# Scaling in financial prices: I. Tails and dependence

## Benoit B Mandelbrot

Sterling Professor of Mathematical Sciences, Yale University, New Haven, CT 06520-8283, USA

## Abstract

The scaling properties of financial prices raise many questions. To provide background—appropriately so in the first issue of a new journal!—this paper, part I (sections 1 to 3), is largely a survey of the present form of some material that is well known yet repeatedly rediscovered. It originated in the author's work during the 1960s. Part II follows as sections 4 to 6, but can to a large extent be read separately. It is more technical and includes important material on multifractals and the 'star equation'; part of it appeared in 1974 but is little known or appreciated—for reasons that will be mentioned. Part II ends by showing the direct relevance to finance of a very recent improvement on the author's original (1974) theory of multifractals.

## Introduction

As usual, a power-law distribution of financial price changes will be written as $\Pr(U > u) \sim u^{-\alpha}$. The key question is whether or not the exponent $\alpha$ is restricted to $\alpha < 2$. Such is the case in the now classical model Mandelbrot (1963) based on Lévy-stable independent increments. In contrast, the multifractal model described in Mandelbrot (1997) corresponds to dependent increments and allows $1 < \alpha < \infty$. In that model, price is a fractional Brownian function of a trading time, which itself is a non-decreasing multifractal function of clock time.

Indeed, a major asset of multifractals, hence of my 1997 model, is that under wide conditions the power-law distribution is provided with one of its very few legitimate derivations. This feature vanishes in the familiar non-random examples of the binomial multifractal and related 'cartoons'. In addition, it did not matter in the best known early applications to physics. In finance, in contrast, it emerges, arguably, as one of the most important features of the original form of multifractals presented in Mandelbrot (1974) and now generalized in Barral and Mandelbrot (2000).

The standard restriction to $\alpha < 2$ can now be inverted to imply that $\alpha < 2$ is required for the price changes to be independent, as long as they have an infinite variance. In contrast, the full multifractal model of Mandelbrot (1997) allows variance to be finite but involves infinite-range dependence.

Related empirical questions are: (i) for which prices is $\alpha$ well defined? (ii) Fama (1936b, 1965) observed instances of power law tails with $\alpha > 2$, which is incompatible with scaling under addition combined with independence. (iii) Officer (1972) made the very important empirical discovery that some financial prices clearly contradict additive scaling combined with independence. These findings became familiar among economists and were later rediscovered by many newcomers to the field; they are obviously important; how should one react to them?

The old and new issues are discussed in this paper informally and in historic sequence. The paper that follows discusses the same issues formally in the context of a fundamental functional equation, called the 'star equation'. It expresses invariance under different successive forms of rescaling (or 'exact renormalizability' in the language of physics). The star equation went through successive stages, in parallel with successive financial models.

The ancient original form, introduced by Cauchy in 1853, was the pioneering form of the concept of scaling and expressed that the distribution of a random variable is invariant under *non-random weighting*, a form of addition. The equation's full solution is a (Lévy) stable distribution, in which the tail probability follows a power law whose exponent is restricted to $\alpha < 2$. Independent Lévy stable price changes are the basis of a model that Mandelbrot (1962, 1963, 1967) applied successfully to the variation of *some* financial prices, namely,

cotton and other *some* commodities, *some* securities, as well as *some* interest rates. Fama (1963b, 1965) promptly extended this model to represent a wider variety of securities prices. A more general star equation was introduced in Mandelbrot (1974). It was applied to prices when chapter E6 of Mandelbrot (1997) generalized scaling to involve exact renormalizability simultaneously in time and price and investigated models of price variation in which the rules of dependence are scaling and take a multifractal form.

'Generically', the multifractal model yields power-law tails and can yield any value $\alpha > 1$. This generic property has remained little known, however, because it was not used much. Also, the heuristic presentation that many scientists follow says nothing directly about the probability distribution. The basic finding underlying the distribution of the values of multifractals contributed to the (correct) reputation that Mandelbrot (1974) is complicated. It took much space to show that in cascade-generated random multifractals the Cauchy scaling invariance under non-random weighting is replaced by invariance under *random weighting* and associated with a power-law distribution having a critical exponent.

However, Mandelbrot (1974, 1997) was necessarily limited to multifractal interdependence with an artificial grid-bound cascade, which is also the source of examples without a power-law distribution. To the multifractals in Barral and Mandelbrot (2000), the star equation does not extend naturally. But they have an important property: for them the unboundedness of $\alpha$ is a generic property.

After the fact, a parallelism exists between this sequence of star equations and the progression of models of price variation from the Brownian, to the 'mesofractal' Mandelbrot (1963) and to the multifractal Mandelbrot (1972, 1997).

Taken together, those examples suffice to establish that all observations of power laws in finance can be accounted for, at least in principle, as long as scaling is redefined in a suitably generalized form.

# 1. Preliminaries

## 1.1. The challenge of 'lifetime' data stated graphically

The challenge this paper faces is that of representing financial price variation by suitable random models. The sequence of the three models I have developed since the early 1960s is rooted in an important and parsimonious concept, an early form of scaling.

However, the underlying motivations and the degree to which they succeed are best assessed by combining the analysis of actual data with the synthesis of model 'data'. In the past, graphics were slow, expensive and inaccurate. Hence, it was of no help and all fields had to compare scientific models and reality via short lists of numerical quantities investigated by analytic statistical methods. With computers, the actual data and simulated samples of the models must continue to be compared analytically. But they can also be displayed side by side and compared visually.
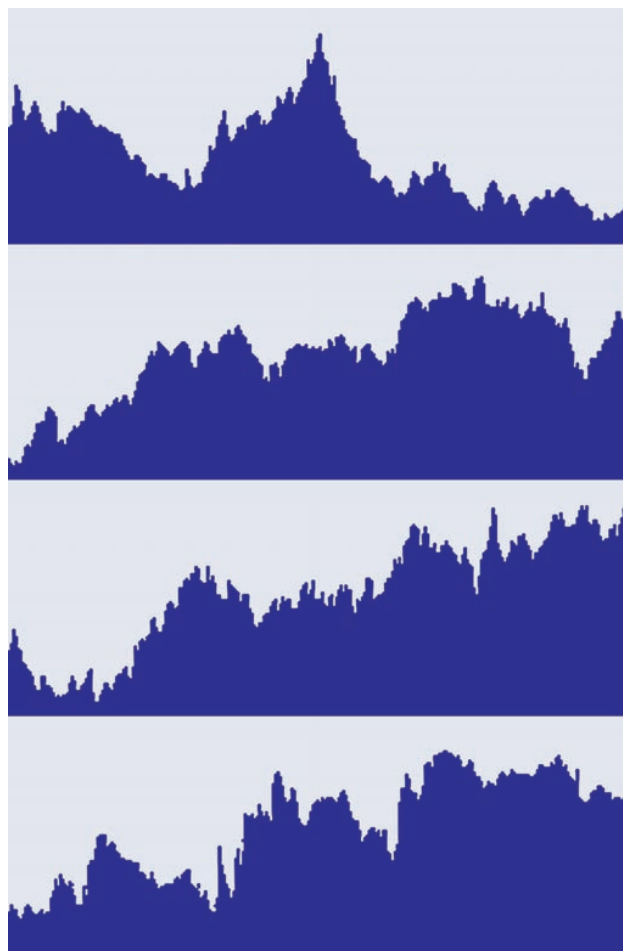


**Figure 1.** A collection of diagrams, illustrating, in no particular order, the behaviour in time of at least one actual financial price and of at least one mathematical model of this behaviour. It would be difficult to identify the models.

**The challenge made graphic by figures 1 and 2.** It is even easier to create confusion or lie by pictures than by words, numbers and statistics. Two major visual exhibits I often use, reproduced here as figures 1 and 2, are specifically devised to illustrate and avoid a certain form of confusion.

The financial press was forward-looking in using graphics, but is accustomed to plotting the price itself. In this spirit, figure 1 intermixes two extremely different synthetic records and two actual price data sets over horizons of a 'lifetime', namely, a few times ten years. The four lines are very hard to tell apart.

One of those lines represents the early and idealized Brownian motion model put forward by Louis Bachelier in 1900 (see section 7 in part II). A standard response is that figure 1 confirms that the actual data are adequately represented by Brownian motion, hence there is no need to search for better models.

Very different conclusions are reached by examining figure 2. Instead of the real or synthetic prices themselves, it plots their 'daily increments'. By design, the 'pen width' chosen in preparing figure 2 is about the same as the time lag
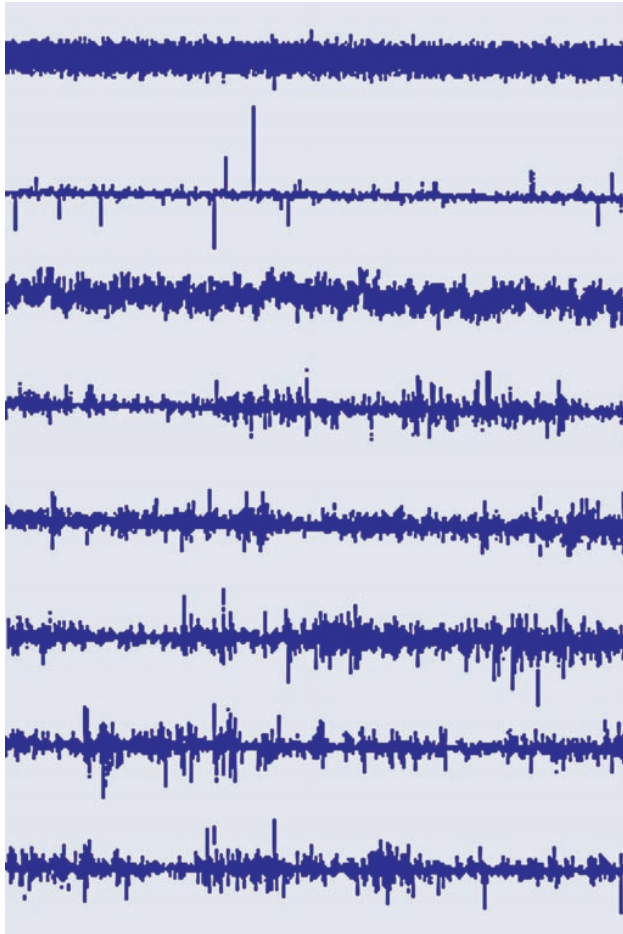
**Figure 2.** A stack of diagrams, illustrating the successive 'daily' differences in at least one actual financial price and some mathematical models. It is obvious that lines 1 and 3 do not report on data but on models; in contrast, to identify the models among the lower five lines is difficult. As to line 2, a referee greatly flattered me by observing that the fact that it illustrates a model is *not* obvious. But it does indeed illustrate the model I introduced in 1963, the best available until a few years ago.

between observations. The resulting 'strip' is an artefact but a very useful one.

The top line illustrates white noise, the sequence of increments of Brownian motion. Lines 2 and 3 illustrate my two early and very imperfect would-be improvements on Bachelier. Because of the main dates associated with them, they will be called the M1963 (Mandelbrot 1963) and M1965 (Mandelbrot 1965) models.

The remainder of figure 2 is a medley whose main point is that the sources of the different lines are difficult to identify either by eye or by algorithm. Told that at least one is a real record and at least one is computer generated, the reader is free to guess which diagrams are real and which ones are forgeries. Here is the answer to the game: the fifth line plots the price of IBM shares and the sixth line, the Dollar–Deutschmark exchange rate. The remaining lines are synthetic records of the latest multifractal model, M1992/1997, first fully described in Mandelbrot (1997). It is hoped that the forgeries will be perceived as effective.
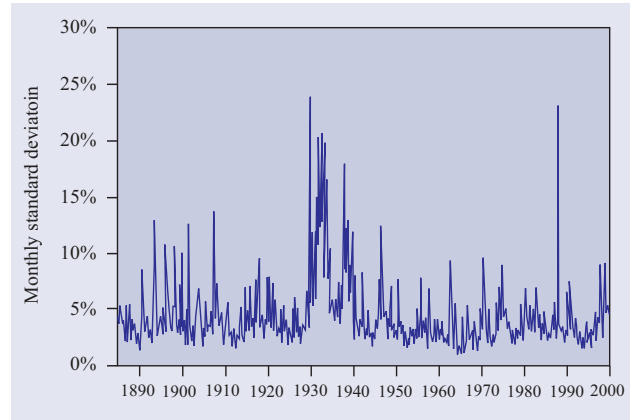


**Figure 3.** G W Schwert's financial index in New York for over a century. This index is pieced together from several sources described in section 1.2. Daily standard deviations are plotted in units of monthly standard deviations.

## 1.2. The challenge of 'secular' data stated graphically

Figure 3, kindly provided by G W Schwert, is a composite of squares of daily returns. It has been 'processed' and, therefore, is not directly comparable to the real data included in figure 2. But it suffices for the point to be made here. The variability of the real data remains extreme over a century and is approximately of the same form as the variability figure 2 reports over a lifetime.

A minor complication is that, by necessity, the underlying returns are, first, of the DJIA, later, of the S&P composite, then of the CRSP value-weighted index, and again of the S&P composite.

A major complication is that this figure incorporates two mean-square averages: it plots the standard deviation for days in units of the monthly standard deviation. Hence, the evenness of the minima is due to the fact that the variability of the variance on the scale of a century has been largely eliminated by averaging. Figure 3 must not be compared directly to the unaveraged figure 2, but averaging has a much lesser effect on the features of figure 3 that attract the strongest and most immediate attention.

## 1.3. Can the challenges of short- and long-range modelling be combined?

The unconventional thinking behind my work on price variation can only emerge gradually through this paper but deserves to be briefly stated here. The natural and usual response is that the data in figures 2 and 3 must be dealt with *separately*. To the extent that price variation follows any rule, it is, indeed, generally taken for granted that changes over a day, a week, a month, a year, a lifetime or a century follow separate sets of rules. That is, each time increment raises a separate and distinct challenge.

Quite to the contrary, my belief is that the great overall similarity between figures 2 and 3 suggests that, in a first approximation, price variation presents very similar features

over a century and a lifetime. A similarity of features not only suggests that the existence of underlying rules is not excluded, but that those rules *may* be the same at all time scales.

The reader may be tempted to stop here and ponder *a priori* whether or not my claims agree with economic thinking or make 'intuitive' sense. I suggest it is more reasonable to go ahead and first see what the multiscale models accomplish. Those models are parsimonious and their parameters have clearcut individual meanings, for a reason that is important to state. My parsimonious models postulate that price records are invariant under scaling or multiscaling—and the parameters are the defining characteristics of those invariances. They may involve unfamiliar mathematics, but they are as intrinsic as can be. In contrast, the alternative approaches that concern each scale separately tend to borrow and generalize some existing formulae addressed to past needs. They inevitably involve large numbers of parameters as well as delicate problems of actual matching. In order to be preferred to the joint models, they will have to present special advantages, which I do not think exist as of now.

## 1.4. Essential characteristics of the price-increment records, in contrast to the ideal coin-tossing hypothesis

**1.4.1. Three basic observations.** Figure 3, as well as the medley at the bottom of figure 2, exhibit many features that are familiar even to those who know little about financial markets.

Firstly, a substantial number of 'spikes' stand out clearly: all correspond to unusually large price changes and many correspond to instantaneous discontinuities.

Secondly, most relative changes other than the spikes merge into a strip. On the top line of figure 2, the strip's width is constant but on the bottom five lines, it constantly varies. In figure 3, it is more or less constant because of the averaging that is part of its construction.

Thirdly, the spikes tend to cluster and occur during periods when the strip is broad. This behaviour brings up the notion of 'volatility'. In the Brownian universe, the volatility can be defined by a single number. It is the width of the strip or the standard deviation $\sigma$ that is roughly four times smaller. In this sense, the five bottom lines of figure 2 exhibit 'variable volatility'. My very different response is that the definition chosen for the very elusive concept of volatility will have to be thoroughly re-examined.

Of those and other characteristics of real markets, the white noise on the top of figure 2 incorporates *none*. The strip is of constant width and no spike stands out. That is, Brownian motion, Gaussianity and independence represent the observed price series $P(t)$ very poorly.

To quantify how poorly, note that for both ideal and real charts, one is free to define the overall strip as leaving outside about 5% of the cases. In the Brownian case, this width is of the order of $4\sigma$. However, many real charts, such as those plotted in figure 2, include many '$10\sigma$' spikes. In an ideal market where $\sigma$ is the standard deviation, those events would have a probability of about $10^{-23}$, a few millionth of a millionth of a millionth of a millionth. This is roughly the inverse of

Avogadro's number. The ideal market completely disregards those spikes—but a realistic model cannot.

The observed extremes deserve to be further documented. IBM saw its stock fall instantaneously by 10% early in 1996, and later in that year rise instantaneously by 13.2%. Concentration with or without discontinuity is striking even in the extensively averaged portfolio based on the Standard & Poor 500 index. Of this portfolio's returns over the 1980s, fully 40% was earned during ten days (0.5% of the number of trading days in a decade.) This concentration contradicts a fundamental theorem about the ideal market, namely that even the most active day makes a negligible contribution.

**1.4.2. Comments on the basic three observations reported in section 1.4.1** Even if it were true (it is not) that an ideal market model represents data correctly 95% of the time, its fit would not be sufficient because the 5% remainder includes most major events. Indeed, it cannot be questioned that, irrespective of the measures chosen for the notion of 'cumulative effect of events', the partial effect of the 5% largest far exceeds 5% of the total effect. As a preview, my models predict that the effect of the largest 5% may dwarf the effect of the remaining 95%. That is, all told, the study of finance cannot be blind to extreme price changes.

A further criticism of the ideal market hypothesis is qualitative but deep. Financial dailies, weeklies and monthlies can thrive because every day in the market is unlike any other day, one week, month or year unlike any other. In an ideal market, in contrast, daily chapters of the history books might vary from one another, but all yearly chapters would seem effectively alike.

## 1.5. The combined challenges of short- and long-term data, continued

While the limitations of the ideal market using $B(t)$ are generally acknowledged, that model became the basis of the very sophisticated 'modern portfolio theory' and of the 'calculus of risk' highlighted by the Black–Sholes–Merton theory. This development was unavoidable and sensibly followed the well-trodden example set by the exploration of matter, which began by inventing and exploring the simplifying concept of perfect gas. But many practitioners and academics have now joined the search for realistic models.

This effort, like every feature of financial markets, pits bulls against bears. To the question, 'can large events be handled quantitatively?', the bearish answer is that this is impossible, the argument being that large events are individual 'acts-of-God' or 'anomalies' that present no conceivable regularity. To the question, 'can the so-called changes in volatility and other long term effects be handled quantitatively?', the bearish answer is, again, that it is impossible, the argument being this time that everything in the financial markets is non-stationary. Other bears assume without even seeking evidence that short- and long-term effects follow different rules.

The bulls disagree, of course, and believe that one must not give up without having tried.

**A question is immediately thrown back. 'Do you view the large events as mostly exogenous or endogenous?'** The presence of universal regularities in the exogenous economic fundamentals would be absurdly far-fetched. Therefore, if changes were exogenous, it would be hard to expect any model to be of wide applicability. The endogenous alternative is that the complex interactions in financial markets are strong and systematic enough to create the very specific structures that are observed. In any event, I take the decidedly bullish position that the variation of financial prices does follow rules of its own, therefore may be modelled on its own. For the sake of the already-informed readers, section 1.7 will sketch the rules I propose today. Details will be provided in later sections.

**Another question is: 'Don't you bulls agree that in the economy the non-stationarity of everything is blindingly obvious?'** My response to this question went through several stages.

*Before the fact,* I answered that this may well be true but 'stationarity' is an example of the broader notion of 'existence of unchanging rules'. Without such rules, there is no science, therefore anyone who agrees with the bears' question agrees that the market's behaviour will forever remain inaccessible to rational description.

*After the fact,* the M1963 model and ever more so the M1972/1997 model prove that the most customary meaning of the word 'stationary' is a poor description of anything in finance. Therefore, section 1.4 became the stimulus for the identification of generalized forms of stationarity and the elaboration of a suitable calculus.

Therefore, my response to the bears' question concerning stationarity takes the following form, which is the key to all my work in this area: 'I disagree that non-stationarity is obvious and do my best to avoid it'.

## 1.6. Quirky joint responses to the challenge stated in section 1.5; sketch of the author's three successive models based on invariances

Denote by $P(t)$ the logarithm of a financial price at time $t$. For reasons hinted at in section 1.3, and to be elaborated momentarily, the only models that I ever considered, respond simultaneously to the challenges posed by data over a century, a lifetime and shorter time spans.

**A classical invariance property of the Wiener ('ordinary') Brownian motion $B(t)$.** It is well known that, if $B(0) = 0$, the function $|\mu|^{-H} B(\mu t)$ has the same distribution for all $\mu \neq 0$. It reduces to its form for $\mu = 1$, which is $B(t)$ itself. Until the 1960s, this special form of scaling was standard but nameless. I put forward the term, 'self-affinity', which was widely accepted as part of the vocabulary of fractal geometry. The Brownian self-affinity exponent is $H = 1/2$.

Self-affine processes are exactly renormalizable (in other words, they provide fixed points) under suitable linear changes applied simultaneously to both the $t$ and $P$ axes.

**Self-affinity as mathematical expression of market folklore.** As is widely known, a fractal is a geometric shape that can be separated into parts such that each part is a reduced-scale version of the whole. To implement this characterization, one must define the notion of 'reduction'. Fractals using isotropic reductions are called self-similar. They have become well known, but prices call for the more novel concept of self-affine fractality.

Self-affinity expresses mathematically the claim that all market charts look alike. The 'whole' chart is usually wider than it is high, but its small parts are higher than they are wide. That is, in order to move from the whole to a part, one must reduce the time scale far more than the price scale. Self-affine fractality makes available many powerful tools of analysis. Some are very new; others are described in Mandelbrot (1982, 1997, 1999, 2001).

## 1.7. Fractional Brownian motion compounded in multifractal trading time: subordination; three examples investigated earlier

My present preferred model of price variation combines two essential notions I had originally introduced for different purposes. Those notions need not be defined until section 5 but those who are familiar with them may welcome an early survey.

The combination of fractional Brownian motion and multifractality first described in Mandelbrot (1997) postulates that $P(t)$ is a 'compound process' of the form $P(t) = B_H[\theta(t)]$, where $B_H(\theta)$ is a fractional Brownian motion (FBM) in terms of an auxiliary variable $\theta$, and $\theta(t)$ is called a multifractal trading time (MTT) and is a multifractal function of the clock time $t$.

FBM and MTT are both self-affine, and self-affinity is preserved when they are compounded.

**Special cases that reduce to models considered in 1900 or the 1960s.** The compound process $B_H\{\theta(t)\}$ is specialized but of great generality. In addition, it has the virtue of including, as very simple special cases, three models advanced in the past.

**The standard Bachelier model, when $H = 1/2$ and $\theta(t)$ reduces to the degenerate limit case $\theta(t) = t$.** Then, the compound function $B_H(\theta)$ reduces to the classical Wiener Brownian motion postulated by the Bachelier model. Once again, its increments are drawn on the top line of figure 2.

**The M1965 model, when $H \neq 1/2$ and $\theta(t) = t$.** Then $P(t)$ reduces to fractional Brownian motion of time and the model falls back on one proposed in Mandelbrot (1965). Once again, its increments are drawn on the third line of figure 2.

**The M1963 model, when $H = 1/2$ and $\theta(t)$ reduces to a 'stable subordinator'.** Mandelbrot and Taylor (1967) reproduced in chapter E21 of Mandelbrot (1997), showed that $P(t)$ reduces to a $L$-stable function of time and the model

falls back on one proposed in Mandelbrot (1963). Its high explanatory power will be described in section 3.2, but it is not general enough. Once again, its increments are drawn on the second line of figure 2.

**Beyond the special cases: general compounding as compared to the special case of subordination.** By a definition that deserves to be left unchanged (that is, not generalized), 'subordinators' are monotone, non-decreasing, random processes with independent increments. P Clark, in Mandelbrot (1972) (chapter E21 of Mandelbrot (1999)), as well as numerous other authors considered non-fractal subordinators. The formalism became complicated but preserved independent increments; that is, did not face the critical presence of strong dependence. In contrast, general compounding allows the increments of the trading time to be statistically dependent. As a result, while preserving self-affinity, the FBM (MTT) model allows $P(t)$ to follow a wide variety of specific behaviours.

**Useful fractal terminology.** The above three special examples are jointly called *fractal*. When emphasis is needed, the Bachelier model is called *Fickian,* the M1965 model is called *unifractal* and the 1963 model, *mesofractal*.

For the sake of symmetry with the M1963 and M1965 models, multifractals are said to define the M1972/1997 model. This construction has also been referred to as BMMT: Brownian motion of multifractal time or (in Mandelbrot, Calvet and Fisher (1997) and Calvet and Fisher (2000)) as MFAR: multifractal model of asset returns.

**In the most useful models, the number of parameters is as small as possible and they have independent significance; uni- and multivolatility.** The Fickian case is unique in its being fully specified by location and scale, and a single parameter, namely, scale, suffices to define and measure volatility. The models offered in Mandelbrot (1963, 1965) are specified by scale and location, but also one or a few additional numbers. Those additional numbers have a clearcut independent significance and must be made part of the measurement of volatility.

**The limit-log-normal model.** Among multifractal measures, families that involve few parameters deserve special attention. The limit log-normal family, studied in Mandelbrot (1972), chapter N14 of Mandelbrot (1999), happens to provide a surprisingly close approximation to the Dollar–Deutschmark exchange rate (see Mandelbrot, Calvet and Fisher (1997) and Calvet and Fisher (2001)).

**Beyond the limit-log-normal model.** In general, the specification of a multifractal involves a 'spectrum' that is in effect a probability distribution function. Its full specification involves parameters in large (theoretically infinite) numbers. Once again, the saving grace is that each of those parameters has a well-defined meaning.

**Levels, or degrees, of stationarity.** To the eye, the top line of figure 2 is clearly stationary, the third line is dubious and all the others appear to be non-stationary. In fact, all are stationary, at least 'conditionally'. This very important issue is addressed in Mandelbrot (1982) starting on page 383.

# 2. Challenges raised by the power laws of exponents $\alpha$ and $H$ that rule the tail and dependence

The discussion began in section 1 will now be resumed in detail.

## 2.1. The power laws

**Personal incomes.** Pareto (1896) discovered purely empirically that the distribution of personal income has a high-income tail that follows a power law $\Pr\{U > u\} \sim u^{-\alpha}$. Papers I published around 1960 interpreted Pareto's power law in terms of Lévy stability (see several chapters of Mandelbrot (1997)). Then I moved on to financial price change data sets and discovered two power laws in their context.

**Fat tails of financial prices.** For all time increments $\Delta t$, the tail distributions of the price increments are 'fat' and follow a power law with an exponent that usually continues to be denoted by $-\alpha$. Mandelbrot (1963) derived this power law through a theory, as a necessary consequence of a form of postulated 'self-affinity', scale-invariance or scaling. The assumption is that, after suitable renormalization, the same distribution holds for the price changes $\Delta P$ over all values of $\Delta t$.

**Infinite dependence: correlation or another suitably defined measure of statistical dependence is 'infinitely long' with a tail that follows a power law with an exponent that is usually denoted by $2H - 2$.** 'Infinitely long' is not a loose wording but a precise technical term. When a correlation $C(s)$ is defined, the condition for infinitely long dependence is that $\sum C(s)$ from $s = -\infty$ to $s = \infty$, is infinite or zero. An example is $C(s) = s^{2H-2}$ with $\frac{1}{2} < H < 1$. This power law introduced in Mandelbrot (1965) involved a different form of theoretical scale invariance. Scaling was postulated in order to account for an empirical discovery called the Hurst puzzle.

## 2.2. The scientific challenges posed by the power laws were recognized instantly in the 1960s and again attract wide attention

- When data sets of tail and dependence properties are examined in isolation from each other, which are the empirically observed ranges of the values of $\alpha$ and $H$?
- For which values of $\alpha$ are power-law tail distributions compatible with an absence of dependence?
- For which values of $H$ is the power-law dependence compatible with the absence of fat tails, for example, with Gaussianity?

- To be realistic, one must examine tail and dependence properties together. When this is done, is it possible to prove mathematically, or at least illustrate by examples, that the tail and dependence power laws are mutually compatible for either some or all values of $\alpha$ and $2H - 2$?
- Last but not least, are the various power laws, with suitable ranges of $\alpha$ and $H$, compatible with scaling? That is, can both power laws be valid simultaneously for all values of $\Delta t$ used in taking price differences?

## 2.3. Empirical power laws first occurred in economics and social sciences

Zipf (1949) is noteworthy in several contradictory ways. The book's principal conclusion was that many social sciences data follow a power-law probability distribution that generalizes Pareto's law of income distributions. This conclusion was sharply attacked and dismissed by most statisticians, but eventually vindicated and expanded. Zipf's broader conclusion was that the distinction between the Gaussian and the power-law distribution coincides with the distinction between the physical and the social sciences. This claim was thoroughly discredited by my work and (implicitly, but thoroughly and definitely) by statistical physics of critical phenomena. Zipf was not satisfied with pioneering curve-fitting but claimed having explained the power-law distribution as resulting from a purely verbal argument he called 'principle of least effort'.

## 2.4. Generalization of power laws: the critical moment exponents

Many examples involve a power law with a slowly varying prefactor, therefore require a generalization of $\alpha$. One says that a random variable $X$ has a finite critical moment exponent $\alpha$ if $EX^q < \infty$ for $q < \alpha$ and $EX^q = \infty$ for $q > \alpha$.

## 2.5. Digression on multiplicative growth models for positive variables with a power-law distribution

Pareto inspired a multitude of authors to try and explain why personal incomes are power-law distributed. The number of models and the fact that many fall into a fairly small number of basic patterns are painstakingly documented in a very useful but inaccessible paper over 100 pages long, Chipman (1976).

Among those models that actually yield a power-law distribution, all too many are old, new, or even brand new variants of the following steps. (i) When $u$ itself satisfies $\Pr\{U > u\} = u^{-\alpha}$, the transform $\log U = V$ satisfies $\Pr\{V > v\} = \exp(-\alpha v)$. (ii) Physics knows a plethora of arguments that look different but ultimately agree in yielding an exponential behaviour for $V$. (iii) This explains the power law behaviour for $\Pr\{U > u\}$. The key ingredient to the scaling output of these models lies in the 'principle of proportionate effect' that is used to justify the transform $\log U = V$.

A skeptic's opinion is described in chapter 10 of Mandelbrot (1997). One reason for skepticism is that seemingly innocuous changes in the assumptions often yield a thouroughly different prediction: the log-normal instead of the power law. For large values of $u$ the power law is correct. In contrast, the log-normal fails. For smaller values of $u$, the log-normal automatically includes a bell, which is a testable prediction. In contrast, the power-law distribution must somehow, be either truncated or smothed off into a bell; to this end, new conditions, beyond multiplicative growth, have to be added.

The scope and significance of the preceding remarks will become manifest in sections 4 and 5 (next paper), where the fractal and multifractal models will be shown to yield a bell and a power law tail simultaneously.

# 3. Extensive informal summary in historical sequence

The responses to the various questions raised in section 2.2 are continually improving. Many partial responses are available but scattered among the reprints and newly-written chapters that are collected in Mandelbrot (1997, 1999, 2001) and also in Mandelbrot (2000) and in forthcoming papers. Now that those questions are again widely asked, this paper begins by bringing some scattered thoughts together, with additional references and substantial new material.

In addition to power laws, the main conceptual tools are the ancient notion of scaling and the notion of renormalization together with the more precise notion of exact renormalizability. For a long time, I was vaguely aware of the uses of those notions in the study of turbulence. Since 1965 for scaling, and 1972 for renormalization, these notions also came to play a central role in the chapter of statistical physics that is concerned with critical phenomena. Those notions' uses in my work on both turbulence and finance were closely interrelated through multifractals, and came before 1965 and 1972.

The key facts are as follows. In the case of independence, a form of scaling is a classical concept in probability theory. It is expressed by a functional equation put forward by Cauchy and generalized by Lévy. In the case of dependence, a generalized form of scaling and the corresponding generalized functional equation are part of the theory of multifractals, which I put forward (in 1969, 1972, 1974) in papers reproduced in Mandelbrot (1999).

From the old form of scaling to its little-known generalizations, the most logical and convenient informal approach happens to follow chronology.

## 3.1. Brief reference to Mandelbrot's early work in the 1950s

My earliest use of scaling, renormalization etc concerned a topic of limited importance, (see Mandelbrot (1999), middle of page 104) but acquired some significance because of its date: it came out in 1956.

## 3.2. Scaling properties reported in Mandelbrot (1962, 1963, 1967)

The original treatment of financial prices with independent increments was the topic of Mandelbrot (1963) and its delayed
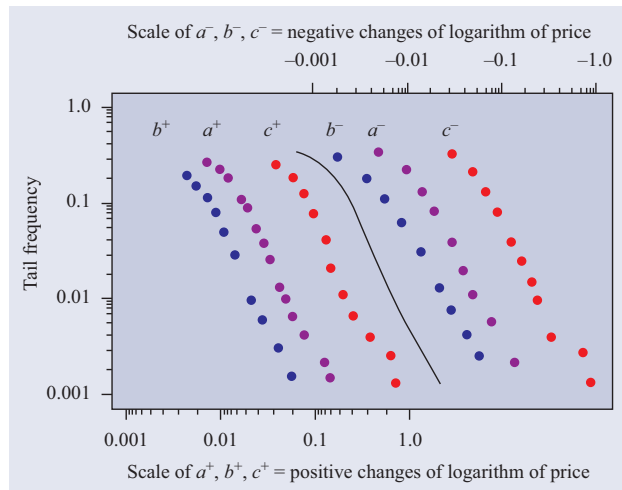
**Figure 4.** The original evidence of scaling in finance/economics. This evidence, reproduced from Mandelbrot (1963), is discussed in section 3.2. The following series of data are plotted, the positive and absolute negative values being treated separately in both cases. (*a*) $X = \log_e Z(t + 1 \text{ day}) - \log_e Z(t)$, where $Z$ is the daily closing price at the New York Cotton Exchange, 1900–1905 (data communicated by the US Department of Agriculture). (*b*) $X = \log_e Z(t + 1 \text{ day}) - \log_e Z(t)$ where $Z$ is an index of daily closing prices of cotton on various exchanges in the US, 1944–1958 (communicated by Hendrik S Houthakker). (*c*) $X = \log_e Z(t + 1 \text{ month}) - \log_e Z(t)$, where $Z$ is the closing price on the 15th of each month at the New York Cotton Exchange, 1880–1940 (communicated by the US Department of Agriculture).

but important follow-up, Mandelbrot (1967). Both papers' contents first came out in an IBM Report, Mandelbrot (1962), which circulated widely. To some of the questions raised in section 2.2, those papers gave partial answers characterized by strong assets and limitations.

The M1963 model's main asset is the fact that the power-law tail behaviour is obtained as an immediate consequence of postulated scaling combined with serial independence. Power-law tails characterize the (Lévy) stable distribution, an analytic (though non-explicit) formula that can be used to fit the whole range of price changes.

**The case of cotton.** The data reported in Mandelbrot (1963) concerned the changes of the spot price of cotton for $t = $ one day and $t = $ one month. Figure 4 combines, (*a*) the cumulated density function of the symmetric scale distribution of exponent $D = 1.7$, which is actually a slightly overestimated value of $D$, with (*b*) the doubly logarithmic plots of tail frequencies. In all cases the ordinate gives the relative frequency of cases where one of the quantities changes by more than $|u|$. For a close examination, make a transparency and the theoretical curve will superimpose on either of the empirical graphs with slight discrepancies. More specifically, I was led to the following inferences.

- *Power law.* For both $\Delta t$, the tail exponent is $\alpha = 1.7$. *Comment.* The M1963 model's most widely noted limitation was that variance was divergent because of $\alpha < 2$. Worse, $\alpha = 2$ plays a complicated role. For the

small samples that were available in practice, values of $\alpha$ close to 2 yield a very narrow power-law range and that range vanishes in the theoretical limit $\alpha = 2$. Fortunately, the $\alpha = 2$ bound did not matter in the first stage of the theory because for cotton $\alpha$ was safely below 2.

- *Asymmetry of the tails.* For both $\Delta t$, the frequency distribution is conspicuously asymmetric requiring stable distributions other than the symmetric only. *Comment.* This asymmetry, not known before 1962, is not great but consistent. It manifested itself by comparison with the symmetric *L*-stable case. The transition between the bell and the left histogram tail exhibits a more pronounced bump and the transition between the bell and right tail, a flattening.

- *Scaling, that is the commonality of rules between the short- and the long-run.* In a first approximation, horizontal translations suffice to superpose the two-tail histograms corresponding to the two $\Delta t$. This method was independently imagined by physicists, who call it 'collapse'.

**Comment (i). First approximation: independence and Cauchy–Lévy stability.** Collapse suggested that the distribution is well-fitted for both $\Delta t$ by the (Lévy) stable distribution. The first approximation involving independence is therefore reasonable and became widely known.

**Comment (ii). Second approximation dependence.** A closer comparison of the prefactors for daily and monthly data showed that the corresponding $\Delta P$ are definitely *not* independent except in a first approximation. Those deviations from independence were duly noted in Mandelbrot (1963). They became well known to readers of a critique—Cootner (1964)—that was not answered publicly until chapter 17 of Mandelbrot (1997). In 1963, this approximation could not be studied for lack of appropriate statistical tools.

**Commodities other than cotton, securities and interest rates.** A first step was taken in Mandelbrot (1962, 1967), but these papers are less well known than Mandelbrot (1963), which they preceded or followed. They dealt with wheat, other commodities, securities from the 19th Century (mostly railroads), and several interest rates. On the topic of the sign and value of $\alpha = 2$, this broader evidence was mixed. This is why my early papers' titles include the cautious terms 'certain' and 'some other'.

Specifically, exponents about $\alpha = 1.7$ were found in diverse cases other than cotton. But wheat prices suggested a power-law exponent $\alpha$ closer to 2 and holding over a broad range of values of the price change. Once again, in contrast, the power-law range of the (Lévy) stable distribution with the same $\alpha$ is short if $\alpha$ is just below 2 and disappears in the limit $\alpha = 2$.

## 3.3. Fama (1963a, b, 1965) and the extension of the applicability of power laws to a large class of then-recent current securities

The step beyond Mandelbrot (1962) was taken by a student of mine. Fama (1963b, 1965) dealt with then-recent securities prices and concluded that the increments are definitely not Gaussian and the M1963 model applies with $\alpha$ close to 2. This raised serious difficulties that were not solved until much later, in my work on multifractals.

Fama's conclusion was confirmed repeatedly in the economics/finance literature. In addition, it started—several years ago—being continually 'rediscovered' on the basis of present data.

## 3.4. The significant year 1972: the 'Officer effect' and the remark concluding chapter N14 of Mandelbrot (1999), written in 1972

The year 1972 was significant for two independent reasons: the scaling I used in 1963 was challenged and I conceived and developed a generalized form of scaling—that was not developed until 1997.

*The attack*: Officer (1972) examined securities data for different values of $\Delta t$. The collapse test had established that scaling holds in the case of cotton. But its application to diverse other financial data sets revealed unquestioned sharp deviations from scaling. This finding was repeatedly confirmed and continues to receive fresh confirmations.

Officer's challenge to the M1963 model immediately became influential in the financial community. It stands out as a milestone because the responses were immediate, strong and diverse. Their diversity becomes best understood by assigning financial models into the three main 'states of variability and randomness' described in section 3.5.

## 3.5. The fundamental new concept of 'states of variability and randomness'

Here is a paraphrase from chapter E5, p 120 of Mandelbrot (1997). 'While a unique theory of physical interactions applies to every form of matter, the detailed consequences of those unique general laws differ sharply and physics has to distinguish between several states of matter. I argue that a similar distinction should be useful in probability theory. Nearly every scientist engaged in statistical modelling 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, namely *wild* and *slow*. Mildness is characterized by an absence of structure recalling a gas. Wildness is characterized by the presence of structure recalling a solid and slow randomness recalls liquids.'

**Practical consequences concerning the effectiveness of risk reduction by diversification.** The proposed distinction between states of variability and randomness is neither philosophical nor simply metaphorical. It concerns directly a foundation of finance, namely, the notion that risk is reduced by averaging. Denote by $N$ the number of items being averaged. Under mild randomness, the central limit theorem says that this reduction proceeds as $1/\sqrt{N}$. In wild randomness the typical reduction is smaller or non-existent. Moreover, the variability around the typical value is greater and can be very large, accounting for the occurrences of discontinuity and concentration in economics. There is no space to dwell further on this effect, which, once again, is the topic of Mandelbrot (1997) and also of Mandelbrot (1999, 2001).

I view the study of wild variability/randomness as being at present one of the proposed frontiers of scientific knowledge. But here it will serve the modest goal of showing that different reactions to the Officer effect proceeded in three altogether distinct directions.

## 3.6. Mild randomness; the claim that the Officer effect establishes that coin tossing and Bachelier can safely be trusted, after all

There is no question that Officer has shown convincingly that the M1963 model (however effective it may be in some cases) *is not* the last word on financial markets. However, the nascent quantitative finance community of the day concluded that Officer also discredited the overwhelming evidence of non-Gaussianity. Surviving witnesses recall that this impression helped overcome the, then, widespread awareness of the M1963 model and unblocked the door to the Brownian 'modern portfolio theory'.

**Critique of the revival argument.** There are several reasons why the step back to Bachelier is *thoroughly unwarranted*. Asymptotically for large $\Delta t$, the move away from long tails to shorter ones only describes a trend but not its limit. The notion that in the long-run everything is safely Brownian is sharply contradicted by figure 3. In any event, portfolio theory deals primarily with short-term price changes, therefore is not primarily influenced by the fact that the distribution of $\Delta P(t)$ has increasingly short tails as $\Delta t$ increases. What matters most is the fact that short-term changes are clearly non-Gaussian. In addition, they are non-independent, a feature that Officer himself and seemingly all those who reproduced his results failed to investigate.

## 3.7. Slow randomness; the 'durable transients' argument: Brownian motion holds for large time increments, but small time increment changes are neither Gaussian nor independent

### 3.7.1. A generic 'durable transients' argument.
The overall theme is that Bachelier is not completely right but not too far off and the details must be fine tuned for small $\Delta t$. But the averaging that underlies diversification proceeds as usual and it continues to be safe to believe that financial reality can be modelled without having to give up any essential tool, such as the law of large numbers and the central limit theorem. In terms of states of randomness, this means that fine-tuning adjustments are needed.

The best developed and known implementation of this attitude is the GARCH model. It immediately involves a large number of parameters.

### 3.7.2. Truncated power-law distribution, an extremely dangerous dead end pioneered in Pareto (1896) that unfortunately is back in fashion.

Instead of fitting the distribution tails, several recent writers trained in physics fitted the central bells and reported extremely small values of $\alpha$ in the range of 1.3 to 1.4. Since those values clearly fail to fit the tails, those authors proposed diverse forms of tail truncation. The details do not matter much, therefore complete references are unnecessary, but some general points deserve to be made.

**The atypical Grand Duchy of Oldenburg.** The underlying issue was identified in Pareto (1896): in Oldenburg, the high income tail was far shorter than in a power-law distribution. Pareto suggested an exponential factor $\exp(-\beta u)$. One may just as well replace the power law by $\exp(-\beta u)$ for large $u$.

A cynic might observe that tiny Oldenburg hardly justified individual attention. Perhaps a few high incomes have been excused from being reported. Be that as it may, the exponential correction became very popular among old-fashioned statisticians for the sole reason that it avoids divergent variance.

**Cut-offs in the physics of Pareto's time.** Quanta did not merely avoid an unpleasant divergence but became the physically real seed of much of 20th Century physics. A second cut-off postulate was meant to avoid certain divergence paradoxes of Newtonian gravitation by 'replacing' the hallowed $1/r^2$ by $\exp(-\beta r)/r^2$. The paradoxes proved non-lethal and the correction unnecessary and short lived.

**Cut-offs in critical phenomena.** The words 'high critical temperature superconductors' hit front page headlines a few years ago. The critical $T_c$ is the value of the temperature $T$, below which a compound is superconducting and above which it is not. The underlying concept is old: for magnets, Pierre Curie observed that magnetism disappears above a certain 'Curie temperature', $T_c$.

The very attractive and important notion of criticality is meaningful only in the presence of a concretely meaningful variable that can be tuned very precisely, like a magnet's temperature. It was shown in the 1960s and 1970s that in the 'critical' case $T = T_c$, magnetism and other phenomena are ruled by power laws. In the neighbourhood of $T_c$, all those laws change; one needs exponential corrections, that are not *ad hoc* but physically real functions of the distance from criticality. Moreover, both the critical exponents and the corrections are obtained by exact analytic arguments.

Unfortunately, turbulence and finance lack an intrinsic tunable parameter. These examples show that power laws and fractality can have sources other than criticality and its variants.

Having participated to some extent in the development of the theory of critical phenomena, I observed a by-product of the existence of physically meaningful cut-offs: the physicists rarely faced infinite moments. Few acquired the necessary specialized skills in probability theory. Few became acquainted with the fine points of multifractals that are needed in finance (see part II). This is perhaps why many physicists proved to be ill-prepared for the study of power laws in finance.

Be that as it may, they suggested one should follow the lead of Pareto and superpose an exponential correction on the Lévy stable distribution. When combined with independent price increments, the exponential correction only holds for one value of $\Delta t$. Moreover, it destroys scaling. Therefore, the evidence that was interpreted as demanding the truncation was also interpreted as marking the limit of validity of scaling.

### 3.8. Wild randomness; financial prices' 'misbehaviour' never settles down to the Brownian; their variation remains 'wild' even on the scale of the century; scaling need not be abandoned, only amended

The third approach to financial data, mine, begins with the notion that the search for transients towards Brownian motion is a thoroughly ill-conceived idea. The evidence exemplified by figure 3 shows that the deviations from the Gaussian and the Brownian are not limited to small $\Delta t$.

Financial reality is *not* mildly variable even on the scale of a century. All things considered, one must adjust to the fact that financial reality is wildly variable. It would be totally unmanageable, unless there is some underlying property of invariance.

My search for an invariance beyond those that led to the M1963 and M1965 models was not triggered by Officer. Well before 1972, I had recognized that the M1963 model had to be amended to include *global long-term* dependence. I was strongly pressed by a very astute observer of finance, William S Morris, an early promoter and pioneer of the use of computers in trading. In describing the M1963 model, Morris suggested that one should be able to model dependence by analogy with Berger and Mandelbrot (1963). This brilliant insight was impossible to implement and the underlying idea had to wait for the discovery of multifractals.

By coincidence, the key to the solution of the Officer effect was also given in the same year 1972, in the last remark of one of my first papers on multifractals (reprinted in Mandelbrot (1999) as chapter N14). That paper was the first in which I showed how multifractals can generate power-law probability distributions. It was not hastened either, which I now regret.

## References

Barral J and Mandelbrot B B 2000 *Multifractal Products of Cylindrical Pulses*

Berger J and Mandelbrot B B 1962 (Reprinted as chapter 6 of Mandelbrot 1999)

Calvet L and Fisher A 2001 *On the Multifractal Model of Asset Returns*

Chipman J 1976 *Revue Européenne des Sciences Sociales et Cahiers Vilfredo Pareto* **XIV** 65–173 (a reprint is being considered and deserves to be encouraged)

Cootner P H (ed) 1964 *The Random Character of Stock Market Prices* (Cambridge, MA: MIT Press)

Durrett R and Liggett T M 1983 *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **64** 275–301

Fama E F 1963a *J. Business* **36** 420–29 (Reprinted in Cootner (1964) and Mandelbrot (1997))

Fama E F 1963b The distribution of daily differences of stock prices: a test of Mandelbrot's stable Paretian hypothesis *PhD Dissertation* Graduate School of Business, University of Chicago (In part published as Fama (1965))

Fama E F 1965 *J. Business* **38** 34–105

Lévy P 1925 *Calcul des probabilités* (Paris: Gauthier–Villars)

Mandelbrot B B 1962 The variation of certain speculative prices *IBM External Research Report NC-87*

Mandelbrot B B 1963 *J. Business* **36** 394–419 (Reprinted in Cootner 1964, Mandelbrot (1997) and several collections of papers on finance)

Mandelbrot B B 1965 *C. R. Acad. Sci., Paris* **260** 3274–7 (Engl. transl. see Mandelbrot (2001))

Mandelbrot B B 1967 *J. Business* **40** 393–413 (Reprinted as chapter 15 of Mandelbrot (1997))

Mandelbrot B B 1972 (Reprinted as chapter 14 of Mandelbrot (1999))

Mandelbrot B B 1974 *J. Fluid Mech.* **72** 401–16 (also Mandelbrot B B 1974 *C. R. Acad. Sci., Paris* **278A** 289–92; 355–8) (reprinted as chapters 15 and 16 of Mandelbrot (1999))

Mandelbrot B B 1982 *The Fractal Geometry of Nature* (San Francisco, CA: Freeman)

Mandelbrot B B 1997 *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk* (Berlin: Springer)

Mandelbrot B B 1999 *Multifractals and 1/f Noise: Wild Self-Affinity in Physics* (Berlin: Springer)

Mandelbrot B B 2000 Cartoons of the variation of financial prices and of Brownian motions in multifractal time *Discussion Paper 1256 of the Cowles Foundation for Economics Yale University New Haven, CT*

Mandelbrot B B 2001 *Gaussian Self-Affinity and Fractals* (Berlin: Springer)

Mandelbrot B B, Calvet L and Fisher A 1997 The multifractal model of asset returns; large deviations and the distribution of price changes; the multifractality of the Deutschmark/US Dollar exchange rate *Discussion Papers of the Cowles Foundation for Economics Yale University New Haven, CT*. Paper 1164: http://papers.ssrn.com/sol3/paper.taf?ABSTRACT_ID=78588; Paper 1165: http://papers.ssrn.com/sol3/paper.taf?ABSTRACT_ID=78606; and Paper 1166 http://papers.ssrn.com/sol3/paper.taf?ABSTRACT_ID=78628

Officer R R 1972 *J. Am. Stat. Assoc.* **67** 807–12

Pareto V 1896 *Cours d'économie politique* (Reprinted in Pareto V 1965 *Oeuvres complètes* (Geneva: Droz))

Zipf G K 1949 *Human Behavior and the Principle of Least-Effort* (Cambridge, MA: Addison-Wesley) (Reprinted Zipf G K 1972 *Human Behavior and the Principle of Least-Effort* (New York: Hefner))