Power laws in economics and elsewhere

J. Doyne Farmer[*] and John Geanakoplos[†]

May 14, 2008

Abstract

We review power laws in financial economics. This is a chapter from a preliminary draft of a book 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 preliminary and not entirely finished – comments at any level are greatly appreciated.

Unfinished manuscript! Contains omissions and typos. Read at your own risk. Comments are appreciated.

Contents

[*]McKinsey Professor, Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe NM 87501
[†]Economics Department, Yale University, New Haven CT

1 Introduction

There is good evidence for the presence of power law distributions in many if not most high frequency economic variables, such as returns, order flow, volume, and liquidity. They are an important regularity of many facets of financial markets that equilibrium theories have so far failed to illuminate. To quote Ijiri and Simon, ".. on those occasions when a social phenomenon appears to exhibit some of the same simplicity and regularity of pattern as is seen so commonly in physics, it is bound to excite interest and attention" [39]. Despite the growing empirical evidence for the existence of power laws and their practical importance, the existence of power laws has received little attention from financial economists. Many aspects of the subject are widely misunderstood. For this reason, and because there is no good comprehensive review of this subject, we devote an entire chapter to it.

Crudely speaking a power law is a relation of the form $f(x) = Kx^\alpha$, where $x > 0$ and K and α are constants. Power laws can appear in many different contexts. The most common are that $f(x)$ describes a distribution of random variables or the autocorrelation function of a random process, but power laws can appear in many different contexts. Although this continues to be controversial, there is now a large body of evidence suggesting that many properties of financial markets are power laws. This has important

Figure 1: When a power law is plotted in double logarithmic scale, it becomes a straight line. In general one expects power law scaling only as an asymptotic property; even if a pure power law is modified by a slowly varying function which alters the scaling at any finite x, but becomes unimportant in the limit, it is still called a power law.

practical consequences for risk management, volatility forecasting, statistical estimation, and derivative pricing. It is also conceptually important because it suggests a different emphasis in economic modeling. While power laws may be consistent with equilibrium theory, it has so far failed to address them.

The property that makes power laws special is that they describe scale free phenomena. To see why, suppose a variable undergoes a scale transformation of the form $x \rightarrow cx$. If $f(x) = Kx^\alpha$, it is transformed as $f(x) \rightarrow Kc^\alpha x^\alpha = c^\alpha f(x)$. Changing the scale of the independent variable thus preserves the functional form of the solution, but with a change in its scale. Power laws are a necessary and sufficient condition for scale free behavior. To see this, consider the condition for scale invariance, which can be written as a functional equation of the form $f(cx) = K(c)f(x)$. For any constant $c > 0$, there exists another constant $K(c) > 0$ such that there is a solution $f(x)$ with $x > 0$. A power law is the only[1] possible solution. Scale free behavior has important scientific implications because it strongly suggests that the same mechanism is at work across a range of different scales.

A power law is just a linear relationship between logarithms, of the form

$$\log f(x) = -\alpha \log x + \log K.$$

We give an example in Figure 1. The quick and dirty test for a power law is to simply plot the data in double logarithmic scale and look for a straight line. The scaling exponent α can be determined by measuring its slope. But when power law scaling is only approximate and data is limited, this can yield ambiguous results. More rigorous statistical testing procedures also have problems. This has caused considerable debate, as discussed in Section **??**.

The crude definition of a power law given above is misleading because power law scaling allows for asymptotically irrelevant variations, such as

---

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

logarithmic corrections. Confusion about this has led to a great deal of misunderstanding in the literature, so it is worth spending some time to discuss this carefully.

The notion of a power law as it is used in extreme value theory [27] is an asymptotic scaling relation. Two functions f and g have equivalent scaling, $f(x) \sim g(x)$, in the limit[2] $x\infty$ if

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

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

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

for any $t > 0$. Examples of slowly varying functions are the constant function and the logarithm.

A power law is defined as any function satisfying $f(x) \sim x^\alpha$. Under this definition, a power law is not a single function, but rather the family of functions that are asymptotically equivalent to a power law. The slowly varying function $L(x)$ can be thought of as the deviation from a pure power law for finite x. For $f(x) = L(x)x^{-\alpha}$, taking logarithms of both sides and dividing by $\log x$ gives $\log f(x)/\log x = -\alpha + \log L(x)/\log x$. Providing $L(x)$ is a slowly varying function, in the limit $x \to \infty$ the second term on the right goes to zero, and this reduces to $\log f(x)/\log x = -\alpha$. See Figure **??**.

In a similar vein, for any $t > 0$ a *regular function* is one that satisfies

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

where $\chi(t)$ is positive and bounded. Unlike a slowly varying function, under a change of scale a regular function is not asymptoticaly invariant. The connection to power laws becomes clear 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 $\chi(t) = t^{-\alpha}$. In the limit $x \to \infty$, any function of the form $L(x)x^{-\alpha}$ satisfies this relation as long as $L(x)$ is slowly varying, making it clear that power laws are regular functions.

To physicists, the apparent prevalence of power laws in financial markets is an important modeling clue. Explaining power laws in financial markets is important for its own sake, and it may also have broader consequences

---

[2]It possible to use any limit, but unless otherwise specified, for convenience we will assume the limit $x \to \infty$, which is the most relevant one for finance.

for economic theory. If an exogenous properties of the market such as the information arrival rate is a power law, under a standard equilibrium model this can cause market properties such as the distribution of price changes to be a power law. However, there is considerable evidence that many power laws are endogenous properties of markets. It is not clear whether this is compatiable with equilibrium. Providing a proper endogenous explanation of power laws may force us to develop nonequilibrium theories.

The importance and ubiquity of scale free behavior was originally pointed out by Mandelbrot [61, 62, 63]. He coined the word "fractals" to describe the nondifferentiable geometric objects that satisfy power law scaling when α is not equal to an integer. Fractals have the property that, by using an appropriate magnifying glass, one sees the same behavior across different scales (in length, time, price, or any other relevant variable with a scale attached to it). Mandelbrot demonstrated that fractals are ubiquitous in nature, describing phenomena as diverse as coastlines, clouds, floods, earthquakes, financial returns, and fundamental inaccuracies in clocks. For coastlines or clouds there is a power law relationship between size and measurement resolution. The coastline of Brittainy wiggles on all scales, is longer when measured accurately than when measured crudely, and increases as a power law as a function of measurement resolution. Similarly, clouds have soft boundaries, so that the volume of what one would consider to be a cloud depends on a water concentration threshold. The volume of clouds increases as a power law, inversely with this threshold. For earthquakes, floods, or financial returns, the probability of a large event greater than a given size decreases as a power law. For clocks ranging from hour glasses to lasers, the fundamental source of inaccuracies is a random process that is correlated in time, with a correlation function that decays as a power law (this is called a long-memory process – see Section 3). Given the previously prevailing assumption that nature should generally be described by smooth functions, the realization that so many diverse phenomena could be modeled based on non-differentiable geometry was a major shift in worldview[3].

Of course, the assumption of power law scaling is always just an approximation, which is only valid across a given range. For most examples there are cutoffs at large and small scales where the scaling assumption ceases to be valid. But when it applies, the power law assumption parsimoniously

---

[3]From a certain point of view, all of the examples given above can be related to sampling from a power law distributed random variable. One can randomly generate a coastline, for example, by constructing a curve whose increments are power law distributed variables with random orientations. As we demonstrate in Section 6.9, power laws are associated with a lack of differentiability

captures an important regularity.

We begin our exposition of power laws by carefully defining power law distributions of random variables and discussing their properties. We then introduce the closely related phenomenon of long-memory random processes. To explain why power laws are important in economics, we list some of the many examples where power laws are claimed to occur, and use important economic problems such as clustered volatility, risk control, statistical estimation, and hyperbolic discounting to illustrate their practical importance in finance. We then address some of the controversy surrounding empirical work on power laws, and give a response to recent criticism [26, 43]. In order to give a better understanding of where power laws come from, and in order to illustrate the kind of models that can explain them, we present a review of mechanisms for generating power laws. Finally we discuss the implications for economic theory.

2   Power law distributions of random variables

Perhaps the most common context in which power laws occur is as probability distributions of random variables. A power law distribution is an example of what is often called a *fat tailed* distribution. The interchangeable 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. The fact that many economic data sets are described by fat-tailed distributions is not controversial. In fact, as we explain more precisely below, any distribution with "sufficiently fat tails" is a power law distribution. Thus, the debate concerning power laws is reduced to the question of just how fat the tails of economic data sets really are.

To make the effect of fat tails more tangible, in Table 1 we compare a normal distribution to a power law, in this case a Student's t-distribution with three degrees of freedom. To calibrate this to a distribution of price movements, we choose both distributions to have the same standard deviation of 3%, which is a typical figure for daily price movements in a financial market. 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. The difference is something a risk manager should take seriously. This becomes even more dramatic when looked at the other way: The probability of observing a fluctuation of 21% (the size of

6

Probability	0.9	0.99	0.999	0.9999	...	10^{-16}
Normal	3.8	7.0	9.2	11	...	21
Student	2.8	7.8	17.7	38.5	...	XXX

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 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 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.

For probability distributions it is standard to express the scaling in terms of the associated exponent of the cumulative distribution $P(x > X) \sim X^{-\alpha}$, where $\alpha > 0$ is called the *tail exponent*. Assuming it exists, from elementary calculus the corresponding probability density function p (defined as $P(x > X) = \int_0^X p(x')dx'$) scales as $p(x) \sim x^{-(\alpha+1)}$. The fact that the scaling exponent of the density function is equal to $\alpha + 1$ is a rule we will use often. For convenience we will assume $x > 0$; the negative values of a double-sided distribution are treated by taking absolute values. In general the positive and negative sides of an asymmetric distribution can obey different power laws, and a distribution might be an asymptotic power law with different values of α in two different limits[4], e.g. $x \to 0$ and $x \to \infty$.

The tail exponent α has a natural interpretation as the cutoff above which moments no longer exist. This is because for a density function with power law scaling $p(x) \sim x^{-(\alpha+1)}$, the mth moment scales as

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

γ is finite when $m < \alpha$ and it is infinite otherwise. The tail exponent thus

---

[4]A good example is the double Pareto distribution, discussed in Section 6.5.

provides a single number summary of "how fat" the tails are – the lower the exponent, the fatter the tails, and the lower the cutoff above which the moments of the distribution no longer exist. This result holds generally for power laws – slowly varying functions cannot change whether or not a given moment exists. In fact, as we will make precise a bit later, all well-behaved distribution functions with moments that do not exist are power laws.

This is clearly vitally important: When a moment doesn't exist, any attempt to compute a statistic based on it will fail to converge to a limiting value, even with an infinite amount of data. As we discuss later, if $\alpha < 2$ then the variance doesn't exist, the central limit theorem no longer applies, and sums no longer converge to a normal distribution. If $\alpha \leq 1$, the mean does not exist. For this reason there is no such thing as an "average flood" - it is only possible to measure flood likelihoods in terms of quantiles, as in the statement "this is a 100 year flood".

The first power law (in any discipline) was discovered by Pareto, who in his book *Cours d'Economie Politique* noted that "in all countries and at all times the the extreme distribution of income and wealth follows a power law behavior". In his honor the pure power law distribution $P(x) = Kx^{-\alpha}$ is often called the Pareto distribution. In Pareto's statement the word *extreme* is important, as it is typically only the tail of the wealth or income distribution that is a power law – the main body of the distribution is usually better described by a log-normal or exponential distribution. The problem of separating the tail and the body of power law distributions has created controversy ever since.

For continuous random variables it is particularly important to stress the asymptotic nature of power laws. For a continuous variable defined on $[0, \infty]$ there is no such thing as a "pure" power law distribution across the entire interval. This is easily seen by contradiction: Suppose there existed a density function of the form $p(x) = Kx^{-(\alpha+1)}$. For $\alpha \leq 0$, $\int_0^\infty p(x)dx = \infty$ due to the upper limit, and similarly, for $\alpha \geq 0$, $\int_0^\infty p(x)dx = \infty$ due to the lower limit. A pure power law distribution on $[0, \infty]$ cannot be normalized for any α. This is of course possible on any restricted domain $[a, \infty]$, where $a > 0$. But more typically one finds distributions such as the Lorentzian distribution $P(X > x) = A/(1 + x)^\alpha$, which can be defined on the whole interval, but which differs throughout from a pure power law by a slowly varying function $L(x) \neq 1$

Distributions of discrete variables, in contrast, do not suffer from the problem of unbounded normalization. This is for the obvious reason that discreteness provides a built in lower bound, so the distribution is always normalizable as long as $\alpha > 0$. This is in a certain sense just a technical

distinction; for example, one might argue that since there is a minimum denomination of currency, wealth is a discrete variable, providing a lower cutoff. Clearly some common sense is needed, and in most cases the central point that a power law is an asymptotic notion still applies.

That said, it is also important to note that some power law distributions converge to their asymptotic behavior much faster than others. There are instances where power law scaling is an excellent approximation across a broad range of values. A good example is the distribution of firm sizes which fits a power law with $\alpha \approx 1$ very well, from the smallest firms with only one employee to the largest firms with 10^6 employees [4]. In biology, the power law scaling of metabolic rate against animal size applies across 24 orders of magnitude [83, 84]. Empirically, rapid convergence makes the power law hypothesis much easier to test with a given amount of data. Theoretically, rapid convergence is important because it implies that scale free behavior applies across a broader range, and gives an important clue about mechanism – some mechanisms tend to yield faster convergence than others.

2.1   Invariance under aggregation

One of the reasons that power laws are ubiquitious is because of their strong invariance under aggregation. The property of being a power law is preserved under addition, multiplication, and polynomial transformation. When two independent power law distributed variables are combined, either additively or multiplicatively, the one with the fattest tail dominates; the tail exponent of the combined distribution is the minimum of the tail exponents of the two distributions being combined. When a power law distributed variable is raised to a (nonzero) power, it remains a power law but with an altered exponent[5] Letting $\alpha(x)$ be the tail exponent of the random variable x, we can write these three transformation rules in the following form:

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

---

[5]These aggregation rules are intuitively easy to understand. Let $z = x + y$, where x and y are both power law distributed. If $\alpha(x) < \alpha(y)$, then in the tails $P(y) \ll P(x)$, and $P(z) \approx P(x)$. Similarly, suppose $z = xy$; after taking logarithms this becomes $\log z = \log x + \log y$. As shown in Section 6.1, the logarithm of a power law variable is an exponential. By the same argument above, the one with slowest decay dominates. The rule for polynomials is obvious from taking logarithms.

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.

In thinking about how a power law affects different time scales it is useful to understand how the whole distribution evolves under aggregation. To take a concrete example, consider a highly idealized model for prices[6]. Let π_t be the price at time t, and $f_t > 0$ be the multiplicative change in price from the previous period, so that $\pi_t = f_t \pi_{t-1}$. By taking logarithms and summing this can be rewritten

$$r_t(\tau) = \log \pi_{t+\tau} - \log \pi_t = \sum_{i=t}^{i=t+\tau} \log f_t. \qquad (7)$$

$r_t(\tau)$ is the logarithmic price return on timescale τ. Assume that $\log f_t$ is a random variable with a symmetric power law distribution with $\alpha > 2$.

How does the distribution of returns, $P(r(\tau))$, change with τ? As τ increases, due to the central limit theorem, the center of the distribution approaches a normal distribution[7]. However, due to the aggregation laws given above, the tails remain power laws, and the tail exponent α is unchanged. There is a competition between these two processes. In the limit as $\tau \to \infty$, the normal distribution wins, but for finite τ the power law is always there. As τ increases the fraction of the distribution that is approximately normal grows, while the fraction with power law scaling shrinks. However, the power law never goes away, even on long timescales; it just describes rarer but more extreme events. It is worth keeping in mind that as one aggregates, the most extreme events grow in size. Thus, though the events in the power law tail may become rare, they may be very large when they happen. See [17] for a more quantitative description.

2.2 Limiting distributions of extrema

One reason that power laws are ubiquitous is that they are one of three possible limiting distributions for extreme values [27]. In a sense made more precise here, any "well behaved" distribution with "sufficiently fat tails" is a power law. Just as the normal distribution emerges as a limiting value of

---

[6]It may help to think about Student's t distribution. The tail exponent α is just the number of degrees of freedom.

[7]Note that when the log-return $r(\tau)$ is normally distributed, the actual return $R_t(\tau) = (\pi_{t+\tau} - \pi_t)/\pi_t$ is log-normally distributed (see Section 6.4). While the log-normal distribution is quite fat-tailed, all its moments exist, and its tails are not fat enough for it to be a power law.

sums of random variables, the power law emerges as one of three possible limiting distributions for extrema, such as the maximum or minimum. These limiting distributions come about because in the limit $x \to \infty$, the typical size of the maximum or the minimum in a large but finite sample effectively determines the behavior of the tail of the distribution.

We will illustrate this for the maximum. For a given distribution $P(x)$, if a limiting extremal distribution for the maximum exists, it can be constructed according to the following procedure:

1. Make independent draws from P of k sequences $(x_1, \ldots, x_n)$, each of length n.

2. Compute the maximum value $M_k(n)$ for each sequence.

3. Compute a rescaled variable $M_k'(n) = (M_k(n) - d_n)/c_n$, where d_n and c_n are centering and normalizing constants that depend on $P(x)$ and n.

4. Take the limit first as $k \to \infty$ and then as $n \to \infty$.

If it is possible to find c_n and d_n so that the distribution of $M_k'(n)$ converges to a limiting distribution, there are only three possibilities (listed here as cumulative distributions, i.e. $P(M' < x)$):

$$
\begin{aligned}
\text{Fréchet:} \quad & \Phi_\alpha(x) &=& \begin{cases} 0, & x \le 0 \\ \exp\{-x^{-\alpha}\}, & x > 0 \end{cases} \quad \alpha > 0. \\
\text{Weibull:} \quad & \Phi_\alpha(x) &=& \begin{cases} \exp\{-(-x)^\alpha\}, & x \le 0 \\ 1, & x > 0 \end{cases} \quad \alpha > 0. \\
\text{Gumbel:} \quad & \Lambda(x) &=& \exp\{-e^{-x}\}, x \in \mathbf{R}.
\end{aligned}
$$

The limiting distribution that emerges depends on the fatness of the tails of $P(x)$. If $P(x)$ has finite support[8], 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, for example normal, exponential, and log-normal distributions, then the limiting distribution is Gumbel. But if the tails die off sufficiently slowly that some higher order moments do not exist, then the limit is a Frechet distribution, which is a power law. This can be seen by expanding the exponential function in a Taylor series. $P(X > x) = 1 - \Phi_\alpha(x) = 1 - \exp\{-x^{-\alpha}\} \approx x^{-\alpha}$. (We have subtracted

---

[8]Finite support means that there exists x_{min} and x_{max} such that $P(x) = 0$ for $x < x_{min}$ and $x > x_{max}$.

this from one to convert to a tail probability). Examples of power law distributions are the Pareto, Cauchy, Student's t, Levy stable, Lorentzian, log gamma, and double Pareto. The three possible limiting distributions are closely related, in the following sense: If a random variable x has a Frechet distribution, then $\log x$ has a

A few caveats: The resulting limiting distributions are only unique up to affine transformations[9]. The criteria for whether or not a limit exists is essentially a continuity condition, but not all distributions have well-defined limits. The Poisson distribution is the most famous counterexample. However, most common distribution functions are sufficiently continuous in their tails that they have well defined limits. It turns out that if it is possible to find a sequence of normalizing and centering constants c_n and d_n that gives a limiting distribution, that sequence is unique. For example, if $P(x)$ is the uniform distribution defined on $(0, 1)$, $d_n = 1$ and $c_n = n^{-1}$. For a Pareto distribution the norming constant is $c_n = (Kn)^{1/\alpha}$

The key point here is that (when it exists), the limiting distribution describes not just the behavior of the maximum, but also the second maximum, third maximum, etc., and in fact the entire order statistics of P for extreme values. In the large n limit it tells us the probability of drawing a value of a given size. It thus fully characterizes the tail of the distribution function.

It can be shown that a distribution $P(x)$ converges to the Frechet distribution if and only if P is a power law. Thus any distribution which is sufficiently continuous to have a limiting distribution for its extremum, and that has a cutoff above which moments do not exist, is a power law. This makes precise our statement above that a power law describes the tail behavior of any "well-behaved" distribution with "sufficiently fat tails".

3 Long-memory processes

The relevance of power laws is not limited to marginal distributions of a single variable. Joint distributions can asymptotically follow power laws, reflected in the scaling properties of moments such as correlation functions. A particularly relevant example for economic time series is the long-memory random process, defined as a random process with a positive autocorrelation function $C(\tau) \sim \tau^{-\beta}$, with $0 < \beta < 1$. Such a strong autocorrelation implies a high degree of long-term predictability, particularly when β is small. Long-

---

[9]The limiting distribution $H(x)$ is equivalent under an affine transformation to $aH(x) + b$, where a and b are constants.

memory also has important consequences for diffusion processes and for the rate of convergence of statistical estimates.

Long-memory processes have been observed in natural and human phenomena ranging from the level of rivers to the temperature of the Earth [12]. Reports of long-memory in economic data span the gamut from macroeconomics to finance [7]. In macroeconomics this includes GNP data [24], the consumer price index and other measures of inflation[8, 38], and the term structure of interest rates [6]. In finance long-memory in price volatility has been observed both for stocks [25, **?**] and exchange rates [37] and in trading volume [49]. Recently long-memory has been observed for time series of order signs (whether or not a new order is to buy or to sell); this is seen in high-frequency data for both the Paris Stock Exchange and the London Stock Exchange [15, 44]. The fact that this is really long-memory can be verified at a very high level of statistical significance [44]. Surprisingly, the autocorrelation function for the signs of individual buy and sell orders (in markets with thousands of orders per day) is significantly positive over lags as long as two weeks. Liquidity, as measured by the volume at the best bid or ask, also shows long-memory [44].

In finance there have also been claims that stock returns display long-memory [58, 36], but the evidence has been disputed [48]. Long-memory in price returns would be remarkable because of its implications for market efficiency, and in any case it is clear that if it exists at all, the long-memory of price returns is very weak. More recent high-frequency studies do not show long-memory in prices, but they do raise the interesting question of how market efficiency coexists with the long-memory of order flow [15, 44].

Long-memory has several important consequences. An obvious one is that it implies a high degree of predictability. This can be made explicit by constructing, for example, an autoregressive model of the form $\hat{x}(t + \tau) = \sum_i^N a_i x(t - i)$. The persistence in the autocorrelation function makes it useful to use a high value of N, which dramatically improves forecasts over longer time horizons τ. Whereas for a normal Markov process the accuracy of forecasts decays exponentially, for a long-memory process it decays as a power law.

Another consequence is that the variance of a long-memory diffusion process grows faster than a normal diffusion process. Consider a discrete diffusion process y built out of the sum of random variables x_t, i.e. $y(N) = \sum_{t=1}^N x_t$. If x_t does not have long-memory, then y behaves in the usual way, and the variance of $y(N)$ increases proportional to the number of steps N. When x_t has long-memory, however, the variance grows as N^{2H}, where H is called the *Hurst exponent*. For a long-memory process the Hurst exponent is

in the range $1/2 < H < 1$. It is related to the scaling of the autocorrelation function by $H = 1 - \beta/2$. When $\beta = 1$, $H = 1/2$, and normal diffusion is recovered. In physics, a random process with $H \neq 1/2$ is often referred to as *anomalous diffusion*[10]. The reason that $\beta = 1$ is the cutoff for long-memory behavior can be seen from the derivation of this result; the rate of diffusion depends on the integral of the autocorrelation function, which when $\beta < 1$ becomes unbounded [?].

In statistical estimation long-memory has the important consequence that it causes sample means to converge very slowly. The standard deviation of the sample mean of a long-memory process converges as $\sigma \sim N^{-H+1}$, where N is the number of data points. $H = 1/2$ gives standard square root convergence of errors, but as H increases the convergence becomes slower, until for $H = 1$ the process become nonstationary, and the mean fails to converge at all. Thus in a certain sense an increase in long-memory can be viewed as making a timeseries less stationary. For economic series with long-memory, it takes a lot more data than one would normally expect to get an answer at a given level of accuracy.

There are many problems in physics that exhibit long-memory, and the question of what causes long-memory has received a great deal of attention [?]. Reviewing this literature is beyond the scope of this paper.

4 Practical importance of power laws in financial economics

Power laws have both practical importance and theoretical implications for financial economics. In this section we begin by briefly reviewing the empirical literature relating to power laws in financial economics. There are a sufficient number of different examples that we can only list most of them. We then discuss a few of them in more detail, in particular the problems of clustered volatility, risk control, option pricing, statistical estimation, and hyperbolic discounting. We should stress that some of claims made here are controversial; in the next section we discuss this in the context of reviewing empirical methods of testing for power laws.

---

[10]Unfortunately the term anomalous diffusion is used in two different senses. The loose sense refers to any random process involving sums of uncorrelated random variables; the strict sense refers to variables that are sufficiently correlated to alter the Hurst exponent.

4.1 Summary of empirical evidence for power laws

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

- *Clustered volatility.* The autocorrelation of the absolute value of price changes is a long-memory process whose autocorrelation function decays as $\tau^{-\beta}$, where β is approximately in the range $0.2 < \beta < 0.5$ [25, 77, 65].

- *Large price changes on short time scales* [57, 30, 74, **?**, 42, 51, 64, 50, 52, 34, 69, 75, 65]. Price changes are measured in terms of log-returns $r(\tau) = \log p(t+\tau) - \log p(t)$, where p can be either a transaction price or the average of the best quoted buying and selling prices. Log returns are generally observed to be a power law, with a tail exponent in the range $1.5 < \alpha < 6$.

- *Hyperbolic discounting.* Psychological evidence [**?**] suggests that people do not discount future utility decays according to an exponential, and that a power law may be a better approximation. This may emerge for good theoretical reasons in circumstances where interest rate variations are not taken into account [5]

- *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 $1 < \alpha < 3$ [21, 80, 39, 73].

- *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 $\alpha \approx 1$ [87, 39, 4].

- *Fluctuations in the width of the distribution of growth rates of companies* [1]. Letting s be the standard deviation in the logarithmic growth rate, $P(s > S) \sim S^{-\alpha}$, with $\alpha \approx 0.2$.

- *The volume of individual transactions* for NYSE stocks [35] has a power law distribution with tail exponent $\alpha \approx 1.7$.

- *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,

15

π_{best} is the best quoted price for buy orders. $\alpha \approx 0.8$ for the Paris Stock Exchange [16], and $\alpha \approx 1.5$ for the London Stock Exchange [88].

- *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 $\gamma \approx 0.4$ [47].

- *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 $\gamma \approx 1$ [46]. A similar relationship for seismometer readings after large earthquakes was observed by Ohmori in the nineteenth century [63].

- *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. This is a long-memory process that decays as $\tau^{-\beta}$, where τ can be either the time or the number of events separating the orders. $\beta \approx 0.2$ for the Paris and $\beta \approx 0.6$ for the London Stock Exchange [78, 45].

- *Autocorrelation of order volume.* For the London Stock Exchange the order volume measured in either shares or pounds is a long-memory process whose autocorrelation function decays as roughly $\tau^{-\beta}$, with $\beta \approx 0.6$ [45].

- *Autocorrelation of liquidity at the best bid and ask.* For the London Stock Exchange the volume at either the best bid or the best ask is a long-memory process whose autocorrelation decays roughly $\tau^{-\beta}$, with $\beta \approx 0.6$ [45].

For a more in-depth discussion of some of these, see Cont [**?**].

4.2   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

16

(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 $0.2 < \alpha < 0.5$, indicating that volatility is a long-memory process [25, 77, 65, 12]. This gives rise to bursts of volatility on timescales ranging from minutes to decades.

Standard equilibrium models predict that the amplitude of price changes is driven solely by the information arrival rate. If the states of nature become more uncertain, then prices respond by fluctuating more rapidly. Indeed, it is well-established that most natural disasters, such as flood, hurricanes, and droughts, are long-memory processes, so this explanation is plausible [12]. Another plausible explanation which is also compatible with standard equilibrium models is that this is due to an innate non-economic property of human interactions that causes people to generate news in a highly correlated way. Under either of these hypotheses, clustered volatility is just a reflection of an exogenous property, which is then passively echoed in the resulting equilibrium.

However, as we have already discussed in Section **??**, this does not appear to be compatible with studies that show a low correlation between news arrival and price movements [22, **?**]. While there are several reasons to believe that exogenous factors influencing news may be long-memory processes, these do not appear to be the principal inputs the market is responding to.

In contrast, clustered volatility emerges endogenously in many agent-based models with bounded rationality, which allow deviations from rational expectations equilibrium [2, 18, 54]. Many of these models also simultaneously 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, which also explain clustered volatility.

While standard equilibrium models do not seem to be compatible with clustered volatility, it is possible that they can be extended in some way to include it. This might come about naturally, for example, in a temporary equilibrium setting. More work is needed to determine whether equilibrium is compatible with clustered volatility, and if so, the necessary and sufficient conditions for it to occur. There are also practical reasons to understand the power law nature of clustered volatility, in particular its role in risk control and option pricing, as discussed in the next section.

17

4.3   Option pricing and risk control

Power laws have important practical implications for both option pricing and risk control. This is 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 larger than one might expect, and the power law hypothesis provides a parsimonious method of characterizing them.

The fat tails in prices returns have received a great deal of study [57, 30, 74, **?**, 42, 51, 64, 50, 52, 34, 69, 75, 65]. Power law behavior is much more evident at short time scales and for large data sets. For returns of individual American stocks on timescales of the order of 15 minutes, for example, power law scaling is a good approximation for about 50 standard deviations (a range of variation of about two orders of magnitude) [75]. Although the first papers by Mandelbrot [57] and Fama [30] 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, though because of the difficulty of producing reliable error bars, this remains a debated point [31]. Power laws have been claimed for returns on timescales as long as a month; insofar as the model for price aggregation given in Section 2.1 is valid, one one would expect this to be relevant on longer times scales as well (though it is harder to measure empirically due to data limitations).

The practical value of the power law hypothesis for risk control is that it results in more efficient extreme risk estimates than standard non-parametric methods. Suppose one wishes to estimate 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 so is more efficient with limited data. This can result in less biased estimates.

Risk control estimates are also affected by the long-memory nature of clustered volatility. As we have discussed in Section 3, when the amplitudes of the increments of a random walk have long-memory, the variance of the process grows faster than it does under a standard random process. This implies greater risk. Understanding the properties of the long-memory (e.g. having a good estimate of the Hurst exponent) makes it possible to estimate risks more accurately.

This is closely connected to the problem of forecasting volatility. The mainstream approach for doing this is with ARCH models and their generalizations [28, 27], which fail to properly capture either long-memory or power law tails. When an ARCH model is used to generate a stochastic volatility random process, the resulting variables have a power law tail. However, when the model is fit to real data, the tail exponent is much too large, i.e. the tails of an ARCH process are too thin to explain the fat tails of prices. More importantly, the ARCH random process is not a long-memory process. One of the main symptoms that result from this is that an ARCH model fit on one timescale does not work well on a different timescale [?]. This is in contrast to models that explicitly take long-memory into account [7, 13, 70, ?]. It appears that long-memory volatility models have substantially more predictive power than standard ARCH models [53], and furthermore that they are more parsimonious in that a single model can be used to successfully forecast volatility on a variety of different timescales.

This has practical implications for derivative pricing. Both the fat tails in price movements and the long-memory of volatility affect option prices. Models that explicitly take this into account are more accurate than the standard Black-Scholes model, and a provide a more parsimonious fit to the data than non-parametric alternatives [17, 14]. The long-memory property of clustered volatility is also important for theoretical reasons, as the scale invariance associated with power law scaling suggests that a similar mechanism may drive fluctuations in the amplitude of price movements, across a spectrum of different timescales ranging from minutes to years.

4.4   Statistical estimation in economics

As listed in Section 4.1, volatility is only one of many economic time series that are long-memory processes with power law tails. This has important consequences for statistical estimation. Power tails and long-memory can substantially increase the error bars associated with statistical estimation. While robust and nonparametric statistical estimation methods attempt to take this into account, they typically lack efficiency. The moving block bootstrap, for example, is a standard method that attempts to cope with clustered volatility. However, the time interval chosen for the bootstrap forces the choice of a specific timescale, a procedure that is inherently unsuited for a scale free long-memory process. Techniques that are specifically designed for long-memory process, such as the variance plot method [12], produce better results. Given the prevalence of long-memory processes in economics, it is surprising that this problem has not received more atten-

tion, and that methods specifically tailored to long-memory and power law tails are not better developed and more widely used.

4.5   Utility discounting

The assumption that future utility is less valuable than present utility is pervasive in economics. It is almost universally assumed that the proper function for weighting utility as a function of time is an exponential, e^{-rt}. The standard argument is that this depends on interest rates. If a dollar today can yield two dollars ten years from now, then a dollar ten years from now is only half as valuable as a dollar today.

Psychological experiments suggest, however, that most people do not use exponential weights in considering future utility. Instead, they place stronger weights on utility in the far future than would be expected by an exponential. It has been suggested that a power law provides a better fit to the empirical data [?].

This can be supported by theoretical arguments [5]. In the real world, interests rates are not constant, but rather vary in an essentially random way. In world of uncertain interest rates, the loss of utility with time must be weighted by the distribution of interest rates, and so is of the form

$$u(t) = \int P(r)e^{-rt}dr.$$

Under the standard assumption that $P(r)$ is a log-normal distribution, $u(t)$ is a power law[11] (see Section 6.5).

The consequences of this have not been carefully studied, and its implications for equilibrium models are not clear. It is possible that this might explain some of the power laws observed empirically. Given the practical importance of utility discounting, and its centrality in economic theory, it seems surprising that this has not received more attention. Perhaps the most surprising thing is that ordinary people apparently intuitively understand this, while mathematical economists do not.

---

[11]Ayres and Axtell originally made this argument assuming $P(r)$ is an exponential function. The log-normal is a better approximation to real interest rates. An even better approximation is that real interest rates have power law tails. All three of these assumptions yield power law utility functions.

5 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 briefly review methods of data analysis for determining whether power laws exist, and discuss some of the criticisms that have been raised.

5.1 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 [55, 3].

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 2.2. 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[12]. There are several examples where these methods have been applied and give highly statistically significant results supporting power laws [?, 42, 51, 50, 52, 69]. 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[13].

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

---

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

[13]A related problem is that of testing for long-memory. The test originally proposed by Mandelbrot [59, 60] 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 [48] 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.

is enough data to be inside the asymptotic regime. As we have already said, some power law 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: A power law is a family of distrubutions, the properties of whose members are not well specified in advance. Testing for membership is more difficult than testing for conformity to a specific distribution. 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. Sensible interpretation of results depends on good judgement. The crude visual inspection method has merit in forcing the reader to use judgement and common sense in interpreting the results [85], and should always be used in tandem with more formal methods.

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 frequency data, often involving millions of observations [69, 75, 16, 45]. Understanding at longer frequencies can be achieved by studying the time aggregation properties of the time series (e.g to make sure that large events are not strongly reverting), and making use of the fact that the power law tails of a distribution are preserved under most aggregation mechanisms [17, 70]. Thus, if one finds a power law in high frequency data, barring a rather unusual time aggregation mechanism, it will still be present at lower frequencies, even if it describes 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 improve our predictive or explanatory power by leading to better models. Self-similarity 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 also have more detailed testable predictions. See the discussion in Section 6.

5.2 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 [43], who showed that he could mimic the power law behavior of a real price series using a model that can be proven to not have power law scaling. He fitted the parameters of a standard volatility model[14] 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 [64, 50] who claimed that the evidence supported power law scaling in prices. LeBaron demonstrated that he could produce price series with scaling properties very similar to those of the real data 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 true asymptotic power law scaling. 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 [26].

The physicist responds by noting that in order to fit this model, LeBaron had to choose very special parameters. In his model the volatility level is driven by a combination of three AR(1) models, one of which is $y_t = 0.999y_{t-1} + n_t$, where n_t is an IID noise process. The parameter 0.999 is very close to one; when it is one, the model does have asymptotic power law behavior. The reason the model has an approximate power law is because it is extremely 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 very rare. In Section 4.1 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

---

[14]The stochastic volatility model he used was not an ARCH model, and does have power law behavior except for certain special parameter values.

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 – while all mixture distributions have fatter tails, they do not all exhibit power laws.

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

6 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. In physics, once it became clear that power laws cannot be explained by linear or (physical) equilibrium models, a great deal of research was undertaken to develop nonlinear and nonequilibrium models. Such a burst of research in this direction has not occurred in economics. Economic equilibrium is very different from physical equilibrium, and there is at least one example illustrating that economic equlibrium can be consistent with power laws [72]. Nonetheless, the existence of power laws suggests a change in the focus of attention in model building.

In this section, with the goal of stimulating future research in economics along these lines, we give a review of mechanisms for generating power laws. This is not a well-developed subject – there are no theorems stating the necessary and sufficient conditions for power laws to occur. Furthermore, there are many levels of description 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. These may all be consistent with each other, but at operating at different levels of explanation, and 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. Cer-

tain themes emerge, such as self-similarity, hierarchy, competition between exponentials, growth, amplification, and long-range interaction. The knowledge that these themes are present may suggest new modeling directions in economics, which may have implications beyond the existence of power laws.

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 similarity 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 review by Mitzenmacher [68], as well as the books by Sornette [82], and Mandelbrot [63], though these sources do not span all the topic discussed here.

6.1   Hierarchies and exponentials

We will begin by constructing a few trivial examples of power laws "by hand", with the goal of illustrating some of the recurring themes of mechanisms for power laws. 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 $i = 0, 1, \ldots, n$, at the ith level of the tree there

are $N(i) = k^{n-i}$ employees with salary $s(i) = \gamma^i s_0$. Eliminating i shows that the number of employees with salary s is $N(s) = k^n(s/s_0)^{-(\alpha+1)}$, where $\alpha + 1 = \log k / \log \gamma$. The cumulative distribution $N(s > S)$ is a power law in S with tail exponent α. Note that if $\log k < \log \gamma$ then the manager at each level makes more than all the employees immediately below her, and in the limit $N \to \infty$ almost all the income is paid to the CEO.

Another trivial example is a Cantor set, which is a simple geometric illustration of the connection between power laws and fractals. A Cantor set can be constructed by removing the middle third of the unit interval and then removing the middle third of each remaining interval ad infinitum. It has the property that, in a certain sense made precise below, the size of a Cantor set is a power law function of the scale of resolution with which it is measured. This is true of any fractal (and indeed can be used to define a fractal).

For the simple example above, we can measure the size of the Cantor set by partitioning the interval into equal subintervals of size ϵ. We can define the coarse grained size of the Cantor set at resolution ϵ as $S(\epsilon) = N\epsilon$, where N is the number of subintervals that contain any part of the Cantor set. If $\epsilon = 3^{-i}$, where i is an integer, then the coarse grained size of the Cantor set scales as as power law $S(\epsilon) = \epsilon^{-D}$, where $D = \log 2 / \log 3$ is the fractal dimension. The scale invariance is obvious from the fact that the Cantor set contains an infinite number of self-similar subsets.

We can construct a more general example of a Cantor set by instead successively dividing each interval into k equal subintervals, each $1/\gamma$ smaller than the previous subinterval. Defining the coarse-grained resolution in this case requires some more care, but with an appropriate generalization it scales as $N(\epsilon) = \epsilon^{-\log k / \log \gamma}$. We have chosen the notation intentially to make the analogy to the previous example of a firm clear: The construction of the Cantor set can be pictured as a tree, and the width of each subinterval is analogous to the salaries of the employees. These examples illustrate how power laws are typically associated with hierarchies, even though the underlying hierarchy is not always obvious.

These two examples also illustrate how power laws involve competition between exponentials. For a power law $y = x^\alpha$ the logarithm of y is a linear function of the logarithm of x. If $\log x$ grows linearly, then x grows exponentially, and y also grows exponentially, albeit at a different rate. The exponent *alpha* gives the relative rate of growth of the two logarithms. A closely related fact is that an exponential transformation of an exponentially distributed variable yields a power law distribution. To show this formally, suppose X and Y are random variables, and X is exponentially distributed

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

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

This states the obvious fact that for a power law distributed function we can always make a logarithmic transformation to coordinates where the power law becomes an exponential function. This is a very useful fact, since there are many mechanisms that generate exponential probability distributions, and there are many situations where exponential transformations are natural. 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 growth in salary moving up the tree. In the Cantor set example it comes from the competition between the exponential proliferation of intervals and their corresponding exponential decrease in size.

One of the great early discoveries in economics nicely illustrates the emergence of a power law through the competition between exponentials. This discovery is the St. Petersburg paradox, which Daniel Bernoulli originally published in the Commentary of the St. Petersburg Academy in 1730 [66]. Consider a fair game in which the original stake is one dollar. Your opponent tosses the coin. If it is heads, she pays you the stake; otherwise the stake doubles. How much would you pay in order to play this game?

Of course the first question you should ask is, "How much can I expect to win?". This is easy to calculate: The probability of getting tails n times in a row is $p(n) = 1/2^n$, while the payoff is 2^{n-1}. Thus the average payoff is $\sum_{i=1}^{\infty} p(n)2^{n-1} = \sum_{i=1}^{\infty} 1/2 = \infty$. The average payoff couldn't be better. Nonetheless, the most likely outcome is that you make only a dollar. The distribution is heavily skewed – even though good outcomes are extremely favorable, such outcomes are rare. Bernoulli argued that the infinite payoff is misleading, and that what one should do instead is add the expected utility. If the marginal utility of wealth is decreasing, the bet is not so attractive. He suggested that a more realistic measure of utility might be the logarithm of wealth, in which case the expected utility is only about four dollars[15].

In fact, the St. Petersburg experiment can never be performed, because no one would ever offer to take the other side of a bet whose expected payoff is infinite. This is made worse by the fact that the distribution of outcomes

---

[15]In fact, this cannot fully explain the paradox. People are not consistently risk averse, as illustrated by the fact that some people purchase lottery tickets, even when the expectation is unfavorable. Utility depends on many other factors, such as the size of the bet relative to one's current wealth, or whether one is allowed to play the game repeatedly [66].

is extremely fat tailed. In fact it is a power law. To see this, note that because the sum $\sum_{i=n}^{\infty} 1/2^i = 2^{n+1}$, the probability of winning 2^n or more is $1/2^n$. This means that the probability of gaining g or more is $1/g$, i.e. the distribution of outcomes is a power law with tail exponent one.

A real casino will never offer the St. Petersburg bet, but they will let you double your bets. Despite the fact that this strategy no longer has an infinite payoff, it does have the same power law tail of outcomes. Thus, the probability of the casino going broke is unacceptably high. By limiting the bets they guarantee that no one can play such a strategy very far into the tail of large outcomes. With repeated non-doubling bets the law of large numbers guarantees a thin-tailed distribution, and the casino can be quite confident that they will not go out of business due to a large fluctuation.

As in the previous examples, the power law tail of the St. Petersburg gamble (or the related doubling strategy) is 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.

6.2 Maximization principles

One way to derive power laws is by maximizing an appropriate function, possibly under constraints. Examples of possible functions that can be maximized include objective functions, such as expected utility, or entropy functions. Constraints can play a critical role in determining the form of the solution. Of course, one must have a good argument for why it is reasonable to maximize a particular function, or impose a particular set of constraints. Because they provide little detailed information about mechanism, maximization arguments are not always fully convincing, and may not be strongly testable. They are often compatible with more detailed explanations. We begin our accounting of mechanisms with maximization arguments because they operate at a high level and are relatively simple.

6.3 Maximum entropy

Maximizing entropy amounts to assuming that something is as random as it can be subject to constraints. The exponential or Gibbs distribution, for example, is the solution that emerges from maximizing entropy subject to the constraint that the mean takes on a fixed value. This assumption, which is the underpinning of classical statistical mechanics, is very natural in

a physical system where energy is conserved, and the behavior is otherwise random. Similarly, if one imposes a constraint on the variance as well as the mean, the solution is a normal distribution.

A power law, in contrast, emerges from maximizing the entropy under a constraint on mean value 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 constraints that $\int (\log x) p(x) dx = C$ and $\int p(x) dx = 1$, where C is a positive constant. Constructing the Lagrangian and setting the functional derivative with respect to the probability distribution to zero gives

$$\frac{\partial}{\partial p(x)} [\int p(x) \log p(x) dx + \lambda (\int \log x p(x) dx - C) - \kappa (\int p(x) dx - 1)] = 0,$$

where λ and κ are Lagrange multipliers. This has the solution $p(x) = Kx^{-\lambda}$. Assuming the power law is defined on the domain $[a, \infty]$, the constant $K = (\lambda - 1)a^{\lambda - 1}$ is determined by the normalization condition, and the scaling exponent $\lambda = 1 + 1/(C - \log a)$ is determined by the constraint on the mean of $\log x$.

Perhaps the earliest use of this explanation for a power law is due to Mandelbrot, who offered it as an explanation of Zipf's law for word frequencies [56]. Zipf's law states that the probability of occurence of a given word is a power law function of its rank, $p_j = Kj^{1/(\alpha - 1)}$ [29, 86]. The rank j of the most common word is one, the second most common two, etc. Mandelbrot's hypothesis is that languages roughly maximize the information communicated for a given cost of transmission. The key point in his argument is that the cost of transmission of a given word is roughly proportional to the logarithm of its rank. Suppose, for example, that we simply encode words by their rank. This code has the advantage that the most common word, encoded by "1", has the shortest sequence, and less common words have longer sequences. In base M, the number of digits required to send a word of rank j is roughly log j. Sending the maximum information is equivalent to maximizing the entropy. Thus, substituting j for x, and replacing the integrals by sums, the argument above implies that p_j should be a power law. Mandelbrot discusses several different efficient codes for encoding words and shows that they all have this basic property.

Non-extensive entropy   As mentioned in the introduction to this chapter, in physics power laws are often associated with nonequilibrium behavior. In contrast, when a system is in physical equilibrium its entropy is at a maximum, and under normal constraints on its energy or variance, the relevant

distributions are exponentials are normals. Non-extensive entropy is a generalization that attempts to characterize systems that are out of physical equilibrium. It naturally results in power laws.

For standard problems in statistical mechanics the entropy is an extensive quantity. This means that the entropy of a system is proportional to its volume – a system of twice the size will have twice the entropy. 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. A good example is a hard sphere gas, where particles interact only when they collide. Such a system comes to physical equilibrium quickly, where the energies of the particles have an exponential distribution.

There are some systems, however, with very long-range interactions, that are extremely slow to fully reach equilibrium. This is because distant parts of the system interact with each other, so it is not possible to assume independence. Such behavior is seen in simulations of particles, such as stars, interacting under the influence of long-range forces, such as gravity. In such cases the approach to equilibrium may be so slow that the assumption of equilibrium fails for any practical purpose. In simulations of galaxy formation, for example, starting from an arbitrary energy configuration, the distribution of the kinetic energy of the stars will quickly approach a power law distribution, and will settle into an exponential distribution only after an extremely long time. The system has a very long-lived transient in which its properties are described by power laws. In some cases the time needed to reach equilibrium is so long that for all practical purposes it never happens.

In circumstances it can be shown that a good description of the statistical properties of the system during its long-lived power law state can be obtained through the use of a generalization of the entropy. This generalization is called the Tsallis entropy,

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

$p(x)$ is the probability density associated with state x, and q is a positive integer that depends on factors such as how long range the interactions are. 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 $q \to 1$ the Tsallis entropy reduces to the standard entropy.

To see why the Tsallis entropy is nonextensive, assume a system can be described by a joint probability density $p(x_1, x_2)$. If x_1 and x_2 are independent, then $p(x_1, x_2) = p(x_1)p(x_2)$, and the normal entropy is additive, i.e. $S_1(p(x_1, x_2)) = S_1(p(x_1)) + S_1(p(x_2))$, due to the additivity of the logarithm of products. When $q \neq 1$, however, this is no longer true. The term "non-extensive" harks back to problems in statistical mechanics where the variables x_1 and x_2 represent the state of a system in two different physically separated regions, but the point about non-additivity under independence holds more generally.

By analogy with the maximum entropy principle, we can also maximize the Tsallis entropy. If we constrain the mean (which is natural e.g. in the case where x represents energy), then the Lagrangian is

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

This has the solution

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

where K 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 $\alpha = 1/(1 - q) - 1 = q/(1 - q)$. In a similar manner to the above calculation, by constraining the variance as well as the mean it is also possible to derive a generalized form of the normal distribution (which also has power law tails). It can be shown that maximizing the Tsallis entropy gives the correct solution for several problems, such as nonlinear diffusion according to a multiplicative Langevin process (Section 6.4), or the solution of some nonlinear differential equations at bifurcation points (Section 6.8). It also gives good empirical fits in many situations, such as the distribution of the energies of stars in a simulation of galaxy formation or the number of transactions in a given length of time in a financial market.

The Tsallis entropy has generated controversy because its underpinnings in most circumstances remain unclear. While the functional forms of the power law generalizations of the exponential and normal distributions seem to empirically fit a broad range of different data sets, their superiority to alternative power law distributions is not always clear.

Maximizing utility 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) [19, 20]. Put in economic terms, their assertion is that under some circumstances, maximizing a risk-neutral utility function results in solutions with power law tails. For example, using an argument that parallels Mandelbrot's derivation of Zipf's law for word frequencies given above, they have proposed that the power law distribution of file sizes observed on the internet maximizes storage efficiency.

Another example considers an idealized model of forest fires [20]. In this model a forester is charged with maximizing the harvest of trees, in a highly idealized setting where the only concern is forest fires. Trees are planted on a grid. Fires are limited by constructing firebreaks, which are built by leaving grid spaces blank. Once in every generation, just before the tree harvest, there is a spark, which lands at a random grid site. It starts a fire that burns all trees that are not protected by a fire break. To contain the fire, the firebreaks must fully enclose a region around the original spark. I.e., the fire will spread as much as it can – while it cannot cross empty sites, it can burn around them unless the firebreak forms a "wall" that it cannot cross. An optimal configuration of firebreaks separates the forest into contiguous regions of trees, each enclosed by a firebreak. The distribution of the size of the fires is precisely the distribution of the size of these regions. The tension is that firebreaks consume space that could otherwise be used for trees. An optimal solution maximizes the number of trees that are harvested, finding the best compromise between the number of trees that are lost to fire and the number of grid sites that are lost to firebreaks.

Carlson and Doyle showed that if the spatial distribution of sparks is continuous but nonuniform, the distribution of fire sizes is a power law. They argue that the characteristics of this solution are quite general: High-performance engineering in complex environments often leads to systems that are robust to common perturbations but fragile to unusual ones. For the forest system, the optimal system of firebreaks achieves good yields on average, but large fires are more frequent than one would expect, due to the power law tails. Furthermore, the system is fragile in the sense that perturbations to the firebreaks or changes in the spark distribution can lead to disastrously sub-optimal performance.

This model has been criticized for several reasons. Although Carlson and Doyle argue that their solution fits the distribution of forest fires very well, it fits the distribution of naturally occuring forest fires, where there is no good argument for the construction of an optimized system of firebreaks. The power law in the forest model derives from the two-dimensional geometry of the problem, in particular the fact the area enclosed by a curve

is proportional to its length. Thus, the power law is built into the problem, and it is not clear that this will generalize to problems without such geometric constraints.

Finally, this model has been criticized because its results depend on risk neutrality. As was pointed out in [71], if one puts risk aversion into the utility function of the forest manager, the power laws disappear. That is, suppose that instead of maximizing the average tree harvest, the forest manager maximizes a risk averse utility function, such as a logarithm or a power law, which gives a strong penalty for huge disasters. The resulting distribution of forest fires no longer has a power law tail. This approach was jokingly called Constrained Optimization with Limited Deviations or "COLD". It suggests that the power law in the original HOT solution is not of such general origin.

The St. Petersburg paradox discussed earlier is another example that involves power laws and utility maximization. While the power law distribution of outcomes is consistent with utility maximization, the power law is really inherent in the setup of the problem, and does not depend on what utility one maximizes in deciding what the bet is worth. The power law came about simply because successively doubling the bet leads to extremely fat tails in the distribution of payoffs. In contrast, the power law in the HOT forest fire model depends on the maximization of yield (but disappears with 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.

6.4   Multiplicative processes

Simple multiplicative random processes can generate fat tailed distributions. Multiplicative random processes naturally occur in many different settings, such as models of feedback, growth, and fracture, and they are a common cause of power laws. A pure multiplicative process gives a log-normal distribution, which is fat tailed but is not a power law. However, small modifications 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

$$x(t + 1) = a(t)x(t) \tag{11}$$

33

where $x(0) > 0$ and $a(t)$ are positive random numbers. If we iterate the process its solution is trivially written as

$$x(t) = \prod_{i=0}^{t-1} a(i)x(0). \qquad (12)$$

Taking logarithms this becomes

$$\log x(t) = \sum_{i=0}^{t-1} \log a(i) + \log x(0). \qquad (13)$$

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

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\log x - \mu)^2/2\sigma^2}. \qquad (14)$$

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

$$\log f(x) = -\frac{(\log x)^2}{2\sigma^2} + (\frac{\mu}{\sigma^2} - 1)\log x + \text{constant terms}. \qquad (15)$$

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. When this happens a lognormal distribution can have approximate power law scaling over many decades, and as a result lognormals are easily confused with power laws.

Note that in general a pure multiplicative process requires a 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 growth or contraction of a multiplicative process this problem is much more severe. If $E[\log a(t)] < 0$ then the distribution exponentially collapses to a spike at the origin, and if $E[\log a(t)] > 0$ it exponentially blows up. To see the lognormal shape of the distribution one must use an appropriately contracting or expanding scale.

In the case where $E[\log a(t)] < 0$ a pure multiplicative random process can be turned from a lognormal into a power law by simply imposing a barrier to repel $x(t)$ away from zero. The power law can be understood by taking

logarithms and taking advantage of classical results on random walks with barriers. distribution is normally distributed. The condition $E[\log a(t)] < 0$ guarantees that $x(t)$ tends to drift to the left. However, in the presence of a barrier it will pile up against it, and from a standard result the normal distribution for an unrestricted additive random walk is replaced by an exponential probability density of the form $P(x) = \mu e^{-\mu(\log x - b)}$, where b is the logarithm of the position of the barrier and $/logx \geq b$. Exponentiating to undo the logarithmic transformation gives a power law with tail exponent $\alpha = \mu + 1$, as shown in equation 8. A model that is essentially equivalent to this was originally introduced by Champernowne to describe income distribution in 1953 [21].

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

$$x(t+1) = a(t)x(t) + b(t), \tag{16}$$

where both $a(t)$ and $b(t)$ are positive random numbers. This is called a Kesten process [41]. 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 collapse to the origin, 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 standard 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, i.e. the asymptotic probability distribution for a Langevin process maximizes the Tsallis entropy.

This last result demonstrates how maximization arguments can be compatible with a more detailed microscopic prescription, and justifies what might otherwise seem like the *ad hoc* nature of the Tsallis entropy. When it is possible, the latter is of course always preferable, since it forces one to identify the source of the nonlinearities, and produces more detailed and hence more testable predictions.

6.5 Mixtures of distributions

A mixture of distributions combines individual distributions with different scale parameters, i.e.

$$f(x) = \int g(\sigma) p_\sigma(x) d\sigma \qquad (17)$$

where σ is the scale parameter (e.g. the standard deviation) of each distribution $p_\sigma(x)$. This is often offered as an explanation for fat tails in the standard equilibrium model of price fluctuations: Since the information arrival rate varies, the standard deviation of price fluctuations varies. Even though the instantaneous distribution at any given time might be a thin-tailed normal distribution, when these are blended together for different times, the result is a fat-tailed distribution. Therefore, according to this explanation, the fat tails of price fluctuations 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. 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.

For instance, what mixture function will combine exponential distributions to get a power law?

$$f(x) = \int g(\sigma) e^{-\sigma x} d\sigma \qquad (18)$$

It is possible to show that the function $g(\sigma)$ that will give a power law with tail exponent α is $g(\sigma) \sim 1/\sigma^{2+\alpha}$. 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 die out sufficiently fast [82]. Therefore this result applies to mixtures of normal distributions, and makes it clear that a power law mixture is required. To explain the power law nature of price fluctuations in terms of variation in the information arrival rate, one needs to explain why information arrival has a power law distribution in the first place.

There do in fact exist non-power law mixtures of thin-tailed distributions that give rise to power laws. An important example is an exponentially weighted mixture of log-normal distributions. This occurs naturally, for example, in the context of a multiplicative process with a distribution of stopping times. Consider the process $x(i+1) = a(i)x(i)$ of Section 6.4, but

now assume that the stopping time t is itself an exponentially distributed random variable with density function $p(t) = \lambda e^{-\lambda t}$. For any fixed t the distribution is lognormally distributed, but when we have a mixture of stopping times we have to weight each log-normal by its stopping time, which also affects the scale parameter of the log-normal. It is straightforward to show that this integral gives is a power law, called the double Pareto distribution [79]. The name is chosen 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 λ of the stopping time[16].

This mechanism can be used to provide another possible explanation of Zipf's law for word frequencies. Suppose that monkeys type randomly on a keyboard with M characters plus a space. Assume they hit the space bar with probability s, and they type non-space characters with probabilities p_i that differ for each character. Then the probability of typing any particular word of length l will be approximately log-normally distributed for large l. The probability that a word has length l is $(1 - s)^l s$, which for large l is roughly exponentially decreasing with l. Thus, approximating the sums by an integral, the distribution of word frequencies is a double Pareto distribution. This argument was originally given by Miller [67] (but under the assumption that all non-space characters have the same probability, which also gives a power law with $\alpha = 1/(1 - \log_M(1 - s))$, where M is the number of non-space characters.

6.6   Preferential attachment

Preferential attachment was originally introduced by Yule to explain the distribution of species within genera of plants, and is perhaps the oldest known mechanism for generating a power law. Yule's idea was 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, giving rise to a very fat tailed distribution. The argument was developed by Simon and proposed as a possible explanation for a variety of other phenomena, including Zipf's law for 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.

---

[16]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 $x^{-\alpha}$.

We will summarize the basic argument in the context of Zipf's law for word frequencies, but with changes of a few details the same argument can be applied to a variety of examples (including more recently the distribution of sites to a given link on the World Wide Web [10, 68]). Consider a partially completed text that is t words long. Assume that with probability λ an author chooses a new word at random, and with probability $1 - \lambda$ she chooses a previously used word, with probability proportional to the previous number of occurrences of that word. Following the argument originally given by Simon [80, 39] this gives a power law distribution of word frequencies. This can be derived via a master equation as follows:

For a text of length t, let $N_j(t)$ be the number of words that occur j times. (We will drop the argument t when it is obvious). For example, if "the", "of", and "to" are the only words that appear one hundred times, then $N_{100} = 3$. What is the probability that N_j will increase, i.e. that $N_j(t+1) = N_j(t)+1$? Since at time $t+1$ the next word occurs j times, at time t it must have occurred $j-1$ times. If the word already exists in the text, the probability that a word that occurs $j-1$ times will be chosen is proportional to the number of words that occur $j - 1$ times, weighted by how much it already occurs. Therefore, when $j > 1$ the probability that N_j increases when the next word is chosen is proportional to $(j - 1)N_{j-1}$. Similarly the probability that N_j decreases (due to a word that previously occurred j times being chosen, so that it now occurs $j + 1$ times) is proportional to jN_j. Putting this together gives

$$E[N_j(t + 1) - N_j(t)] = K(t)((j - 1)N_{j-1}(t) - jN_j(t)). \qquad (19)$$

where $K(t)$ is a normalization constant that turns these into proper probabilities. The case $j = 1$ has to be handled separately. In this case the rate of change is just the probability of a new word being chosen, times the probability that a word that occurs once at t is chosen (so that it now appears twice, decreasing N_1). This gives

$$E[N_j(t + 1) - N_j(t)] = \lambda - K(t)N_1(t).$$

The normalization constant can be computed from the condition that the probability that a previously used word is chosen is $1 - \lambda$. It therefore satisfies the normalization condition $\sum_i = 1^k K(t)jN_j = K(t)\sum_i = 1^k jN_j = 1 - \lambda$. Since $jN_j(t)$ is the number of words that appear j times, multiplied by the number of times each word occurs, the sum over all j is just the total number of words in the text, i.e. $\sum_i = 1^k jN_j = t$. This implies $K(t) = (1 - \lambda)/t$.

Suppose we make the steady state assumption that for large t the word frequencies converge to constant values r_j, i.e. words that occur in the text

N_j times constitute a fraction r_j of the total number of words in the text. This means that the average number of occurrences of each word grows as $E[N_j(t)] = r_j t$, which implies that $E[N_j(t+1) - N_j(t)] = r_j$. With some rearranging of terms, after plugging in the expression for the normalization constant $K(t)$, equation 19 becomes

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

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

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

This has the solution $r_j = r_0 j^{-(2-\lambda)/(1-\lambda)}$, which is a power law with tail exponent $\alpha = (2-\lambda)/(1-\lambda) - 1 = 1/(1-\lambda)$. When λ is small this gives a tail exponent a little bit greater than one, which matches the empirical result.

6.7 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 [23, 81, ?]. 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 [11]. 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 [83, 84].

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 [23, 81]. 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, which is discussed in more detail in Section ??. 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 rate δ, 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

$$E[s] = \frac{\mu}{\alpha} \Phi_s(\frac{\sigma}{p_c}, \frac{\Delta p}{N_c}), \tag{20}$$

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

$$D = \frac{\mu^2 \delta}{\alpha^2} \Phi_D(\frac{\sigma}{p_c}, \frac{\Delta p}{N_c}). \tag{21}$$

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 [32]. 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 [33, 32, 31].

6.8   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[17] 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

$$x(t) = At^{1/(\beta-1)}. \tag{22}$$

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 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 [76], 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

---

[17]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.

the deterministic mapping

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

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 [40]. 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 [9]. 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 22 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 [75]. However, subsequent studies seem to suggest that there are statistically significant variations in the tail exponents of different assets [31].

6.9 "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 $\alpha = n$ [82].

As already discussed in Section 2, 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.

47

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 6.8, 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.

7 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 [72]. 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. [33] 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 6.4). 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.

8   References

References

[1] N. Amaral, L. A, S. V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M. A. Salinger, H. E. Stanley, and M. H. R. Stanley. Scaling behavior in economics: I. empirical results for company growth. *Physica A*, 244(1-4):1–24, 1997.

[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] D. Avnir, O. Biham, D. Lidar, and O. Malcai. Is the geometry of nature fractal? *Science*, 279:39–40, 1998.

[4] R. Axtell. Zipf distribution of U. S. firm size. *Science*, 293:1818–1820, 2001.

[5] R. U. Ayres and R. Axtell. Foresight as a survival characteristic: when (if ever) does the long view pay? *Technological Forecasting and Social Change*, 51:209–235, 1996.

[6] D. K. Backus and S. E. Zin. Long-memory inflation uncertainty: Evidence from the term structure of interest rates. *Journal of Money, Credit and Banking*, 25:681–700, 1993.

[7] R. T. Baillie. Long-memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73:5–59, 1996.

[8] R. T. Baillie, C-F. Chung, and M. A. Tieslau. Analyzing inflation by the fractional integrated arfima-garch model. *Journal of Applied Economics*, 11:23–40, 1995.

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

[10] A.-L. Barabasi, R. Albert, and H. Jeong. Mean-field theory for scale-free random networks. *Physica A*, 272:173–189, 1999.

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

[12] J. Beran. *Statistics for Long-Memory Processes*. Chapman & Hall, New York, 1994.

[13] T. Bollerslev and H-O. Mikkelsen. Modeling and pricing long-memory in stock market volatility. *Journal of Econometrics*, 73, 1996.

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

[15] J-P. Bouchaud, Y. Gefen, M. Potters, and M. Wyart. Fluctuations and response in financial markets: The subtle nature of "random" price changes. *Quantitative Finance*, 4(2):176–190, 2004.

[16] 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.

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

[18] 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 Economic and Financial Models*, pages 3–41. Wiley, New York, 1997.

[19] 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.

[20] J. M. Carlson and J. Doyle. Highly optimized tolerance: robustness and design in complex systems. *Physical Review letters*, 84(11):2529–2532, 2000.

[21] D. Champernowne. A model of income distribution. *Economic Journal*, 63:318–351, 1953.

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

[23] M. G. Daniels, J. D. Farmer, L. Gillemot, G. Iori, and D. E. Smith. Quantitative model of price diffusion and market friction based on trading as a mechanistic random process. *Physical Review Letters*, 90(10):108102–108104, 2003.

[24] F. X. Diebold and G. D. Rudebusch. Long-memory and persistence in aggregate output. *Journal of Monetary Economics*, 24:189–209, 1989.

[25] 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(1):83–106, 1993.

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

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

[28] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

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

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

[31] J. D. Farmer, L. Gillemot, F. Lillo, S. Mike, and A. Sen. What really causes large price changes? *Quantitative Finance*, 4(4):383–397, 2004.

[32] J. D. Farmer and F. Lillo. On the origin of power laws in financial markets. *Quantitative Finance*, 314:7–10, 2004.

[33] 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.

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

[35] 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.

[36] M. Greene and B. Fielitz. Long-term dependence in common stock returns. *Journal of Financial Economics*, 4:339–349, 1977.

[37] A. C. Harvey. Long-memory in stochastic volatility. Working paper, London School of Economics, 1993.

[38] U. Hassler and J. Wolters. Long-memory in inflation rates: International evidence. *Journal of Business and Economic Statistics*, 13:37–45, 1995.

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

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

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

[42] 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.

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

[44] F. Lillo. Agents in the market. 2004 in preparation.

[45] F. Lillo and J. D. Farmer. The long memory of the efficient market. *Studies in Nonlinear Dynamics & Econometrics*, 8(3), 2004.

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

[47] 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.

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

[49] I. N. Lobato and C. Velasco. Long-memory in stock-market trading volume. *Journal of Business & Economic Statistics*, 18:410–427, 2000.

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

[51] 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 Finance*, 1(2):211–248, 1994.

[52] 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.

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

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

[55] O. Malcai, D. Lidar, O. Biham, and D. Avnir. Scaling range and cutoffs in empirical fractals. *Physical Review E*, 56(3):2817–2828, 1997.

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

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

[58] B. Mandelbrot. When can price be arbitraged efficiently? a limit to the validity of the random walk and martingale models. *Review of Economics and Statistics*, 53:225–236, 1971.

[59] B. Mandelbrot. Statistical methodology for non-periodic cycles: from the covariance to r/s analysis. *Annals of Economic and Social Measurement*, 1:259–290, 1972.

[60] B. Mandelbrot. Limit theorems on the self-normalized range for weakly and strongly dependent processes. *Z. Wahrscheinlichkeitstheorie verw, gebiete*, 31:271–285, 1975.

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

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

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

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

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

[66] R. Martin. The st. petersburg paradox, 2001.

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

[68] 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.

[69] 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.

[70] J-F. Muzy, J. Delour, and E. Bacry. Modelling fluctuations of financial time series: from cascade process to stochastic volatility model. *European Physical Journal B*, 17(3):537–548, 2000.

[71] M. E. J. Newman, M. Girvan, and J. d. Farmer. Optimal design, robustness and risk aversion. *Physical Review Letters*, 89(2):028301, 2002.

[72] M. Nirei. Threshold behavior and aggregate fluctuation. Working paper 03-09-049, Santa Fe Institute, 2003.

[73] M. Nirei and W. Souma. Income distribution dynamics: A classical perspective. 2003.

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

[75] 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.

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

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

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

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

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

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

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

[83] 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.

[84] 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.

[85] S. T. Ziliak and D. N. McCloskey. Size matters: the standard error of regressions in the american economic review. *Journal of Socio-Economics*, 33(5):523–675, 2004.

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

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

[88] 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