Power laws in finance and their implications for economic theory

J. Doyne Farmer* and John Geanakoplos†

April 28, 2004

Abstract

The following presents a review of power laws in financial economics. It is a chapter from a preliminary draft of a very long review article called "Beyond equilibrium and efficiency". While some of the discussion is specific to economics, most of it applies to power laws in general – the nouns may change, but the underlying questions are similar in many fields. This draft is still highly preliminary – comments at any level are greatly appreciated.

Contents

*McKinsey Professor, Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe NM 87501

†Economics Department, Yale University, New Haven CT

1 Power laws

Crudely speaking a power law is a relation of the form $f(x) \approx Kx^{\alpha}$, where K and α are constants, and $x > 0$. Although their existence and interpretation is controversial, there is considerable evidence suggesting that many of the statistical properties of financial markets are asympototically described by power laws. This has important practical consequences for risk management, volatility forecasting, and derivative pricing. It is also conceptually important because it suggests a different emphasis in economic modeling.

Power laws correspond to scale free phenomena. To see this, consider a scale change $x \to cx$. A power law is the only[1] the scale invariance equation $f(cx) = Kf(x)$. If $f(x) = Kx^{\alpha}$, it is transformed as $f(x) \to Kc^{\alpha}x^{\alpha} = c^{\alpha}f(x)$. Changing scale thus preserves the form of the solution with only a change of scale. Thus power laws are a necessary and sufficient condition for scale free behavior.

The importance and ubiquity of such scale free behavior was pointed out by Mandelbrot [43, 44, 45], who coined the word "fractals" for geometric objects with power law scaling with α not equal to an integer, and demonstrated that such fractals are ubiquitous in nature, describing phenomena as diverse as coastlines, clouds, floods, earthquakes, fundamental inaccuracies in clocks, and financial returns. Fractals have the property that, by using an appropiate magnifying glass, one sees self-similar behavior across different scales. Of course, this is always just an approximation, which is only valid across a given range. To produce fractals, the underlying generating mechanism should also reflect this self-similar behavior. To physicists, the apparent prevalence of power laws in financial markets is an important clue about how to model them. Explaining power laws is important for its own sake, but it is also likely to have broader consequences in other respects.

[1] $f(x) = 0$ or $f(x) = 1$ are also scale-invariant solutions, but these are just power laws with exponents $\alpha = -\infty$ or $\alpha = 0$.

The existence of power laws may or may not be compatible with economic equilibrium, but in either case, it suggests a different emphasis.

We begin by defining what a power law is, and explaining why power laws are important, using the phenomenon of clustered volatility as an illustration. After reviewing empirical observations of power law scalings, we address some of the controversy surrounding this question, and give a response to some of recent criticism [16, 31]. We discuss the practical consequences of power laws for risk management and options pricing, to make it clear why they are important for financial engineering, We then review a few of the mechanisms for generating power laws, and discuss how addressing this might either alter the approach to making equilibrium models, or challenge their appropriateness as a description of reality.

1.1 What is a power law?

The crude definition of a power law we have given above is misleading, for two reasons. First, we need to be more precise about what we mean by "approximately", since the notion of power law scaling allows for asymptotically irrelevant variations, such as logarithmic corrections. Second, a power law is inherently an asymptotic notion, exactly valid only in a specified limit. Confusion on these two points has led to a great deal of misunderstanding in the literature, so it is worth spending some time to discuss these issues carefully.

The notion of a power law as it is used in extreme value theory [17] is an asymptotic scaling relation. Two functions f and g can be defined to have equivalent scaling, $f(x) \sim g(x)$, in the limit as $x \to x_0$ if

$$\lim_{x \to \infty} L(x) f(x)/g(x) = 1, \tag{1}$$

where $L(x)$ is a slowly varying function. A *slowly varying function* $L(x)$ satisfies

$$\lim_{x \to \infty} L(tx)/L(x) = 1. \tag{2}$$

for any $t > 0$. Examples are $L(x) = K$, where K is a constant, and $L(x) = \log x$. A function f has power law scaling in a given limit if $f(x) \sim x^{\alpha}$. For the applications discussed here, unless otherwise stated, we are typically interested in large fluctuations, i.e. in the limit $x \to \infty$. For a distribution functions $P(x > X)$ is it standard to express the scaling in terms of the cumulative as $P(x > X) \sim X^{-\alpha}$, where $\alpha > 0$ is called the *tail exponent*. The corresponding distribution function scales as $p(x) \sim x^{-(\alpha+1)}$.

For convenience we will assume $x > 0$; the negative values of a double-sided distribution are treated by taking absolute values.

Power laws emerge naturally as a limit law for random processes in several different contexts. One of these is the central limit theorem: A sum of IID random variables whose second moments are bounded is described by a normal distribution, but if the second moment is unbounded, it is described by a Levy stable distribution, which has power law tails with $0 < \alpha < 2$. As we will discuss in Section 1.4, there are other random processes, such as mixtures of sums and products of IID random variables with appropriate constraints, which can lead to power laws.

Power laws also emerge as one of three possible limiting distributions in extreme value theory [17]. A power law describes the extrema for independent draws from any distribution with sufficiently fat tails'[2]. For example, consider a sequence of n IID random variables x_i drawn from a distribution $P(x)$. Is there any universality in the limiting behaviors of the extrema? For example, consider the maximum $M_n = \max(x_1, \ldots, x_n)$ in the limit as $n \to \infty$. It turns out that after normalizing by location and scale parameters, and taking the limit of large n, there are only three possibile limiting distributions for M_n, whose names are Weibull, Gumbel, and Frechet. There are of course examples in which $P(x)$ has no limiting behavior at all, but these are pathological – well-behaved distributions are described by one of these three behaviors.

These three limiting distributions describe the behavior of not just the maximum, but also the second maximum, third maximum, etc. For any fixed m, for sufficiently large n, if a limit exists, the m^{th} maximum is described by one of these three distributions. Thus these distributions also describe the asymptotic order statistics for a given probability distribution, and thus provide a means of characterizing the distribution of extreme values.

Which limiting distributions is appropriate depends on how fat the tail of $P(x)$ is. If it has finite support ($P(x) = 0$ for $x > x_{max}$), then the limiting distribution is Weibull. If it has infinite support but the tails decrease sufficiently fast so that all the moments of $P(x)$ exist, then the limiting distribution is Gumbel. (Examples are normal and exponential distributions). The Frechet distribution $\Phi_\alpha(x) = 1 - \exp(-x^{-\alpha}) \sim x^{-\alpha}$ describes the limiting behavior of distributions with tails that die off sufficiently slowly that not all their moments exist. In fact, it is possible to show that the set of

[2] The terms "fat tails", "heavy tails", and "long tails" are loose designations for any distribution whose measure at extreme values is greater than that of a "thin-tailed" reference distribution, typically a normal or an exponential.

distributions that are asymptotically equivalent to the Frechet distribution are *regular functions*, i.e. those with the property that

$$\lim_{x \to \infty} h(tx)/h(x) = \chi(t), \tag{3}$$

where $\chi(t)$ is positive and bounded[3]. This is a large and important class of functions. A power law, then, describes the scaling behavior of the extremal values of any non-pathological distribution with sufficiently fat tails.

The tail exponent α has a natural interpretation as the cutoff above which moments no longer exist. To see this, consider the density function $p(x) \sim x^{-(\alpha+1)}$. The m^{th} moment is

$$\gamma = \int x^m p(x) dx \sim \int x^m x^{-(\alpha+1)} dx. \tag{4}$$

γ is finite when $m < \alpha$ and it is infinite otherwise. The tail exponent thus provides a single number summary of the fatness of the tails.

We want to stress that power law behavior is an asymptotic scaling property, defined in a particular limit. In fact, there is no such thing as a "pure" power law distribution on $[0, \infty]$. This is because of integrability: To see this suppose there were a density function $p(x) = Kx^{-(\alpha+1)}$. For $\alpha < 0$, $\int_0^\infty p(x)dx = \infty$ due to the upper limit, and similarly, for $\alpha > 0$, $\int_0^\infty p(x)dx = \infty$ due to the lower limit. Thus, when we say that "a distribution is a power law", we are necessarily referring to an approximate behavior that becomes exact only in a given limit. This makes testing for power laws challenging, because there is always ambiguity as to whether one has enough data to be in the asymptotic limit. We will say more about this later.

The relevance of power laws is not limited to marginal distributions of a single variable. Joint distributions can asymptotically follow a power law, which can be reflected in the scaling properties of moments, such as the autocorrelation function. A particularly interesting case is that of a long-memory random process, which has an autocorrelation function that scales as $\tau^{-\alpha}$, with $\alpha < 1$. Long-memory processes appear to be surprisingly common in economic time series.

One of the reasons that power laws are ubiquitous is because of their invariance under aggregation: The property of being a power law is conserved under addition, multiplication, and polynomial transformation. When two

[3]The connection to power laws becomes obvious by writing $h(tsx)/h(t) = (h(tsx)/h(tx))(h(tx)/h(x))$. This implies that $\chi(ts) = \chi(t)\chi(s)$, which has the solution $\chi(t) = t^\alpha$.

power law distributed variables are combined, either additively or multiplicatively, the one with the fattest tailed distribution, dominates, i.e., the tail exponent of the new distribution is the minimum of the tail exponents of the two distributions being combined. Furthermore, when a power law distributed variable is raised to a (nonzero) power, it remains a power law but with an altered exponent. Letting $\alpha(x)$ be the tail exponent of the random variable x, we can write these three transformation rules in the following form:

$$\begin{aligned} \alpha(x+y) &= \min(\alpha(x), \alpha(y)) \\ \alpha(xy) &= \min(\alpha(x), \alpha(y)) \qquad (5) \\ \alpha((x^k) &= \alpha(x)/k \qquad (6) \end{aligned}$$

The first two rules state that under addition or multiplication the fattest tailed distribution dominates. Under a polynomial transformation, the lowest order term of the polynomial will dominate.

Of course, these simple aggregation properties are only valid in the tail, and may not be a good description for what happens to the rest of the distribution. For example, consider an idealized model for price aggregation: Let π_t be the price at time t, and $f_t > 0$ be the multiplicative change in price from the previous period, $\pi_t = f_t \pi_{t-1}$. This can be rewritten as

$$\log \pi_t = \log \pi_{t-1} + \log f_t. \qquad (7)$$

The logarithmic price increments $\log f_t$ can be approximated as having a symmetric distribution whose positive and negative tail exponents are typically in the range $1.5 < \alpha < 6$. (See Section 1.2.1). It may help to think about Student's t distribution, which has symmetric positive and negative power law tails whose tail exponent α is the number of degrees of freedom. If we sum together n independently drawn price increments $\log f_t$, providing $\alpha \geq 2$, as n increases the center of the distribution becomes more and more normal-looking, due to the action of the central limit theorem, but the tails remain power laws with the same tail exponent α. As n increases the functional form of the tail is preserved, but the power law scaling region becomes smaller. Thus, on short timescales a power law describes a large fraction of events, but on longer timescales normality is a good approximation for the center of the distribution. The power law never goes away, even on long timescales; it just describes rarer and more extreme events. See the discussion of risk analysis in Section 1.2.3.

1.2 Practical importance of power laws in financial economics

Power laws are relevant for financial economics because they have both practical importance and theoretical implications. In this section we begin by briefly reviewing the literature reporting power laws in financial economics, listing some of the diverse phenomena that are reported to have power law properties. We then discuss several of these in more detail, explaining how an understanding of power law properties has both practical and theoretical importance for problems such as clustered volatility, risk control, and option pricing. In Section 1.3 we present a summary of the debate surrounding this literature.

1.2.1 Empirical evidence for power laws

Power laws in have been reported for a wide variety of different phenomena in financial markets. Some examples are:

- *Large price changes on short time scales*, e.g. a month or less [42, 20, 53, 1, 30, 37, 46, 36, 38, 25, 50, 54, 47]. Price changes are measured in terms of log-returns $\Delta p = \log p(t+\tau) - \log p(t)$, where p can be either a transaction price average of the best quoted buying and selling prices; τ is the timescale. The tail exponent is typically in the range $1.5 < \alpha < 5$. For individual American stocks, for example, the exponents for positive and negative returns are roughly the same, and power law scaling is a good approximation for almost the entire range [54]. Although the first papers by Mandelbrot [42] and Fama [20] gave $\alpha < 2$, suggesting that the second moment did not exist, most later work reports $\alpha > 2$. There are probably real variations in the tail exponent across different assets.

- *Clustered volatility.* The autocorrelation of the absolute value of price changes decays as roughly $\tau^{-\gamma}$ for large τ, with $\gamma \approx 0.2$ [15, 56, 47].

- *The volume of individual transactions* for NYSE stocks [26]. The scaling exponent $\alpha \approx 1.7$.

- *Fluctuations in the width of the distribution of growth rates of companies* [**?**]. Letting s be the standard deviation in the logarithmic growth rate, $P(s > S) \sim S^{-\alpha}$, with $\alpha \approx 0.2$.

- *Firm size.* The size s of large firms measured by a variety of different methods, e.g. market capitalization or number of employees has a tail exponent $\alpha \approx 1$ [67, 27, 3].

- *The prices for limit order placement* measured relative to the best price. Let the relative limit price be $\Delta = |\pi - \pi_{best}|$, where π is the price where a new limit order is placed, and π_{best} is the best quoted price for orders of the same type, e.g. if the limit order is a buy order, π_{best} is the best quoted price for buy orders. $\alpha \approx 0.8$ for the Paris Stock Exchange [7], and $\alpha \approx 1.5$ for the London Stock Exchange [68].

- *The price impact as a function of market capitalization.* Price impact is defined as the difference between the average of the bid and ask quotes immediately before and after a transaction. Even after a normalization dividing the trading volume by the average trading volume for the given stock, the price impact scales as C^γ, where C is the market capitalization and $\gamma \approx 0.4$ [34].

- *The cumulative sum of negative returns following a crash.* Following a large downward move in prices, all subsequent downward price movements that exceed a given threshold are accumulated. The cumulative sum increases as t^γ, where t is the time since the crash, and $\gamma \approx 1$ [33]. A similar relationship for seismometer readings after large earthquakes was observed by Ohmori in the nineteenth century [45].

- *The autocorrelation of signs of trading orders.* Let the sign of a buy order be $+1$, and the sign of a sell order be -1. The autocorrelation of signs decays as $\tau^{-\gamma}$, where τ is the number of events separating the orders. $\gamma \approx 0.2$ for the Paris and $\gamma \approx 0.6$ for the London Stock Exchange [57, 32].

- *Autocorrelation of order volume.* For the London Stock Exchange the autocorrelation function of order sizes measured in either shares or pounds decays in event time as roughly $\tau^{-\gamma}$, with $\gamma \approx 0.6$ [32].

- *Autocorrelation of liquidity at the best bid and ask.* For the London Stock Exchange the autocorrelation function of the volume at either the best bid or the best ask decays in event time as roughly $\tau^{-\gamma}$, with $\gamma \approx 0.6$ [32].

- *Distribution of income or wealth.* The distribution of income or wealth has a power law tail. The exponent varies from country to country and epoch to epoch, with the tail exponent in the range $1 < \alpha < 3$. This was the first power law ever discovered, and for that reason power law distribution are sometimes also called Pareto distributions [**?**].

For a more in-depth discussion of many of these, see Cont [12].

1.2.2 Clustered volatility

Rational expectations equilibrium predicts that prices should be uncorrelated in time. This is observed to good approximation in real prices. However, even though *signed* price changes are uncorrelated, their amplitude or *volatility* is strongly positively correlated. This is called *clustered volatility*. That is, if the market makes a big move on a given day, it is likely to make a big move on the next day, even though the sign remains unpredictable (at least from the point of view of a linear model). Studies of price time series show that the autocorrelation of absolute price returns asymptotically decays as a power law of the form $\tau^{-\alpha}$, where $\alpha \approx 0.2$. The fact that this is a power law with an exponent less than one is important because it implies a long-memory process [**?**]. That is, its autocorrelation function is nonintegrable, so that events in the far past have a substantial influence on the present. This long-memory gives rise to bursts of volatility on timescales ranging from minutes to decades. This implies that it follows a random process very different from standard Markov processes.

Equilibrium models of the type that we have discussed here predict that the amplitude of price changes is driven soley by the information arrival rate. If the states of nature become more uncertain, then prices respond by fluctuating more rapidly. Thus, clustered volatility is just a reflection of an exogenous property of the economy. This could be due to physical driving forces such as natural disasters, or it could be due to some innate non-economic property of human interactions that causes people to generate news in a highly correlated way. In fact, it is well-established that most natural disasters, such as flood, hurricanes, and droughts, are long-memory processes, so the physical explanation is plausible [**?**].

Measuring the arrival rate of news quantitatively is more difficult, but studies that attempt to correlate news arrival with large market moves seem to generate results that suggest that the correlation is not very strong. For example, Culter, Poterba and Summers [13] examined the largest 100 daily price movements in the S & P index during a 40 year period, and showed that most of the largest movements occur on days where there is no discernable news, and conversely, days that are particularly newsworthy do not typically correspond with large price movements. This suggests that a substantial fraction of price changes are driven by factors unrelated to information arrival. Furthermore, price volatility when markets are closed, even on non-holidays, is much lower than when the market is open [**?**]. These studies seem to suggest that news arrival and market movements are not closely correlated. It appears that the market makes its own news.

It is noteworthy that clustered volatility emerges more or less automatically in many agent-based models with bounded rationality, which allow deviations from a rational expectations equilibrium [2, 9, 40]. Many of these models also capture the property that signed price series are uncorrelated. Thus, while the lack of correlation in prices is often cited as a validation of equilibrium theory, the same prediction is also made by models with weaker assumptions. Furthermore, these models also display clustered volatility, a feature that is not present in current equilibrium models. This suggests that the nonequilibrium models contain aspects of realism not captured by their equilibrium counterparts.

We currently do not know whether it is possible to make equilibrium models that can make their own news, and spontaneously generate clustered volatility. This might come about naturally in a temporary equilibrium setting, whose finite planning horizon is a form of bounded rationality. More work is needed to determine the necessary and sufficient conditions for clustered volatility.

There are also practical reasons to focus on clustered volatility due to its role in risk control and option pricing, as discussed in the next section.

1.2.3 Option pricing and risk control

Power laws have important practical implications for both option pricing and risk control. This comes about both because of the fat tails of the marginal distribution of price changes and because of clustered volatility.

Power laws are important for risk control because extreme price movements are more than one might expect, and the power law hypothesis provides a parsimonious method of characterizing them . To make the effect of fat tails more tangible, in Table 1 we compare a normal distribution to a power law distribution. To calibrate this to price distributions, we choose both distributions to have the same standard deviation. We use Student's t-distribution as proxy for a price distribution, and choose it to have a tail exponent $\alpha = 3$, comparable to daily price returns. This table makes it clear that there is little difference in the typical fluctuations one expects to observe every ten or one hundred days, but the typical 1/1000 event is twice as large for a power law and the 1/10,000 event is three and a half times as large, something a risk manager might want to take seriously. This becomes even more dramatic when looked at the other way: The probability of observing a fluctuation of 25% (the size of the famous negative S&P return on October 19, 1987) under the normal hypothesis is less than 10^{-16}, whereas the probability under the power law distribution is 0.08%. Under

Probability	0.9	0.99	0.999	0.9999
Normal	3.8	7.0	9.2	11
Student	2.8	7.8	17.7	38.5

Table 1: A comparison of risk levels for a normal vs. a power law tailed distribution. Student's t distribution with three degrees of freedom, which has a tail exponent $\alpha = 3$, is chosen as a proxy for daily price returns. Both distributions are normalized so that they have a standard deviation of 3%, a typical value for daily price fluctuations. We assume that returns on successive days are independent. The top row gives the probability associated with each quantile, and the values in the table are the size of the typical events for that quantile, in percent. Thus, the first column corresponds to typical daily returns that one would expect to see every ten days, and the last column events one would expect every 10,000 days, i.e. every 40 years.

the normal distribution it is essentially impossible that this event could ever have occurred, whereas under a power law distribution such an event is to be expected.

The practical value of the power law hypothesis is that it results in better extreme risk estimates. Consider the problem of estimating the future risk of extreme events from an historical sample of past returns. Commonly used nonparametric methods, such as the empirical bootstrap, work well for interpolating risk levels that have already been experienced in the sample. However, when used to extrapolate risk levels that are not contained in the sample, they will consistently underestimate risk. The power law hypothesis, in contrast, is more parsimonious, and can result in more accurate and less biased estimates.

In addition, risk control estimates are affected by clustered volatility. Power law scaling of volatilty implies that prices obey a random walk with anomalous diffusion. For a random walk the variance of price fluctuations on timescale τ grows at τ^{2H}, where H is called the Hurst exponent. For a normal random walk $H = 0.5$, but for a superdiffusive random walk, $H > 0.5$. There is evidence suggesting that prices obey a superdiffusive random walk (which is equivalent to saying that volatilty is a long-memory random process). This implies that prices can make much larger excursions than one would expect if their size were uncorrelated.

Power laws also have practical importance for estimating volatility. The mainstream approach for understanding clustered volatility is in terms of ARCH models and their generalizations [18, 17]. The ARCH family of models are linear time series models with characteristic length scales, and fail to capture the power law autocorrelation structure of real data. Several models have been proposed that use the power law hypothesis to forecast volatility [**?**, **?**]. The results so far suggest that such models have substantially more predictive power than standard ARCH models [39]. The fact that volatility time series show scaling across a wide range of timescales suggests that a similar mechanism may be a work. The origin of such a regularity is an interesting for its own sake.

Power laws also have practical implications for derivative pricing. This due to both fat tails in prices and power law scaling of clustered volatility, both of which affect option prices. Models that incorporate the power law tails of real prices provide a better characterization of option prices than the standard Black-Scholes models, and a more parsimonious fit to the data than non-parametric alternatives [8, 6].

1.3 The empirical debate

Many economists have been quite sceptical about power laws, and whether power laws exist at all in economics has been a subject of debate. In this section we review methods of data analysis for determining whether power laws exist, and discuss some of the criticisms that have been raised.

1.3.1 Testing the power law hypothesis

The most common procedure used to test for the existence of a power law is visual inspection. In a typical paper, the authors simply plot the data in double logarithmic scale and attempt to fit a line to part of it. If the line provides a good fit over a sufficiently wide range, hopefully at least two orders of magnitude, then the authors suggest that the data obey a power law with an exponent equal to the slope of the line. This has many obvious problems: First, there is no objective criterion for what it means to be a “good fit”, and second, the choice of a scaling range creates worries about overfitting. Not surprisingly, the subjectivity of this procedure has engendered criticism in economics and elsewhere [**?**].

A quantitative approach to hypothesis testing makes use of extreme value theory to reduce this to a statistical inference problem. This takes advantage of the fact that there are only three possible extremal limiting distributions, as described in Section 1.1. The testing procedure uses each of the three limiting distributions as a null hypothesis. If the Weibull and Gumbel hypotheses are strongly rejected, but the Frechet hypothesis is not, then there is good evidence for a power law distribution[4]. There are several examples where these methods have been applied and give highly statistically significant results supporting power laws [1, 30, 37, 36, 38, 50]. These methods, however, are not fully satisfying. There are several problems. One is that these tests assume the data are IID, whereas price returns have clustered volatility and are so are not IID. It is an open problem to develop a test that properly takes this into account[5].

Testing for power laws is inherently difficult due to the fact that a power law is an asymptotic property, and in a real data set one can't be sure there is enough data to be inside the asymptotic regime. Some power law behaviors converge very quickly, so that for most of the regime the power law is a good approximation, while others converge very slowly. It is quite easy to construct distributions that will fool any test unless there is a very large sample of data. This is a reflection of a broader problem: Testing for a power law is inherently more difficult than testing for conformity to a specific distribution. This is because the power law is a property of a family of distributions, and so requires testing for membership in an equivalence class whose members have properties that are not well specified in advance. This is further complicated by the fact that in many cases boundary constraints dictate inherent cutoffs to power law scaling. The magnitude of earthquakes, for example, displays clear power law scaling across many orders of magnitude, but there is an obvious cutoff due to the physical constraint that there is an upper bound on the amount of energy that can be stored in the earth's crust. Thus, while a power law is an asymptotic behavior, for real applications there are always limits imposed by finite size. These issues combine to make testing for power laws inherently difficult. Sensible interpretation of results depends on good judgement. In this context, the crude but commonly used visual inspection method is valuable, and may have merit in forcing the reader to use judgement and common sense in interpreting the results, rather than obscuring them with formal methods whose results may or may not be meaningful [65].

The simplest method for improving the fidelity of tests for power laws is to use more data. Recent studies have achieved this by studying high

[4] Alternatively, one can show that the posterior odds of the Frechet hypothesis are much higher than either of the alternatives.

[5] A related problem is that of testing for long-memory. The test originally proposed by Mandelbrot [**?**, **?**] is too weak (in that it often fails to reject long-memory even when it is not present), while a revised test proposed by Lo [35] is too strong (it often rejects long-memory even when it is known to be present). This is another area where improved hypothesis testing would be very useful.

frequency data, often involving millions of observations [50, 54, 32]. Understanding at longer frequencies can be achieved by making assumptions about time aggregation, and making use of the fact that the power law tails of a distribution are preserved under most aggregation mechanisms [8, ?]. Thus, if one finds a power law in high frequency data, barring rather unusual time aggregation mechanisms, it will still be present at lower frequencies, even if it describes only rarer events.

Data analysis should always be viewed as a first step whose primary importance is in guiding subsequent modeling. The real test is whether power laws can be demonstrated to improve our predictive or explanatory power. There is already some evidence for this, e.g. in risk control, in predicting volatility, and explaining option prices. Self-similiarity is such a strong constraint that, even if only an approximation over a finite range, it is an important clue about mechanism. Ultimately, the best method to demonstrate that power laws are applicable is to construct theories that explain their existence, whose validity can be demonstrated through the details of the underlying mechanism. See the discussion in Section 1.4.

1.3.2 The critical view

Because of the problems with hypothesis testing discussed above there has been considerable debate about whether power laws exist at all in economics. One of the often-cited studies is by Lebaron [31], who showed that he could mimic the claimed power law behavior of a real price series using a model that does not have power law scaling. He fitted the parameters of a standard stochastic volatility model to match the price statistics of a Dow-Jones index proxy. The data set contains daily prices averaged over the 30 largest U.S. companies for a period of about a century, with roughly $30,000$ observations. This price series was studied by several authors [46, 36] who claimed that the evidence supported power law scaling in prices. Lebaron demonstrated that he could produce similar results using a stochastic volatility model with three time timescales. This is significant since it can be shown that the model he used does not have asymptotic power law scaling, even though it might be mistaken for power law scaling on finite data sets (of the size of the proxy Dow Jones series). Thus, he suggests, the power law found in the data may only be an illusion. This study has been cited as raising grave doubts about the whole question of power laws in finance and economics [16].

The physicist responds by noting that in order to fit this model, Lebaron had to choose very special parameters. In his model the logarithmic volatility level is driven by a combination of three AR(1) models, one of which is $y_t = 0.999 y_{t-1} + n_t$, where n_t is an IID noise process. The parameter 0.999 is very close to one; when it is one, the model has asymptotic power law behavior. The reason the model has an approximate power law is because it is very close to a model with a true power law.

This is a reflection of a broader issue: For the family of volatility models Lebaron uses, under random variations of parameters, those that mimic power laws are rare. In Section 1.2.1 we listed twelve different aspects of markets where the evidence suggests power laws. While it might be possible that a few of these are better described in other terms, it seems unlikely that this could be true of all of them.

Furthermore, there is the important issue of parsimony: Why use a model with three parameters when one can describe the phenomena as well or better using a model with one or two? To fit the series Lebaron has to choose three timescales which have no natural a priori interpretation. The scale free assumption is both more parsimonious and more elegant.

A common statement by economists is that power law scaling is easily explained in terms of mixture distributions. This statement derives from the fact that mixtures of distributions, for example a linear combination of normal distributions with different standard deviations, have fatter tails than any of the individual distributions by themselves. However, the key point that often seems to go unrecognized is that this is not sufficient to get asympototic power law behavior – for this to be true the mixture has to be sufficiently inhomogeneous. In the case of price fluctuations, there is now good evidence that this is not the correct explanation [23, 62].

The critiques certainly make the valid point that better and more careful testing is needed, and that too much of the data analysis relies on visual inspection alone. However, there is a substantial body of evidence suggesting that power law behaviors exist in economics, at least as an approximation. Either we need to do more work to reconcile this with equilibrium models, or we need to find entirely new approaches.

1.4 Mechanisms for generating power laws

Physicists view the existence of power laws as an important modeling clue. It seems this clue has so far been largely ignored by financial economists. For physical systems, for example, it is clear that power laws cannot be explained by linear or equilibrium models (equilibrium in the physics sense). This is not necessarily true in economics – indeed, there is at least one example illustrating that economic equlibrium can be consistent with power laws

[52]. Nonetheless, the existence of power laws suggests a change in the focus of attention in model building.

In this section we give a review of mechanisms for generating power laws. The requirements for generating power laws is not a well-developed subject – there are no theorems stating the necessary and sufficient conditions. Furthermore, there are many levels of detail on which models can be constructed, and these levels are not necessarily mutually exclusive. Thus, the same phenomenon might be explained in terms of a maximization argument, a nonlinear random process, and a more detailed deterministic dynamics, all of which might be consistent with each other, but at different levels of explanation (revealing different aspects of the underlying phenomenon). There is a large body of modeling lore concerning the types of mechanisms that can generate power laws, which we have collected together here. Certain themes emerge, such as hierarchy, competition between exponentials, growth, amplification, and long-range interaction. self-similarity, competition between exponent. These themes are potentially instructive about features of financial markets that are not addressed in mainstream contemporary models, which way have implications in other domains.

From our discussion of how power laws emerge from extreme value theory, it seems that the generation of power laws should not be a difficult task. Any process with sufficiently fat tails will generate a power law, so all we have to do is create large extremal values. However, it should be born in mind that some power laws are "purer" than others, i.e. some processes converge to a power law quickly, while others do so slowly. Furthermore, some processes, such as pure multiplicative processes (which have a log-normal as their solution) can mimic power laws for a range of values, and then fail to be power laws asymptotically. While this may be confusing, an examination of the underlying mechanisms for generating power laws makes it clear how this comes about.

The self-similarity associated with power laws is an important and potentially simplifying clue about model construction. For example, the apparent fact that price volatility scales as a power law on scales ranging from minutes to years suggests that the mechanism generating this scaling is the same across these scales. The alternative is that it is just a coincidence: there are different processes on different scales, that just happen to have the same scaling exponent. While possible, this seems unlikely, although of course *how unlikely* depends on the degree of accuracy to which the dynamics are self-similar.

The discussion presented here draws on the reviews of mechanisms for producing power laws by Mitzenmacher [49], as well as the books by Sornette

[61], and Mandelbrot [45].

1.4.1 Hierarchies and exponentials

A simple example that illustrates a common mechanism for producing a power law distribution was originally given by Simon [59]. Imagine a company whose organizational chart is a tree with k branches at each node of the tree. Furthermore, suppose that the salaries of the employees increase by a constant multiplicative factor $\gamma > 1$ at each node as we move up the tree. Thus, if employees at the bottom of the tree have salary s_0, moving up the tree the salaries are $\gamma s_0, \gamma^2 s_0, \ldots, \gamma^N s_0$, where N is the depth of the tree. If we label the management levels in the company from the bottom as $i = 0, 1, \ldots, N$, there are k^{N-i} employees with salary $\gamma^i s_0$. Plotting the logarithm of the number of people with a given salary against the logarithm of the salary gives a straight line with slope $-\log k/\log\gamma$, and the number of employees with salary s is $N(s) = s_0 s^{-\log k/\log\gamma}$. Providing $\log k > \log\gamma$, the cumulative distribution of incomes $N(s > S) \sim s^{\log k/log\gamma - 1}$, i.e. it is a power law with tail exponent[6] $\log k/\log\gamma$.

A geometric example which is essentially identical to the example above is the size of a Cantor set as a function of the resolution at which it is measured. Suppose we poke $k-1$ holes in the unit interval, in such a way that we divide it into k equal subintervals each of size $1/\gamma$. If we repeat this process for each remaining subinterval indefinitely the result is a Cantor set. At the i^{th} level of construction there are k^i subintervals of size γ^{-i}. As before, plotting the number of intervals against their size on double logarithmic scale gives a line of slope $\log k/log\gamma$; in this case, the constraint of geometry ensures that $\log k \geq \log\gamma$. Alternatively, we can measure the coarse-grained size of the Cantor set by dividing the interval into l equal increments, and using the rule that if an increment has any part of the Cantor set in it, it contributes $1/l$ to the total size $S(l)$. In the limit $l \to \infty$ the coarse-grained size scales as $S(l) \sim l^{-\log\gamma/\log k}$.

These two examples illustrate how power laws are often associated with hierarchies, and with the non-differentiable geometric objects called fractals, which have an intrinsic notion of hierarchy built into them. They also illustrate how power laws emerge naturally from the competition between exponentials. An exponential transformation of an exponentially distributed variable yields a power law. That is, suppose X and Y are random variables,

[6] Note that if $\log k < \log\gamma$ then the manager at each level makes more than all the employees immediately below her combined, and in the limit $N \to \infty$ almost all the income is paid to the CEO.

and X is exponentially distributed with $P(X > x) \sim e^{-ax}$, if $Y = e^{bX}$ then

$$P(Y > y) = P(e^{bX} > y) = P(X > \log y/b) = y^{-a/b} \tag{8}$$

In some sense this is a trivial statement, since for a power law distributed function we can always make a logarithmic transformation to coordinates where the power law becomes an exponential function. However, this connection is important to bear in mind since there are many mechanisms that generate exponential distributions, and there are many processes, e.g. growth and death, that naturally give rise to exponential transformations. In the example of the hierarchical firm, for instance, the power law comes from the competition between the exponential growth in the number of employees moving down the tree and the exponential increase in salary moving up the tree, and in the Cantor set example it comes from the competition between the exponential proliferation of intervals and the rate of their exponential decrease in size.

The St. Petersburg paradox provides another example of the competition between exponentials. Consider a fair coin-toss game in which heads doubles your money and tails loses it. Your strategy is to let your winnings ride for up to n tosses. If you are lucky enough to get heads n times in a row your sequence of bets is $1, 2, 4, \ldots, 2^{n-1}$. With probability $1/2^n$ you win 2^n and with probability $1 - 1/2^n$ you lose all your money. This is an even bet – on average you win what you originally wagered. But it does have a very broad distribution of possible outcomes – for large n you almost certainly lose, but the tail on the winning side is nonetheless very fat, in fact it is a power law tail with exponent one. Because of this, if you decide to let n be arbitrarily large and play until you either break the bank or go broke, the house should indeed be afraid of you. To see this, note that the probability of winning 2^n or more is $1/2^n$, i.e. the probability of gaining g or more is $1/g$. In order to prevent their probability of ruin from being unacceptably high, it is essential that the casino have a maximimum bet size to truncate the power law. With enough customers playing simulataneously, due to the cutoff, the sum converges to a Gaussian distribution.

As in the previous examples, the power law comes about due to the competition between the exponentially decreasing probability of being eliminated and the exponentially increasing payoff if not eliminated. In this case the exponential rate of increase is equal to the rate of decrease, and so the exponent of the power law is one.

1.4.2 Maximization principles

One way to derive power laws is in terms of maximization of an appropriate function, possibly under constraints. The function that is maximized can be an objective function, such as information transmission or utility, or it can be the entropy. Maximizing the entropy amounts to assuming that something is as random as it can be subject to the constraints. This is the basic assumption underlying statistical mechanics. The exponential or Gibbs distribution, for example, is the solution that emerges from maximizing the entropy subject to the constraint that the mean takes on a fixed value. So, for example, in a physical system where energy is conserved, absent any other constraints the distribution of energies will be an exponential function. Similarly, if one imposes a constraint on the variance as well as the mean, the solution is a normal distribution.

A power law emerges from maximizing the entropy when there is a constraint on mean of the logarithm. This can be demonstrated via the method of Lagrange multipliers. We are seeking the probability distribution $p(x)$ with $x > 0$ that maximizes the entropy $\int p(x) \log p(x) dx$ subject to the constraint that $\int (\log x)^\alpha p(x) dx = C$, where C and α are constants. Constructing the Lagrangian and setting the functional derivative with respect to the probability distribution to zero gives

$$\frac{\partial}{\partial p(x)}[\int p(x) \log p(x) dx + \lambda \int ((\log x)^\alpha - C) p(x) dx] = 0,$$

where λ is the undetermined Lagrange multiplier. This has the solution $p(x) = Ax^{-\lambda}$, where A is a normalization constant[7]. Of course, to use this as an argument to explain a power law, one must have a plausible argument for why the logarithm of a variable should be constrained.

For standard problems in statistical mechanics the entropy is an extensive quantity. This means that as the volume of a system varies, the entropy increases proportional to volume. For this to be true it is necessary that different regions of the system be independent, so that the probability of a state in one region is independent of the probability of a state in another region. This is true for any system in equilibrium (in the physics sense). Physical systems with short range interactions come to equilibrium quickly, and there are many circumstances where extensivity is a good assumption.

[7] As already mentioned, due to normalization problems a power law cannot be defined over the interval $[0, \infty]$. For $a > 0$ the above derivation assumes that either $\alpha > 1$ and the domain is $[a, \infty]$, or $\alpha < 1$ and the domain is $[0, a]$, or $\alpha = 1$ and the domain is $[a, b]$, where $b > a$.

There are some systems, however, with very long-range interactions, that are very slow to come to equilibrium due to the fact that distant parts of the system continue to interact for a long time and so it is not possible to assume independence. Thus, extensivity is a good assumption for a hard sphere gas, where the particles interact only when they collide. The system comes to equilibrium quickly, and the energies of the particles have an exponential distribution. This is not a good assumption for particles (such as stars) interacting under the influence of gravitational forces, which have a very long range. Under long range interactions the approach to equilibrium may be so slow that it fails for any practical purpose. Thus in simulations of galaxy formation, the distribution of the energy of stars does not approach an exponential distribution, but rather has a more complicated distribution with a power law tail.

A heuristic method of dealing with this problem, which seems to work very well in many cases, is to introduce a nonextensive entropy function. The most successful of these is the Tsallis entropy

$$S_q = \frac{1 - \int p(x)^q dx}{q - 1}. \tag{9}$$

where $p(x)$ is the probability density at x and q is a positive integer that depends on factors such as how long range the interactions are and how far from equilibrium the system is. When $q > 1$, raising the probability density to the power q gives more weight to high probability regions and less weight to improbable regions, and *vice versa* when $q < 1$. In the limit $q \to 1$ this reduces to the standard entropy.

In the same vein as the maximum entropy calculation above, we can also optimize the Tsallis entropy. If we constrain the mean (which is natural e.g. in the case where x represents energy), then the Lagrangian is

$$\frac{\partial}{\partial p(x)}[\frac{\int (p(x) - p(x)^q)dx}{q-1} + \lambda \int xp(x)dx - C] = 0.$$

This has the solution

$$p(x) = A(1 - (1-q)\lambda x)^{-\frac{1}{1-q}} \tag{10}$$

where A is a normalization constant. In the limit $q \to 1$ this reduces to an exponential distribution, but otherwise it is a power law distribution with a tail exponent $\alpha = 1/(1-q) - 1 = q/(1-q)$. This has been shown to give a remarkably good fit to many situations, such as the distribution of the energies of stars in a simulation of galaxy formation and the number of transactions in a given length of time in a financial market. In a similar manner to the above calculation, by constraining the variance as well as the mean it is also possible to derive a power law generalization of the normal distribution. This gives a good fit to the distribution of price returns.

A mechanism for generating power laws which is in the spirit of the maximum entropy principle, though without using it explicitly, is the explanation of Zipf's law of word frequencies based on monkeys typing randomly, due to Miller [48]. Zipf's law states that the frequency of word usage as a function of rank forms a power law with slope minus one [19, 66]. In this case rank is the ordering of the frequencies of word usage, i.e. the fifth most used word has rank five. This explanation is as follows: Suppose that monkeys type randomly on a keyboard with M characters plus a space. Assume they hit the space bar with probability p, and the non-space characters with probability $(1-p)/M$. Then the probability that they will type a particular word of length l is the probability that they hit l characters followed by a space, i.e.

$$p(l) = (\frac{1-p}{M})^l p = pe^{l \log \frac{1-p}{M}} \tag{11}$$

There are M^l words of length l, and there are $M + M^2 + \cdots + M^{l-1}$ (more common) words of shorter length. About half of the words of the same length are likely to have greater frequencies (we can assume either sample fluctuations or slight variations in the probability of characters), so that to compute the rank of a typical word of length l we should also add another term of size $M^l/2$. For M large it is a good approximation to take $M+M^2+\cdots+M^{l-1}+M^l/2 \approx M^l/2$. Thus the rank of a typical word of length l is the order of $r = M^l$. Putting this together with equation 11 and eliminating l gives

$$p(l) = pr^{-\frac{1}{\alpha}},$$

where

$$\frac{1}{\alpha} = 1 - \log_M(1-p).$$

(We have written the exponent as $1/\alpha$ because the exponents of rank ordering relations are the reciprocal of the tail exponent.) Note that this can once again be viewed as a competition between two exponentials: Both the word frequency and the rank decrease exponentially with length.

While this may seem like an obvious explanation of word frequency, it is worth noting that the original explanation due to Mandelbrot used a maximization principle [41]. He assumed that languages are designed to be efficient to transmit, in the sense of cost per bit of information transmitted, under the assumption that the cost is proportional to the logarithm of word length. This also gives a power law. Interestingly, as we will discuss later, Simon has also offered quite a different explanation based on preferential attachment. In his 1997 book, Mandelbrot supports the random monkey explanation [45]. While all three of these explanations give a power law, it is not obvious which is correct. This reflects a common problem that needs more attention in the literature: The existence of a power law *per se* is typically not sufficient to distinguish competing theories; there are many ways to produce power laws, and more testing based on the details of the mechanism is required to determine which is correct.

There has recently been a revival of maximization principles to explain power laws by Carlsen and Doyle, via a mechanism they have dubbed Highly Optimized Tolerance (HOT) [10, 11]. Using an argument that is similar to Mandelbrot's orginal derivation of Zipf's law for word frequencies, they have proposed that the power law distribution of file sizes on the internet is a side-effect of maximizing efficiency of storage. In another example they consider an idealized model of forest fires [11]. In this model a forester is charged with finding the optimal distribution of trees on a grid so as to maximize tree harvest in the face of occasional fires that burn complete connected clusters of trees and are started by sparks that arrive with a given spatial distribution. They find that optimizing the harvest, or yield, for the model gives rise to a segmented forest consisting of contiguous patches of trees separated by firebreaks, and that the resulting distribution of fire sizes usually follows a power law. While this type of configuration typically achieves good yields, the system is also fragile in the sense that perturbations to the firebreaks or changes in the spark distribution can lead to disastrously sub-optimal performance (due to the power law tail for the distribution of large fires). They argue that these are pervasive phenomena: high-performance engineering leads to systems that are robust to stresses for which they were designed but fragile to errors or unforeseen events. The power law in the forest model derives from the geometry of the constraints, i.e. from the geometrical constraints on constructing firebreaks in two dimensions.

Although not cast in a traditional economic framework, the power law in the HOT forest fire model comes from maximization of utilty. Thus, it connects to one of the central precepts in economics, and thus seems well suited to potentially explain power laws in economics within the mainstream canon. However, as was pointed out in [51], if one puts risk aversion into the utility function, the power laws disappear. (This approach was jokingly called Constrained Optimization with Limited Deviations or “COLD”. This seems to cast doubt whether this approach can explain power laws, since it is at least widely believed that humans display some level of risk aversion. While maximization principles offer an intriguing possibility to explain the pervasive nature of power laws in economics, the details of how this would be done, and whether or not it is economically plausible, remains to be investigated.

1.4.3 Multiplicative processes

Multiplicative processes generate fat tailed distributions, and since multiplicative processes occur for a broad class of nonlinear dynamics, e.g. feedback effects, they repesent a natural candidate for causing power laws. A pure multiplicative process gives a log-normal distribution, which is fat tailed but is not a power law, but small modifications of this process, such as the inclusion of a barrier or an additive term, give rise to true power law distributions. Thus, log-normal and power law distributions are closely related.

Consider a simple multiplicative process of the form

$$x(t+1) = a(t)x(t) \tag{12}$$

where $x(0) > 0$ and $a(t) > 0$. This is a sensible model for growth or fracture. If we iterate the process its solution is trivially written

$$x(t) = \prod_{i=0}^{t-1} a(i)x(0). \tag{13}$$

If we take logarithms this becomes

$$\log x(t) = \sum_{i=0}^{t-1} \log a(i) + \log x(0). \tag{14}$$

Providing the second moment of $\log a(i)$ exists and the $a(i)$ are sufficiently independent of each other, in the large time limit $\log x(t)$ will approach a normal distribution, i.e. $x(t)$ will approach a log-normal distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\log x - \mu)^2/2\sigma^2}. \tag{15}$$

μ and σ^2 are the mean and variance of the associated normal process. Taking logarithms this becomes

$$\log f(x) = -\frac{(\log x)^2}{2\sigma^2} + (\frac{\mu}{\sigma^2} - 1)\log x + \text{constant terms}. \tag{16}$$

In the limit $x \to \infty$ the quadratic term dominates, so this distribution is of Gumbel type – it does not have a power law tail. However, if the variance is sufficiently large, then the coefficient of the quadratic term is small while the coefficient of the linear term is of order one. Thus the lognormal distribution will exhibit approximate power law scaling, and may do so over many decades if σ is sufficiently large.

Note that in general a pure multiplicative process requires some normalization of scale for its log-normality to become apparent. This is of course already true for an additive random walk, but because of the exponential the problem is much more severe. If $E[\log a(t)] < 0$ then the process exponentially collapses to the origin, and if $E[\log a(t)] > 0$ it expands exponentially. Thus, at first approximation the asymptotic distribution appears to be either a spike at the origin, or to blow up. To view the lognormal fluctuations, one must use an appropriately contracting or expanding scale, which depends on the number of iterations of the process.

The pure multiplicative process can be turned into a power law by simply imposing a barrier to repel $x(t)$ away from zero. Providing $E[\log a(t)] < 0$ the process is stable and so will tend to be attracted to the origin. The emergence of the power law can once again be easily understood by taking logarithms. For the pure multiplicative process the asymptotic distribution is normally distributed. Providing $E[\log a] < 1$, it will tend to drift to the left. However, in the presence of a barrier it will pile up against the barrier, and the normal distribution for a random walk will be replaced by an exponential probability density of the form $P(x) = \mu e^{-\mu x}$. In general the exponent μ depends on the details of the random process. Undoing the logarithmic transformation gives a power law with tail exponent $\alpha = \mu + 1$.

Another small modification that results in a power law is the addition of an additive term

$$x(t+1) = a(t)x(t) + b(t), \tag{17}$$

where $b(t) > 0$. This is called a Kesten process [29]. It is power law distributed providing $E[\log a] < 1$ and there are values of t with $a(t) > 1$. Intuitively, the first condition ensures that the process is attracted to the origin. The inclusion of the additive term makes sure that the process does not simply collapse to the origin, and and the condition that occasionally $a(t) > 1$ creates intermittent bursts that form the fat tail. Thus we see that this is closely related to the pure multiplicative process with a barrier. The tail exponent of the Kesten process depends on the relative sizes of the additive and multiplicative terms. Processes of this type are very common, describing for example random walks in random environments, a model of cultural evolution, and a simple stochastic model for the distribution of wealth. The Kesten process is nothing but a discrete time special case of the Langevin equation, which is a widely used model in statistical physics. In fact, Tsallis and XXX have shown that under fairly broad conditions, Langevin equations (in continuous time) give equation 10 as a solution.

1.4.4 Mixtures of distributions

It is well known that mixtures of non-fat-tailed distributions can be fat tailed. A mixture of distributions mixes together distributions with different scale parameters, i.e.

$$f(x) = \int g(\sigma) p_\sigma(x) d\sigma \quad (18)$$

where σ is the scale parameter of the distribution $p_\sigma(x)$. This is often offered as the explanation for fat tails in prices: Since the information arrival rate varies, the standard deviation of price fluctuations varies. Thus even though the instantaneous distribution might be a thin-tailed normal distribution, when distributions of many standard deviations are blended together, the result is a fat-tailed distribution. Therefore, according to this explanation the fat tails of prices come entirely from non-uniformity in information arrival, creating a mixture of different volatilities in price changes.

This explanation misses the mark in several ways. First, as mentioned already, there is good evidence that other factors are more important than information arrival in determining the volatility of prices. But in addition, it is incomplete; while any mixture will fatten the tails, not all mixtures do so sufficiently to create a power law. In general the condition that a mixture function $g(\sigma)$ generates a particular target function $f(x)$ is quite restrictive. (And of course one must explain why this mixture function $g(x)$ takes on the particular form required).

For instance, we can ask what mixture function can combine exponential distributions to get a power law?

$$f(x) = \int g(\sigma) e^{-\sigma x} d\sigma \quad (19)$$

It is possible to show that the function $g(\sigma)$ that will give a power law with tail exponent α is $g(\sigma) \sim 1/\sigma^{2+\alpha}$. Thus, to get a power law by combining exponentials it is necessary for the mixture function to be itself a power law. Sornette has shown that this result applies to any function with tails that dies out sufficiently fast [61]. Thus this result also applies to mixtures of normal distributions, and indicates that a power law mixture is required.

Thus, to explain the power law nature of price fluctuations in terms of variation in rates of information arrival, one needs to explain why information arrival has a power law distribution.

One can ask whether there are non-power law mixtures of thin-tailed distributions that give rise to power laws? The answer is yes. An important example is provided by an exponential mixture of log-normal distributions. This occurs naturally in the context of a multiplicative process with a distribution of stopping times, i.e. consider the process $x(i+1) = a(i)x(i)$, but now assume that the stopping time t is exponentially distributed, $p(t) = \lambda e^{-\lambda t}$. For fixed stopping time the distribution is lognormally distributed, but for exponential stopping time the result is an exponentially weighted mixture of lognormals, and is a power law, called the double Pareto distribution [58]. This name is appropriate because this distribution actually has a power law tail in both limits $x \to 0$ and $x \to \infty$, though with different tail exponents (which solves the normalization problem). The exponents depend on the parameters of the multiplicative process, as well as the scale of the stopping time[8]. With unequal frequencies of the non-space characters, the example given earlier of monkeys typing randomly provides a good illustration: While the frequency of words of fixed length is lognormally distributed, the length of words is exponentially distributed, so that the result is a power law. Other proposed applications include the distribution income, number of pages in web sites, size of human settlements with indeterminate growth times, particle size under an interminate number of fractures, and oil field size.

1.4.5 Preferential attachment

Preferential attachment was originally introducted by Yule to explain the distribution of species within genera of plants, is perhaps the oldest known mechanism for generating a power law. The basic idea is that mutuations are proportional to the number of species, so a genus with more species has more mutuations and thus grows at a faster rate. The argument was developed by Simon and proposed as a possible explanation for a variety of other phenomena, including the distribution of word frequencies,the distribution of numbers of papers that a scientist publishes, the distribution of city sizes, and the distribution of incomes.

We will summarize the basic argument in the context of word frequencies. Consider a partially completed text containing t different words. Assume

[8]The tail exponents are the roots α and $-\beta$ of the equation $\sigma^2 z^2 + (2\mu - \sigma^2)z - 2\lambda = 0$, where $\alpha, \beta > 0$. The tail at zero scales as x^β, and the tail at infinity as $x^{-\alpha}$.

that with probability λ an author chooses the next word randomly from the dictionary and with probability $1-\lambda$ she chooses a previously used word, with probability proportional to the previous number of occurrences of the word. Let N_j be the number of different words that occur exactly j times. For $j \geq 1$, the probability of choosing a word that occurs N_j times is

$$\lambda N_{j-1}/t + (1-\lambda)(j-1)N_{j-1}/t.$$

The first term is the probability of choosing a new word at random, and the second is the probability of choosing a word that is already in the text. The probability of choosing a word that already occurs N_j times (and therefore decreasing N_j due to a word being added to N_{j+1}) is the same expression, but substituting j for $j-1$. Therefore the expected rate of growth of N_j is

$$E[N_j(t+1) - N_j(t)] = \frac{\lambda(N_{j-1} - N_j) + (1-\lambda)((j-1)N_{j-1} - jN_j)}{t}. \quad (20)$$

Suppose we now make the steady state assumption that for large t the word frequencies converge to constant ratios r_j, so that the number of occurrences of each word grows as $N_j(t) = r_j t$. This implies that $E[N_j(t+1) - N_j(t)] = r_j$. With some rearranging of terms, equation 20 becomes

$$\frac{r_j}{r_{j-1}} = \frac{(\lambda + (1-\lambda)(j-1))}{(1+\lambda+j(1-\lambda))}.$$

If we assume that j is large and expand the denominator to first order (neglecting terms of size $1/j^2$ and smaller), this can be approximated as

$$\frac{r_j}{r_{j-1}} \approx 1 - \frac{(2-\lambda)}{(1-\lambda)}(\frac{1}{j}).$$

This has the solution $r_j = r_0 j^{-(2-\lambda)/(1-\lambda)}$, which is a power law with tail exponent $\alpha = (2-\lambda)/(1-\lambda) - 1 = 1/(1-\lambda)$.

1.4.6 Dimensional constraints

There are many cases where dimensional constraints, such as the geometry of space, dictate the existence of power laws. This can be understood in terms of dimensional analysis, which is based on the principle that scientific laws should not depend on arbitrariness that is inherent in the choice of units of measurement. It shouldn't matter whether we measure lengths in meters or yards – while changing units will affect the measurement of any quantity that is based on length, this dependence is trivial, and anything that doesn't depend on length should remain the same. The basic form of of a physical law does not depend on the units. While this may seem like a trivial statement, in fact it places important restrictions on the space of possible solutions and can sometimes be used to get correct answers to problems without going through the effort of deriving a solution from first principles. Although dimensional analysis has normally used in engineering and the physical sciences, recent work has shown that dimensional analysis can also be useful in economics [14, 60, 22]. Since dimensional analysis is essentially a technique exploiting scale invariance, it is not surprising that dimensional constraints naturally give power laws.

The connection between power laws and the constraints of dimensionality can be derived from the requirement that their is no distinguished systems of units, i.e. that there is no special unit of measure that is intrinsically superior to any other [5]. Assume that we choose a system of fundamental quantities, such as length, mass and time in physics, such that by using combinations of them they are sufficient to describe any quantity ϕ that we wish to measure. We can now consider how ϕ will change if we use units that differ by factors of L, M or T from the original units. The *dimension function* $[\phi]$, which is traditionally denoted by brackets, gives the factor by which ϕ will change. For example, for the velocity v the dimension function $[v] = L/T$.

The reason that power laws emerge naturally from dimensional constraints is because the dimension function is always a power law monomial. To see why, suppose there is a quantity that has a value a_0 in an original system of units. Now compare its values in two other systems of units differing by factors (L_1, M_1, T_1) and (L_2, M_2, T_2), where it takes on values $a_1 = a_0\phi(L_1, M_1, T_1)$ and $a_2 = \phi(L_1, M_1, T_1)$. Thus

$$\frac{a_1}{a_2} = \frac{\phi(L_1, M_1, T_1)}{\phi(L_2, M_2, T_2)}.$$

Since no system of units is preferred, we can equivalently assume that system 1 is the original system of units, in which case it is also true that

$$a_2 = a_1\phi(L_2/L_1, M_2/M_1, T_2/T_1).$$

Combining these two equations gives the functional equation

$$\frac{\phi(L_1, M_1, T_1)}{\phi(L_2, M_2, T_2)} = \phi(L_2/L_1, M_2/M_1, T_2/T_1).$$

Assuming that ϕ is differentiable it is possible to show that the only possible solutions are of the form

$$\phi = L^{\alpha} M^{\beta} T^{\gamma}$$

where α, β and γ are constants. That this is not obvious can be demonstrated by assuming that there is a preferred system of units, which leads to an functional equation that does not have power law monomials as its solution.

This relationship has important consequences in generating power laws, as becomes evident from the fundamental theorem of dimensional analysis, called the Π theorem. Consider some quantity a that is a function of n parameters. A set of parameters $(a_1, \ldots, a_k)$ are said to have independent dimensions if none of them has dimensions that can be represented in terms of a product of powers of the dimensions of the others. It is always possible to write a function of n parameters in the form

$$a = f(a_1, \ldots, a_k, a_{k+1}, \ldots, a_n),$$

where the first k parameters have independent dimensions, and the dimensions of parameters $a_{k+1}, \ldots, a_n$ can be expressed as products of the dimensions of the parameters $a_1, \ldots, a_k$, and $0 \leq k \leq n$. Then, by making a series of transformations to dimensionless parameters, it can be shown that this can generally be rewritten in the form

$$f(a_1, \ldots, a_n) = a_1^p \cdots a_k^r \Phi\left(\frac{a_{k+1}}{a_1^{p_{k+1}} \cdots a_n^{r_{k+1}}}, \ldots, \frac{a_n}{a_1^{p_n} \cdots a_k^{r_n}}\right).$$

The sequence of positive constants $p, \ldots, r$ of length k is chosen in order to make the product $a_1^p \cdots a_k^r$ have the same dimensionality as f, and the sequences of positive constants $\{p_{k+1}, \ldots, r_{k+1}\}, \ldots, \{p_n, \ldots, r_n\}$, which are also each of length k, are chosen to make the transformed parameters $a_{k+1}/(a_1^{p_{k+1}} \cdots a_n^{r_{k+1}}), \ldots, a_n/(a_1^{p_n} \cdots a_k^{r_n})$ dimensionless.

This relation demonstrates that any quantity that describes a scientific law expressing relations between measurable quantities possess the property of generalized homogeneity. The product $a_1^p \cdots a_k^r$ trivially reflects the dimensionality of f, and Φ is a dimensionless function that contains all the nontrivial behavior. If we move the product $a_1^p \cdots a_k^r$ to the left hand side of the equation, then it makes it clear that the effect has been to transform a dimensional relationship into a dimensionless relationship, confirming by construction our initial requirement that sensible scientific laws should not depend on arbitrariness in the choicd of units. This representation also reduces the dimensionality and hence the complexity of the solution. In the best circumstances $k = n$ and Φ is a constant. More typically $k < n$, but this is still extremely useful, since it reduces the dimensionality of the solution from n to $n - k$. For a problem such as fluid flow in a pipe, where $n = 4$ and $k = 3$, this can dramatically simplify analysis.

The product $a_1^p \cdots a_k^r$ is a power law in each of the variables a_i. If Φ is a slowly varying function of all of its arguments in one or both of its limits then this gives a power law in each of the variables $a_1 \ldots a_k$. This happens, for example, when all the variables have independent dimensions ($k = n$) and thus $\Phi =$ constant. Of course, it is also possible that Φ is not a slowly varying function, in which case the power law behavior will be broken (e.g. if Φ is an exponential) or modified (if Φ is a power law).

The power laws that are generated by dimensional constraints of simple geometric quantities typically have exponents $p, \ldots, r$ that are integers or ratios of small integers. This is because the quantities we are interested in are usually constructed from the fundamental units in simple ways, e.g. because quantities like volume or area, are integer powers of the fundamental unit of distance. However, for problems with more complicated geometry, e.g. fractals, the powers can be more complex. For example, recent work has shown that the 3/4 power that underlies the scaling of metabolic rate vs. body mass can be explained in terms of the hierarchical fractal geometry of the cardiovascular system [63, 64].

Although dimensional analysis is widely used in the physical sciences and engineering, economists have typically never heard of it. Recently, however, it has been shown to be useful for financial economics in the context of the continuous double auction, for understanding the bid-ask spread or the volatility as a function of order flow [14, 60]. For this problem the fundamental dimensional quantities were taken to be price, shares, and time, with corresponding scaling factors P, S, and T. There are five parameters in the model, three of which have independent dimensions. The three order flow parameters are market order rate μ, with $[\mu] = S/T$, limit order rate α, with $[\alpha] = S/(PT)$, and order cancellation rateδ, with $[\delta] = 1/T$. The two discreteness parameters are the typical order size σ and the tick size Δp. Quantities of interest include the bid-ask spread s, defined as the difference between the best selling price and the best buying price, and the price diffusion rate, defined as the diffusion rate for the random walk underlying prices, which is the driver of volatility. The bid-ask spread s, for example, has dimensions of price. As a result, by expressing the dimensional scaling in terms of the three order flow parameters and applying the Π theorem the average value of the spread can be written in the form

$$E[s] = \frac{\mu}{\alpha}\Phi_s(\frac{\sigma}{p_c}, \frac{\Delta p}{N_c}),$$

where $p_c = \mu/\alpha$ is the unique characteristic price scale that can be constructed from the three order flow parameters and $N_c = \mu/\delta$ is the unique characteristic quantity of shares. The use of dimensional analysis thus reduces the number of free parameters from five to two, and makes the arguments of Φ_s nondimensional. Through more complicated analysis and simulation it can be shown that Φ_s depends more strongly on σ/p_c than $\Delta p/N_c$, and that in the limit $\Delta p/N_c \to 0$ and $\sigma/p_c \to 0$ it approaches a constant. Thus, in this limit the spread is described by a power law, albeit a simple one.

Similarly for the price diffusion rate D, which has dimensions $[D] = P^2/T$, can be written as

$$D = \frac{\mu^2\delta}{\alpha^2}\Phi_D(\frac{\sigma}{p_c}, \frac{\Delta p}{N_c}).$$

In this case, through simulation it is possible to demonstrate that Φ_D also depends more strongly on σ/p_c than $\Delta p/N_c$. In the limit $p/N_c \to 0$ and $t \to 0$ (describing price diffusion on short time scales), Φ_D is a power law of the form $\Phi_D = (\sigma/p_c)^{-1/2}$. As a result, in this limit the diffusion rate is a power law function of its arguments, of the form $\Phi_D \sim \mu^{5/2}\delta^{1/2}/(\alpha^2\sigma^{1/2})$.

These relations have been tested on data from the London Stock Exchange and shown to be in remarkably good agreement [21]. This demonstrates that dimensional analysis is useful in economics, demonstrates how some power laws might be explained in economics, and perhaps more importantly, shows the power of new approaches to economic modeling. Note, though, that this does not explain the power law tails of prices, which seems to be a more complicated phenomenon [24, 21, 23].

1.4.7 Critical points and deterministic dynamics

The dynamical mechanisms for producing power laws that we have discussed so far are stochastic processes, in which noise is supplied by an external source and then amplified and filtered, e.g. by a simple multiplicative process or a growth process such as preferential attachment. Under appropriate conditions it is also possible to generate power laws from deterministic dynamics. This occurs when the dynamics has a critical point. This can happen at a bifurcation, in which case the power law occurs only for the special parameter values corresponding to the bifurcation. But there are also more robust mechanisms such as self-organized criticality, which keep a system close to a critical point for a range of parameters. Critical points can amplify noise provided by an external source, but the amplification is potentially infinite, so that even an infinitesimal noise source is amplified to macroscopic proportions. In this case the properties of the resulting noise are independent of the noise source, and are purely properties of the dynamics.

Critical points occur at the boundary between qualitatively different types of behavior. In the classic examples in physics critical points occur at the transition between two states of matter, such as the transition from a solid to a liquid or a liquid to a gas. Critical points also occur more generally in dynamical systems where there is a transition from locally stable to locally unstable motion, such as the transition from a fixed pont to a limit cycle or a limit cycle to chaos. To see why critical points give rise to power laws, consider a nonlinear dynamical system of the form $dx/dt = F(x, c)$, where c is a control parameter that continuously changes the functional form of a smooth nonlinear function F. Suppose that for some parameter interval there is a stable fixed point $F(x_0) = 0$, which is an attractor of the dynamics. For small perturbations of the solution near the fixed point we can get a good approximate solution by expanding F in a Taylor's series around x_0 and neglecting everything except the leading linear term. This gives a solution which in one dimension[9] is of the form $x(t) = ae^{\lambda t}$. As long as $\lambda \neq 0$, the linear solution is the leading order solution, and will provide a reasonable approximation in the neighborhood of x_0. However, suppose c is varied to a critical value c_0 where the dynamics are no longer linearly stable. In this case the linear approximation to $F(x)$ vanishes, so that it is no longer the leading order term in the Taylor approximation of $F(x)$. To study the stability of the dynamics at this point we are forced to go to higher order, in which case the leading order approximation to the dynamics is generically of the form $dx/dt = \alpha x^\beta$, where $\beta > 1$. This has a solution of the form

$$x(t) = At^{1/(\beta-1)}. \tag{21}$$

Thus, whereas when the system is either stable or unstable the leading order solution is an exponential, at the critical point the leading order solution is a power law. This is the underlying reason why critical points play an

[9] We are being somewhat inconsistent by assuming one dimension, since chaotic behavior in a continuous system requires at least three dimensions. The same basic discussion applies in higher dimensions by writing the solutions in matrix form and replacing λ by the leading eigenvalue.

important role in generating power laws.

An important special property of the critical point is the lack of a characteristic timescale. This is in contrast to the stable or unstable case, where the linearized solution is $x(t) = ae^{\lambda t}$. Since the argument of an exponential function has to be dimensionless, λ necessarily has dimensions of $1/(time)$, and $1/|\lambda|$ can be regarded as the characteristic timescale of the instability. For the critical point solution, in contrast, the exponent is $1/(\beta - 1)$, and β is dimensionless. The solution $x(t) = At^{1/(\beta-1)}$ is a power law, with no characteristic timescale associated with the solution.

One of the ways power laws manifest themselves at critical points is in terms of intermittency. This was demonstrated by Pomeau and Manneville [55], who showed how at a critical point a dynamical system could display bursts of chaotic behavior, punctuated by periods of laminar (nearly periodic) behavior of indeterminant length. This can be simply illustrated with the deterministic mapping

$$x_{t+1} = (1 + \epsilon)x_t + (1 - \epsilon)x_t^2 \pmod 1$$

For $epsilon > 0$ this map displays chaotic behavior. However, near $x_t = 0$ the quadratic term is small, and so $x_{t+1} \approx (1 + \epsilon)x_t$. When ϵ is small, it is also the case that $x_{t+1} \approx x_t$. Thus, starting from an initial condition close to the origin, subsequent iterations of the map change very slowly, and may spend many iterations almost without changing. This is called the laminar phase. The length of time the laminar phase persists depends on the value of ϵ, and also on how close the initial condition is to zero. When x_t finally gets far enough away from the origin it experience a burst of chaotic behavior, but eventually (as if by chance) a new value close to zero will be generated, and there is another laminar phase. When $\epsilon = 0$ Manneville showed that the length τ of the laminar phase are distributed as a power law of the form $P(\tau) \sim 1/\tau$. As a consequence of this, the power spectrum $S(f)$ (the average of the square of the absolute value of the Fourier transform of x_t) behaves in the limit $f \to 0$ $S(f) \sim 1/f$, where f is the frequency of the Fourier transform. Such power law behavior occurs for a bifurcation of any dynamical system in which the eigenvalue becomes positive by moving along the real axis.

Critical points thus provide a mechanism for generating power law behavior in a dynamical system, but this mechanism is limited by the fact that it pertains only near bifurcations. Bifurcations typically occur only at isolated points in parameter space, and form a set of measure zero. A set of parameters drawn at random is unlikely to yield a critical point, and variations of the parameters will typically the power law associated with the critical point disappear. Thus, in order to explain power laws in terms of critical points, it is necessary to find mechanisms that make the critical point robust, i.e. that maintain it at through a wide range of parameter values, at least as an approximation.

One example of this is due to spatio-temporal intermittency, and was discovered by Keeler and Farmer [28]. In the system of coupled maps that they studied the dynamics organizes itself into regions of high frequency chaotic behavior and regions of low frequency laminar behavior, like the laminar and chaotic regions in Pomeau-Manneville intermittency, except that they coexist at the same time, but at different points in space – it is as though there were a smoothly varying "local" parameter determining the dynamics in each region, with small variations of the value of that parameter around the critical point. The fronts separating these regions move, but their motion is extremely slow. As a result, there is an eigenvalue associated with the motion of these fronts that is very near zero. This behavior persists across a wide range of parameter values. As a result, the system has a robust power law, with a power spectrum that behaves as $1/f$ for frequencies f near zero. Such behavior is also observed in many situations in fluids near the transition to turbulence.

Another mechanism for making fixed points robust, called *self-organized criticality*, was introduced by Bak, Tang, and Weisenfeld [4]. Basic idea is that some phenomena, by their very nature maintain themselves near a critical point The classic example is a sandpile. Consider a thin stream of sand falling vertically, for example in an hourglass. A sandpile will build up underneath, and its sides will steepen until it becomes too steep, and then there is an avalanche. It will then steepen again until there is another avalance, and so on. The sandpile maintains itself near a critical state, through dynamics that are inherent to the physical constraints of the situation. Bak, Tang and Weisenfeld build a deterministic model of the sandpile in terms of a cellular automaton, and showed that it displayed approximate power law tails. Though this was later shown to not be a true power law, more detailed models of the sandpile show true power laws, and models of power law behavior in many other systems have been found based on this mechanism.

The suggestion has been made that arbitrage efficiency may be a self-organising critical mechanism. The basic idea is that arbitrageurs tend to drive a financial economy to an efficient state. However, once it gets too close to efficiency, profits become very low, and in the presence of negative fluctuations there can be avalanches of losses driving many arbitrageurs out of business. After an avalanche, arbitraguers re-enter the market and once again more the market toward efficiency. Under this theory the power laws are thus explained as fluctuations around the point of market efficiency. We will describe such a scenario in more detail in Section **??**.

One of the reason that physicists find power laws associated with critical points particularly interesting is because of *universality*. There are many situations, both in dynamical systems theory and in statistical mechanics, in which many of the properties of the dynamics around critical points are independent of the details of the underlying dynamical system. For example, bifurcations can be organized into groups, and the exponent β at the critical point in equation 21 may be the same for many systems in the same group, even though many other aspects of the system are different. One consequence of this is that the tail exponents of the associated power laws take on a value that is the same for many different dynamical systems. It has been suggested, for example, that the exponent of price fluctuations may have a tail exponent near three [**?**]. However, more detailed studies seem to suggest that there are statistically significant variations in the tail exponents of different assets [23].

1.4.8 "Trivial" mechanisms

We should not conclude our review of mechanisms for generating power laws without mentioning a few "trivial" ways to make power laws. These mechanisms are obvious (e.g. transforming by a power law) or inadequate (e.g.

One obvious way to make a power law is through a power law transformation. Suppose, for example, that x is a variable with a density function $p_x(x)$ that approaches a constant in the limit $x \to 0$, i.e. $\lim_{x\to 0} p_x(x) = K$. Let y be a power law transformation of x, of the form $y = f(x) = x^{-\beta}$. Then under conservation of probability, $p_x(x)dx = p_y(y)dy$,

$$p_y(y) = \frac{p_x(f^{-1}(y))}{dy/dx} = \frac{p_x(y^{-1/\beta})}{\beta y^{1+1/\beta}} \approx \frac{K}{\beta y^{1+1/\beta}}.$$

This is a power law with tail exponent $\alpha = 1/\beta$. Note that a little algebra shows that in the case where $p(x)$ is a power law this is consistent with the transformation rule for tail exponents given in equation 5.

It is not a surprise that a power law transformation can create a power distributed variable, and for this reason we have labeled it as "trivial". At the same time, this mechanism generates power laws in many different physical problems, and cannot forgotten. The existence of a power law transformation is not always obvious; a good example is Student's t distribution with n degrees of freedom, which is a power law with tail exponent $\alpha = n$ [61].

As already discussed in Section 1.1, sums of random variables converge to the Levy stable distribution, which is a power law, when the second moments of the variables fail to exist. This is often given as a mechanism for generating power laws. However, this mechanism doesn't really generate a power law, since the fact that the second moment does not exist implies that the tail exponent of the random variable being combined already has a tail exponent $0 < \alpha < 2$. Thus, by definition it is already a power law, with a tail exponent equal to that of the Levy distribution.

Another simple mechanism for making a power law is the ability of a dynamical system to act as a low pass noise filter with a power law cutoff. Consider a dynamical system with added noise, of the form

$$\frac{dx}{dt} = f(x(t)) + n(t)$$

where f is a smooth function and $n(t)$ is a white noise process. Suppose we Fourier transform both sides of the equation. Letting $X(\omega)$ be the Fourier transform of $x(t)$, where ω is the frquency, the Fourier transform of the derivative dx/dt is $i\omega$. The power spectrum is the average of the square of the absolute value of the Fourier transform. Since f is a smooth function, in the limit $\omega \to \infty$ its power spectrum decreases faster than a power law, whereas since the noise is white, its power spectrum is constant. Thus, in the high frequency limit

$$\omega^2 \langle |X(\omega)|^2 \rangle = \text{constant}.$$

This implies that the power spectrum $S(\omega) = \langle |X(\omega)|^2 \rangle$ falls off as $1/\omega^2$ in the high frequency limit. This can be extended for differential equations of order m to show that in the general case the power spectrum scales as $1/\omega^{2m}$ in the high frequency limit.

The argument above is the basic idea behind the method used to design filters, such as those used in audio equipment to reduce high frequency noise. A power law in the high frequency behavior is not very interesting, as it has no dramatic effects. Power laws at low frequencies, such as those discussed in Section 1.4.7, as more dramatic, since they correspond to very low frequency motions, such as intermittency or long-memory processs that can easily be mistaken for nonstationarity. It is possible to construct high pass noise filters, e.g. using a dynamical system with a critical point, or by explicitly making a power law transformation.

The argument for why the power spectrum of an analytic function decreases rapidly at high frequencies is instructive concerning how power laws are related to discontinuities. If f is a smooth function, then by definition all its derivatives are bounded. Furthermore, analyticity implies that there is a limit to how much any of the derivatives can change in any given period of time. Thus, there is also an upper bound $B \geq 0$ to the square of the modulus of the Fourier transform at any given frequency. Thus, the fact that the Fourier transform derivative of $d^m x/dt^m$ is $i^m \omega^m X(\omega)$ implies that

$$\omega^{2m} |X(\omega)|^2 \leq B.$$

Thus the power spectrum of any smooth function f falls off faster than any power in the limit $\omega \to \infty$. To get a power law, then, requires some discontinuity, either in the form of added noise (which is inherently discontinuous) or compounded nonlinearities that produce effective discontinuities.

1.5 Implications for economic theory

Once one accepts that power laws indeed occur in economics, then it becomes necessary to ask whether they can be explained within the equilibrium framework. Of course, there is always the possibility that power laws are imposed by factors that are exogenous to the economy, e.g. if information arrival is a power law, then this will explain why clustered volatility scales according to a power law. But this seems to be simply avoiding the problem, and as already discussed, does not seem to fit the facts.

So far it seems that there is only moderate interest by economists in verifying whether or not power laws exist, and very little work trying to reconcile them with equilibrium. The only model that we are aware of along these lines is a general equilibrium model for the business cycle proposed by Nirei [52]. This is an SOC model in which the power law behavior is driven by the granularity of the production mechanism. Many industries require production facilities and infrastructure of at least a certain size. When a new production facility is built or an old one is retired, production makes a discrete jump, and the supply function is discontinuous. Such changes in production can affect equilibrium allocations, driving the system from one metastable equilibrium to another. The granularity of production sizes causes a distribution of earnings with a power law distribution with $\alpha = 1.5$. Although this is a macroeconomic phenomenon, it is conceivable that fluctuations in earnings could drive other power laws, for example in price changes. More detailed empirical testing is needed.

In agent based models allowing non-equilibrium effects, in contrast, power laws are common, even if there is still no good understanding of the necessary and sufficient conditions for them to occur. The minority game provides a simple illustration (see Section **??**). The prevalence of power laws in such models suggests that the explanation may be a manifestation of non-equilibrium behavior. Much of the modeling by physicists has so far has been focused on trying to find models of financial markets capable of generating power laws, but these models are still qualitative and it is still not possible to claim that any of them explain the data in a fully convincing manner.

The origin of power laws is a property of financial markets whose explanation may have broader consequences in economics. For example, a proposed explanation by Gabaix et al. [24] suggests that power laws in prices are driven by power law fluctuations in transaction volume, which they suggest are driven by a power law distribution of wealth, is caused by a Gibrat-style multiplicative process mechanism (see Section 1.4.3). The conversion of tail exponents from transaction volumes to price fluctuations is postulated to depend on a square root law behavior of the market impact function, which relates trade size to changes in prices. This is derived based on an argument involving minimization of transaction costs by financial brokers. In contrast, other theories have suggested that the market impact function is an inherent statistical property of the price formation dynamics which can be explained by zero or low intelligence models. This is described in more detail in the next section. In any case, it seems that power laws are a ubiquitous feature of economic systems, and finding the correct explanation for them is likely to be illuminating about other aspects of the financial economy.

2 References

References

[1] V. Akgiray, G. G. Booth, and O. Loistl. Stable laws are inappropriate for describing german stock returns. *Allegemeines Statistisches*, 73(2):115–121, 1989.

[2] W. B. Arthur, J. H. Holland, B. LeBaron, R. Palmer, and P. Tayler. Asset pricing under endogenous expectations in an artificial stock market. In W. B. Arthur, S. N. Durlauf, and D. H. Lane, editors, *The Economy as an Evolving Complex System II*, pages 15–44. Addison-Wesley, Redwood City, 1997.

[3] R. Axtell. Zipf distribution of u. s. firm size. *Science*, 293:1818–1820, 2001.

[4] P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality - an explanation of 1/f noise. *Physical Review Letters*, 59(4):381–384, 1987.

[5] G.I. Barenblatt. *Dimensional Analysis*. Gordon and Breach, New York, 1987.

[6] L. Borland. Option pricing formulas based on a non-gaussian stock price model. *Physical Review Letters*, 89(9), 2002.

[7] J-P. Bouchaud, M. Mezard, and M. Potters. Statistical properties of the stock order books: empirical results and models. *Quantitative Finance*, 2(4):251–256, 2002.

[8] J-P. Bouchaud and M. Potters. *Theory of Financial Risk:From Statistical Physics to Risk Management*. Cambridge University Press, Cambridge, 2000.

[9] W. A. Brock and C. H. Hommes. Models of complexity in economics and finance. In C. Hey, J.M. Schumacher, B. Hanzon, and C. Praagman, editors, *System Dynamics in Eco-nomic and Financial Models*, pages 3–41. Wiley, New York, 1997.

[10] J. M. Carlson and J. Doyle. Highly optimized tolerance: A mechanism for power laws in designed systems. *Physics Review E*, 60(2):1412–1427, 1999.

[11] J.M. Carlson and J. Doyle. *Physical Review Letters*, 84:2529, 2000.

[12] R. Cont and J-P. Bouchaud. Herd behavior and aggregate fluctuations in financial markets. *Macroeconomic Dynamics*, 4(2):170–196, 2000.

[13] D. M. Cutler, J. M. Poterba, and L. H. Summers. What moves stock prices? *The Journal of Portfolio Management*, 15(3):4–12, 1989.

[14] M. G. Daniels, J. D. Farmer, L. Gillemot, G. Iori, and E. Smith. Quantitative model of price diffusion and market friction based on trading as a mechanistic random process. *Physical Review Letters*, 90(10), 2003.

[15] Z. Ding, C. W. J. Granger, and R. F. Engle. A long memory property of stock returns and a new model. *Journal of Empirical Finance*, 1:83, 1993.

[16] S. N. Durlauf. Complexity in economics. Working paper 03-02-014, Santa Fe Institute, 2003.

[17] J. P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modeling Extremal Events*. Springer-Verlag, Berlin, 1997.

[18] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united-kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

[19] J. B. Estoup. *Gammes Stenographiques*. Institut Stenographique de France, Paris, 1916.

[20] E. F. Fama. The behavior of stock-market prices. *The Journal of Business*, 38(1):34–105, 1965.

[21] J. D. Farmer and F. Lillo. On the origin of power laws in financial markets. *Quantitative Finance*, 4(1):7–10, 2004.

[22] J. D. Farmer, P. Patelli, and Ilija Zovko. The predictive power of zero intelligence, 2003.

[23] J.D. Farmer, L. Gillemot, F. Lillo, S. Mike, and A. Sen. What really causes large price changes? *Quantitative Finance*, to appear in August, 2004.

[24] X. Gabaix, P. Gopikrishnan, V. Plerou, and H. E. Stanley. A theory of power-law distributions in financial market fluctuations. *Nature*, 423:267–270, 2003.

[25] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge. Turbulent cascades in foreign exchange markets. *Nature*, 381(6585):767–770, 1996.

[26] P. Gopikrishnan, V. Plerou, X. Gabaix, and H. E. Stanley. Statistical properties of share volume traded in financial markets. *Physical Review E*, 62(4):R4493–R4496, 2000. Part A.

[27] Y. Ijiri and H. A. Simon. *Skew Distributions and the Sizes of Business Firms*. Studies in Mathematical Economics. North- Holland, Amsterdam, 1977.

[28] J. D. Keeler and J. D. Farmer. Robust space-time intermittency and 1/f noise. *Physica D*, 23(1-3):413–435, 1986.

[29] H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, CXXXI:207–248, 1973.

[30] K. G. Koedijk, M. M. A. Schafgans, and C. G. de Vries. The tail index of exchange rates. *Journal of International Economics*, 29(1-2):1–197, 1990.

[31] B. LeBaron. Stochastic volatility as a simple generator of apparent financial power laws and long memory. *Quantitative Finance*, 1(6):621–631, 2001.

[32] F. Lillo and J. D. Farmer. The long memory of the efficient market. 2003.

[33] F. Lillo, J. D. Farmer, and R. N. Mantegna. Econophysics - master curve for price-impact function. *Nature*, 421(6919):129–130, 2003.

[34] F. Lillo and R. N. Mantegna. Power-law relaxation in a complex system: Omori law after a financial market crash. *Physical Review E*, 68(1), 2003. Part 2.

[35] A. W. Lo. Long-term memory in stock market prices. *Econometrica*, 59(5):1279–1313, 1991.

[36] F. M. Longin. The asymptotic distribution of extreme stock market returns. *The Journal of Business*, 69(3):383–408, 1996.

[37] M. Loretan and P. C. B. Phillips. Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets. *Journal of Empirical Fnance*, 1(2):211–248, 1994.

[38] T. Lux. The stable paretian hypothesis and the frequency of large returns: an examination of major german stocks. *Applied Financial Economics*, 6(6):463–475, 1996.

[39] T. Lux. Turbulence in financial markets:the surprising explanatory power of simple cascade models. *Quantitative Finance*, 1(6):632–640, 2001.

[40] T. Lux and M. Marchesi. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397(6719):498–500, 1999.

[41] B. Mandelbrot. An informational theory of the statistical structure of languages. In W. Jackson, editor, *Communication Theory*, pages 486–502. Butterworth, London, 1953.

[42] B. Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419, 1963.

[43] B. Mandelbrot. *Fractals: Form, Chance, and Dimension*. W. H. Freeman & co., San Francisco, 1977.

[44] B. Mandelbrot. *The Fractal Geometry of Nature*. W.H. Freeman & co., San Francisco, 1982.

[45] B. Mandelbrot. *Fractals and Scaling in Finance*. Springer-Verlag, New York, 1997.

[46] R. N. Mantegna and H. E. Stanley. Scaling behavior in the dynamics of an economic index. *Nature*, 376(6535):46–49, 1995.

[47] R. N. Mantegna and H. E. Stanley. *Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press, Cambridge, 1999.

[48] G. A. Miller. Some effects of intermittent silence. *American Journal of Psychology*, 70:311–314, 1957.

[49] M. Mitzenmacher. A brief history of generative models for power law and log normal distributions. In *Proceedings of the 39th Annual Allerton Conference on Communication,Control, and Computing*, pages 182–191. University of Illinois, Urbana-Champagne, 2001.

[50] U. A. Muller, M. M. Dacorogna, and O. V. Pictet. Heavy tails in high-frequency financial data. In R. J. Adler, R. E. Feldman, and M. S. Taqqu, editors, *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, pages 55–78. Springer-Verlag, Berlin, 1998.

[51] M.E. Newman, M. Girvan, and J.D. Farmer. Optimal design, robustness and risk aversion. *Physical Review Letters*, 89:2, 2002.

[52] M. Nirei. Threshold behavior and aggregate fluctuation, 2003.

[53] R. R. Officer. Distribution of stock returns. *Journal of the American Statistical Association*, 67(340):807–812, 1972.

[54] V. Plerou, P. Gopikrishnan, L. A. N. Amaral, M. Meyer, and H. E. Stanley. Scaling of the distribution of price fluctuations of individual companies. *Physical Review E*, 60(6):6519–6529, 1999. Part A.

[55] Y. Pomeau and P. Manneville. Intermittent transition to turbulence in dissipative dynamical systems. *Communications in Mathematical Physics*, 74:189–197, 1980.

[56] S-H. Poon and C. W. J. Granger. Forecasting volatility in financial markets: a review. *Journal of Economic Literature*, 41(2):478–539, 2003.

[57] M. Potters and J-P. Bouchaud. More statistical properties of order books and price impact. *Physica A*, 324:133–140, 2003.

[58] W. J. Reed. The double pareto-lognormal distribution- a new parametric model for size distribution, 2001.

[59] H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.

[60] E. Smith, J. D. Farmer, L. Gillemot, and S. Krishnamurthy. Statistical theory of the continuous double auction. *Quantitative Finance*, 3(6):481–514, 2003.

[61] D. Sornette. *Critical Phenomena in Natural Sciences*. Springer-Verlag, Berlin, 2000.

[62] P. Weber and B. Rosenow. Large stock price changes: volume or liquidity?, 2004.

[63] G.B. West. Scale and dimension from animals to quarks. In N.G. Cooper and G.B. West, editors, *Particle Physics*. Cambridge University Press, Cambridge, UK., 1988.

[64] G.B. West, J.H. Brown, and B.J. Enquist. The fourth dimension of life: Fractal geometry and allometric scaling of organisms. *Science*, 276:122, 1997.

[65] S. T. Ziliak and D. N. McCloskey. Size matters: the standard error of regressions in the american economic review. *Journal of Socio-Economics*, 2004.

[66] G. Zipf. *Selective Studies and the Principle of Relative Frequency in Language*. Harvard University Press, Cambridge, 1932.

[67] G. Zipf. *Human Behavior and the Principle of Least Effort*. Addison-Wesley, Cambridge, 1949.

[68] I. Zovko and J. D. Farmer. The power of patience; a behavioral regularity in limit order placement. *Quantitative Finance*, 2(5):387–392, 2002.

Acknowledgements

We would like to thanks Legg-Mason and Bill Miller for supporting the workshop that stimulated this paper. We would also like to thank Credit Suisse, the McDonnell Foundation, the McKinsey Corporation, Bob Maxfield, and Bill Miller for their support.