of changes or of the "correlation." For the latter, much of my work centered around R/S analysis, whose asset is extreme robustness with respect to non-Gaussian margins. Soon after 1965, as seen in this part, I began to face diverse classes of prices in which long-tailedness and long-dependence are intertwined indissolubly. This stage continues to this day, as seen in Parts I and II, and this book did not wait for it to be completed.

&&&&&&&&&&&&&&&&&&&&&&&&&&&&

Journal of Business: 39, 1966, 242-255.

E19

Nonlinear forecasts, rational bubbles, and martingales

• *Chapter foreword.* Two terms are found in the title of this reprint, but not of the originals, namely, "nonlinear" and "rational bubble." They express the two main points of this paper in words that were not available to me in 1966.

The main substantive finding was that rational behavior on the part of the market may lead to "wild" speculative bubbles of the form illustrated in Figure 1. The randomness of these bubbles is called "wild" in my present vocabulary, because they can be extremely large, and their sizes and duration follow a scaling distribution. This distribution is closely akin to the L-stable distribution introduced in the model of price variation presented in M 1963b{E14}.

Today, the existence of very large bubbles in actual market records is known to everyone: both to the economists and to the newspaper-reading public from New York to Tokyo. But it may be interesting that neither the term *bubble* nor the behavior it describes were known to me when I wrote this paper. Hence, my argument can be viewed as a *genuine prediction.*

A basic mechanism that is part of my argument was rediscovered in Blanchard & Watson 1982. But the bubbles examined by these authors are very different from those I predicted. They are not "wild" at all and can even be called "mild" because the distribution of size and duration is not scaling but exponential. (This is a possibility this paper considers briefly but then dismisses.) The duration of exponentially distributed bubbles would have a small scatter, and the largest among such bubbles would not be sufficiently distinct to be observed individually. In addition, Blanchard & Watson 1982 predicts that all bubbles end by a rise or a fall of the same size. This conclusion is closely akin to Bachelier's original idea that price changes are Gaussian, but is in disagreement with reality, which is altogether different.

Now we move on to the term "nonlinear" that was also added to the title. It does not concern the economics substance, but a mathematical technique. Wild bubbles appear when forecasting is nonlinear. Linear forecasting yields altogether different results investigated in M 1971e{E20}.

Given that both themes in this paper were ahead of their time, one hears with no surprise that it was originally viewed with trepidation and a certain fear. It was submitted to the *Journal of Political Economy*, but they pressed me to allow it to be transferred to the end of a special issue of the *Journal of Business*. The editor of that special issue, James H. Lorie, contributed *Comments* that remain interesting, and include statements that can serve as substitute to the *Abstract* that was lacking in my paper. •

✦ **In lieu of abstract: Comments by J. H. Lorie.** " [This] interesting article ... is almost purely theoretical and has no direct application to the selection of investments or the management of portfolios; however, it should prove to be very important. In the last few years a significant controversy has developed over whether the prices of stocks follow a "random walk." The proponents of this view – primarily academicians – have presented an impressive body of evidence, although it is by no means definitive. If they are to be believed, knowledge of the history of

movements of the prices of stocks is of no value. Strongly opposing this view are the technicians who believe that knowledge of the movement of stock prices, properly interpreted and usually considered in conjunction with information on volume, can yield extraordinarily high profit. The technicians have to believe that "tomorrow's" prices depend to some extent and in some way on "yesterday's" prices. There is some evidence of weak dependence. The random walkers say that it is too weak to be meaningful, while the technicians would assert the contrary ... Mandelbrot [shows] that it is theoretically possible for dependence to exist without knowledge of such dependence being valuable or capable of producing a profit.

"Of the work in progress on [diverse aspects of finance], much has had the effect of discrediting beliefs – and even some relatively sophisticated ones – about the behavior of security prices. Much of the work now in progress centers on the careful testing of more such beliefs, and I feel safe in predicting that the majority of the findings will be of the same general sort." ✦

THE BEHAVIOR OF SPECULATIVE PRICES has always been a subject of extreme interest. Most past work, including M 1963b{E14} emphasized the statistics of price changes. The present paper goes one step farther, and relates my earlier findings concerning the behavior of prices to more fundamental economic "triggering" quantities. This effort is founded on an examination, one that is simplified but demanding and detailed, of the roles that anticipation and expected utility play in economics.

1. INTRODUCTION AND SUMMARY OF EARLIER INVESTIGATIONS

My findings will depend on both the behavior of the underlying "triggering" variable and the relationship between the "triggering" variable and the price. It is possible to conceive of models where the price series follows a pure random walk, that is, price changes are *independent.* It is also possible to conceive of models where successive price changes are *dependent*. When prices do not follow a random walk, but the dependence cannot be used to increase expected profits, probabilists say that prices follow a "martingale process." Before exploring these intriguing possibilities, however, it is appropriate to begin with a brief review of the current state of affairs in the field.

When examining prices alone, one assumes implicitly that all other economic quantities are unknown and that their effects on the development of the price series $Z(t)$ are random. The stochastic mechanism that will generate the future values of $Z(t)$ may, however, depend on its past and present values. Insofar as the prices of securities or commodities are concerned, the strength of this dependence has long concerned market analysts and certain academic economists, and remarkably contradictory conclusions have evolved.

Among the market analysts, the technicians claim that a speculator can considerably improve his prospects of gain by correctly interpreting certain telltale "patterns" that a skilled eye can help him extract from the records of the past. This naturally implies that the future development of $Z(t)$ is greatly, although not exclusively, influenced by its past. It also implies that different traders, concentrating on different portions of the past record, should make different estimates of the future price $Z(t+T)$.

Academic economists tend to be skeptical of systematic trading schemes. An example is found in Section VI of M 1963b{E14} and its continuation in Fama & Blume 1966. These economists like to emphasize that, even if successive price changes were generated by tossing a fair coin, price series would include spurious "patterns." One should therefore expect that more elaborate probabilistic generating mechanisms could account for some other patterns as well and possibly even for all patterns. As a result, the basic attitude of economists is that the significance of any pattern must always be evaluated in the light of some stochastic model.

The earliest stochastic treatment of price behavior is found in the 1900 dissertation of Louis Bachelier. Bachelier 1900 conceived several models, of varying generality and complexity. His most general and least developed model states that the present price is an unbiased estimator of the price at any moment in the future. Bachelier's second-level model asserts that, for every t and T, the increment $Z(t+T)-Z(t)$ is independent of the values of Z up to and including time t. This assumption is best referred to as the "random walk." Bachelier's third-level model, the only one to be fully developed, asserts that $Z(t+T)-Z(t)$ is a Gaussian random variable with zero mean and a variance proportional to T. The present term for such a $Z(t)$ is "Brownian motion."

Despite its popularity, the Gaussian model clearly is contradicted by the facts. First of all, M 1963b{E14} focussed on the distribution of price changes and showed that price increments which are stable-Lévy accounts surprisingly well for many properties of extremely long price series. The present paper strives for an even better model, one generated by an explicit economic mechanism. The marginal distribution of price increments will be scaling, but the increments *will not* be independent. To be perfectly honest, an assumption of independence will creep in by the back door, through the hypotheses that will be made concerning the intervals between the instants when the weather changes. It would be easy to make less specific probabilistic assumptions but very hard to carry out their implications.

We shall find that the sample variation of price exhibits a variety of striking "patterns" but that these not benefit the trader, on the average.

The stochastic process $Z(t)$ to be examined will be a martingale. To define this concept, denote by t, $t+T$ and $\tilde{t}_i$ the present instant of time, a future instant and an arbitrary set of past instants. $Z(t)$ is a martingale if

$$E[Z(t+T), \text{ given the values of } Z(t) \text{ and of all the } Z(\tilde{t}_i)] = Z(t).$$

One immediate result of this definition is that $E[Z(t+T)$, given the value of $Z(t)] = Z(t)$. However, *much more* is implied in the martingale equality. It demands statistical independence between future anticipations and *all* past values of Z. Thus, one can define a martingale in two stages. (A) It is possible to speak of a single value for $E[Z(t+T) \mid Z(t)]$, without having to specify by which *past* values this expectation is conditioned. (B) One has $E[Z(t+T) \mid Z(t)] = Z(t)$. This two-stage definition should underline the central role that martingales are likely to play in the problem to which the present work is devoted: that of the usefulness of a knowledge of past prices for purposes of forecasting.

It should also be stressed that the *distribution* of $Z(t+T)$, conditioned by known values of $Z(t)$ and of the $Z(\tilde{t}_i)$, may very well depend upon the past values $Z(\tilde{t}_i)$. The expectation is unique in being unaffected by the $Z(\tilde{t}_i)$.

The application of martingales to price behavior gives meaning to the loose idea that prices are somehow "unbiased." This idea goes back at least to Bachelier, in whose mind "unbiasedness" meant that price determination in active speculative markets is governed by a linear utility function.

However, let us consider some nonlinear function F of the price. In general the expectation of $F[Z(t+T)]$ will *not* equal the present value $F[Z(t)]$. This means that, if our speculator's private utility function is not linear in Z, playing on Z may be advantageous or disadvantageous for him. Moreover, individual speculators need not be led by the same utility as the market considered as a whole. They may, for example, either seek or avoid a large dispersion of possible future prices $Z(t+T)$. Even in the case of a martingale, an increasingly detailed knowledge of the past may be useful for such purposes.

Similarly, if $\log Z$ is a martingale, playing on Z will be advantageous to speculators having a linear utility function. The fact that unbiasedness is linked to a choice of scale for Z is well known to mathematical statisticians.

Interest in martingales among pure probabilists is such that an immense variety of martingale processes has been described. If we dealt with a single economic series, namely the price, the choice among this wealth of possibility could only be directed by purely mathematical criteria – a notoriously poor guide. Hence, the present step beyond the random walk was undertaken only within the context of a "fundamental analysis," in which the price attempts to follow "value." That is, the present price $Z(t)$ is a function of past prices, and of the past and present values of the exogenous trigger $Y(t)$. In the present paper, the process generating value will be such that, as T increases, the expectation of $Y(t+T)$ will tend fairly rapidly toward a limit. Taking that limit to be the present price $Z(t)$, will achieve two results. (1) Price and value will occasionally coincide. (2) Price will be generated by a martingale stochastic model in which the present $Z(t)$ is an unbiased estimator of $Z(t+T)$. Moreover, for large enough values of T, $Z(t)$ is an unbiased estimator of $Y(t+T)$.

If, however, the process generating Y has other properties, the forecast future value $E[Y(t+\text{infinity})]$ need not be a martingale. An example to the contrary is given in Section IIG. Therefore, the fact that forecasting the value leads to a martingale in the price tells us something about the structure of the value as well as the structure of the market mechanism. If the forecasted value does not follow a martingale, price could follow a martingale only if they fail to follow forecasted value.

The above considerations are linked with the often-raised question of whether one can divide the speculators into several successive groups where members of the first group know only the present and past values of Z; members of the next group also know the present and past values of the single series Y, and know how the price will depend upon the variation of Y; members of further groups also know the temporal evolution of various series that contribute to Y, and again know how these series affect the price. In the model of the present paper knowing anything beyond the present $Z(t)$ brings no advantage, on the average.

Martingales are naturally closely related to other techniques of time-series analysis that involve conditional expectations, such as regression theory, correlation theory, and spectral representation. In particular, if $Z(t)$ is a martingale, its derivative is spectrally "white" in the sense that the covariance $C(\tau)$ between $Z'(t)$ and $Z'(t+\tau)$ vanishes if $\tau \neq 0$. It follows that the expected value of the sample spectral density of $Z'(t)$ will be a constant independent of frequency. A market that can associate such a series $Z(t)$ with the exogenous $Y(t)$, can be called a "whitener" of the derivative $Y'(t)$. However, one must keep in mind that spectral methods are concerned with measuring *correlation* rather than statistical *dependence*. Spectral whiteness expresses lack of correlation, but it is *not* synonymous with independence, except in one important but atypical case: when the marginal distribution and the joint distributions of prices at different times are Gaussian. Clearly, the examples I have constructed for this paper, are *not* Gaussian. In fact, whiteness is even weaker than the martingale identity.

II. THE FORECASTING FUNCTION OF EXCHANGE MARKETS AND THE PERSISTENCE OF PRICE MOVEMENT IN AGRICULTURE COMMODITIES

II.A. Statement of the problem

The present section will be devoted to the series of *equilibrium prices* for an agricultural commodity. Consideration of fluctuations around this series, due to temporal scatter of supply and demand, will be postponed until Section III. Here, the price $Z(t)$ will be equal to the expected value of the future crop, which in turn only depends upon past and future weather, according to the following five rules: (1) Weather can only be good, bad, or indifferent. (2) One is only interested in deviations of the price from some "norm," so that it is possible to neglect the price effects of indifferent weather. (3) When there were g good days and b bad days between moments t' and t'' within the growing season, the size of crop will have increased by an amount proportional to $g-b$. (4) Under the conditions of rule (3), the "value" $Y(t)$ of a unit quantity of the crop will have decreased by an amount proportional to $g-b$. (5) At any instant t, there is a single price of a unit quantity for future delivery, equal to

$$\lim_{T \to \infty} E[Y(t+T)].$$

Units will be so chosen that the price will increase by 1¢ when the ultimate expected value $Y(t)$ increases by the effect upon the crop of one day's bad weather. These rules are very simplified, and they do not even take into account the effect upon future prices of the portions of past crops that are kept in storage.

The total problem is so complex, however, that it is best to begin by following up each of its aspects separately.

It is readily acknowledged that the rules would be much more realistic if they referred to $\log Z$ instead of Z, and similarly to the logarithms of other quantities. This transformation was avoided, however, in order to avoid burdening the notation. The interested reader can easily make the transformation by himself.

Our rational forecast of Y naturally depends upon the weather forecast, i.e., upon the past of Y, the probability distribution of the lengths of the weather runs, and the rules of dependence between the lengths of successive runs.

The crudest assumption is to suppose that the lengths of the runs of good, bad, or indifferent weather are ruled by statistically independent exponential variables – as is the case if weather on successive days is determined by independent random events. Then the future discounted with knowledge of the past is exactly the same as the future discounted without knowledge of the past. In particular, if good and bad days are equally probable, the discounting of the future will not change the prices based upon the present crop size. This means that the process ruling the variation of $Z(t)$ is the simplest random walk, with equal probabilities for an increase or a decrease of price by 1¢.

Our "intuition" about the discounting of the future is of course based upon this case. But it is not necessary that the random variable U, designating the length of a good or bad run, be exponentially distributed. In all other cases, some degree of forecasting will be possible, so that the price will be influenced by the known structure of the process ruling the weather. The extent of this influence will depend upon the conditional distribution of the random variable U, when it is known that $U \geq h$. The following subsection will therefore discuss this problem.

II.B. The distribution of random variables conditioned by truncation

Exponential random variables. To begin with, let us note that the impossibility of forecasting in the exponential case can be restated as being an aspect of the following observation: Let U be the exponential random variable for which $P(u) = \Pr\{U \geq u\} = \exp(-bu)$, and let $U(h)$ designate the conditioned random variable U, conditioned by $U \geq h > 0$. The Bayes theorem, then, yields the following results: If $u < h$, one has $\Pr\{U(h) \geq u\} = 1$; if $u > h$, one has

$$\Pr\{U(h) \geq u\} = \Pr\{U \geq u \mid U \geq h\} = \frac{\exp(-bu)}{\exp(-bu)} = \exp[-b(u-h)].$$

This means that $U(h) - h$ is a random variable independent of h, but having a mean value $1/b$ determined by the original scale of the unconditioned U.

Uniformly scaling random variables. Assume now that the distribution of U is scaling. That is, two positive parameters σ and α are given. If $u < \sigma$, one has $\Pr\{U \geq u\} = 1$; if $u > \sigma$, one has $\Pr\{U \geq u\} = (u/\sigma)^{-\alpha}$. In the present case, Bayes's theorem yields the following results: If $h < \sigma$, one has $\Pr\{U(h) \geq u\} = \Pr\{U \geq u\}$; if $\sigma < u < h$, one has $\Pr\{U(h) \geq u\} = 1$; finally, if $\sigma < h < u$, one has

$$\Pr\{U(h) \geq u\} = \Pr\{U \geq u \mid U \geq h\} = \frac{(u/\sigma)^{-\alpha}}{(h/\sigma)^{-\alpha}} = (u/h)^{-\alpha}.$$

It is clear that the typical values of $U(h)$, such as the quantiles or the expectation, are proportional to h. For example, $hq^{-1/\alpha}$ gives the value of $U(h)$ that is exceeded with the probability q.

The mean of $U(h)$ is finite only if $\alpha > 1$. In that case, one has

$$E[U(h)] = \int_h^\infty \alpha h^\alpha u^{-\alpha} du = \frac{\alpha h}{(\alpha - 1)}.$$

$$E[U(h) - h] = \frac{h}{(\alpha - 1)} = \frac{E[U(h)]}{\alpha}.$$

This last quantity is smaller or greater than h according to whether α is smaller or greater than 2; if $\alpha = 2$, one finds $E[U(h) - h] = h$.

As to the marginal probability that $h < U < h + dh$ (knowing that $h < U$) it is equal to $\alpha h^{-(\alpha+1)} dh / h^{-\alpha} = \alpha dh/h$, which decreases with h.

In order to fully assess the above findings, it is helpful to contrast them with the result valid in the Gaussian case. As a simplified intermediate case, consider the random variable U for which

$\Pr\{U \geq u > 0\} = \exp(-bu^2)$. Then the arguments developed above show that, for $u > h$, one has

$$\Pr\{U(h) \geq u\} = \Pr\{U \geq u \mid U \geq h\} = \exp[-b(u^2 - h^2)]$$
$$= \exp[-b(u+h)(u-h)].$$

It follows that all the typical values of $U(h) - h$, such as the expected value or the quantiles, are smaller than the mean, and are smaller than the quantiles of an auxiliary exponential variable $W^0(h)$ such that $\Pr\{W^0(h) \geq w\} = \exp(-2hbw)$. This shows that the mean of $U(h) - h$ is smaller than $1/2hb$, and therefore tends to zero as h tends to infinity.

Results are very similar in the Gaussian case, but the algebra is complicated and need not be given here.

An important property of the present conditioned or truncated variable $U(h)$ is that it is *scale-free* in the sense that its distribution does not depend upon the original scale factor σ. One may also say that the original scaling law is self-similar. Self-similarity is very systematically exploited in my studies of various empirical time series and spatial patterns. In particular, runs whose duration is scalingly distributed provide a very reasonable approximation to the "trend" component of a number of meteorological time series; and this is, of course, the motivation for their use in the present context.

Proof that the property of self-similarity uniquely characterizes the scaling distribution. Indeed, it means that the ratio $\Pr\{X \geq u\}/\Pr\{X \geq h\}$ be the same when X is the original variable U or the variable U divided by any positive number k. For this condition to be satisfied, the function $P(u) = \Pr\{U \geq u\}$ must satisfy $P(u)/P(h) = P(ku)/P(kh)$. Let $R = \log P$ be considered as a function of $v = \log u$. Then, the above requirement can then be written as

$$R(v) - R(v^0) = R(v + \log k) - R(v^0 + \log k).$$

This means that $R = \log P$ must be a linear function of $v = \log u$, which is, of course, the definition of the scaling law through doubly logarithmic paper, in the manner of Pareto.

II.C. Prices based upon a forecast crop size

Keeping the above preliminary in mind, let us return to the crop-forecasting problem raised in Section IIA, and assume that the lengths of successive weather runs are statistically independent random variables following the scaling distribution. It is clear that a knowledge of the past now becomes useful in predicting the future. The results become especially simple if one modifies the process slightly by assuming that weather alternates between "passive runs" or indifferent behavior, and "active runs" when it can be good or bad with equal probabilities. Then, as long as one is anywhere within a "passive run," prices will be unaffected by the number of indifferent days in the past. But if there have been h good or bad days in the immediate past, the same weather is likely to continue for a further period whose expected value is $h/(\alpha - 1)$. (Things are actually slightly more complicated, as seen in the *Comment* at the end of this Subsection.)

Recall that the crop growth due to one day of good weather decreased the price by 1¢. A good day following h other good days will then decrease the price by the amount $[1 + 1/(\alpha - 1)]$¢ $= [\alpha/(\alpha - 1)]$¢, in which the $1/(\alpha - 1)$¢ portion is due to revised future prospects. But, when good weather finally turns to "indifferent," the price will go *up* by $h/(\alpha - 1)$, to compensate for unfulfilled fears of future bounties. It should be noted that $h/(\alpha - 1)$ *is not* a linear function of the known past values of Y. This implies that the best linear forecast in not optimal.

As a result, the record of the prices of our commodity will appear as a random alternation of three kinds of period, to be designated as "flat," "convex," and "concave," and defined as follows. During flat periods, prices will very little and "aimlessly." During concave periods, prices will go up by small equal amounts every day, yet, on the last day of the period they will fall by a fixed proportion of their total rise within the whole period. Precisely the opposite behavior will hold for convex periods.

Examples of these three kinds of periods have been shown in Figure 1. If a run of good weather is interrupted by a single indifferent day, the pattern of prices will be made up of a "slow fall, a rapid rise, a slow fall, and a rapid rise." Up to a small day's move, the point of arrival will be the same as if there had been no indifferent day in between; but that single day will "break" the expectations sufficiently to prevent prices from falling as low as they would have done in its absence.

Near the end of the growing season, the above forecasts should of course be modified to avoid discounting the weather beyond the harvest. If the necessary corrections are applied, the final price will precisely correspond to the crop size, determined by the difference between the number of days of good and bad weather. These corresponding corrections will not, however, be examined here.

Comment. Let us return to $h/(\alpha - 1)$ for the expected value noted at the end of the first paragraph of this subsection. In fact a positive scaling random variable must have a minimum value $\sigma > 0$; therefore, after an active run as started, its expected future length jumps to $\sigma/(\alpha - 1)$ and stays there until the actual run length has exceeded σ. Such a fairly spurious jump will also appear in the exponential case if good weather could not follow bad weather, and conversely. One can, in fact, modify the process so as to eliminate this jump in all cases, but this would greatly complicate the formulas while providing little benefit.

It is also interesting to derive the forecast value of $E[Y(t+T) - Y(t)]$, when T is finite and the instant t is the hth instant of a bad weather run. One readily finds that

$$\frac{1}{h} E[Y(t+T) - Y(t)] = \alpha(\alpha - 1)^{-1}[1 - (1 + T/h)^{1-\alpha}] - 1.$$

This shows that the convergence of $E[Y(t+T)]$ to its asymptote is fast when h is small, and slow when h is large.

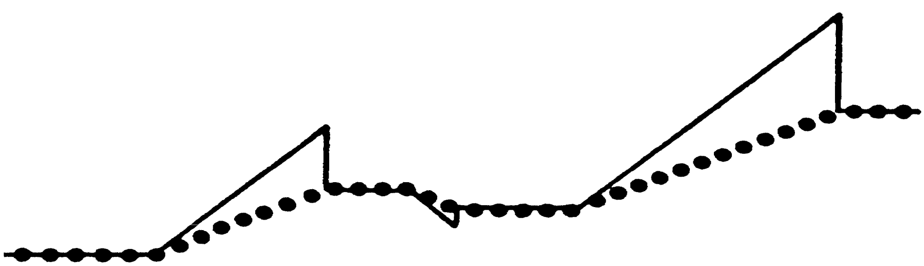


FIGURE E19-1. The abscissa is time for both the dotted and the bold lines; the ordinate is $Y(t)$ for the dotted line; The ordinate is $Z(t)$ for the bold line.

II.D. The martingale property of forecasted prices

The random series $Y(t)$ is *not* a martingale. To prove this fact, it suffices to exhibit one set of past values of Y for which the martingale property is not verified. We shall show that there is a nonvanishing conditioned expectation

$$E[Y(t+1) - Y(t), \text{ knowing the number } h \text{ of past good days}].$$

Proof. $Y(t+1) - Y(t) = 0$ if and only if the run of good weather breaks today. The probability of that event is $[h^{-\alpha} - (h+1)^{-\alpha}]/h^{-\alpha} \sim \alpha/h$. Otherwise, $Y(t+1) - Y(t) = -1$. Thus, the expectation of $Y(t+1) - Y(t)$ equals the probability that $Y(t+1) - Y(t) = -1$, which is $1 - \alpha/h$, a nonvanishing function of the past weather (whose history is fully represented for our present purpose by the duration of the current good weather run).

The price series $Z(t)$ is a martingale. To begin with, let us assume that h is known and evaluate the following conditioned increment:

$$-Z(t) + E[Z(t+1), \text{ given the value of } Z(t) \text{ and given that the number of preceding good days was exactly } h].$$

Let h be sufficiently large to avoid the difficulties due to the existence of a lower bound to the duration of a weather run. Two possibilities arise.

If weather continues to be good today, the price will go down by an amount equal to $\alpha/(\alpha-1)$. This event has a probability of $(h+1)^{-\alpha}/h^{-\alpha} \sim 1 - \alpha/h$.

If today's weather is indifferent, an event of probability α/h, the good-weather run is over and that the advance discounting of the effect of future weather was unwarranted. As a result, the price will climb up abruptly by an amount equal to $h/(\alpha-1)$.

The expected price change is thus approximately

$$\left(1 - \frac{\alpha}{h}\right)\frac{\alpha}{\alpha-1} - \frac{\alpha}{h}\frac{h}{\alpha-1} = \frac{1}{h}\alpha^2(\alpha-1),$$

which is approximately zero. The more involved rigorous derivation of the expected price change yields a value exactly (and not just approximately) equal to zero.

Now take into account the fact that one's actual knowledge of the past is usually not represented by the value of h but by some past values $Z(\tilde{t}_i)$ of $Z(t)$. The number of good days in the current run is then a random variable H, and $D(h) = \Pr\{H < h\}$ is a function determined by the values of the $Z(\tilde{t}_i)$. It follows that

$$E[Z(t+1), \text{given the value of } Z(t) \text{ and the past prices } Z(\tilde{t}_i)]$$

$$= \int dD(h) E[Z(t+1), \text{given the value of } Z(t) \text{ and the value of } h]$$

$$= \int dD(h) Z(t) = Z(t).$$

One shows very similarly that $E[Z(t+T)] = Z(t)$ when T exceeds 1, showing that $Z(t)$ *is indeed a martingale process.*

***Variance of* $\mathbf{Z(t+1) - Z(t)}$.** If the number of preceding good days was exactly h, this variance is equal to

$$\left(1 - \frac{\alpha}{h}\right)\left(\frac{\alpha}{\alpha - 1}\right)^2 + \frac{\alpha}{h}\left(\frac{h}{\alpha - 1}\right)^2,$$

which becomes proportional to h when h is large. Now suppose that h is not a known number, but is generated by a random variable H, that is conditioned by some known past prices $Z(\tilde{t}_i)$. Then the variance of $E(t+1) - Z(t)$ is proportional to $E(H)$.

If no past price is known, and $1 < \alpha < 2$, one can show that $E(H) = \infty$, and one falls back upon the infinite-variance property in M 1963b{E14}.

Comment. We have reached the climax of this story, and this is appropriate to comment again upon some observations made in the Introduction. If the price Z were generated by a random walk, then, whichever measure of risk has been adopted, no knowledge of the past should influence estimates of the risks involved in trading in Z. If, on the contrary, Z is generated by the present martingale, then the only "risk" that is not influenced by the past is constituted by the expectation of Z. A martingale is thus a "fair game." But, as h increases, so do the expected deviations from the expectation of $Z(t+1)$, and so do all other measures of "risk." This was to be expected, since, as h increases, so does the relative contribution to Z of anticipated changes in Y. Clearly, all traders, both risk-seekers and risk-avoiders, will want to know how the market value of a crop is divided between its present value and the changes anticipated before harvest time!

Also note the following: If the market is influenced more by risk-avoiders than by risk-seekers, the martingale equality should be replaced by $E[Z(t+T)] - Z(t) > 0$, the difference increasing with the variance of $Z(t+T)$. As a consequence, prices would increase in time, on the average, especially during the periods of high variance. However, this "tendency toward price increase" would be of significance only for traders who seek risk more than does the average trader on the market.

II.E. The distributions of price changes

This distribution is symmetric, thus it will suffice to derive it when $\Delta Z = Z(t+1) - Z(t)$ is positive or zero. We will denote by W' the mean duration of an indifferent weather run, and by W'' the mean duration of a good or bad weather run. Moreover, (for simplicity's sake) it will be assumed that W' and W'' are both large when measured in days.

The most significant price changes are those that satisfy $\Delta Z > \alpha/(\alpha - 1)$. These occur only on the last days of good weather runs, so that their total probability is $1/2(W' + W'')$. Their precise distribution is obtained by simply rescaling the law ruling the duration of good weather runs. Therefore, for $z > \alpha(\alpha - 1)$, (one has)

$$\Pr\{\Delta Z \geq z\} = \left\{ \frac{z+1}{\alpha/(\alpha-1)} \right\}^{-\alpha} \frac{1}{2(W' + W'')} .$$

Next, consider the probability that $Z' = 0$. This event occurs when t is anywhere within a run of indifferent weather, so that its probability is $W'/(W' + W'')$.

Finally, consider $\Delta Z = \alpha/(\alpha - 1)$ when the instant t is within a bad weather run but is not the last instant in that run. This event has the probability $(W'' - 1)/2(W' + W'')$.

The overall distribution of daily price changes is thus a "bell" with two scaling tails. It is shaped very much like a L-stable distribution, in this sense, the present model may be considered to provide a further elaboration of the process first proposed in M 1963b{E14}.

It is now safe to mention that the martingale property of forecast prices holds independently of the distribution $P(u)$ of bad weather runs, as long as runs are statistically independent. However, any non-scaling form of $P(u)$ would predict a marginal distribution of price change that is in conflict with the evidence brought forth in M 1963b{E14}.

II.F. A more involved agricultural commodity

Although still very crude, the preceding model seems more realistic than could have been expected. It can be further improved by taking into account the possibility of crop destruction by a natural calamity, such as hail. I have found that at least some natural calamities have scaling distributions. The extent of such calamities is presumably known only gradually, and they may therefore give rise to "patterns" similar to those we have studied above. The main interest of a mixture of several exogenous variables, however, is that it is unrealistic to believe that there is a proportionality between the distribution of large price changes and that of the time intervals between them. Such a proportionality holds in the case of a single trigger $Y(t)$, but not in the case of many triggers.

II.G. Best linear forecasts cannot be used to define prices

The following results, which I state without proof can be omitted without interrupting the continuity of the present work. Let us suppose that, instead of being ruled by the process $Y(t)$ that we have described, the value is ruled by a process $\tilde{Y}(t)$ with the following properties: $\Delta\tilde{Y}(t) = Y(t+1) - Y(t)$ is a stationary Gaussian process whose covariance function is equal to that of $\Delta Y(t)$. If so, the best extrapolate $\tilde{E}[\tilde{Y}(t+T)$, knowing $\tilde{Y}(t)$ for $s \leq t]$ is linear, and identical to the best linear extrapolate of $\tilde{Y}(t+T)$. As $T \rightarrow \infty$, this extrapolate tends to infinity therefore cannot define a price series $Z(t)$.

The above example suffices to show that, in order that the price based upon a forecast value *need not* be martingale. {P.S. 1996: This topic is discussed in M 1971e{E20}.}

III. PERSISTENCE OF PRICE MOVES FOR INDUSTRIAL SECURITIES

III.A. First approximation

The arguments of Section II can be directly translated into terms of "fundamental analysis" of security prices. Suppose, that the market value of a corporation is equal to the expected value of its future size X, computed while taking account of current and past values of its size $Y(t)$. If the rules of growth are of the form that we shall presently describe, it is meaningful to specify "the" expected future size by a single number, independent of the moment in the future to which one refers, and independent of the elements of the past history available for forecasting. The resulting theory is again greatly simplified (note the omission of all reference to current yield).

Our rules of growth are such that the lengths of periods of growth and decline are random, independent, and scaling. Thus, the longer a company has grown straight up, the more the outsiders should justifiably expect that it will grow in the future. Its market value Z should therefore justifiably increase by the multiple $1+1/(\alpha-1)$ of any additional growth actually observed for $Y(t)$. If, however, the growth of Y ever stops sufficiently for everyone to perceive it, one should observe a "break of confidence" and a fall of Z equal to the fraction $1/\alpha$ of the immediately preceding rise. If the growth of Y is stopped by "breathing spells," the growth of Z will have a sawtooth pattern. If a long growth period of Y, ending on a breathing spell, is modified by the addition of an intermediate breathing spell, the ultimate value of the company would be unchanged. But a single big tooth of Z would be replaced by two teeth, neither of which attains equally dizzying heights. In the absence of breathing spells, the price can go up and up, until the discounted future growth would have made the corporation bigger than the whole economy of its country, necessitating corrections that will not be examined in this paper.

Most of the further developments of this model would be very similar to those relative to the commodity examined in Section IIC. There is, however, a difference in that, if α is small, the expected length of the further growth period may be so long that one may need to discount the future growth at some nonvanishing rate.

III.B. Second approximation

Let us now examine the case of an industrial security whose fundamental value $X(t)$ follows a process of independent increments: either Bachelier's process of independent Gaussian increments, or the process in which the increments are L-stable (M 1963b{E14}. In both models, the rate of change of X may sometimes be very rapid; in the latter model it may even by instantaneous. But it will be assumed that the market only follows $X(t)$ through a smoothed-off form $Y(t)$ for which the maximum rate of change is fairly large, but finite. (In some cases, the establishment of an upper bound $\tilde{u}$ to the changes of Y may be the consequence of deliberate attempts to insure market continuity.)

In order to avoid mathematical complications, we will continue under the simplifying assumption that time is discrete. (The continuous time case is discussed at the end of this subsection.) In addition, assume that the maximum rate of change $\tilde{u}$ is known. It is clear that, whenever the market observes $Y(t) - Y(t-1) < \tilde{u}$, it will be certain that there was no smoothing off at time t and that $X(t) = Y(t)$. If $\tilde{u}$ is large enough, the equality $X = Y$ will hold for most values of t. Thus the market price $Z(t)$ will be equal most of the time to the fundamental value $X(t)$. Every so often, however, one will reach a point of time where $Y(t) - Y(t-1) = \tilde{u}$, a circumstance that may be due to any change $X(t) - X(t-1) \geq \tilde{u}$. At such instants, the value of $X(t) - X(t-1)$ will be greater than the observed value of $Y(t) - Y(t-1)$, and its conditional distribution will be given by the arguments of Section IIB; it will therefore critically depend upon the distribution of $X(t) - X(t-1)$.

If the distribution of the increments of X is Gaussian, and $\tilde{u}$ is large, the distributions of $X(t) - X(t-1)$, assuming that it is at least equal to $\tilde{u}$, will be clustered near $\tilde{u}$ as will the distribution of $X(t+1) - X(t-1)$. Hence, there will be a probability very close to 1 that $X(t+1) - X(t-1)$ will be smaller than $2\tilde{u}$ and $X(t+1) - X(t)$ will be smaller than $\tilde{u}$. As a result, $Y(t+1)$, will equal $X(t+1)$ and $Z(t+1)$ will be matched to $X(t+1)$. In other words, the mismatch between Z and Y will be small and short-lived in most cases.

Suppose now that the distribution of ΔX has two scaling tails with $\alpha < 2$. If $Y(t) - Y(t-1) = \tilde{u}$, while $Y(t-1) - Y(t-2) < \tilde{u}$, one knows that $X(t) - X(t-1) > \tilde{u}$ has a conditional expectation that is independent of the scale of the original process equal to $\alpha\tilde{u}/(\alpha - 1)$. The market price increment $Z(t) - Z(t-1)$ should therefore amplify, by the factor $\alpha/(\alpha - 1)$, the increment $Y(t) - Y(t-1)$ of the smoothed-off fundamental value.

Now, proceed to time $t+1$ and distinguish two cases: If $Y(t+1) - Y(t) < \tilde{u}$, the market will know that $X(t+1) = Y(t+1)$. Then the price $Z(t+1)$ will equal $X(t+1) = Y(t+1)$, thus falling from the inflated anticipation equal to $X(t-1) + \alpha\tilde{u}/(\alpha - 1)$. But if $Y(t+1) - Y(t) = \tilde{u}$, the market will know that $X(t+1) - X(t-1) = Y(t-1) + 2\tilde{u}$.

It follows that the conditioned difference $X(t) - X(t-1)$ will be close to following a scaling distribution truncated to values greater than $2\tilde{u}$. Thus the expected value of $X(t+1)$ – which is also the market price $Z(t+1)$ – will be

$$Z(t-1) + 2\tilde{u}\alpha/(\alpha - 1) = Z(t) + \tilde{u}\alpha/(\alpha - 1).$$

After $Y(t)$ has gone up n times in succession, in steps equal to $\tilde{u}$, the value of $Z(t+n-1)-Z(t-1)$ will approximately equal $n\tilde{u}\alpha/(\alpha-1)$. Eventually, however, n will reach a value such that $Y(t+n-1)-Y(t-1)<n\tilde{u}$, which implies $X(t+n-1)-X(t-1)<n\tilde{u}$. The market price $Z(t+n-1)$ will then crash down to $X(t+n-1)$, losing all its excess growth in one swoop.

As the size of the original jump of X increases, the number of time intervals involved in smoothing also increases, and correction terms must be added.

Let us now discuss qualitatively the case where the value of the threshold $\tilde{u}$ is random. After a change of $Y(t)$, the market must attempt to determine whether it is a fully completed change of fundamental conditions, equal to a change of $X(t)$, or the beginning of a large change. In the first case, the motion need not "persist," but in the second case, it will. This naturally involves a test of statistical significance: A few changes of Y in the same direction may well "pass" as final moves, but a long run of rises should be interpreted as due to a "smoothed-off" large move. Thus, the following, more complicated pattern will replace the gradual rise followed by fast fall that was observed earlier. The first few changes of Z will equal the changes of Y, then Z will jump to such a value that its increase from the beginning of the rise equals $\alpha/(\alpha-1)$ times the increase of Y. Whenever the rise of Y stops, Z will fall to Y; whenever the rise of Y falters, and then resumes, Z will fall to Y and then jump up again.

In a further generalization, one may consider the case where large changes of Y are gradually transmitted with probability q and very rapidly transmitted with probability $1-q$. Then the distribution of the changes of Z will be a mixture of the distribution obtained in the previous argument and of the original distribution of changes of Y. The scaling character is preserved in such a mixture, as shown in M 1963e{E3}.

Remark on continuous-time processes. Let us return to the condition of discrete time made earlier in this subsection. Continuous-time processes with independent increments were considered in M 1963b{E14}. It was shown that in the L-stable case, $X(t)$ is significantly discontinuous, in the sense that if it changes greatly during a unit time increment, this change is mostly performed in one big step somewhere within that time. Therefore, the distribution of large jumps is practically identical to the distribution of large changes over finite increments. In the Gaussian case, to the contrary, the interpolated process is continuous. More generally, whenever the process $X(t)$ is interpolable to continuous time and its increments have a finite variance, there is a great difference between the distributions of its jumps (if any) and of its changes over finite time increments. This shows that the case of infinite variance – which in practice means the scaling case – is the only one for which the restriction to discrete time is not serious at all.

III.C. More complex economic models

There may be more than one "tracking" mechanism of the kind examined so far. It may, for example, happen that $Z(t)$ attempts to predict the future behavior of a smoothed-out form Y of $X(t)$, while $X(t)$ itself attempts to predict the future behavior of $\tilde{X}(t)$. This would lead to zigzags larger than those observed so far. Therefore, for the sake of stability, it will be very important in every case to examine the driving function $Y(t)$ with care: is it a smoothed-off fundamental economic quantity, or is it already influenced by forecasting speculation.

Suppose now that two functions $Z_1(t)$ and $Z_2(t)$ attempt to track each other (with lags in each case). The zigzags will become increasingly amplified, as in the divergent case of the cobweb phenomenon. All this hints at the difficulty of studying in detail the process by which "the market is made" through the interactions among a large number of traders. It also underlines the necessity of making a detailed study of the role that the SEL assigns to the specialist, which is to "insure the continuity of the market."

IV. COMMENTS ON THE VALUATION OF OIL FIELDS

Another illustration of the use of the results of Section IIB is provided by the example of oil fields in a previously unexplored country.

"Intuitively," there is a high probability that the total oil reserves in this country are very small; but, if it turns out to be oil-rich, its reserves would be very large. This means that the a priori distribution of the reserves X is likely to have a big "head" near $x=0$ and a long "tail"; indeed, the distribution is likely to be scaling. Let us now consider a forecaster who only knows the recognized reserves $Y(t)$ at time t. As long as the reserves are not completely explored, their expected market value $Z(t)$ should equal $\alpha Y(t)/(\alpha-1)$. The luckier the explorers have been in the past, the more they should be encouraged to continue digging and the more they should expect to pay for digging rights in the immediate neighborhood of a recognized source. Eventually, $Y(t)$ will reach X and will cease to increase; at this very point, $Z(t)$ will tumble down to $Y(t)$, thus decreasing by $Y(t)/(\alpha-1)$.

If the distribution of X were exponential, $Z(t)$ would exceed $Y(t)$ by an amount independent of $Y(t)$: the market value of the entirely unexplored territory. If $Y(t)$ were a truncated Gaussian variable, the premium for expected future findings would rapidly decrease with $1/Y(t)$.

It would be interesting to study actual forecasts in the light of those three possibilities. But the main reason for considering oil fields was to demonstrate how the variation of prices can be affected by *unavoidable* delays in the transmission of information about the physical world.

Acknowledgement

An earlier version, titled "Speculative Prices and a 'Martingale' Stochastic Model" was privately circulated in February, 1964. On December 29, 1964, it was presented at the Annual Winter Meeting of the Econometric Society, held in Chicago. I am greatly indebted to Professor Eugene F. Fama for many discussions that helped clarify the presentation. The present text also takes account of comments by Professor Charles Ying.

&&&&&&&&&&&& ANNOTATIONS &&&&&&&&&&&&

Editorial changes. The footnotes of the original were either incorporated in the text or placed as digressions at the end of subsections. A last section was moved to Chapter E3, for emphasis.

Annotation 6 Section IV. I discussed the valuation of oil fields in an IBM Report that was honored over thirty years later by coming out in print, as M 1995b. I understand that M1995b has become quite influential.

The Review of Economics and Statistics, 53, 1971, 225-236.

E20

Limitations of efficiency and of martingale models

✦ **Abstract, added to the reprint in 1996.** In the moving away process

$$C(t) = \sum_{s=-\infty}^{t} L(t-s)N(s),$$

the quantities $N(s)$, called "innovations," are random variables with finite variance and are orthogonal (uncorrelated) but are not necessarily Gaussian. Knowing the value of $C(s)$ for $s < t$, that is, knowing the present and past "innovations" $N(s)$, the optimal least squares estimator of $C(t+n)$ is the conditional expected value $E_cC(t+n)$. In terms of the $N(s)$,

$$E_cC(t+n) = \sum_{s=-\infty}^{t} L(t+n-s)N(s),$$

which is a linear function of the $N(s)$ for $s \leq t$. This paper that the large n behavior of $E_cC(t+n)$ depends drastically on the value of $\Lambda = \sum_{m=0}^{\infty} L(m)$.

When $\Lambda < \infty$, then $\lim_{n\to\infty} E_cC(t+n)$ is defined and non-degenerate.

When $\Lambda = \infty$, then $\lim_{n\to\infty} E_cC(t+n)$ does not exist.

This paper interprets $C(t)$ as being a price series that remains to be arbitraged. Hence, $\lim_{n\to\infty} E_cC(t+n)$ is a linear least square extrapolation over an infinite horizon using zero interest rate.

When $\Lambda < \infty$, this extrapolate can serve to define a fully arbitraged price that is a martingale.

When $\Lambda = \infty$, this form of arbitraging diverges; it cannot produce a martingale. ✦

A COMPETITIVE MARKET OF SECURITIES, commodities or bonds may be considered efficient if every price already reflects all the relevant information that is available. The arrival of new information causes imperfections, but it is assumed that every such imperfection is promptly arbitraged away. When there is no risk aversion and the interest rate is zero, it can be shown that the arbitraged price must follow a "martingale random process." $P(t)$ will designate a price at time t, so $P(t+s)-P(t)$ is the random price change between the present time t and the future time $t+s$. The *martingale model* asserts that, knowing the present price and/or any number of past prices, the conditional expectation of $P(t+s)-P(t)$ vanishes. This simply expresses that no policy exists for buying and selling that has an *expected return* above the average return of the market.

However, as I propose to show in this paper, there exists a class of important cases where useful implementation of arbitraging is *impossible*. The principal purpose of this paper is to show why these cases are interesting. Numerous related issues will be considered along the way.

In addition, to avoid extraneous complications, I assume, first, that the process of arbitraging starts with a single, well-defined price series $P_0(t)$ – which is not itself a martingale. The idea is that $P_0(t)$ summarizes the interplay of supply, demand, etc., in the absence of arbitraging. Specifically, I shall assume that the increments of $P_0(t)$ form a stationary finite variance process. Further *I assume that the purpose of arbitraging is to replace $P_0(t)$ by a different process $P(t)$ that is a) a martingale and b) constrained not to drift from $P_0(t)$ without bound.* Had not $P(t)$ and $P_0(t)$ been constrained in some such way, the problem of selecting $P(t)$ would have been logically trivial and economically pointless. Our goal will be to achieve the smallest possible mean square drift: the variance of $P(t)-P_0(t)$ must be bounded for all t's and be as small as possible. In addition, we shall assume that the martingale $P(t)$ is linear, that is, related to $P_0(t)$ linearly (we shall explain how and why).

This being granted, this paper makes three main points:

1) A necessary and sufficient condition for the existence of the required $P(t)$ is as follows: as the large s increases, the strength of statistical dependence between $P_0(t)$ and $P_0(t+s)$ must decrease "rapidly," in a sense to be characterized later on. If an arbitraged price series $P(t)$ exists, its intertemporal variability depends upon the process P_0, and may be either greater or smaller than the variability of $P_0(t)$. In the majority of cases, arbitraging is "destabilizing," but under certain circumstances, it can be stabilizing. Note that a market specialist, in order to "insure the conti-

nuity of the market," must stabilize the variation of price. Under the circumstances under which perfect arbitraging would thus be destabilizing, the specialist prevents arbitraging from working fully and prevents prices from following a martingale.

2) When the strength of statistical dependence of $P_0(t)$ decreases very slowly, the situation is very different: the common belief is that perfect arbitraging is possible and leads to a martingale, but this belief is unfounded. Contrary to what one might have thought, such cases are much more than mathematical curiosities. Indeed, most economic time series exhibit a "Joseph Effect," a colorful way of saying that they seem ruled by a hierarchy of "cycles" of different durations (see Figure 1.) (M & Van Ness 1968, M & Wallis 1968, M & Wallis 1969a, M & Wallis 1969b.)

The simplest way of characterizing such processes is to assume that their spectral density at zero frequency is infinite, or at least extremely large. When the spectral density of $P_0(t)$ at zero frequency is infinite, it can be shown that the distribution of the daily changes of the arbitraged price $P(t)$ would have to be some known distribution scaled by an infinite constant; this outcome is absurd and demonstrates the impossibility of arbitraging.

When the spectral density of $P_0(t)$ at zero frequency is finite but very large (i.e., its memory is long, in a sense to be described, but not global,) a finite $P(t)$ can be defined, but it would not be acceptable because the variance of $P(t) - P_0(t)$ would have to increase without bound.

An interesting feature of the above syndrome ("Joseph Effect") is that it is intimately related to the "infinite variance syndrome" ("Noah Effect") which M1963b{E14} shows to be characteristic of price changes. A full empirical description of prices must take account of both Effects.

3) Imperfect arbitraging never leads to prices following a martingale. We shall gradually increase imperfection, and describe the effect on arbitraged prices, especially when perfect arbitraging is impossible.

INTRODUCTION

Classical preliminary definitions. The *random-walk model* asserts that the probability distribution of $P(t+s) - P(t)$ is independent of the current price $P(t)$ and of all past prices. That is, in selecting investments, knowledge of past prices is of no assistance. One additional assumption which is almost always made, states that the expectation $E[P(t+s) - P(t)]$ vanishes. If it does not, one speaks of a "random-walk with a drift."

The *martingale model* is less demanding: it does allow the *actual distribution* of $P(t+s)-P(t)$ to depend on past and present prices, and therefore *it does not deny* that past and present prices can serve in the selection of portfolios of different desired degrees of riskiness. For example, the martingale model allows for buying and selling policies which have much better than an even chance of being superior to the average of the market, but also have a highly appreciable chance of being enormously worse.

A correct distinction between the concepts of martingale and random-walk is made in Bachelier 1900, where one finds an informal statement of the modern concept that price in an efficient market is a martingale, and a near definitive statement of the Gaussian random-walk. A correct distinction is also made in Mandelbrot (M 1963b{E14}, M 1966b{E19}, M

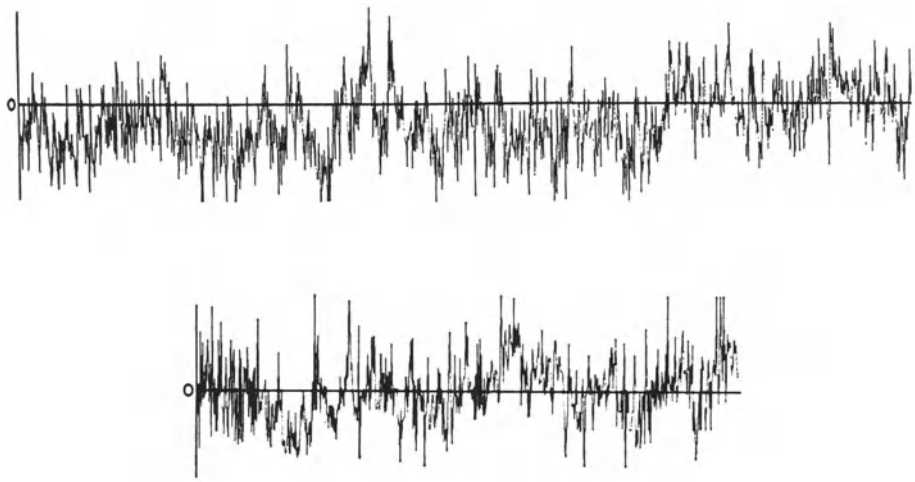


FIGURE E20-1. Either of these graphs may be a record of precipitation (M & Wallis 1969b), a computer simulated fractional Gaussian noise (M & Wallis 1969a), or an economic record interpretable as a function of causes.

This uncertainty underlines conveniently the striking resemblance that exists between those possible sources. The low frequency components show in each case through an astonishing wealth of "features" of every kind. These functions oscillate, either irregularly or near-periodically. In the latter case, the apparent wavelength of the "slow cycles" is about a third of the sample size: about 10 in a sample of 30, but 100 in a sample of 300, etc.

If fractional Gaussian noise is indeed the source of these graphs, the clear-cut cycles can only be viewed as "perceptual artifacts."

1967j{E15}), in Samuelson 1965, and in Fama 1969. Fama 1965 had claimed that evidence supports the random-walk model, but this was unwarranted.

Less special than random-walks but more special than general martingales are processes with uncorrelated ("orthogonal") increments. When $E[P(t+s)-P(t)]^2$ is finite for all t and s, and $P(t)$ is a martingale, price increments are uncorrelated and spectrally "white." If in addition, the process $P(t)$ is Gaussian, orthogonality becomes synonymous with independence and a *Gaussian martingale can only be a Gaussian random-walk.*

Combining the last result with the definitions that precede it, it is clear that *every random-walk without drift is a martingale.* The converse is also true in a Gaussian universe. But these results do not exhaust the problem of the relation between random-walks and martingales.

Attainment of complete or approximate market efficiency. One reason why the problem remains open is that market efficiency is an aspect of economic equilibrium. It is widely agreed among economists that it does not suffice to affirm that equilibrium must be realized and to study its properties; one must also show how equilibrium is either achieved or approached. Compromising between generality and tractability, studies of this type must be addressed to some fully specified model of a competitive market, which combines two assumptions. (A) An assumption about the prices that would have prevailed in the absence of arbitraging. These prices are thought to be determined by the exogenous variables. We shall assume they would have followed a finite variance stationary process – not necessarily Gaussian but nondeterministic. (B) An assumption about the chosen criterion of arbitraging. We shall assume the martingale is linear and the mean square drift is minimized. We shall only briefly comment upon other methods. Now that a model has been specified, several questions arise.

First, as was already mentioned, in all the cases where our form of perfect arbitraging does lead to well-defined prices, such prices necessarily follow a martingale. But it does *not* follow that a specific method of arbitraging *necessarily* leads to well-defined prices. Roughly speaking, fully arbitraged prices are well defined, if and only if, price changes before arbitraging $P_0(t)$ satisfy a certain special condition expressing that statistical dependence decreases rapidly. In addition, the drift of the fully arbitraged prices around $P_0(t)$ has a finite variance, if and only if, $P_0(t)$ satisfies a second special condition: rapidly decreasing dependence.

One must also investigate the partly arbitraged prices prevailing when anticipation is less than perfect. Assuming linear least squares arbitraging with a finite horizon, one would expect that arbitraging is generally less than perfect. And indeed, the changes of the arbitraged prices generally remain correlated, so prices do not follow a martingale. As anticipation improves, the correlation between successive arbitraged price changes decreases and the changes come ever nearer to being uncorrelated, but the process does not necessarily have a finite limit.

Increments of a martingale process are spectrally "white," so perfect anticipation can be called "spectrally whitening." I increasingly anticipated prices to be increasingly close to whiteness. But we shall see that, in general, an improvement in the perfection of the anticipation leads to an increase in the variance of price changes. Such a "variance increasing" transformation can be considered "destabilizing."

Last, but not least, it is important to investigate the actual process of arbitraging, but I am not qualified for this task.

Market efficiency and the syndrome of infinite variance and global dependence. The preceding reasons for being concerned about the approach to efficiency through arbitraging lie in the mainstream of conventional finite variance or Gaussian econometrics. But there is another, more personal reason for my interest. I wish to find out what arbitraging can tell us about the relations between two syndromes in which I am greatly interested: the Infinite Variance Syndrome (the Noah Effect) and the Global Dependence Syndrome (the Joseph Effect).

The term Joseph Effect is, of course, inspired by the Biblical story of the seven fat years followed by the seven lean years (*Genesis* 6:11-12). Pharaoh must have known well that yearly Nile discharges stay up and then down for variable and often long periods of time. They exhibit strong long-run dependence and a semblance of "business cycles," but without either visible or hidden sinusoidal components (Figure 1). The total size of crops in Egypt closely depends on the Nile levels. Thanks to the Biblical Joseph's ability to forecast the future by interpreting Pharaoh's dream and to arbitrage through storage, crop prices did not plummet during the seven fat years, and did not soar during the seven lean years. Unfortunately, political economists lack Joseph's gift, so the question arises, how does perfect anticipation perform when the exogenous variables and the resulting nonarbitraged prices exhibit the kind of long-run dependence described by the familiar intuitive notion of "nonsinusoidal business cycles"? I showed elsewhere (M & Wallis 1968a) that such dependence expresses a very slow decay of statistical dependence between $P_0(t)$ and $P_0(t+s)$. In this paper, I will show that, if anticipation is perfect least squares, the distribution of the arbitraged price changes rescaled by an infinite constant, for example, is a Gaussian with divergent variance, *which is absurd*. Therefore, under the stationary finite variance model of unanticipated prices, perfect linear least squares arbitraging leads nowhere.

One way to resolve this problem is to be content with imperfect (finite horizon) least squares arbitraging. This option will be explored in this paper. A different route was investigated in M 1966b{E19}, which assumed a scaling exogenous variable with a finite variance. In that case, I showed that perfect least squares arbitraging is not linear and that absurd divergence of $P(t)$ is avoided. That is, perfectly arbitraged prices are well defined and their changes follow a nondegenerate non-Gaussian distribution with an *infinite* variance, exhibiting a "Noah Effect." This last result brings us to the finding in M 1963b{E14} that the actual distributions of price increments tend to be L-stable, with infinite variance. In this paper, further pursuit of this line of thought would be out of place.

My original discovery of the Noah Effect for prices resulted from theoretical insight and from empirical study of commodity prices (cotton, wheat and other grains), security prices (rails) and various interest and exchange rates. Fama 1965 extended the L-stable model to a case study of the thirty securities of the Dow Jones index. The empirical evidence in favor of the reality of the infinite variance syndrome has continued to broaden considerably since then.

Digression concerning the use of the logarithm of price. Both the random-walk and the martingale models of price variation conflict with the basic fact that a price is necessarily positive. Perhaps the most obvious such conflict is that a price that follows a stationary random-walk would almost surely eventually become negative, which is an impossibility. Another conflict involves the martingale model itself. Suppose it were rigorously true that price itself is positive, and follows the martingale model. If so, the "martingale convergence theorem" which is described in Doob 1953, p. 319, allows one to conclude that such a price would almost surely eventually converge. A commodity or security such that its price converges must eventually cease to fluctuate, hence cease to be the object of speculation. That feature is acceptable in the special case of finite horizon commodity futures (Samuelson 1965), but in general is too stringent.

The first of the above examples of conflict is well known and has suggested to many authors that the random-walk model should not be applied to price itself, but rather to some nonlinear function of price that can tend to either plus or minus infinity – usually the logarithm of price.

This function also avoids the second conflict. But reliance on log price raises many issue. In particular, one can write price = exp (log price) and the exponential function is convex, so when log price is a martingale, price itself increases on the average. How could all the prices in an economy increase on the average?

The issues relative to log price are very interesting, but are entirely distinct from those tackled in this paper. Therefore, for the sake of notational simplicity, all arguments will be carried out in terms of price itself.

PERFECT ARBITRAGING

Nonarbitraged prices and the function of causes. In order to study the mechanism of arbitraging, we shall assume that the price $P_0(t)$ that would have prevailed in the absence of arbitraging is well defined. This assumptions is admittedly artificial. The function

$$\Delta P_0(t) = P_0(t) - P_0(t-1) = C(t)$$

will be called the "function of causes", it is supposed to summarize in dollar units all the effects of supply and demand and of anything else that can conceivably affect price – with the exception of arbitraging itself. For a heavily traded security, $C(t)$ may be dominated by significant information. For a lightly traded security, timing of large block sales and a variety of other comparatively insignificant circumstances we may call "market noise," may be dominant. In order to avoid mathematical complications, our discussion will be carried out under the assumptions that the cause $C(t)$ appears in discrete integer-valued time, and that the price change $\Delta P(t) - P(t) - P(t-1)$ follow immediately.

Independent causes and the random-walk of prices. If successive causes are independent, the arbitraged price $P(t)$ satisfies $P(t) = P_0(t)$ and $\Delta P(t) = C(t)$. Successive increments of the price $P(t)$ are independent and $P(t)$ follows a random-walk.

Dependent causes with finite variance and no deterministic component. In general, successive causes of price change cannot be assumed independent. At any moment, "something" about the future development of $C(t)$ – although of course, not everything – may be extrapolated from known past and present values. But an efficiently arbitraged market should eliminate any possibility that a method of buying and selling based on such extrapolation be systematically advantageous. When setting up prices, everything that is extrapolable from present values of the causes should be taken into account. To study such extrapolation, assume that the process $C(t)$ is generated as a moving average of the form

$$C(t) = \sum_{s=-\infty}^{t} L(t-s)N(s).$$

The quantities $N(s)$ in this expression, called "innovations," are random variables with finite variance and are orthogonal (uncorrelated) but are not necessarily Gaussian. The "lagged effect kernel" $L(m)$ must satisfy

$$\sum_{m=0}^{\infty} L^2(m) < \infty,$$

which implies that $L(m) \to 0$ as $m \to \infty$. If $L(m) = 0$ for $m > m_0$, the moving average is finite. (N is the initial of "new," and L, of "lagged.")

Moving average processes are less special than it might seem, since every "purely nondeterministic" stationary random process is of this form (see Fama 1965, Section 12.4.). The definition of "nondeterministic" is classical and need not be repeated. Our assumption about $C(t)$ implies only that deterministic elements have been removed from $C(t)$. We write

$$\Lambda = \sum_{m=0}^{\infty} L(m) \text{ and } V = \sum_{u=1}^{\infty}\left[\sum_{m=u}^{\infty} L(m)\right]^2.$$

For the random function $C(t)$, define $E_cC(t+n)$ as the conditional expected value of $C(t+n)$, knowing $C(s)$ $s<t$, that is, knowing $C(s)$ for $s<t$, in other words knowing the present and past causes. $E_cC(t+n)$ is also known to be an optimal "least squares" estimator of $C(t+n)$. Wold has shown that, in terms of the $N(s)$,

$$E_cC(t+n) = \sum_{s=-\infty}^{t} L(t+n-s)N(s)$$

which is a linear function of the $N(s)$ for $s \le t$.

Wiener and Kolmogorov have given an alternative expression of $E_cC(t+n)$ as a linear function of the past and present values of $C(t)$ itself.

However, the Wiener-Hopf technique used in implementation requires that $\Lambda < \infty$; this assumption of convergence is not innocuous; in fact, the most interesting case to be studied in this paper is when $\Lambda < \infty$. "Control-theoretical" tools have begun to draw the economists' attention. Their basic ideas are borrowed from the Kolmogorov-Wiener theory.

We shall now study the effect of this $E_cC(t+n)$ on arbitraging.

Search for the arbitraged price series P(t). A linear function of the values of $P_0(t)$ or $\Delta P_0(t)$ for $s < t$ can always be expressed as a linear function of the past values of $N(s)$, and conversely. Therefore, the prices series $P(t)$ we seek must be such that $\Delta P(t)$ is a linear function of the values of $N(s)$ for $s < t$. For $P(t)$ to be a finite variance martingale, and a linear function of past $P_0(t)$, is it necessary that $\Delta P(t)$ be proportional to $N(t)$. We shall now present an indirect argument that yields the value of this coefficient of proportionality.

Formalism of infinite horizon linear least squares arbitraging that yields a martingale. At time t, potential arbitragers will know that, for all time instants in the future, $E_cC(t+n) = E_cP_0(t+n) - E_cP_0(t+n-1)$. We suppose an infinite arbitragers' horizon, zero interest rates and no risk aversion.

The fact that $E_cC(t+n)$ is non-zero for some n implies that prices are expected to go up or down. On the average, arbitragers will bid so at to make expected arbitraged prices changes vanish. One may argue that an arbitrager should take account of the value of $E_cC(t+n)$ at any instant in the future as if it were a current cause. That is, he should add up the expected future lagged effects of each innovation $N(t)$. Clearly, the total lagged effect of $N(t)$ is $N(t)\Lambda$, where

$$\Lambda = \sum_{m=0}^{\infty} L(m).$$

That is, our arbitrager will attempt to achieve prices that satisfy

$$(*) \qquad \Delta P(t) = N(t)\Lambda.$$

Many questions arise: Can this attempt be successful; in other words, does the preceding formal expression have meaning? If it has meaning, then $P(t)$ is a martingale, but how does $P(t) - P_0(t)$ behave with increasing time? Is the mean square price drift $E[P(t) - P_0(t)]^2$ bounded for $t \rightarrow \infty$, – and among martingales – does $P(t)$ minimize this drift? And a more basic but less urgent question: are the assumptions of the present discussions realistic? All the answers will be shown to be in the affirmative if the moving average is finite, that is, if $L(m) = 0$ for large enough values of m. Otherwise, the answers depend upon the rapidity of the decrease of $L(m)$ as $m \to \infty$. Three cases must be distinguished:

- The general case where $V = \Sigma_{s=1}^{\infty} L^2(s) < \infty$.
- The more restricted but classical case where $V = \infty$, but $|\Lambda| < \infty$.
- The special case where $|\Lambda| = \infty$.

The classical case defined by $|\Lambda| < \infty$**.** This condition is necessary and sufficient to make (*) meaningful and finite. If so, the succession of price changes $\Delta P(t)$ is a sequence of orthogonal random variables with zero expectation and a finite and positive variance. These properties characterize the most general martingale for which the increments have a variance. In summary: *Under infinite horizon least squares anticipation in a finite variance universe, arbitraged prices ordinarily follow a martingale whose increments have finite variance, hence they are orthogonal.*

Subcase of the classical case: Gaussian causes. In a Gaussian universe, orthogonality is synonymous with independence. Therefore, *infinite horizon least squares anticipation in a Gaussian universe ordinarily generates prices that follow the prototype martingale, namely, the Gaussian random-walk without drift.*

Observe that in the special case $\Lambda = 0$, $P(t)$ is identically constant, hence a degenerate martingale.

Mutual price drift $P(t) - P_0(t)$ ***in the classical case.*** For all t,

$$P(t) - P_0(t) = \sum_{s=-\infty}^{t} N(s)\Lambda - \sum_{s=-\infty}^{t} N(s) \sum_{m=0}^{t-s} L(m)$$
$$= \sum_{s=-\infty}^{t} N(s) \sum_{m=t-s+1}^{\infty} L(m).$$

As a result, $E[P(t) - P_0(t)]^2$ is independent of t and equal to the previously defined quantity V, with the following definition (note that the dummy variable $t - s + 1$ is rewritten as u):

$$V = \sum_{s=1}^{\infty} L^2(s).$$

This expression introduces a second criterion.

If $V < \infty$, the martingale $P(t)$ wanders on both sides of $P_0(t)$, but does remain in a band of finite variance. If $P(t)$ is replaced by any other martingale, that is, by any martingale proportional to $P(t)$, the mean square drift is increased, which shows that if $P(t)$ is defined, it is a linear lease squares martingale.

If $V = \infty$, on the contrary, $P(t) - P_0(t)$ will drift away without bound, which according to our criteria is not admissible.

The "nonclassical case" defined by $\Lambda = \infty$**.** In this case, the perfectly arbitraged price changes should have an infinite variance, because the total price changes triggered by each innovation should be infinite. This conclusion means that perfect arbitraging case is impossible.

Discussion: role of the fractional noises. Of the two conditions $V < \infty$ and $|\Lambda| < \infty$, the condition $V < \infty$ is the more demanding one. And both are more demanding than the condition $\sum L^2(m) < \infty$ that $L(m)$ must satisfy in order to be acceptable as a kernel. One might ask, however, why go through this complicated series of conditions? Will not every decent and useful $L(m)$ satisfy any condition one might demand? The answer is no: for example, there exist processes called fractional noises (see M & Van Ness 1968), for which either or both conditions fail, and which are encountered widely.

One specific subfamily of these processes is called discrete fractional Gaussian noises (dfGn) and is characterized by the covariance

$$C_H(s) = \frac{1}{2}\{|s+1|^{2H} - 2|s|^{2H} - |1|^{2H}\},$$

where the parameter H lies between 0 and 1. The value $H = 0.5$ corresponds to the independent Gauss process, so the interesting cases are H between 0 and 0.5, and H between 0.5, and 1. If $P_0(t)$ is a dfGn with $0.5 < H < 1$, $L(m)$ is such that $\Lambda = \infty$, so perfect arbitraging is impossible even if the drift is allowed to be infinite. If $P_0(t)$ is a dfGn with $0 < H < 0.5$, then it follows that $\Lambda = 0$ but $V = \infty$, so perfect arbitraging is possible only if the drift is allowed to be infinite.

Since nonarbitraged prices are – by definition – not directly observable, it is impossible to verify the claim that any actual $P_0(t)$ function behaves like the above mentioned dfGn. But there is much indirect evidence of such behavior not only in economic time series, but also in branches of physics such as meteorology, hydrology, etc., which provide many among the more important exogenous economic variables. An excellent example is provided by the fluctuations in the level of the Nile River. Although they are devoid of sinusoidal components, the series I have in mind typically exhibit a multitude of different cycles of different apparent wavelengths: short cycles, middle cycles, and long cycles whose wavelengths have the same order of magnitude as the total sample durations. In many fields, the most economical model for such behavior is dfGn, as shown in M 1970e, M & Van Ness 1968{H}, M & Wallis 1968{H}, 1969b{H}). To be on the safe side, let me simply say that all this suggests that economic $P_0(t)$ series resembling fractional noise behavior are not exceptional. Hence, the fulfillment of the conditions Λ and $V < \infty$ is not trivial. Exceptions to martingale behavior for $P(t)$ should be expected.

IMPERFECT ARBITRAGING

The need for discounting of the distant future. Even in the classical case $\Lambda < \infty$, long-range predictions are so risky that infinite horizon least squares arbitraging would give excessive weight to future lagged effects of past innovations. Unless $L(m)$ vanishes or becomes negligible when m is still small, one must assume that the interest rate is positive and that the horizon is finite. The horizon decreases with increased risk aversion. Let us now show that under these more restrictive conditions, market efficiency must cease to coincide with the martingale condition.

Finite horizon anticipation. In the present section, the lagged effect of each innovation in the causes $C(t)$ will be followed up to some finite horizon, beyond which it will be neglected. This expresses that, for every past innovation $N(s)$, one only adds up its lagged effects up to time $t+f$, with t designating the present and f the depth of the future. Thus, the total effect of the innovation $N(s)$ will be considered equal to

$$N(s) \sum_{n=0}^{t+f-s} L(n).$$

The resulting price $P_f(t)$ satisfies

$$\Delta P_f(t) = \sum_{s=-\infty}^{t} N(s) \sum_{m=0}^{t+f-s} L(m) - \sum_{s=-\infty}^{t-1} N(s) \sum_{m=0}^{t-1+f-s} L(m)$$
$$= N(t) \sum_{m=0}^{f} L(m) + \sum_{s=-\infty}^{t-1} N(s) L(t+f-s).$$

Since $\lim_{n\to\infty} L(n) = 0$, it is easy to verify that as $f \to \infty$, $\Delta P_f(t) \to \Delta P(t)$ and $P_f(t) - P_f(0) \to P(t) - P(0)$.

This behavior expresses that the martingale process $P(t)$ of the preceding section can be considered identical to $P_\infty(t)$. But for finite f, $\Delta P_f(t)$ is a new moving average of the form

$$\Delta P_f(t) = \sum_{s=-\infty}^{t} N(s) L_f(t+1-s).$$

The function $L_f(n)$ is defined as follows

$$L_f(n) = \sum_{m=0}^{f} L(m) \qquad \text{for } n = 0;$$

$$L_f(n) = L(f+n) \qquad \text{for } n \geq 1.$$

The relationship between the two kernels $L(n)$ and $L_f(n)$ is illustrated in figure 2. The formula for $L_f(n)$ shows that the effect of finite horizon anticipation takes different forms depending upon whether or not the lagged effect function becomes strictly zero for large enough lags.

Suppose that the after-effects have a finite span f_0, meaning that $L(n)$ vanishes for all lags n satisfying $n > f_0$. Then we have

$$L_f(n) = L(f+n) = 0 \text{ for all } n \geq 1 \text{ and } f > f_0.$$

Therefore, $f > f_0$ suffices for $P_f(t)$ to become identical to the martingale $P(t) = P_\infty(t)$.

Recall that the assumption of nearly independent causes appears most reasonable when such causes are dominated by "market noise." This suggests that, among arbitraged markets, those closest to efficiency are of two kinds. In some, the anticipatory horizon is infinite. In others, the "market noise" is so overwhelming that prediction is impossible and the assumption of efficiency cannot be disproved!

Suppose now that lagged effects continue indefinitely, meaning that – however large the value of f – there exists at least one value of $n > f$ such that $L(n) \neq 0$. Then, *the arbitraged price* $P_f(t)$ *is not a martingale.* That is, it remains possible to forecast that the price will increase or decrease on the average. Good market analysts would obviously know of such instances, and they could trade accordingly, but they will be tempted to do so only if their horizon of forecasting exceeds that of the rest of the market; that is, only if the degree of risk they find acceptable – and hence the resources available to them – exceeds those of the rest of the market.

A danger is that one may proceed to partial arbitraging of an already arbitraged price; this would involve a longer horizon and greater risks than are wished. (One is reminded of Keynes's sarcastic remark about competitive prices being based on expectations about expectations, or on expectations about expectations about expectations.)

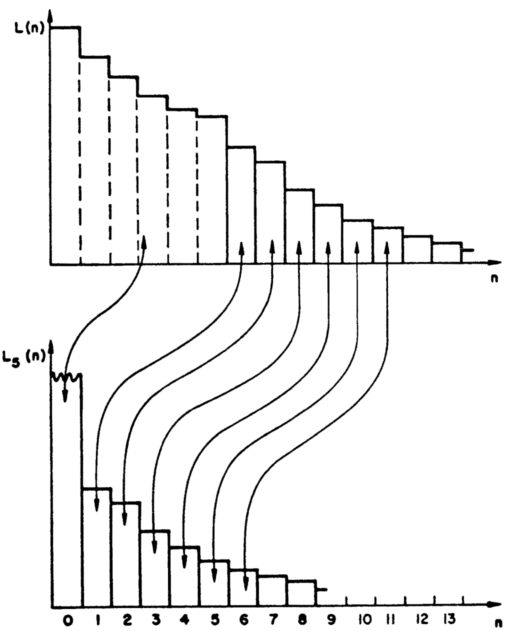


FIGURE E20-2. This is an illustration of the relationship between an original lag effect kernel $L(n)$, and the lag effect kernel $L_5(n)$ corresponding to finite horizon arbitraging of horizon $f = 5$. The heights of the bars equal the values of the kernels; areas of contours linked by arrows are identical.

Mutual drift of $P_f(t)$ versus $P_0(t)$. For all t,

$$P_f(t) - P_0(t) = \sum_{s=-\infty}^{t} N(s) \sum_{m=0}^{t-s} L_f(m) - \sum_{s=-\infty}^{t} N(s) \sum_{m=0}^{t-s} L(m)$$

$$= \sum_{s=-\infty}^{t} N(s) \sum_{m=0}^{t-s} [L_f(m) - L(m)].$$

Thus, $E[P_f(t) - P_0(t)]^2$ is again independent of t and equal to

$$V_f = \sum_{u=1}^{\infty} \left\{ \sum_{m=0}^{u-1} [L_f(m) - L(m)] \right\}^2 .$$

If $u > f$, the u^{th} term of this infinite series reduces to $[\Sigma_{m=u}^{u+f-2} L(m)]^2$, which is also the u^{th} term of the infinite series yielding $E[P_0(t+f) - P_0(t)]^2$. For every acceptable $L(m)$, that is, for every $L(m)$ satisfying $\Sigma L^2(m) < \infty$, the latter series converges, so that one has $V_f < \infty$. This proves that the drift of $P_f(t)$ from $P_0(t)$ is bounded without any additional assumption. Of course, as $f \to \infty$, $V_f \to V$, a quantity we know may be either finite or infinite.

Alternative forms of finite-horizon anticipation. One could also consider the lagged effects of all past and present innovations up to a lag of f. This leads to a price series $P_f^*(t) = N(s)\Sigma_{m=0}^{f} L(m)$, which means that $P_f^*(t)$ is a martingale. But as $t \to \infty$, the mutual drift, now defined as $P_f^*(t) - P(t)$, increases without bound for every $L(m)$.

A third form of imperfect anticipation may attribute a decreased weight $W(f)$ – for example, an exponential discount factor – to the lagged effect the innovation $N(s)$ will have at the future instant $t+f$. Then, at time t the innovation $N(s)$ has a total weighted effect equal to

$$N(s) \left\{ \sum_{n=1}^{t-s} L(n) + \sum_{f=1}^{\infty} L(t-s+f) W(f) \right\}.$$

In this case, $P(t)$ does not diverge even if $\Sigma L(m) = \infty$. With this exception, the spirit of the conclusions in preceding subsections remains unchanged.

EFFECT OF ARBITRAGING UPON VARIANCES, CORRELATIONS AND SPECTRA

The effect of arbitraging on the variance of price changes. Under the three basic conditions, namely infinite horizon least square anticipation, finite horizon anticipation, and absence of anticipation, the prices $P_\infty(t) = P(t)$, $P_f(t)$ and $P_0(t)$ satisfy

$$E[\Delta P(t)]^2 = \left\{\sum_{n=0}^{\infty} L(n)\right\}^2 E(N^2),$$

$$E[\Delta P_f(t)]^2 = \left\{\sum_{n=0}^{\infty} L_f^2(n)\right\} E(N^2), \text{ and}$$

$$E[\Delta P_0(t)]^2 = \left\{\sum_{n=0}^{\infty} L^2(n)\right\} E(N^2).$$

These formulas show that the effect of arbitraging depends on the shape of the function $L(n)$.

Case where the lagged effect kernel L(n) remains positive for all values of n and decreases monotonically to 0. If so, arbitraging is variance increasing and can be called "destabilizing."

On the one hand the variance of finite horizon anticipatory price changes $\Delta P_f(t)$ increases from a minimum for $P_0(t)$ ($f = 0$) to an asymptotic maximum for $f = \infty$. Indeed,

$$\sum_{n=0}^{\infty} L_f^2(m) = \sum_{m=0}^{\infty} L^2(m) + 2 \sum_{0 \le p < q < 0} L(p)L(q).$$

The second term on the right-hand side takes the form of a sum to which new elements are added as f increases.

On the other hand, the *lag correlation of $P_f(t+1) - P_f(t)$ decreases monotonically with f*; that is, price increments become less and less strongly interdependent as the horizon of forecasting lengthens. The proof is tedious and those not interested will skip to the end marked by QED.

In terms of the covariance

$$Cov_f(1) = E[P_f(t+1) - P_f(t)][P_f(t) - P_f(t-1)],$$

the lag correlation is written as

$$\frac{Cov_f(1)}{Cov_f(0)} = \frac{Cov_f(1)}{E(\Delta P_f)^2}.$$

We have already shown that the denominator $Cov_f(0)$ increases, therefore, it suffices to prove that the numerator $Cov_f(1)$ decreases. Observe that

$$Cov_f(1) = \sum_{n=0}^{\infty} L_f(n)L_f(n+1) = \left\{\sum_{m=0}^{f} L(m)\right\} L(f+1) + \sum_{n=f+1}^{\infty} [L(n)L(n+1)],$$

where

$$Cov_f(1) - Cov_{f-1}(1) = \left\{\sum_{m=0}^{f} L(m)\right\} L(f+1) - \left\{\sum_{m=0}^{f-1} L(m)\right\} L(f)$$

$$+ \sum_{n=f+1}^{\infty} L(n)L(n+1) - \sum_{n=f}^{\infty} L(n)L(n+1).$$

Rearranging the terms, the preceding expression equals $A + B$, where

$$A = \left\{\sum_{m=0}^{f-1} L(m)\right\} [L(f+1) - L(f)], \text{ and}$$

$$B = \left\{L(f)L(f+1) + \sum_{n=f+1}^{\infty} L(n)L(n+1) - \sum_{n=f}^{\infty} L(n)L(n+1)\right\}.$$

Term B vanishes, and term A is proportional to $L(f+1) - L(f)$, which is negative, as asserted; QED.

An important distinction. When $\Lambda < \infty$, the limit of $P_f(t)$ for $f = \infty$ is an independent process, but when $\Lambda = \infty$, $P_f(t)$ remains forever correlated.

Therefore, one should expect to find that many actual price series – even on actively arbitraged markets – are correlated. This prediction represents one of the main results of this paper.

Case where $\sum_{n=0}^{\infty} L(n)$ ***is smaller than*** $L(0)$. If so, arbitraging is variance-decreasing and can be called "stabilizing." Price variability is decreased by infinite horizon anticipation, but as f increases from 0 to ∞, the variance of $\Delta P_f(t)$ does not vary monotonically, it oscillates up and down. An example where $\sum_{n=0}^{\infty} L(n) < L(0)$ occurs is when the lagged effects of innovation begin with a nearly periodic "seasonal" before they decay for larger lags. A high seasonal is then present in unarbitraged prices, but – as one would hope – perfect arbitraging eliminates it.

Special example. When $\sum_{n=0}^{\infty} L(n) = 0$, perfect arbitraging cancels price variability completely. If unavoidable effects (like spoilage of a seasonal commodity) enter and impose a finite horizon, the seasonal effects in price variance are attenuated but not eliminated.

The viewpoint of "harmonic" or "spectral" analysis. The preceding results can be expressed in terms of spectra. The spectral density of $P_f(t+1) - P_f(t) = C(t+1)$ is known to be equal to

$$S(\lambda) = \left| \sum_{n=0}^{\infty} L_f(n) \exp(2\pi n \lambda i) \right|^2 .$$

In particular, its value $S'(0)$ at zero frequency λ is equal to $[\sum_{n=0}^{\infty} L_f(n)]^2$. Now observe that the definition of $L_f(n)$ from $L(n)$ (see Figure 2) implies that $\sum_{n=0}^{\infty} L_f(n) = \sum_{n=0}^{\infty} L(n)$, independently of the value of f. It follows that the spectral density of $P_f(t+1 - P_f(t))$ at $\lambda = 0$ is independent of the value of f.

Recall that a process is called "white" if its spectral density is independent of the frequency. The values of a white process are mutually orthogonal; if the process is Gaussian, they are independent. Now examine $P_f(t+1) - P_f(t)$ for f varying from 0 to ∞. We start with $P_0(t+1) - P_f(t) = C(t+1)$, which was assumed nonindependent, and hence nonwhite. We end up with $P(t+1) - P(t) = P_\infty(t+1) - P_\infty(t)$, which is independent (white). Hence, *perfect arbitraging whitens the spectral density.* But the value of the spectral density at $f = 0$ is invariant, and thus constitutes a kind of "pivot point." As f increases from 0 to ∞, and anticipation improves, the spectral density of $P_f(t+1) - P_f(t)$ becomes increasingly flat. But arbitraging can do absolutely nothing to the amplitude of the very low frequency effect. If $\Sigma L(m) = \infty$, the spectrum of the arbitraged price should be expected to remain unavoidably "red," that is, to include large amounts of energy within very low frequency components.

Let us now consider some of our special cases more closely. When $L(n) > 0$ for every value of n and $L(n)$ decreases as $n \rightarrow \infty$, the spectral density of $P_f(t+1) - P_f(t)$, considered for $\lambda > 0$, increases monotonically with f. In other words, the only way in which arbitraging can decrease the correlation of $P_f(t+1) - P_f(t)$ is by making its high frequency effects *stronger*. This is what makes prices more variable. Some authors have proposed to call the expression $\int \lambda S'(\lambda) d\lambda / \int S'(\lambda) d\lambda$ the average frequency of a process. In this case, we see that this quantity *increases* with improved anticipation.

When $\Sigma L(m) < L(0)$, improving arbitraging *decreases* the high frequency effects, and the average frequency decreases. In particular, in the special example where $\Sigma_{n=0}^{\infty} L(n) = 0$, the spectral density of $P_f(t+1) - P_f(t)$ for $\lambda \geq 0$ tends to zero as $f \rightarrow \infty$, though not necessarily monotonically. (For $\lambda = 0$, we know that $S'(\lambda)$ is identically zero for all f.)

***Time increments T longer than* 1.** The spectral density of $P_f(t+T) - P_f(t)$ at $\lambda = 0$ equals $T^2[\Sigma L(m)]^2$, also independently of f. This means that the argument about the origin as "pivot point" continues to hold. But otherwise things are too complicated to be worth describing here.

Alternative definitions of imperfect arbitraging. In all the instances I have examined, the above argument about the pivot at $\lambda = 0$ continues to hold true.

PRICE CONTINUITY AND THE ROLE OF THE SPECIALIST

Continuity and discontinuity are often invoked in the study of prices, but of course these mathematical terms should not be interpreted literally. Transactions occur at discrete instants of time and are quoted in discrete units, so mathematically speaking a price series is never continuous. But a series that only moves by small steps may be interpolated by a continuous function without violence, while series that move by big steps cannot. So the concepts of continuous and discontinuous price variation – if tackled cautiously – are useful. *Roughly speaking, one can say that improvement in anticipation*, through the resulting increase in high frequency energy, *makes price variation less smooth, less "continuous."*

Let us now turn briefly to the role of the specialist. If he attempts to insure the "continuity of the market" (to use the words of the Securities Exchange Commission (S.E.C.)), he will necessarily try to smooth away the high frequency price fluctuations. This acts against good arbitraging, as we have seen. In other words, the S.E.C.'s prescription creates opportunities for systematic gain.

At this time, it may be appropriate to recall the effect of price smoothing by the specialist in those cases where price changes are not Gaussian, but have infinite variance. M1963b{E14} shows that the big expected gains promised by Alexander's "filter method" hinged entirely on the assumption that price is "continuous," from which it follows that one can buy or sell at any price one has prescribed in advance. My L-stable model of price behavior predicts, to the contrary, that price is violently discontinuous. If combined with the smoothing effect of the specialist, my model predicts that every so often prices will go up or down very steeply. Alexander assumed that one could buy or sell, at a price prescribed in advance, during these periods of steep variation, but of course this possibility is not open to ordinary buyers and sellers, so Alexander's results are not in contradiction with market efficiency.

The role of the specialist deserves more study along the above lines.

Acknowledgement. This paper was read at the Annual Winter Meeting of the Econometric Society on December 28, 1969. Later, I became aware that considerable literature – including contributions by M. Friedman, W. J. Baumol, and L. Telser – concerns the question of whether or not speculation is stabilizing. Between these works and my own there are obvious similarities and differences, but this is not the place to survey the field. This text has benefited greatly from discussions with Murad Taqqu.

&&&&&&&&&&&& ANNOTATIONS &&&&&&&&&&&&

In the original title of this paper, the word *limitation* unaccountably came out as *limit*. This was corrected when the title was shortened. The original footnotes were incorporated in the text or an Appendix.

Operations Research: 15, 1967, 1057-1062;
Econometrica: 41, 1973, 157-159.

E21

Self-affine variation in fractal time (section 1 is by H. M. Taylor)

• *Chapter foreword.* Since the number of transactions in any time period is random, different distributions are needed to represent price changes over fixed numbers of transactions and fixed time periods. It appears that the former follow a Gaussian distribution, while the latter follow a L-stable distribution. M & Taylor 1967 shows that those two distributions are by no means contratictory: a scenario based on a fractal subordination time is proposed by Taylor (Section 1), then shown by Mandelbrot (Section 2) to be intimately related to an earlier discussion of the specialists' function of "ensuring the continuity of the market." Note that this scenario is only compatible with the M 1963 model restricted to a symmetric distribution of price changes. Section 3 – reproducing M 1973c – elaborates by responding to Clark 1973.

The "Panorama" provided by Chapter E6 describes steps that led from the M 1967 model to a fuller model in which "trading time" changes from being fractal to being multifractal. This change is far-reaching and is extensively discussed in *Annotations* at the end of this chapter. •

THERE ARE AT LEAST FOUR SCHOOLS OF THOUGHT concerning the stochastic models for stock prices.

In terms of number of followers, by far the most popular approach is that of the so-called "technical analyst" phrased in terms of short term trends, support and resistance levels, technical rebounds, and so on.

Rejecting this technical viewpoint, two other schools agree that sequences of prices describe a random walk, where price changes are statistically independent of previous price history, but these schools disagree in their choice of the appropriate probability distributions and/or in their choice of the appropriate "time" parameter (the physical time – days, hours – or a randomized operational time ruled by the flow of transactions). Some authors find price changes to be Gaussian (Bachelier 1900, Brada et al. 1966, Laurent 1959 and Osborne 1959), while the other group find them to follow a L-stable distribution with infinite variance (M 1963b{E14}, 1967j{E15}, Fama 1963b{E16}, 1965). Finally, a fourth group (overlapping with the preceding two) admits the random walk as a first-order approximation but notes recognizable second-order effects. (M 1966b{E19}, 1969e, Niederhoffer 1959 and Osborne 1962.)

Sections 1 and 2 of this chapter show that Brownian as motion as applied to transactions is compatible with a symmetric L-stable random walk as applied to fixed time intervals. Section 3 elaborates.

Review of L-stable distributions. Let $\{Z(t),\ t \geq 0\}$ be a stochastic process with stationary independent increments, that is, random walk. $Z(t)$ follows a L-stable distribution if (Gnedenko & Kolmogorov 1954, p. 164) its characteristic function takes the form:

$$\varphi_{Z(t)}(u) = E[\,\exp\{iuZ(t)\}] = \exp\{i\delta tu - \gamma t\,|\,u\,|^{\alpha}[1 + i\beta(u/\,|\,u\,|\,)w(u,\alpha)]\},$$

where $|\beta| \leq 1,\ 0 < \alpha \leq 2,\ \gamma > 0$ and

$$w(u,\alpha) = \tan(\pi\alpha/2) \ \text{ if } \alpha \neq 1, \ \text{ and } \ w(u,\alpha) = (2/\pi)\log|u| \ \text{ if } \alpha \neq 1.$$

In general, α is called the *characteristic exponent* of the L-stable distribution (Gnedenko & Kolmogorov 1954, p. 171.) When $\alpha = 2$, one gets the Gaussian, and $Z(t)$ is a Brownian motion process. The Cauchy corresponds to $\alpha = 1$ and $\beta = 0$. When $1 < \alpha < 2$, one has a finite mean but infinite variance and is the L-stable distribution. It is positive if and only if $\alpha < 1$, $\beta = 1$, and $\delta \geq 0$, and it is symmetric when $\beta = 0$.

1. THE "SUBORDINATION" RELATION BETWEEN THE GAUSSIAN AND THE L-STABLE DISTRIBUTIONS (BY HOWARD M. TAYLOR)

Let $\{X(v),\ v \geq 0\}$ be a Gaussian stochastic process with stationary independent increments, $E[X(v)] = 0$ and $E[X(u)X(v)] = \sigma^2 \min\{u, v\}$. The characteristic function is given by

$$\varphi_{X(t)}(\xi) = E[\exp\{i\xi X(t)\}] = \exp\{-\frac{1}{2}\xi^2\sigma^2 t\}.$$

Let $\{T(t),\ t \geq 0\}$ be a positive L-stable stochastic process with characteristic function

$$\varphi_{T(t)}(\eta) = \exp\{-\gamma t|\eta|^{\alpha}[1 + i(\eta/|\eta|)\tan(\pi\alpha/2)]\},$$

where $0 < \alpha < 1$ and we have taken $\delta = 0,\ \beta = 1$ in the general form, equation (1). Define a new process $Z(t) = X[T(t)]$. Following Solomon Bochner, this process is said to be *subordinated* to $\{X(v)\}$ and the process $\{T(t)\}$ is called the *directing process* (Feller 1950, Vol. II, p. 335).

We interpret $\{X(v),\ v \geq 0\}$ as the stock prices on a "time scale" measured in volume of transactions, and consider $T(t)$ to be the cumulative volume or number of transactions up to physical (days, hours) time t. Then $Z(t)$ is the stock price process on the physical time scale. The key fact is that $Z(t)$ is a L-stable process with independent increments and with characteristic exponent $2\alpha < 2$ (Feller 1950, Vol. II, p. 336, example *(c)*). This property may be shown by computing the characteristic function

$$\varphi_{Z(t)}(\xi) = E[\exp\{i\xi X(T(t))\}] = E\Big[E[\exp\{i\xi X(T(t))\}\,|\,T(t)]\Big]$$
$$= E[\varphi_{X(T(t))}(\xi)] = E\Big[\exp\{-\frac{1}{2}\xi^2\sigma^2 T(t)\}\Big].$$

Formally, this becomes

$$\varphi_{Z(t)}(\xi) = \varphi_{T(t)}(\frac{1}{2}i\xi^2\sigma^2) = \exp\{-\gamma(\sigma^2/2)^{\alpha}t|\xi|^{2\alpha}[1 - \tan(\pi\alpha/2)]\}$$
$$= \exp\{-\hat{\gamma}t|\xi|^{2\alpha}\},$$

where $\hat{\gamma} = \gamma(\sigma^2/2)^{\alpha}[1 - \tan(\pi\alpha/2)]$. We have thus obtained the symmetric L-stable distribution with characteristic exponent $\alpha < 2$ for which the β and δ terms in equation (1) are zero. (The preceding step is called formal because it substitutes a complex variable in a characteristic function formula developed for a real argument. But this substitution step is readily justified, see Feller 1950, Vol. II.

2. THE DISTRIBUTION OF STOCK PRICE DIFFERENCES

2.1. Introduction

As in M 1963b{E14}, successive price changes over fixed time periods are approximately independent and their distribution is approximately L-stable. This means in particular that their population variance is infinite. Defining $T(t)$ to be cumulative number of transaction up to t, one can write $Z(t) = X[T(t)]$. Taylor notes in Section 1 that, when the distribution of $Z(t)$ is symmetric and $T(t)$ is a special random function called "subordinator," the observed behavior of $Z(t)$ is compatible with the assumption that price changes between transactions are independent and Gaussian. (This representation of Z appears due to S. Bochner.)

Other authors have recently shown interest in the relations between price changes over fixed numbers of transactions and over fixed time increments. However, Granger 1966 only points out that it is conceivable that $X(T)$ be Gaussian even when $Z(t)$ is extremely long-tailed, without noting that this requires $T(t)$ to be approximately a subordinator. Brada et al. 1966 belabor the fact that the price changes between transactions are short-tailed and approximately Gaussian, a feature that is ensured by the S.E.C.'s instructions to the specialists.

As H. M. Taylor noted in Section 1, the subordinator is itself a non-Gaussian L-stable random function, with an infinite mean and a foriori an infinite variance. His remark, therefore, does not belong to the category of attempts to avoid L-stable distributions and to "save" the finite variance. Basically, he suggests an alternative way of introducing the infinite-moment hypothesis into the study of price behavior. This alternative is illuminating although restricted to the symmetric case, and as yet devoid of direct experimental backing. It deserves to be developed. Another alternative was developed in M 1966b{E19}.

$T(t)$ being an integer, it cannot be exactly a subordinator, but Taylor's conclusion is practically unaffected if one quantizes T. In the first approximation, we shall assume the jumps of T have been quantized but not smoothed out. Similarly, $X(T)$ must be a process in discrete time. Its independent Gaussian increments will be assumed to have zero mean and unit variance.

2.2. Actual and "virtual" transactions

The subordinator function is (almost surely almost everywhere) discontinuous and varies by positive jumps whose size U is characterized by the following conditional distribution

$$\text{for each } h > 0 \text{ and } u > h, \Pr\{U \geq u \mid U \geq h\} = \left(\frac{u}{h}\right)^{-\alpha/2}.$$

These jumps mean that Taylor implicitly suggests that, if $U > 1$, transactions are performed in "bunches." Let the last transaction in the bunch be called "final" and the other "virtual."

If the amounts traded in "virtual" transactions were negligible, the price change between successive final transactions would be the sum of U independent reduced Gaussian variables. The variance of the price change would be equal to the expectation of U, which is infinite; more precisely, one can easily verify that the distribution of price changes between final transactions is identical to the distribution of the sizes of the jumps of the infinite variance L-stable process.

Thus, Bochner's representation of the L-stable process brings nothing new unless the amounts are traded on the so-called "virtual" transactions are nonnegliglible.

2.3. Specialists' trades

Section VI.B of M 1963b{E14} pointed out that the discontinuities of the process $Z(t)$ were unlikely to be observed by examining transaction data. They are *either* hidden within periods when the market is closed or the trading interrupted, *or* smoothed away by specialists who, in accordance with S.E.C. instructions, "ensure the continuity of the market" by performing transactions in which they are party.

It is tempting to postulate that virtual transactions and the specialists' transactions are identical, though the latter presumably see where the prices are aimed and can achieve the desired ΔZ in less than U independent Gaussian steps. Thus, the Bochner representation is plausible and suggests a program of empirical research of the role of specialists.

The method of filters. The distribution of price changes between transactions has a direct bearing upon the "method of filters," discussed in Section VI.C of M 1963b{E14}.

Observe that the specialist can interpret the "continuity of the market" in at least two ways. *First*, smooth out small "aimless" price drifts, so that the expression $Z_{i+1} - Z_i$ is equal to zero more often than would have been the case without him. *Second*, replace most large price changes by runs of small or medium-sized changes, whose amplitudes will be so interdependent that almost all will have the same sign. On some markets, this is even ensured by the existence of legal "limits." (These are imposed on the price change within a day, but they naturally also impose an upper bound on price changes between transactions.) On other markets, limits are not fixed by distribution. Suppose, however, that right after transaction is performed at time t_i with price Z_i, a major piece of information dries out all supply of that security below a price Z^0. Then the specialist is supposed to sell from his own holdings and he will be one of the parties in the next few or many transactions. As a by-product of the "time for reflection" thus provided, such smoothing will surely affect the distribution of the quantity $Z_{i+1} - Z_i$ by eliminating most large values. The only large values that will surely remain are those corresponding to the cases where t_i is either the instant of the last transaction within a session, or the instant of the last transaction before an interruption of trading. Although such cases are extremely important, their effect upon the statistics of $Z_{i+1} - Z_i$ will be swamped out by the huge number of transactions performed within any trading session.

2.4. Difficulty of using in practice the data relative to price changes between successive transactions

Let Z_i and Z_{i+1} be successive quotations. The fact that the distribution of $Z_{i+1} - Z_i$ is short-tailed (or even Gaussian) is now seen to be fully compatible with the L-stable behavior of $Z(t+T) - Z(t)$. However, even if individual investors had "transaction clocks," which they do not, they cannot insure that their transactions would be the next to be executed on the market, or the 50th or 100th after the next. A buy or sell order registered when price is Z_i will be executed at a price Z^0 such that $Z^0 - Z_i$ is some complicated mixture (carried over all values of j) of quantities of the form $Z_{i+j} - Z_i$. Mixtures of Gaussian distributions are known to have fatter tails than the Gaussian, and these tails can be very fat indeed.

Similarly, it is difficult to see how transaction data can be used by economists who are not primarily concerned with the activities of the specialists. For example, if stock price changes are just one of the variables in a time series model, it would not make sense to measure the price changes over a transaction interval (and thus a variable time interval) if the other variables are measured over fixed intervals of time.

The criticism expressed in the preceding paragraph should not be interpreted as implying that the problem of trading on the NYSE is fully solved by describing the function $Z(t)$ in natural (uniformly flowing) time, because the instant when one's transaction will be executed is also impossible to fix in advance. Therefore, some kind of mixture will again appear. In Bachelier's original model, $Z(t+T)-Z(t)$ being a Gaussian random variable, one encounters the same difficulties as when mixtures of $Z_{i+j}-Z_i$ are considered. On the other hand, the L-stable scaling model has the interesting property that the asymptotic behavior of the distribution is the same for a mixture of $Z(t+T)-Z(t)$ (carried over a set of values of T), and for each $Z(t+T)-Z(t)$ taken separately. Thus, mixing has no effect upon considerations related to the asymptotic behavior.

3. COMMENTS ON A SUBORDINATED STOCHASTIC MODEL WITH FINITE VARIANCE ADVANCED BY PETER K. CLARK

✦ **Abstract.** Clark 1973 reports on experimental findings about price changes, which I view as useful, novel, and worth pursuing. It also advances yet possible alternative to the M 1967 model, in particular to my infinite variance hypothesis. This is a response to Clark. ✦

3.1. Introduction

Both M & Taylor 1967 and Clark 1973 represent price variation by stochastic processes subordinated to Brownian motion. The subordinators follow different formulas, but happen to be surprisingly close numerically. Consequently, adopting Clark's viewpoint would bring no real difference in most concrete predictions.

However, Clark raises again an issue of scientific judgment I have often encountered. On the one hand, he agrees with numerous earlier critics in believing that infinite variance is undesirable per se. In order to paper it over, he thinks the economist should welcome a finite-variance reformulation, even when otherwise undesirable features are present. On the other hand, I believe that scaling is both convenient mathematically and illuminating of reality. In order to achieve it, I believe that economists should learn to live with infinite variance. Clark's scholarly effort brings up the issue from a novel angle, which deserves a fairly detailed discussion.

3.2. The use of subordination

M & Taylor and Clark agree that price change would reduce to the familiar Brownian motion, if only it were followed in an appropriate "trading time," different from "clock time." Taylor and I had thought that trading time might coincide with transaction time, while Clark links it with volume. However, it should be kept in mind that if price variation is to proceed smoothly in trading time, then trading time itself must flow at a random and highly variable rate. Consequently, as long as the flow of trading time remains unpredictable, concrete identification of the applicable trading time leaves the problems of economic prediction unaffected.

3.3. Two possible subordinators

There is no longer any disagreement among economists that the distribution of speculative price changes is extremely long-tailed. Applied statisticians have long noted that such distributions can be described by using mixtures of Gaussian distributions with different variances. Price change distributions being nearly symmetric, these Gaussian components can have zero means. The concept of subordination of stochastic processes is "merely" a concrete interpretation of such mixing. In the present instance, it is agreed that the mixing distribution of the variance is unimodal, with low probability for very small values of variance and appreciable probability for very large values of variance. The question is, which is the precise form of this mixing distribution?

The 1963 model was shown by M & Taylor to be strictly equivalent to postulating that the mixing distribution is the "positive L-stable distribution of exponent $\alpha/2$" – with α usually near 2. Clark proposes to replace it with a lognormal distribution.

The conflict between the scaling and lognormal distributions. The preceding observation rephrases my disagreement with Clark in terms of a very old story, since the positive L-stable and the lognormal distributions have already clashed repeatedly in the past, notably in such contexts as the distribution of personal incomes, of firm sizes, and of city populations. In this instance, Clark 1973 (Footnote 14) concedes that "it is not clear theoretically why operational time should be lognormally distributed." As might be suspected from the very fact that each alternative has found sensible statisticians to defend it, their actual graphs look alike. In fact, different choices of fitting criteria may leave the field to different contenders. Therefore, not only do I have numerous reservations about Clark's statistical techniques – especially about the estimation of α, but his statistical scores also leave me unimpressed. They will be mentioned again, but dwelling upon them would distort priorities.

Scientific model-making is not *primarily a matter of curve-fitting.* When the statistician chooses to fit distinct aspects of the same phenomenon separately, the best fits tend to be both mutually incompatible, and unmanageable outside of their domain of origin. Hence, a combination of simultaneous fits of several aspects is unavoidable. For example, theories centered upon the Gaussian distribution are used in physics even in cases where it is recognized that each specific aspect of the same phenomenon could be fitted better by some alternative. The reason is that the Gaussian has special virtues: it is linked with the classical central limit theorem; it is L-stable, meaning the sum of any number of independent Gaussians is itself Gaussian; and it is analytically simple. Therefore, a Gaussian theory is considered satisfactory as long as it fits everything reasonable well.

While I have often argued against specific Gaussian theories, I do consider the above stance to be correct, because the virtues listed are real. Stability is linked with the potent scientific method of "invariances". In other words, I believe one should make it a principle not to give up the listed virtues of the Gaussian without necessity.

An added virtue of the Gaussian is that its moments are finite. This should not matter, though, because after all, moments are an acquired taste. In this light, let me run through several comparisons between my model and Clark's alternative.

3.4. Motivations for the two subordinators

A motivation for the M 1963 model. My original motivation can be translated in terms of mixing distributions, even though it was originally expressed in terms of the resulting price change distributions. Among mixing distributions of the right general shape, the positive L-stable is the only one to be (i) L-stable – meaning that the mixing distributions corresponding to time spans T of one day, one week, etc., are identical except for scale – and (ii) related to a nonclassical but usable form of the central limit theorem. Since the implementation of fast Fourier transform methods, L-stable probability densities have become easy to calculate.

Motivation for the lognormal alternative. Clark 1973 (in the already cited footnote 14) concedes that he has none. Many candidates were tried, and the lognormal scored best. If this guess is correct, the tests used by Clark would be questionable, because they were designed to test between two hypotheses, not between a hypothesis and the best of several alternatives.

3.5. Statistical tests of the two subordinators

Tests of the M 1963 model. They have been fairly rough and graphical (less so in recent work than in M 1963b{E14}), but covered diverse prices over diverse time spans.

Test of the lognormal subordinator alternative. It is less rough than in M 1963b{E14}, but is limited to one series, the prices of cotton futures averaged over different dates of delivery. If fit is indeed improved by using Clark's model for this particular series, the largest changes in this series would be smaller than predicted by my model. Even if this finding is confirmed, it may be of limited validity, since the government imposes a ceiling on maximum change of the prices of cotton futures. Therefore, cotton futures are not good material for a study of large price changes.

3.6. Finite variance – pro and con

I suspect that the above arguments would have settled the issue in favor of the M 1963 model, were it not for the matter of finiteness of moments. The variance is infinite in the L-stable distributions applicable to the price changes, and even the mean is infinite for the positive L-stable distribution used in the subordinator. In Clark's alternative, all moments are finite. Infinite variance was once called a "radical departure," but I do not believe it should continue to be shunned. The issues are discussed at length in my papers and those of E.F. Fama, but may bear repetition.

Since practical sample sizes are never very large, practical sample moments are only affected by the bulk of the distribution and not by its tail. Since we know that in their central portions all reasonable alternatives are very close to each other, the same should be true of the corresponding distributions of sample moments. If sample sizes could grow indefinitely, then in one case (L-stable) the distribution of sample moments would drift out to infinity, while in the other case (lognormal) it would thin down to a point. But those familiar asymptotics of the lognormal tell us nothing about distributions of small sample moments. When the lognormal is very skewed, its "small sample" properties are extremely erratic, and calculations of sample distributions are very complicated. Thus, the reputed "tameness" of the lognormal is entirely based on its asymptotics, and in practice it has little value. The L-stable distribution's asymptotics, to the contrary, are a precise image of its finite sample behavior. As to the moment behavior of the data (which a careful student cannot avoid and which he must describe), it *is* indeed very erratic, so the reputed "wildness" of the L-stable behavior is realistic.

3.7. Comment on Clark's empirical power law

Clark shows that the daily increments of trading time vary proportionately to the daily volume raised to the power of 2.13. This empirical discovery seems very interesting and deserves careful thought.

&&&&&&&&&&&& ANNOTATIONS &&&&&&&&&&&&

Editional changes. The first few sentences of Section 3 were rephrased to provide an Abstract. Section numbers and titles were added for the sake of clarity. "Local time" was changed to "trading time."

How Sections 1 and 2 came to be written. A very brief mathematical paper was submitted by Howard M. Taylor to the journal *Operations Research.* Called to act as referee, I strongly endorsed it and proposed to the same journal a follow-up paper concerned with the economic aspects of the problem. The editors suggested that we combine the two texts, and we agreed: they are reprinted here as Sections 1 and 2. Not until several years later did the co-authors actually meet in person.

Technical comments on Section 3. A fresh look at Clark 1973 confirms that it was a valuable follow-up to M & Taylor 1967. Unfortunately, reading that paper again revived an old disappointment since this direct filiation is acknowledged near-invisibly and incompletely.

In any event, M & Taylor 1967 and Clark 1972 barely scratch the problem of subordination and compounding. I intended to return to this problem with a paper describing the M 1972 model, which was conceived at about the same time as M 1973c. However, my attention was drawn away from finance to multifractals and fractals in physics. The present status of the M 1972 model is sketched in Chapter E6, and the comments that follow are intended to help tie it up that chapter with the present one.

A) *In terms of empirical analysis, there is no head-on disagreement.* Clark 1973 deals with cotton *futures,* while M 1963b{E14} deals with cotton *spot* prices. Trading limits set an upper bound on future price changes, therefore shorten the distribution tails. They also introduce serial dependence, a topic that was not tackled by Clark but will be examined momentarily.

B) *The availability of explicit formulas, or lack thereof.* This issue deserves comment. In the M 1967 representation, both the subordinator and the compounding function are L-stable, hence no explicit formula is available for either. However, a great deal of knowledge is available.

In Clark 1973, the subordinator has a very simple lognormal density, but the distribution of the compound increment is given by an unmanageable formula, about which hardly anything is known. For periods longer than one day, Clark's proposal becomes even more unmanageable. Against this background, the L-stable looks almost straightforward, contrary to the criticism it often receives.

C) *A subordinator must allow interpolation.* Clark reproduces many standard results and formulas from the literature, but fails to check whether or not the lognormal is suitable as a subordinator in continuous time. In fact, it is, but only due to a fact that was not discovered until later, namely, the proof in Thorin 1977 that the lognormal is infinitely divisible. However, the variables that add to a lognormal are even more unmanageable than sums of lognormals.

D) *Subordination requires the subordinator's increments to be independent.* While the M 1963 model deals with sums of a *fixed* number N of independent random variables, Clark quotes many papers on sums of a random number N of addends. However, *the assumption of independence,* already made by M & Taylor, is preserved, as required by the definition of subordination.

To the contrary, the key step from the M 1967 to the M 1972 model is to replace subordination by a form of compounding that implies a specific form of dependence.

E) *A possible remote relation between lognormality and multifractal compounding; how Clark's evidence might be reclaimed and reintrepreted.* The lognormal distribution plays a central but very delicate role in the study of multifractals. This is why the end of Section 8.6 of Chapter E1 uses a sequence of lognormals to demonstrate that, as the differencing interval increases, the tails of the distribution of price changes can become shorter.

An historical near-coincidence must now be injected. The lognormals postulated in that illustrative example were also put forward in the context of turbulence, in Kolmogorov 1962. Unfortunately, that paper hastily implied that sums of lognormals are themselves lognormal. Correcting this error proved difficult: it required the extensive work presented in M 1972j{N14}, M 1974f{N15} and M 1974c{N16}, and was the source of the notion of multifractal. Lognormality plays a role in M 1972j{N14b}, but only as a rough first approximation. However, more importantly, independence is replaced by strong serial dependence; hence, subordination is replaced by a more general form of compounding.

The data in Clark 1973 may provide unwitting evidence for the M 1972 compounding model. However, a fresh look at those old data would be less valuable than extensive tests carried on better data from scratch.

An earlier short version of Section 3, published as M 1968i. "Granger 1968, Section 3, attempted to represent the long-tailed distribution of price changes by arguing that the price changes between transactions are short-tailed, but the number $N(t)$ of transactions during the time t is a very long-tailed random variable.

"Regrettably, Granger's argument only shifts the point at which a non-Gaussian limit theorem is required.

"To represent price data correctly, the instants of transaction must indeed be such that Granger's function $N(t)$ remains a widely scattered random variable even when t is large. To account for my observation that $Z(t+T)-Z(t)$ has infinite variance, $N(t)$ must have an infinite expectation. But $N(t)/E[N(t)]$ rapidly tends to 1 whenever the Gaussian central limit theorem applies to the instants of transaction. Thus, those instants cannot satisfy the Gaussian central limit theorem. A process for which $N(t)$ remains widely scattered even when t is large is studied in M & Taylor 1957; it is identical to, therefore in no way less "strange" than, the infinite-variance L-stable processes.

"Economists interested in applying the Gaussian central limit theorem to dependent variables may ponder the following, still timely, quotation from Grenander & Rosenblatt 1957 (p. 181): "... the experimentalist would argue that in most physically realizable situations where a stationary process has been observed during a long time interval, compared to time lags for which correlation is appreciable, the average of the sample would be asymptotically normally distributed. ...Unfortunately none of the extensions of the central limit theorem of dependent variables seems to answer this problem in terms well adapted for practical interpretation."

One may add that, if the interdependence between addends is sufficiently (but not unrealistically) strong, their weighted sum will *not* converge to a Gaussian limit. For example, M 1969e describes a process with a Gaussian marginal distribution whose long-term average is not Gaussian but L-stable – as the M 1963 model postulates in the case of prices.

EB

Cumulative Bibliography, including copyright credits

Foreword to the Bibliography. ***Contents.*** *This list puts together all the references of the reprinted and new chapters in this book. The sources being very diverse and some being known to few readers, no abbreviation is used and available variants are included.*

In this list, the Selecta *volumes are flagged by being preceded by a mention of the form *N, which refers to Volume N. Publications reprinted in this or other* Selecta *volumes are flagged by being preceded by a mention of the form *N16, which refers to Volume N, Chapter 16. Those items are followed by the name of the original copyright holder. In references to publications scheduled for reprint in future* Selecta, *chapter indications are tentative or omitted. Finally, the papers and unpublished reports that are not reprinted as such but whose contents are incorporated in a chapter of this book are marked by that chapter's number preceded by an asterisk *.*

Style. *Once again, in absolutely no way does this book claim to provide a balanced exposition of the publications that are based on my work: either challenge it, or praise it by adopting and expanding its conclusions. Extensive discussions with many references are found in Mirowski 1990, 1996 and McCulloch 1996, but I never managed to read more than a few of those articles.*

I can only beg those denied adequate credit for understanding and forgiveness, and ask them to educate me when they have a chance.

ADELMAN, I. 1958. A stochastic process analysis of the size distribution of firms. *Journal of the American Statistical Association*: **58**, 893-994.

ADELMAN, I. 1965. Long cycles: fact or artifact? *American Economic Review*: **60**, 444-463.

ADLER, R., FELDMAN, R. & TAQQU, M. (eds) 1997. *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy Tailed Distributions.* Boston: Birkhäuser.

AITCHISON, J. & BROWN, J. A. C. 1957. *The Lognormal Distribution.* Cambridge: Cambridge University Press.

ALEXANDER, S. S. 1961. Price movements in speculative markets: trends or random walks. *Industrial Management Review of M.I.T.*: **2**, Part 2, 7-26. Reprinted in Cootner 1964: 199-218.

ALEXANDER, S. S. 1964. Price movements in speculative markets: No. 2. *Industrial Management Review of M.I.T.*: **4**, Part 2, 25-46. Reprinted in Cootner 1964: 338-372.

AROV, D. Z. & BOBROV, A. A. 1960. The extreme members of samples and their role in the sum of independent variables. *Theory of Probability and Its Applications*: **5**, 415-435 of the Russian original (377-396 in the English translation.)

ARMITAGE, P. 1952. The statistical theory of bacterial populations subject to mutation. *Journal of the Royal Statistical Society*: **B 14**, 1-40.

ARMITAGE, P. 1953. Statistical concepts in the theory of bacterial mutation. *Journal of Hygiene*: **51**, 162-184.

ASTRÖM, K. J. 1962. *Analysis of a First-Order Linear System...* IBM Nordiska Laboratorier Report: TN 18.073.

AUERBACH, F. 1913. Das Gesetz der Bevölkerungskonzentration. *Petermans Mitteilungen*: **59**, 74-76.

AYRES, H. 1963. Risk aversion in the warrant markets. *Industrial Management Review*: **5**, 45-54. Reprinted in Cootner 1964: 497-505.

BACHELIER, L. 1900. Théorie de la spéculation. (Doctoral disertation in Mathematical Sciences, Faculté des Sciences de Paris, defended March 29, 1900). *Annales Scientifiques de l'École Normale Supérieure*: **III-17**, 21-86. Reprinted in 1995, followed by a sequel (Paris: Gabay). Translated into English in Cootner 1964: 17-78.

BACHELIER, L. 1914. *Le jeu, la chance et le hasard.* Paris: Flammarion.

BAILLIE, R. T. 1996. Long memory processes and fractional integration in econometrics. *Journal of Economics*: **73**, 5-59.

BARNDORFF-NIELSEN, O. & BLAESILD, P. 1983. Hyperbolic distributions, *Encyclopedia of Statistical Sciences*: New York: Wiley, **3**, 700-707.

BARNEY, W. 1964. *An Investigation of Parametric Variation in a Moving Average Investment Rule.* Unpublished S. M. Thesis, M.I.T.

BARTLETT, M. S. 1966. *An Introduction to Stochastic Processes* (Second edition). Cambridge: Cambridge University Press.

BARTON, C. C. & LAPOINTE, P. R. (eds) 1995. *Fractal Geometry and its Uses in the Geosciences and Petroleum Geology.* New York: Plenum.

BATEMAN, H. 1954. *Tables of Integral Transforms*: **2**. New York: McGraw-Hill.

BAUER, J. 1964. *A Diffusion Index as Applied to Price Movement in the Stock Market.* Unpublished S. M. Thesis, M.I.T.

BEALE, R. E. 1966. Review of Cootner 1964. *Journal of Political Economy*: **74**, 219.

BENCKERT, L.-G. & STERNBERG, I. 1957. An attempt to find an expression for the distribution of fire damage amount. *Transactions 15th International Congress of Actuaries*: **2**, 288-294. New York.

BERAN, J. 1994. *Statistics for long-memory processes.* London: Chapman & Hall.

***N5** BERGER, J. M. & MANDELBROT, B. B. 1963. A new model for the clustering of errors on telephone circuits. *IBM Journal of Research and Development*: **7**, 224-236.

BIENAYMÉ, J. 1853. Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés. *Comptes Rendus* (Paris): **37**, 309-329.

BILLINGSLEY, P. 1968. *Convergence of Probability Measures.* New York: Wiley.

BLANCHARD, O. J. & WATSON, M. W. 1982. Bubbles, rational expectations, and financial markets. *Crises in the Economic and Financial Structure.* P. Wachtel (Ed.). Lexington, MA: Lexington Books.

BLATTBERG, R. & GONEDES, N. 1974. A comparison of Student's and stable distributions as statistical models of stock prices. *Journal of Business*: **47**, 244-280.

BLUME, M. 1968. *The Assessment of Portfolio Performance: An Application of Portfolio Theory.* University of Chicago Dissertation.

BLUMENTHAL, R. M., GETOOR, R. V. & RAY, D. On the distribution of first hits for the symmetric stable process. *Translation of the American Mathematics Society*: **99**, 540-554.

BONESS, A. J. 1962. *A Theory and Measurement of Stock Option Value.* Ph.D. Dissertation, University of Chicago.

BONESS, A. J. 1962. Some evidence on the profitability of trading in put and call options. Cootner, 1964: 475-498.

BOWLEY, A. L. 1933. The action of economic forces in producing frequency distribution of income prices and other phenomena: suggestion for a study. *Econometrica*: **1**, 358-372.

BRADA, J., ERNST, H. & VAN TASSEL, J. 1966. The distribution of stock price differences: Gaussian after all? *Operations Research*: **14**, 334-340.

BRAS, R. F. & RODRIGUEZ-ITURBE. 1993. *Random Functions and Hydrology.* New York: Dover.

CALVET, L., FISHER, A. & MANDELBROT, B. B. 1997. Large deviations and the distribution of price changes (to appear).

CANTELLI, F. P. 1921. Sulla deduzione delle leggi de frequenza da considerazioni di probabilità. *Metron*: **1**, 83-91.

CASTELLANI, M. 1951. On multinomial distributions with limited freedom: a stochastic genesis of Pareto's and Pearson's curves. *Annals of Mathematical Statistics*: **21**, 289-293.

CAUCHY, A. 1853. Sur les résultats les plus probables. *Comptes Rendus* (Paris): **37**, 198-206.

CHAMPERNOWNE, G. 1953. A model of income distribution. *Economic Journal*: **63**, 318-351. Reprinted as Appendix 6 to Champernowne, G. *The Distribution of Income Between Persons.* Cambridge, U.K.: Cambridge University Press, 1973, 245-283.

CHANDRASEKHAR, S. 1943. Stochastic problems in physics and astronomy. *Reviews of Modern Physics*: **15**, 1-89. Reprinted in Wax 1954,

CIOCZEK-GEORGES, R. & MANDELBROT, B. B. 1995. A class of micropulses and antipersistent fractional Brownian motion. *Stochastic Processes and their Applications*: **60**, 1-18.

CIOCZEK-GEORGES, R. & MANDELBROT, B. B. 1996. Alternative micropulses and fractional Brownian motion (with Renata Cioczek-Georges). *Stochastic Processes and their Applications*:

CIOCZEK-GEORGES, R. & MANDELBROT, B. B. 1996. Stable fractal sums of pulses: the general case.

CIOCZEK-GEORGES, R., MANDELBROT, B. B., SAMORODNITSKY, G., & TAQQU, M. S. 1995. Stable fractal sums of pulses: the cylindrical case. *Bernoulli*: **1**, 201-216.

CLARK, P. K. 1973. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*: **41**, 135-155.

COOTNER, P. H. 1960. Returns to speculators: Telser vs. Keynes. *Journal of Political Economy*: **68**, 396-404, and Rejoinder, ibid, 415-418.

COOTNER, P. H. 1962. Stock prices: Random walks vs. finite Markov chains. *Industrial Management Reviews of M.I.T.*: **3**, 24-45. Reprinted in Cootner 1964: 231-252.

COOTNER, P. H. (Ed.) 1964. *The Random Character of Stock Market Prices.* Cambridge, MA: M.I.T. Press.

COWLES, A., 1933. Can stock market forecasters forecast? *Econometrica*: **1**, 309-324.

COWLES, A. 1938. *Common Stock Indexes 1871-1937.* Bloomington, IN: Principia Press, 499.

COWLES, A. 1960. A revision of previous conclusions regarding stock price behavior. *Econometrica*: **28**, 909-915. Reprinted in Cootner 1964: 132-138.

COWLES, A. & JONES, H. E. 1937. Some a posteriori probabilities in stock market action. *Econometrica*: **5**, 280-294.

DARLING, D. A. 1952. The influence of the maximum term in the addition of independent random variables. *Transactions of the American Mathematical Society*: **73**, 95-107.

DARLING, D. A. 1956. The maximum of sums of stable random variables. *Transactions of the American Mathematical Society*: **XX**, 114-169.

DAVIS, L. E. 1960. The New England textile mills and the capital market: a study of industrial borrowing 1840-1860. *Journal of Economic History*: **20**, 1-30.

DAVIS, L. E. & HUGHES, J. R. T. 1960. A dollar-sterling exchange 1803-1895. *Journal of Economic History Review*: **13**, 52-78.

DEFRANCE, G. 1948. *Distribution et Dispersion des Prix et Salaires en Belgique.* Thesis of the Institute of Statistics, University of Paris.

DOOB, J. L. 1942. The Brownian movement and stochastic equations. *Annals of Mathematics*: **43**, 351-369. Reprinted in Wax 1954.

DOOB, J. L. 1953. *Stochastic Processes.* New York: Wiley.

EINSTEIN, A. 1956. *The Theory of Brownian Motion.* New York: Dover (reprint).

ELLIOTT, R. N. 1994. *R. N. Elliott's Masterworks: the Definitive Collection.* Gainesville, GA: New Classics Library.

EVERTSZ, C. J. G. 1995. Fractal geometry of financial time series. *Fractals*: **3**, 609-616.

EVERTSZ, C. J. G., PEITGEN, H. O., & VOSS, R. F. 1996. *Fractal Geometry and Analysis: The Mandelbrot Festschrift, Curaçao 1995.* Singapore: World Scientific.

FAMA, E. F. 1963a. *The Distribution of Daily Differences of Stock Prices: a Test of Mandelbrot's Stable Paretian Hypothesis.* Doctoral dissertation, Graduate School of Business, University of Chicago. Published as Fama 1965.

***E16** FAMA, E. F. 1963b. Mandelbrot and the stable Paretian hypothesis. *Journal of Business* (Chicago): **36**, 420-429. Reprinted in Cootner 1964: 297-306. © *University of Chicago Press.*

FAMA, E. F. 1965. The behavior of stock-market prices. *Journal of Business* (Chicago): **38**, 34-105.

FAMA, E. F. 1970. Efficient capital markets: a review of theory and empirical work. *Journal of Finance*: **25**, 383-417. Also in *Frontiers of quantitative economics.* (Ed.) M. D. Intrilligator. Amsterdam: North Holland, 1971, 309-36.

FAMA, E. F. & BLUME, M. 1966. Filter rules and stock-market trading. *Journal of Business* (Chicago): **39**, 226-241.

FAMILY, F. & VICSEK, T. (eds) 1991. *Dynamics of Fractal Surfaces.* Singapore: World Scientific.

FEDER, J. 1988. *Fractals.* New York: Plenum.

FELLER, W. 1936. Zur theorie der stachastischen prozesse. *Mathematische Annalen*: **113**, 119-160.

FELLER, W. 1940. On the integro-differential equations of purely discontinuous Markoff processes. *Transactions of the American Mathematical Society*: **48**, 488-515.

FELLER, W. 1950. *An Introduction to Probability Theory and its Applications.* New York: Wiley. *NOTE.* The two volumes of this classic textbook went through several editions. Volume 1, first edition came out in 1950. Therefore, all those books are identified as Feller 1950, particulars being added when necessary.

FELLER, W. 1951. The asymptotic distribution of the range of sums of independent random variables. *Annals of Mathematical Statistics*: **22**, 427-.

FERMI, E. 1949. On the origin of the cosmic radiation. *Physical Review*: **75**, 1169-1174.

FIELITZ, B. D. & ROZELLE, J. P. 1983. Stable distributions and the mixtures of distributions hypotheses for common stock returns. *Journal of the American Statistical Association*: **78**, 28-36.

FISHER, A., CALVET, L. & MANDELBROT, B. B. 1997. The multifractality of the Deutshmark / US Dollar exchange rates (to appear).

FRÉCHET, M. 1927. Sur la loi de probabilité de l'écart maximum. *Annales de la Société Polonaise de Mathématiques* (Cracovie): **6**, 93.

FRÉCHET, M. 1941. Sur la loi de répartition de certaines grandeurs géographiques. *Journal de la Société de Statistique de Paris* : **82**, 114-122.

FRÉCHET, M. 1959. Sur la répartition des revenus. *Comptes Rendus* (Paris): **249**, 837-9.

FRIEDMAN, J. B. 1974. The architect's compass in creation miniatures of the later middle ages. *Tradition, Studies in Ancient and Medieval History, Thought, and Religion*: 419-429.

GIBRAT, R. 1932. *Les inégalités économiques.* Paris: Sirey.

GINGERICH, O. 1993. *The Eye of Heaven: Ptolemy, Copernicus, Kepler.* Washington DC: American Institute of Physics Press.

GLEICK, J. 1986. *Chaos, the Birth of a New Science.* New York: Viking.

GNEDENKO, B. V. & KOLMOGOROV, A. N. 1954. *Limit Distributions for Sums of Independent Random Variables.* English translation by K. L. Chung. Reading, MA: Addison Wesley.

GODFREY, M. D., GRANGER, C. W. J., and MORGENSTERN, O. 1964. The random walk hypothesis of stock market behavior. *Kyklos*: **17**, 1-30.

GRANGER, C. W. J. 1966. The typical spectral shape of an economic variable. *Econometrica*: **34**, 150-61.

GRANGER, C. W. J. 1968. Some aspects of the random walk model of stock market prices. *International Economic Review*: **9**. 253-257.

GRANGER, C. W. J. & MORGENSTERN, O. 1963. Spectal analysis of New York Stock Exchange prices. *Kyklos*: **16**, 1-27. Reprinted in Cootner 1964: 162-188.

GRANVILLE, J. E. 1960. *A Strategy of Daily Stock Market Timing for Maximum Profit.* Englewood Cliffs, NJ: Prentice-Hall.

GREENE, M. T. & FIELITZ, B. D. 1979. The effect of long-term dependence on risk-return models of common stocks. *Operations Research*: **27**, 944-951.

GRENANDER, U. & ROSENBLATT, M. 1957. *Statistical Analysis of Stationary Time Series.* New York: Wiley.

GUMBEL, E. J. 1958. *Statistics of Extremes.* New York: Columbia University Press.

HAGSTROEM, K.-G. 1960. Remarks on Pareto distributions. *Skandinavisk Aktuarietidskrift*: **33**, 51-71.

HADAMARD, J. 1898. Les surfaces à courbures opposées et leurs lignes géodésiques. *Journal de Mathématiques pures et appliquées*: **5 IV**, 27-73.

HADAMARD, J. 1945. *The Psychology of Invention in the Mathematical Field.* Princeton University Press. (Reprinted, New York: Dover, 1954.)

HALBFASS, W. 1922. *Die Seen der Erde.* Ergänzungsheft Nr. 185 zu *Petermanns Mitteilungen,* Gotha: Justus Perthes.

HALL, P. 1981. A comedy of errors: the canonical form for a stable characteristic function. *Bulletin of the London Mathematical Society:* **13**, 23-27.

HANNA, F. A., PECKHAM, J. A. & LERNER, S. M. 1948. *Analysis of Wisconsin Income:* **9**, *Studies in Income and Wealth.* New York: National Bureau of Economic Research.

HANNAN, E. J. 1960. *Time Series Analysis.* London: Methuen & New York: Wiley.

HART, P. E. 1902. The size and growth of firms. *Economica:* New series, **29**, 29-39.

HART, P. E. & PRAIS, S. 1956. The analysis of the business concentration: a statistical approach. *Journal of the Royal Statistical Society:* **A 119**, 150-181.

HELPERN, S. R. 1962. *An Analysis of Stock Price Movements Using Stop Orders.* Unpublished S. B. Thesis, M.I.T..

HEYDE, C. C. 1963. On a property of the lognormal distribution. *Journal of the Royal Statistical Society* : **25**, 392-393.

HEYDE, C. C. & SENETA, E. 1977. *I. J. Bienaymé: Statistical Theory Anticipated.* New York: Springer.

HOLT, D. R. & CROW, E. L. 1973. Tables and graphs of the stable probability density functions. *Journal of Research of the National Bureau of Standards* (Mathematical Sciences): **77B**, 143-198.

HOLTSMARK, J. 1919. Über die Verbreiterung von Spektrallinien. *Annalen der Physik:* **58**, 577-630.

HOUTHAKER, H. S. 1959. Education and income. *Review of Economics and Statistics:* **41**, 24-27.

HOUTHAKKER H. S. 1959. The scope and limits of future trading. *Allocation of Economic Resources, Essays in honor of Bernard Haley.* M. Abramovitz (Ed.): 134-159.

HOUTHAKKER, H. S. 1961. Systematic and random elements in short-term price movements. *American Economic Review:* **51**, 164-172.

HURST, H. E. 1951. Long-term storage capacity of reservoirs. *Transaction of the American Society of Civil Engineers:* **116**, 770-808.

HURST, H. E. 1955. Methods of using long-term storage in reservoirs. *Proceedings of the Institution of Civil Engineers.* Part I, 519-577.

HURST, H. E., BLACK, R. P., and SIMAIKA, Y. M. 1965. *Long-Term Storage, an Experimental Study.* London: Constable.

ITÔ, K. 1951. On stochastic differential equations. *Memoirs of the American Mathematical Society:* **7**.

JANICKI, A. & WERON, A. 1994. *Simulation and Chaotic Behavior of α-Stable Stochastic Processes.* New York: Marcel Dekker.

JOHNSON, N. L., KOTZ, S. & BALAKRISHNAN, N. 1994-5. *Continuous Univariate Distributions* (2nd ed.) New York: Wiley.

KATZ, R. 1963. The profitability of put and call option writing. *Industrial Management Review:* **5**, 56-69.

KENDALL, D. G. 1952. Les processus stochastiques de croissance en biologie. *Annales de l'Institute Henri Poincaré*: **13**, 43-108.

KENDALL, M. 1953. The analysis of economic time-series – Part I. *Journal of the Royal Statistical Society*: **A 116**, 11-25. Reprinted in Cootner 1964: 85-99.

KEYNES, J. M. 1939. Professor Tinbergen's method. *Economic Journal*: **49**, 558-568.

KEYNES, J. M. 1970. On the method of statistical business research. *Economic Journal*: **50**, 145-156.

KOLMOGOROV, A. N. 1933. *Foundation of Probability Theory*. Berlin: Springer.

KOLMOGOROV, A. N. 1941. The local structure of turbulence in incompressible viscose fluid for very large Reynolds numbers. *Doklady Akademii Nauk SSSR*: **30**, 301-305. Reprinted in Friedlander & Topper 1961, 159-161.

KOLMOGOROV, A. N. 1962. A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. *Journal of Fluid Mechanics*: **13**, 82-85. Original Russian text and French translation in *Mécanique de la Turbulence.* (Colloque International de Marseille, 1961), Paris: Editions du CNRS. 447-458.

KOOSIS, P. 1988-92. *The Logarithmic Integral.* Cambridge University Press.

KORCAK, J. 1938. Deux types fondamentaux de distribution statistique. *Bulletin de l'Institut International de Statistique*: **3**, 295-299.

KRIGE, G. D. 1960. On the departure of ore value distributions from the lognormal model in South African Gold Mines. *Journal of the South African Institute of Mining and Metallurgy*: 231-244.

KRUIZENGA, R. J. 1956. *Put and Call Options: a Theoretical and Market Analysis*, Ph.D. Dissertation, M.I.T..

KRUIZENGA, R. J. 1964. Introduction to the option contract. In Cootner 1964: 377-391.

KRUIZENGA, R. J. 1964. Profit returns from purchasing puts and calls. In Cootner 1964: 392-411.

LAMPERTI, J. 1966. *Probability: a Survey of the Mathematical Theory.* Reading, MA: W. A. Benjamin.

LANGE, O. 1932. *Die Preisdispersion als Mittel zur Statistischen Messung.* Wirtschaftlicher Gleichgewichtsstorungen, Leipzig.

LARSON, A. 1960. Measurement of a random process in future prices. *Food Research Institute Studies*: **1**, 313-324. Reprinted in Cootner 1964: 219-230.

LAURENT, A. G. 1957. A stochastic model for the behavior of a set of price index-numbers and its application. *Operations Research*: **5**, 600-.

LAURENT, A. G. 1959. Comments on Brownian motion in the stock market. *Operations Research*: **7**, 806-807.

LEA, D. E. & COULSON, C. A. 1949. The distribution of the number of mutants in bacterial populations. *Journal of Genetics*: **49**, 264-285.

LEVINE, S. 1962. *Heuristic Determination of Optimum Filters for Use in a Rule for Speculative Market Action*, S. M. Thesis, M.I.T..

LÉVY, P. 1925. *Calcul des probabilités.* Paris: Gauthier-Villars.

LÉVY, P. 1937-54. *Théorie de l'addition des variables aléatoires.* Paris: Gauthier-Villars. (Second edition, 1954).

LÉVY, P. 1948-1965. *Processus stochastiques et mouvement brownien.* Paris: Gauthier-Villars.

LÉVY, P. 1970. *Quelques aspects de la pensée d'un mathématicien.* Paris: Blanchard.

LÉVY-VÉHEL, J. & WALTER, C. 1997. *Les marchés fractals: efficience, ruptures et tendances sur les marchés financiers.* Paris: Presses Universitaires de France. (Collection "Finance.")

LOÈVE, M. 1955. *Probability Theory.* New York: Van Nostrand. (More recent editions were published by Springer.)

LOTKA, A. J. 1925. *Elements of Physical Biology.* Baltimore, MD: Williams & Wilkins Reprinted as *Elements of Mathematical Biology.* New York: Dover, 1956.

LOVEJOY, A. D. 1936. *The Great Chain of Being.* Cambridge, MA: Harvard University Press.

LOVEJOY, S. & MANDELBROT, B. B. 1985. Fractal properties of rain, and a fractal model. *Tellus* A: **37**, 209-232.

LURIA, S. E. & DELBRÜCK, M. 1943. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*: **28**, 491-511.

LYDALL, H. F. 1959. The distribution of employment income. *Econometrica*: **27**, 110-115.

MACAULAY, F. R. 1922. Pareto's laws and the general problem of mathematically describing the frequency of income. *Income in the United States, Its Amount and Distribution, 1909-1919*: **2**. New York: National Bureau of Economic Research.

MACAULAY, F. R. 1932. *The Smoothing of Economic Time Series.* New York: National Bureau of Economic Research.

MACAULAY, F. R. 1936. *Some Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields, and Stock Prices in the United States since 1856.* New York: National Bureau of Economic Research.

MACHLUP, F. 1960. The supply of inventors and inventions. *Welwirtschafliches Archiv*: **85**, 210-254.

MALKIEL, B. G. 1973. *Random Walk Down the Street.* New York: W. W. Norton. (Sixth edition, 1996).

MANDELBROT, B. B. 1951. Adaptation d'un message à la ligne de transmission, I & II. *Comptes Rendus* (Paris): **232**, 1638-1740 & 2003-2005.

MANDELBROT, B. B. 1953i. An informational theory of the statistical structure of language, in *Communication Theory.* (Ed.) W. Jackson. London: Butterworth, 486-504.

MANDELBROT, B. B. 1955b. On recurrent noise limiting coding. *Information Networks*, E. Weber (Ed.). New York: Interscience: 205-221.

MANDELBROT, B. B. 1956g. *Memorandum.* University of Geneva Mathematical Institute.

MANDELBROT, B. B. 1956m. A purely phenomenological theory of statistical thermodynamics: canonical ensembles. *IRE Transactions on Information Theory*: **112**, 190-203.

MANDELBROT, B. B. 1957p. Linguistique statistique macroscopique. In *Logique, Langage et Théorie de l'Information,* by L. Apostel, B. B. Mandelbrot and R. Morf. Paris: Presses Universitaires de France, 1-80.

MANDELBROT, B. B. 1957r. *Application of Thermodynamical Methods in Communication Theory and Econometrics.* Memorandum of the University of Lille Mathematics Institute.

MANDELBROT, B. B. 1958p. Les lois statistiques macroscopiques du comportement (rôle de la loi de Gauss et des lois de Paul Lévy). *Psychologie Française*: **3**, 237-249.

***E10** MANDELBROT, B. B. 1959p. Variables et processus stochastiques de Pareto-Lévy et la répartition des revenus, I & II. *Comptes Rendus* (Paris): **249**, 613-615 & 2153-2155.

MANDELBROT, B. B. 1959s. A note on a class of skew distribution functions. Analysis and critique of a paper by H. A. Simon: *Information and Control*: **2**, 90-99.

***E10** MANDELBROT, B. B. 1960i. The Pareto-Lévy law and the distribution of income. *International Economic Review*: **1**, 79-106. © *International Economic Review.*

MANDELBROT, B. B. 1961s. Final note on a class of skew distribution functions (with a postscript). *Information and Control*: **4**, 198-216 & 300-304.

MANDELBROT, B. B. 1961b. On the theory of word frequencies and on related Markovian models of discourse. *Structures of language and its mathematical aspects.* R. Jakobson (Ed.). New York: American Mathematical Society: 120-219.

***E11** MANDELBROT, B. B. 1961e. Stable Paretian random functions and the multiplicative variation of income. *Econometrica*: **29**, 517-543. © *The Econometric Society.*

***E12** MANDELBROT, B. B. 1962q. Paretian distributions and income maximization. *Quarterly Journal of Economics*: **76**, 57-85. © *Harvard University.*

***E14,15** MANDELBROT, B.B. 1962i. *The Variation of Certain Speculative Prices.* IBM Research Report NC-87, March, 1962.

***E1; FEP** MANDELBROT, B.B. 1962c. Sur certains prix spéculatifs: faits empiriques et modèle basé sur les processus stables additifs de Paul Lévy. *Comptes Rendus* (Paris): **254**, 3968-3970.

***E14** MANDELBROT, B. B. 1963a. Abstract: *Econometrica*: **31**, 757-758. © *The Econometric Society.*

***E14** MANDELBROT, B. B. 1963b. The variation of certain speculative prices. *Journal of Business*: **36**, 394-419. Reprinted in Cootner 1964: 297-337. © *University of Chicago Press.*

***E3** MANDELBROT, B. B. 1963e. New methods in statistical economics. *Journal of Political Economy*: **71**, 421-440. © *University of Chicago Press.* Reprinted in *Bulletin of the International Statistical Institute,* Ottawa Session: **40** (2), 669-720.

***E8** MANDELBROT, B. B. 1963g. *A Survey of Growth and Diffusion Models of the Law of Pareto.* IBM External Research Note **NC-253**.

***E10** MANDELBROT, B. B. 1963i. The stable Paretian income distribution when the apparent exponent is near two. *International Economic Review*: **4**, 111-115. © *International Economic Review.*

***E10** MANDELBROT, B. B. 1963j. *The Stable Paretian Income Distribution when the Apparent Exponent is Near Two.* Memorandum appended to a joint reprint of M 1960i and M 1963i.

***E13** MANDELBROT, B. B. 1963o. *Oligopoly, Mergers, and the Paretian Size Distribution of Firms.* IBM Research Note **NC-246**, March 1963.

***E8** MANDELBROT, B. B. 1964. Random walks, fire damage and related risk phenomena. *Operations Research* **12**, 1964, 582-585. © *O.R.S.A.*

MANDELBROT, B. B. 1964t. Derivation of statistical thermodynamics from purely phenomenological principles. *Journal of Mathematical Physics*: **5**, 164-171.

MANDELBROT, B. B. 1965b. On recurrent noise limiting coding. *Information Networks, the Brooklyn Polytechnic Institute Symposium,* 205-221. New York: Interscience.

***H** MANDELBROT, B. B. 1965h. Une classe de processus stochastiques homothétiques à soi; application à la loi climatologique de H. E. Hurst. *Comptes Rendus* (Paris): **260**, 3274-7.

***E8** MANDELBROT, B. B. 1965m. Very long-tailed probability distributions and the empirical distribution of city sizes. *Mathematical Explanations in Behavioral Science.* F. Massarik & P. Ratoosh (Eds.). Homewood: Richard D. Irwin, 322-332.

***FE** MANDELBROT, B. B. 1966r. Nouveaux modèles de la variation des prix (cycles lents et changements instantanés). *Cahiers du Séminaire d'Économétrie*: **9**, 1966, 53-66.

*E19 MANDELBROT, B. B. 1966b. Forecasts of future prices, unbiased markets, and "martingale" models. *Journal of Business*: **39**, 242-255. © *University of Chicago Press.* Chapter 19 takes account of important errata in a subsequent issue of the same *Journal.*

MANDELBROT, B. B. 1967b. Sporadic random functions and conditional spectral analysis; self-similar examples and limits. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*: **3**, 155-179.

*E15 MANDELBROT, B. B. 1967j. The variation of some other speculative prices. *Journal of Business*: **40**, 393-413. © *University of Chicago Press.*

*E21 MANDELBROT, B. B. 1968i. Some aspects of the random walk model of Stock Market prices: comment. *International Economic Review*: **9**, 258.

MANDELBROT, B. B. 1969e. Long-run linearity, locally Gaussian process, H-spectra and infinite variances. *International Economic Review*: **10**, 82-111.

MANDELBROT, B. B. 1970a. Long-run interdependence in price records and other economic time series. *Econometrica*: **38**, 122-123.

MANDELBROT, B. B. 1970e. *Statistical dependence in prices and interest rates.* Papers of the Second World Congress of the Econometric Society, Cambridge, England.

MANDELBROT, B. B. 1970n. Analysis of long-run dependence in time series: the *R/S* technique. *Fiftieth Annual Report of the National Bureau of Economic Research*, 107-108.

MANDELBROT, B. B. 1970y. *Statistical Self-Similarity and Non-Laplacian Chance.* 1970 Trumbull Lectures, Yale University.

*E20 MANDELBROT, B. B. 1971e. When can price be arbitraged efficiently? A limit to the validity of the random walk and martingale model. *Review of Economics and Statistics*: **53**, 225-236. © *Harvard University.*

MANDELBROT, B. B. 1971n. Statistical dependence in prices and interest rates. *Fifty-first Annual Report of the National Bureau of Economic Research*, 141-142.

MANDELBROT, B. B. 1971q. Analysis of long-run dependence in economics: the *R/S* technique. *Econometrica*: **39** (July Supplement), 68-69.

*E14 MANDELBROT, B. B. 1972b. Correction of an error in "The variation of certain speculative prices" (M 1963b). *Journal of Business*: **40**, 542-543. © *University of Chicago Press.*

MANDELBROT, B. B. 1972c. Statistical methodology for nonperiodic cycles: from the covariance to the R/S analysis. *Annals of Economic and Social Measurement*: **1**, 259-290.

*N14 MANDELBROT, B. B. 1972j. Possible refinement of the lognormal hypothesis concerning the distribution of energy dissipation in intermittent turbulence. *Statistical models and turbulence.* M. Rosenblatt & C. Van Atta (Eds.). Lecture Notes in Physics **12**, New York: Springer, 333-351.

*E21 MANDELBROT, B. B. 1973c. Comments on "A subordinated stochastic process with finite variance for speculative prices," by Peter K. Clark. *Econometrica*: **41**, 157-160. © *The Econometric Society.*

*FE2.1 MANDELBROT, B. B. 1973f. Formes nouvelles du hasard dans les sciences. *Economie Appliquée*: **26** 307-319.

*FE3.2 MANDELBROT, B. B. 1973j. Le problème de la réalité des cycles lents, et le syndrome de Joseph. *Économie Appliquée*: **26**, 349-365.

*FE3.1 MANDELBROT, B. B. 1973v. Le syndrôme de la variance infinie et ses rapports avec la discontinuité des prix. *Économie Appliquée*: **26**, 321-348.

*E8 MANDELBROT, B. B. 1974d. A population birth and mutation process, I: Explicit distributions for the number of mutants in an old culture of bacteria. *Journal of Applied Probability*: **11**, 437-444. (Part II was distributed privately).

*N16 MANDELBROT, B. B. 1974c. Multiplications aléatoires itérées, et distributions invariantes par moyenne pondérée. *Comptes Rendus* (Paris): **278A**, 289-292 & 355-358.

*N15 MANDELBROT, B. B. 1974f. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *Journal of Fluid Mechanics*: **62**, 331-358.

MANDELBROT, B. B. 1975h. Limit theorems on the self-normalized range for weakly and strongly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie*: **31**, 271-285.

MANDELBROT, B. B. 1975o. *Les objets fractals: forme, hasard et dimension.* Paris: Flammarion.

MANDELBROT, B. B. 1980n. Fractal aspects of the iteration of $z \rightarrow \lambda z(1-z)$ for complex λ and z. *Non-Linear Dynamics* (New York, 1979). R. H. G. Helleman (Ed.). *Annals of the New York Academy of Sciences*: **357**, 249-259.

*E14 MANDELBROT, B. B. 1982c. The variation of certain speculative prices. *Current Contents*: **14**, 20. © *Institute of Scientific Information.*

MANDELBROT, B. B. 1982F *(FGN). The Fractal Geometry of Nature.* New York: Freeman.

MANDELBROT, B. B. 1982t. The many faces of scaling: fractals, geometry of nature and economics. *Self-Organization and Dissipative Structures.* W. C. Schieve & P. M. Allen: 91-109 (Eds.).

MANDELBROT, B. B. 1986l. Self-affine fractals and fractal dimension. *Physica Scripta*: **32**, 257-260.

MANDELBROT, B. B. 1986t. Self-affine fractal sets, I: The basic fractal dimensions, II: Length and area measurements, III: Hausdorff dimension anomalies and their implications. *Fractals in Physics (Trieste, 1985).* Ed. L. Pietronero and E. Tosatti. Amsterdam: North-Holland, 3-28.

MANDELBROT, B. B. 1987r. Toward a second stage of indeterminism in science (preceded by historical reflections), *Interdisciplinary Science Reviews*: **12**, 117-127.

MANDELBROT, B. B. 1995b. Statistics of natural resources and the law of Pareto. *Fractal Geometry and its Uses in the Geosciences and in Petroleum Geology.* Edited by C. C. Barton & P. La Pointe. New York: Plenum, 1-12. First appeared as IBM: NC-146, June 29, 1962.

MANDELBROT, B. B. 1995f. Measures of fractal lacunarity: Minkowski content and alternatives. *Fractal Geometry and Stochastics.* C. Bandt, S. Graf & M. Zähle (Eds.). Basel and Boston: Birkhäuser: 12-38.

MANDELBROT, B. B. 1995l. The Paul Lévy I knew. *Lévy Flights and Related Phenomena in Physics* (Nice, 1994). M. F. Schlesinger, G. Zaslawsky, & U. Frisch (Eds.) (Lecture Notes in Physics). New York: Springer, 1995, ix-xii.

MANDELBROT, B. B. 1995n. Introduction to fractal sums and pulses. *Lévy Flights and Related Phenomena in Physics.* Edited by G. Zaslawsky, M. F. Shlesinger & U. Frisch (Lecture notes in Physics). New York: Springer, 110-123.

*H MANDELBROT, B. B. 1997H. *Fractals & Self-Affinity: R/S, 1/f, Global Dependence, Relief & Rivers.* New York: Springer-Verlag.

*N MANDELBROT, B. B. 1997N. *Multifractals & 1/f Noise.* New York: Springer-Verlag.

*FE MANDELBROT, B. B. 1997FE. *Fractales, hasard et finance.* Paris: Flammarion.

MANDELBROT, B. B., FISHER, A. & CALVET, L. 1997. The multifractal model of asset returns (to appear).

*E21 MANDELBROT, B. B. & TAYLOR, H. M. 1967. On the distribution of stock price differences. *Operations Research*: **15**, 1057-1062. © *O.R.S.A.*

*H MANDELBROT, B. B. & VAN NESS, J. W. 1968. Fractional Brownian motions, fractional noises and applications. *SIAM Review*: **10**, 422-437.

*H MANDELBROT, B. B. & WALLIS, J. R. 1968. Noah, Joseph and operational hydrology. *Water Resources Research*: **4**, 909-918.

*H MANDELBROT, B. B. & WALLIS, J. R. 1969a. Computer experiments with fractional Gaussian noises, Part 1, 2 & 3. *Water Resources Research*: **5**, 228-267.

*H MANDELBROT, B. B. & WALLIS, J. R. 1969b. Some long-run properties of geophysical records. *Water Resources Research*: **5**, 321-340.

*H MANDELBROT, B. B. & WALLIS, J. R. 1969c. Robustness of the rescaled range *R/S* in the measurement of noncyclic long-run statistical dependence. *Water Resources Research*: **5**, 967-988.

MANDELBROT, B. B. & ZARNFALLER, F. 1959. *Five-place Tables of Certain Stable Distributions.* IBM. Research Technical Report RC-421.

• NOTE: See also under Berger, Calvet, Cioczek-Georges, Fisher, Lovejoy.

MANSFIELD, E. 1962. Entry, Gibrat's law, innovation, and the growth of firms. *American Economic Review*: **52**, 1023-1051.

MARSHALL, A. 1890. *Principles of Economics.* New York: Macmillan. The ninth (variorum) edition, with a second volume of annotations by C. W. Guillebaud, appeared in 1961.

MATHERON, G. 1962. *Traité de géostatistique appliquée.* **1**, Paris: Technip.

MCCULLOCH, J. H. 1996. Financial applications of stable distributions, *Statistical Methods in Finance* G. S. Maddala & C. R. Rao (Eds.). (*Handbook of Statistics*: **14**). Amsterdam: Elsevier: 393-425.

MCKIE, J. W. 1960. Market structure and uncertainty in oil and gas exploration. *Quarterly J. of Economics*: **74**, 543-571.

MEIDELL, B. 1912. Zur Theorie des Maximums (with thorough English and French summaries). *Comptes-Rendus du Septième Congrès d'Actuaires, Amsterdam*: **1**, 85-99

MEIDELL, B. 1937. Zur Theorie und Praxis des Maximums in Lebensversicherung. *Comptes-Rendus du Onzième Congrès d'Actuaires, Paris*, **1**, 469.

MILLER, H. P. 1955. *The Income of the American People.* New York: Wiley.

MILLS, F. C. 1927. *The Behavior of Prices.* New York: National Bureau of Economic Research.

MIROWSKI, P. 1990. From Mandelbrot to chaos in economic theory. *Southern Economics Journal*, **57**, 289-307.

MIROWSKI, P. 1996. Mandelbrot's economics after a quarter century. *Fractals*, **3**, 581-196. Reprinted in Evertsz & al 1996: 177-196.

MITCHELL, W. C. 1915. The making and using of index numbers. Introduction to *Index Numbers and Wholesale Prices in the United States and Foreign Countries.* U.S. Bureau of Labor Statistics Bulletin 173. Reprinted in 1921 as Bulletin 284.

MOORE, A. 1962. *A Statistical Analysis of Common Stock Prices.* Unpublished Ph.D. dissertation, Graduate School of Business, University of Chicago.

MOORE, H. L. 1911. *The Laws of Wages.* New York: Macmillan.

MORRIS, W. S. 1962. *Towards a Workable Theory of Markets.* New York: W.S. Morris & Co..

MORLAT, G., BILLIET & BERNIER 1956. Les crues de la Haute Durance. *Symposia Darcy*, Dijon.

NIEDERHOFFER, V. 1959. Clustering of stock prices. *Operations Research*: **13**, 258-265.

NOLAN, J. P. 1996a. Numerical calculation of stable densities. Preprint.

OFFICER, R. R. 1972. The distribution of stock returns. *Journal of the American Statistical Association*: **67**, 807-812.

THE OIL AND GAS JOURNAL 1958. Journal survey of proved reserves: Where the nation's oil is stored. *The Oil and Gas J.*, January 27, 1958: 163-168.

OLIVIER, M. 1927. *Les nombres indices de la variation des prix*. Paris.

OLSEN 1996. This global reference calls for a large collection of preprints and hard-to-find articles from Olsen & Associates in Zürich. The contributing authors include the following, listed alphabetically: B. Chopard, M. M. Dacarogna, R. D. Davé, C. G. de Vries, C. L. Gauvreau, D. M. Guillaume, C. Jost, C. Morgenegg, U. A. Müller, R. J. Nagler, M. Oudsaidene, O. V. Pictet, R. Schirru, M. Schwarz, M. Tomassini, J. E. von Weizsäcker, J. R. Ward. Their works can be printed from the "web."

OSBORNE, M. F. M. 1959. Brownian motion in the stock market. *Operations Research*: **7**, 145-173. Reprinted in Cootner 1964: 100-128. Reply to Comments. *Operations Research*: **7**, 807-811.

OSBORNE, M. F. M. 1962. Periodic structure in the Brownian motion of stock prices. *Operations Research*: **10**, 345-379. Reprinted in Cootner 1964: 262-296.

PARETO, V. 1896-1965. *Cours d'économie politique.* The second date refers to the reprint in Pareto: *Oeuvres complètes.* Geneva: Droz.

PARIKH, C. 1991 *The unreal life of Oscar Zariski.* New York: Academic Press.

PARZEN, E. 1960. *Modern Probability Theory and its Applications.* New York: Wiley.

QUÉTELET, A. 1835. *Sur l'homme et le développement de ses facultés: Essai de physique sociale.* Paris: Bachelier.

RACHEV, S. T. 1996. *Modeling Financial Assets with Alternative Stable Models.* New York: Wiley.

RACHEV, S. T. & MITTNIK, S. 1997. *Stable Modelling in Finance.* Special Issue of the journal *Mathematical and Computer Modelling.*

RAY, D. 1958. Stable processes with an absorbing barrier. *Transactions of the American Mathematical Society*: **89**, 16-24.

RÉMÉRY, R. 1946. Distribution et Dispersion des Prix Francais depuis 1929. Thesis of the Institute of Statistics, University of Paris, *Bulletin de la Statistique Générale de la France.*

RICHARDSON, L. F. 1922. *Weather prediction by numerical process.* Cambridge University Press. The Dover reprint contains a biography as part of the introduction by J. Chapman.

RICHARDSON, L. F. 1960. Statistics of deadly quarrels. A chapter of *Arms and Insecurity.* Chicago: Quadrangle Books.

ROBERTS, H. V. 1959. Stock market 'patterns' and financial analysis: methodological suggestions. *Journal of Finance*: **14**, 1-10. Reprinted in Cootner 1964: 7-16.

ROLL, R. 1970. *Behavior of Interest Rates: the Application of the Efficient Market Model to U.S. Treasury Bills.* New York: Basic Books.

ROY, A. D. 1951. Some thoughts on the distribution of earnings. *Oxford Economic Papers* : **3**, 135-146.

ROZELLE, J. & FIELITZ, B. 1980. Skewness in common stock returns. *Financial Review*: **15**, 1-23.

RUTHERFORD, R. S. G. 1955. Income distributions: a new model. *Econometrica*: **23**, 277-294.

RVACEVA, E. L. 1954. On the domain of attraction of multidimensional distributions. *Ucenye Zapiski Lvovskogo Gosudarstvennogo Universiteta, Ser. Mekh. Mat.*: **6**, 5-44.

SAMORODNITSKY, G. & TAQQU, M. S. 1994. *Stable Non-Gaussian Random Processes*. New York: Chapman & Hall.

SAMUELSON, P. A. 1957. Intertemporal price equilibrium: A prologue to the theory of speculation. *Weltwirtschaftliches Archiv.*, **79**, 181-221.

SAMUELSON, P. A. 1965. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*: **6**, 41-49.

SAMUELSON, P. A. 1967. Efficient portfolio selection for Pareto-Lévy investments. *Journal of Financial and Quantitative Analysis*: **2**, 107-122.

SHOHAT, J. A. & TAMARKIN, J. D. 1943. *The Problem of Moments*. Providence, RI: American Mathematical Society.

SICHEL, H. S. 1952. New methods in the statistical evaluation of mine sampling data. *Transactions of the Institute of Mining and Metallurgy*: **61**, 261-288.

SIMON, H. A. 1955. On a class of skew distribution functions. *Biometrika*: **42**, 425-440.

SIMON, H. A. & BONINI, C. 1958. The size distribution of business firms. *American Economic Review*: **48**, 607-617.

SKOROHOD, A. V. 1954. Asymptotic formulas for stable Paretian laws. *Doklady Akademie Nauk SSSR*: 731-35. Or *Selected Translations in Mathematics Statistics and Probability (American Mathematical Society)*: 1961, 157-161.

STEIGER, W. 1964. A test of nonrandomness in stock price changes. In Cootner 1964: **98**, 254-61.

STENT, G. 1972. Prematurity and uniqueness in scientific discovery. *Scientific American*: **227** (December) 84-93. Also see his book, *Paradoxes of progress*, New York: W.H. Freeman.

STIELTJES, T. J. 1894. Recherches sur les fractions continues. *Annales de la Faculté des Sciences de Toulouse*: **8**, 1-122 and **9**, 1-47. Reprinted in Stieltjes 1914.

STIELTJES, T. J. 1914. *Oeuvres complètes de Thomas Jan Stieltjes*; 2 vols. Groningen: Noordhoff. Reprinted further in *Collected Papers*. G. Van Dijk (Ed.). New York: Springer, 1993.

SWIFT, J. 1733. On Poetry, a Rhapsody.

TAQQU, M. 1970. Note on evaluations of *R/S* for fractional noises and geophysical records. *Water Resources Research*: **6**, 349-50.

TAKANO, K. 1954. On some limit theorems of probability distributions. *Annals of Statistical Mathematics*: **6**, 37-113.

TAKANO, K. 1955. Central convergence criterion in the multidimensional case. *Annals of the Institute of Statistical Mathematics*: **7**.

TELSER, L. 1958. Future and the storage of cotton and wheat. *Journal of Political Economy*: **65**, 233-255.

TELSER, L. 1960. Reply, *Journal of Political Economy*: **68**, 404-415.

THÉBAUT, J. Y. 1961. Distribution lognormale de certains caractères de quelques phénomènes géologiques et ses applications. *Revue de Statistique Appliquée*: **9**, 37-87.

THORIN, O. 1977. On the infinite divisibility of the lognormal distribution. *Scandinavian Actuarial Journal*: **BUG** 121-148.

THURSTONE, L. L. 1947. *Multiple-Factor Analysis*. Chicago: University of Chicago Press.

TINBERGEN, J. 1956. On the theory of income distribution. *Welwirtschaftliches Archiv*: **77**, 155-175. Also in Tinbergen, J. 1959, *Selected Papers*. Amsterdam: North-Holland, 243-263.

TINTNER, G. 1940. *The Variate-Difference Method*. Bloomington, IN: Principia Press.

VON SAVITSCH & BENKTANDER, G. 1953. On the variation of the risk premium with the dimension of the house within fire insurance. *Skandinavisk Aktuarietidskrift*: **26**, 203-214.

VOSS, R. F. 1992. $1/f$ noise and fractals in economic time series. *Fractal Geometry and Computer Graphics*. J. L. Encarnacao, H.-O. Peitgen, G. Sakas, & G. Englert (Eds.). New York: Springer, 45-52.

WALSH, J. 1949. Review of Zipf 1949. *Scientific American*.

WALTER, C. 1994. *Les structures des hasard en économie: efficience des marchés, lois stables, et processus fractals*. (Ph.D. Thesis in Economics, Institut d'Etudes Politiques).

WALTER, C. 1995. Lévy-stability-under-addition and fractal structure of markets: implications for the actuaries and emphasized examination of MATIF national contract. *Proceedings of the 5th AFIR Colloquium* (Bruxelles): **3**, 1285-1330. To be reprinted in Rachev & Mittnik 1997.

WAX, N. (Ed.) 1954. *Noise and Stochastic Processes*. New York: Dover, 1954.

WEISS, H. K. 1963. Stochastic model for the duration and magnitude of a deadly quarrel. *Operations Research*: **11**, 101-121.

WESTERGAARD, H. L. & NYBOELLE, H. C. 1928. *Grundzüge der Theorie der Statistik*. Jena: G. Fischer.

WILLIS, J. C. 1922. *Age and Area*. Cambridge: Cambridge University Press.

WOLD, H. 1938. *A Study in Stationary Time Series*. Stockholm: Almqvist & Wicksell. (Second edition, 1954).

WOLD, H. & WHITTLE, P. 1957. A model explaining the Pareto distribution of wealth. *Econometrica*: **25**, 591-595.

WORKING, H. 1933. Price relations between July and September wheat futures at Chicago since 1885. *Wheat Studies*: **9**, 187-238.

WORKING, H. 1934. A random-difference series for use in the analysis of time series. *Journal of the American Statistical Association*: **29**, 11-24.

WORKING, H. 1934. Prices of cash wheat and futures at Chicago since 1883. *Wheat Studies of the Stanford Food Institute*: **2**, 75-124.

WORKING, H. 1948. Theory of the inverse carrying charge in future markets. *Journal of Farm Economics*: **30**, 1-28.

WORKING, H. 1960. Note on the correlation of first differences of averages in a random chain. *Econometrica*: **28**, 916-918. Reprinted in Cootner 1964: 129-131.

YOUNG, M. S. & GRAFF, R. A. 1995. Real estate is not normal: a fresh look at real estate return distribution. *Journal of Real Estate Finance and Economics*: **10**, 225-259.

YOUNG 1969. Winter meeting of the Econometric Society, New York, 1969.

YOUNG, W. E. 1971. Random walk of stock prices; a test of the variance time function. *Econometrica*: **34**, 797-812.

YULE, G. U. 1924. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F. R. S. *Philosophical Traditions of the Royal Society* (London): **213B**, 21-87.

ZAJDENWEBER, D. 1976. *Hasard et prévision* (Ph.D. Thesis in Economics, Faculté de Droit de Paris, December 1972). Paris: Economica.

ZAJDENWEBER, D. 1994. Self-similarity properties of the CAC 40 index (in French, with English summary). *Revue d'Économie Politique*: **104**, 407-434.

ZAJDENWEBER, D. 1995a. Business interruption insurance, a risky business. A study on some paretian risk phenomena. *Fractals*: **3**, 601-608.

ZAJDENWEBER, D. 1995b. Extreme values in business interruption insurance, *The Journal of Risk and Insurance*: **63**, 95-110.

ZIPF, G. K. 1941. *National Unity and Disunity*. Bloomington, IN: Principia Press.

ZIPF, G. K. 1949. *Human Behavior and the Principle of Least-effort*. Cambridge, MA: Addison-Wesley. Reprint, New York: Hefner, 1972.

ZOLATAREV, V. 1985. *Stability Problems for Stochastic Models: Proceedings Uzhgorod, 1984*. Lecture Notes in Mathematics: **1155**. New York: Springer.

ZOLOTAREV, V. 1986. *One-Dimensional Stable Distributions*. Providence, RI: American Mathematical Society.

Index