UNIVERSITY *of* LIMERICK

OLLSCOIL LUIMNIGH

University of Limerick

Final Year Project

Power-Law Distributions and Stock Market Returns

Enda Carroll
13039407

supervised by
Prof. James Gleeson

March 23, 2016

Contents

1 Introduction

The power-law distribution has gained significant attention in the scientific community in recent years for a variety of reasons. Instances where they occur span a wide variety of disciplines. Particular examples of where it arises are the population of cities [22], the number of calls received by customers of AT&T's long distance telephone service in the United States, the severity of terrorist attacks, the intensity of solar flares, book sales in the United States and the number of citations scientific papers receive [3]. The distribution also possesses interesting characteristics that merit research, such as its *scale invariance* and the underlying mechanisms that give rise to its existence in real world phenomena.

A power-law can be described by the relative change in one quantity varying inversely proportional to another. This law or regularity can also be referred to by the Pareto Distribution or Zipf's Law. The distribution is a member of the family of 'fat' or 'heavy' tailed distributions which describe systems where rare or extreme events happen more often than systems described by the Gaussian or Normal distribution. The Gaussian model assigns very little probability to values far from the typical value and therefore underestimates the occurrences of rare incidents. As a result the approach to describing Normal behaviour cannot be adapted to describe data that follow a power-law. The ability to describe the frequency of extreme observations has obvious practical benefits and is another reason for the popularity of power-law in academic research.

The study of power-law distributions hold considerable interest for those in the field of model fitting or model selection. As we will discuss in this paper it is difficult to accurately detect power-law behaviour in empirical data. Power-law behaviour is typically identified in the tail of a distribution and as a result experimental data of such are usually very noisey [1]. Subsequently correct detection and quantification can be difficult.

One area that has seen an increasing interest for study and research for mathematicians and physicists alike is that of finance and the financial markets. The financial markets today are incredibly complex systems with many interacting elements. As a consequence it is very difficult to gain an understanding of the nature of these interacting elements and the factors that drive the system. As such it has become custom to look towards historical empirical evidence to gain an understanding of these structures in an attempt to develop governing fundamentals. Attempts to examine the stochastic nature of returns, in particular, have shown evidence of power-law behaviour [2, 18]. These results are in contrast to how theoretical finance treats the distribution of returns. The Gaussian distribution is used for its simplifications in analytical calculation. In fact, one of the main assumptions of the Black Scholes Option Pricing formula is that the distribution of returns is Gaussian.

The aim of this paper is to gain an understanding of the power-law distribution. We explore the mathematical techniques developed in other papers for testing empirical data, with a particular emphasis on financial market data, for power-law behaviour. In section 2 we examine previous discussions on power-law distributions by reviewing literature on this topic from various authors with backgrounds in different fields. In section 3 we introduce the financial data we examine in this paper, outline how it was acquired and carry out some initial descriptive analysis including testing our data for Gaussian behaviour. In section 4 we discuss the characteristics of a power-law by examining the mathematics behind the distribution. In section 5 we look at ways of identifying power-laws discussed in [22], including plotting data on log-log scales producing the linear relationship which has become synonymous with identifying the distribution. In section 6 we discuss approaches to fitting power-law distributions to empirical data. Section 7 deals with how to determine the goodness-of-fit of a power-law model to empirical data before applying these methods to our financial data. Lastly, we conclude our findings and discuss possible avenues for further work. What we find is that it is very difficult to confidently determine power-law behaviour, especially for small sample data sets - we find not normal for small timescales and approaching gaussian as increasing timescale

2 Literature Review

In this section we review and compare work that has been completed previously in other papers and journals on power-law distributions. The main papers discussed here are *Power laws, Pareto distributions and Zipf's law* (2005) by Newman [22], *Power-Law Distributions in Emperical Data* (2009) by Clauset et al. [3], *Parameter estimation for power-law distributions by maximum likelihood methods* by Bauke [1], *Critical Truths About Power Laws* by Stumpf and Porter [27] and *Scaling of the Distribution of Fluctuations of Financial Market Indices* (1999) by Gopikrishnan et al. [13].

Origins of power-law behaviour have been a topic of discussion in the scientific community for the last few decades. In [22], Newman examines some examples of empirical data that are believed to follow a power-law distribution and the theories to explain why. He begins by discussing methods to measure power-law data and their potential pitfalls. He mentions a standard strategy involving taking the histogram of power-law distributed data on logarithmic scales which produces the straight line commonly associated with the power-law. He discusses another method called *logarithmic binning* before going on to discuss a superior method of plotting the *complementary cumulative distribution function* of the data. He continues by giving real world examples of distributions following power-law behaviour before discussing the mathematics of the power-law distribution itself. He then finishes by examining the mechanisms that can create power-law behaviour, namely, the *Yule Process* [25, 30] and the concept of self-organized criticality.

We will discuss in section 3, power-law distributions are hard to establish and detect as the behaviour only occurs over a certain range. In [3], Clauset et al. attempt to provide a framework of statistical techniques where one can identify and quantify power-law behaviour. They claim that given the vast span of disciplines where power-law behaviour has been reported to occur (physics, biology, social science, computer science and statistics to name but a few) rigorous methods for analysing and detecting power-law data have not always been used. This, as they prove in some cases, leaves open the possibility that empirical data previously conjectured to follow a power-law do not. They examine methods used in the past which can produce inaccurate results in most cases, such as least-squares regression for estimating the scaling exponent α. They establish more well-founded mechanisms like maximum-likelihood fitting, goodness-of-fit tests by way of the Kolmogorv-Smirnov statistic and likelihood ratios to test for power-law characteristics. They test these procedures on a wide range of empirical data conjectured to have power-law behaviour to which they both confirm and disproves these theories.

In [1], Bauke discusses methods for determining the exponent of a power-law distribution using linear least squares regression and maximum likelihood methods. He examines the performance of a least squares fit to accurately estimate α and compares it to the maximum likelihood method. Bauke shows that the accuracy depends enormously on the choice of graphical method. He also shows that these techniques are all inferior to the maximum likelihood method which he develops for the discrete power-law case before discussing numerical methods to implement this approach.

Given the fact that power-laws have been reported in numerous systems from a wide variety of fields, the proper consideration and approach to declaring and identifying power-law behaviour has not always been adhered to. The discussion by Stumpf and Porter in [27] follows in this vain. They argue that some instances claimed to follow power-law behaviour lack the statistical backing and some lack sufficient empirical evidence. Even in the cases where power-law behaviour is supported by sound statistics, without a solid underlying generative mechanism or theory to support it, the claim is of no real use. They voice caution when reporting power-law behaviour and offer a rule of thumb that should be taken into consideration: the linear relationship produced by plotting a power-law on logarithmic scales should extend to at least two orders of magnitude in both axes.

In [13], Gopikrishnan et al. investigate power-law behaviour in the financial markets. They analyse stock market returns on different time scales for power-law characteristics. A stock market return is the return on a purchase of a stock i.e.the much money I will make/lose if I invest my money in a stock now and sell it later. They test a selection of stock market returns data, all of various time scales from the S&P 500 index. The time scales or intervals looked at by Gopikrishnan et al. vary from one minute (i.e. return on the purchase of a stock, then sold one minute later) to daily and monthly intervals. They find for a time scale of less than four days a power-law distribution occurs and for intervals greater then four days they observe a slow convergence to Normal behaviour. They check these results against other markets such as the NIKKEI and the Hang-Seng index to confirm the robustness of their findings. They also discover that randomizing the time series of returns, essentially removing the time dependencies by shuffling the data points up, causes the series of returns to rapidly approach normal behaviour for increasing time scales. For the analysis in this paper, we follow Gopikrishnan's et al. definition of a return and test for power-law behaviour in financial markets on similar time periods.

3 Financial Data

Mathematicians and physicists have been attracted to the study of financial markets, in particular economic time series analysis, of of late for several reasons. Economic time series, such as stock market indices are driven by a large number of interacting systems, and so is an example of a complex system. It is hoped that the experience of studying complex physical systems might lead to new discoveries in economics. Also the availability of large amounts of high frequency financial data aided by increased computing power allow for research producing accurate results [11, 16]. Recent studies have shown that that the probability distribution of stock price changes display pronounced tails indicating power-law behaviour [9, 13], hence the motivation for this paper.

In this section we describe the financial data sets analysed in this paper. We define what we mean by a return on a stock or market index and the time scales we study them on. As well as investigating the distribution of a return, we also look at the distribution of the volume of trades for a given time interval. Gopikrishnan et al. demonstrate that stock market returns on the S&P 500 to tend toward normality for large time scales in their paper [13]. We will also look into methods of testing our financial data for normality.

In this paper we examine returns on two indices: the FTSE 100 index from 04/04/1986-18/03/2016 and the S&P 500 index from 03/01/1950-18/03/2016. The FTSE 100 is a collection of 100 of the top performing companies listed on the London Stock Exchange and began in 1984 at a base of 1000 (see figure (1)). The S&P 500 is a collection of the top 500 companies listed on the New York Stock Exchange. The index in its current form was first introduced in 1957 but originally began in 1950. In figure (1) we see the performances of both indices over their life spans. Initially data was pulled from the 'Yahoo Finance' website for both indices and preliminary testing was carried out. A few discrepancies were discovered from this data such as: days were included on the FTSE 100 that, upon inspection of the British calender were bank holidays where no trading had occurred; some days on the FTSE 100 had only a closing price quoted and not an opening price and visa

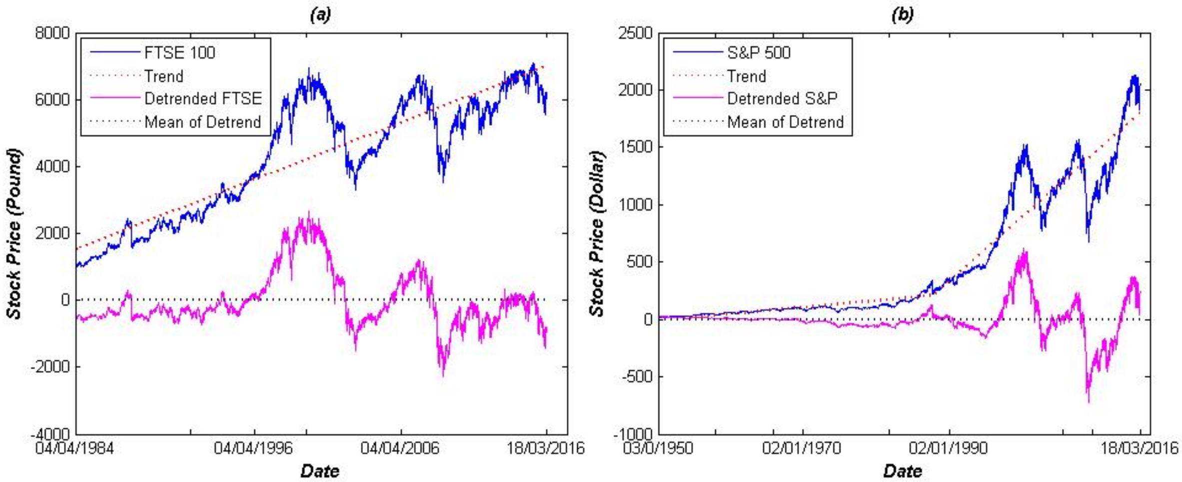


Figure 1: Performance of (a) the FTSE 100 index over the 32-year period 1984-2016 and (b) S&P 500 index during the 62-year period 1962-2014. Both indices display positive linear trend growth over the span of their lifetimes. The S&P 500 displays a shallower growth period in the beginning but around the 1980's it began to trend more upward. Also shown are the trend lines and detrended prices for both indices, giving a better indication of the price fluctuations throughout lifetime of the indices.

versa (similar problem with the S&P 500); and for the daily volume of trades on the FTSE 100 there appeared to be an anomaly in the data (see figure (2)). At first glance there appears to be some sort of haircut on the volume amounts around the 1500 to 2500 mark. Initially it was thought that there may have been some sort of cap on the volumes of trades for this period but could not find anything to confirm this after some background research. Also the figures quoted for the weekly and monthly volume of trades were the average volumes for that period and not the actual values, for example the weekly volume figures would divide in exactly to the sum of the respective daily volume figures. We were also restricted in the data available. There was no intraday or tick (minimum upward or downward movement in the price of a security. Tick = 1c movement) data available on the internet, at least free of charge.

After experiencing these problems we concluded that the data could not be trusted for our an analysis and attempted to acquire data from a different source. We managed to gain access to a Bloomberg terminal and gathered the data from there. The data sourced from here did not contain any of the issues we experience with the Yahoo Finance data (see figure (2)). The Bloomberg terminal also provided us access to intraday data but were limited to how far we could go back depending on the size of the time interval, e.g. we could only go back as far as 6 months for a 30 minute period. Nonetheless we were now able to calculate the returns on these indices.

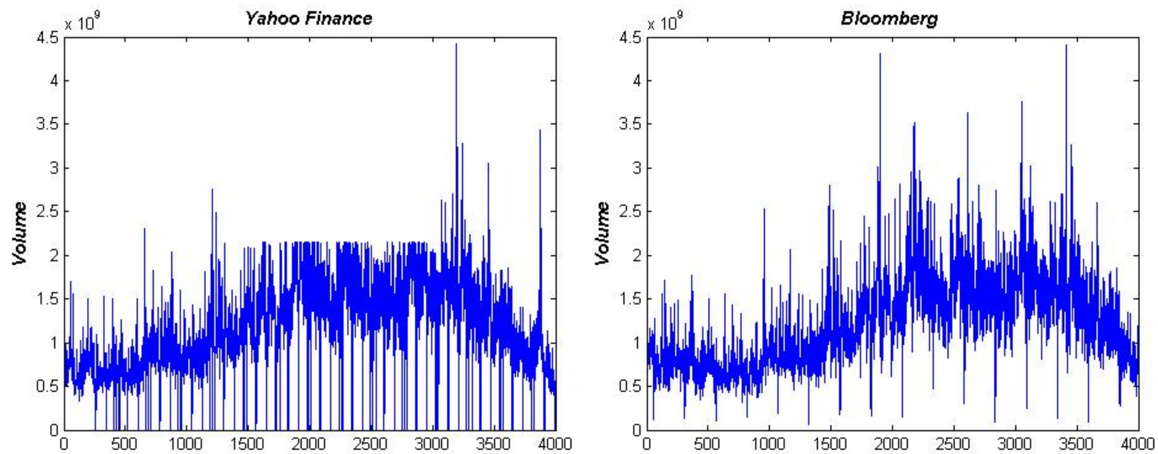


Figure 2: FTSE volume

We examine returns on both indices over different time intervals δt, where δt varies from 5, 15 and 30 minute, to 1 and 5 day, to 1 and 2 month periods. A return for a time series $S(t)$ of a stock price or market index is defined as

$$G_{\delta t}(t) = ln(S + \delta S) - ln(S), \tag{1}$$

where $G_{\delta t}(t)$ is the return over the time scale δt, and $S(t + \delta t)$ is the price of the stock at some time δt into the future. With a little manipulation and use of the Taylor's series expansion for $ln(1 + x)$ we find

$$G_{\delta t}(t) = ln(S + \delta S) - ln(S) = ln\left(\frac{S + \delta S}{S}\right) = ln\left(1 + \frac{\delta S}{S}\right) = \frac{\delta S}{S} + \mathcal{O}\left(\left(\frac{\delta S}{S}\right)^2\right). \tag{2}$$

For small δS, i.e. small changes in the stock price $S(t)$, the higher order terms becomes negligible and a return can approximated as

$$G_{\delta t}(t) \approx \frac{\delta S}{S} = \frac{S(t + \delta t) - S(t)}{S(t)}. \tag{3}$$

In our case we have $\delta t = 5m, 15m$ and so on. For each interval we looked at an opening and closing price for the respective intervals were quoted. We can generalise the formula to the closing price of the respective time interval minus the opening price all over the opening to get the percentage return on the market

$$Return = \frac{Closing - Opening}{Opening}. \tag{4}$$

Since we are dealing time series data it would make sense to apply some theoretical time series analysis. In time series analysis it is typical to want to work with *stationary* data. A stationary series is one that has a constant mean value throughout the series and its autocovariance depends on the time-lag only. The most common departure from stationarity is the presence of *trend* in a time series. If we can somehow measure the trend in a series and remove we can then work on the detrended series. Removing a trend from the data enables you to focus your analysis on the fluctuations in the data about the trend. A linear trend typically indicates a systematic increase or decrease in the data [5]. Consider the time series of stock prices S_t, we can write

$$S_t = \mu_t + X_t \tag{5}$$

$$\Rightarrow \hat{X}_t = S_t - \underbrace{\mu_t}_{\text{trend}}, \tag{6}$$

where $\hat{X}_t$ is are the residuals leftover from removing the trend μ_t. There are several ways to remove trend from a series, the most common of which is *differencing* or sometimes called *integration*. Integrating a time series is simply computing the differences between consecutive time points of the series. For a time series S_t a δ-order integrated series $\nabla^\delta S_t$, is as follows

$$\nabla^\delta S_t = S_t - S_{t-\delta}. \tag{7}$$

If you were to divide the right-hand-side by the original time series S_t you would get the formula we found for our return in equation (3). This is known as a percentage change transformation of a time series and is often used, particularly in financial settings, where the series evolves as a small percentage change P_t

$$S_t = (1 + P_t)S_{t-1}. \tag{8}$$

So in essence calculating the returns on a time series of an asset removes the linear trend of that time series. To confirm this we have plotted the daily returns and monthly returns of the FTSE 100 in figure (3). Observe that there is no obvious trend, compared to that of panel (a) in figure (1) where the time series of the FTSE 100 displays a positive linear trend.

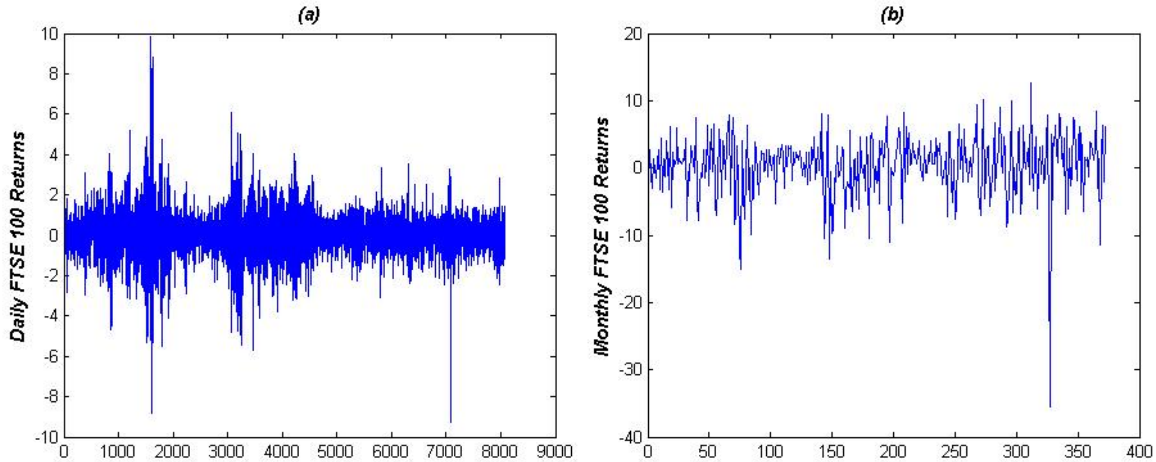


Figure 3: Returns on FTSE 100

Also included in our study are the volumes of trades on the FTSE 100 and S&P 500 for a given time interval. The volume of trades is defined as the total quantity of contracts bought and sold on a particular asset during a trading day. This is not to be confused with the number of trades on an asset. For example a broker could enact 8 trades in one day, each trade selling 100 contracts of Google stocks. The number of trades performed by the broker would be 8 but the volume of trades would be $8x100$, 800. The volume of trade numbers are usually reported as often as once an hour throughout the current trading day but these are only are estimates. Final, actual figures are reported the following day. This is why we were limited to volume of trades amounts for daily, weekly and monthly intervals only as no figures were quoted for intraday intervals. Volume tells investors about the market's liquidity. Higher volume means higher liquidity and better order execution. Volume also tends to be higher near the market's opening and closing times, and on Mondays and Fridays. It tends to be lower at lunchtime and before a holiday [9]. The volume of trades for a particular stock for a fixed interval have been reported to follow power-law behaviour also [12].

3.1 Testing For Normality

The Gaussian or Normal distribution is a fundamental assumption widely used in theoretical finance. The most famous case is that of the Black-Scholes Option Pricing formula where it assumes that the percentage change in the stock or return on the stock is normally distributed [14]. Therefore it makes sense to check if our financial data follows this assumption. In this section we will test our financial data for normality. There are many tools for testing normality some of which I will introduce in the subsequent subsections. There is no defined protocol when it comes to testing for normality as each test has its own strengths and weaknesses. The most common procedures however, are using graphical methods (histogram or Q-Q plots) and formal goodness-of-fit tests [24]. According to [10], it is preferable to use both visual tools and tests achieve a strong conclusion. We will find that for smaller time intervals our data does not follow normal behaviour but tends to normal behaviour as the length of the time interval increases.

3.1.1 Kolmogrov-Smirnov Test

One way to test if a particular sample data set follows normal behaviour is the Kolmogorov-Smirnov test (K-S or KS test for short). It was developed by Andrey Nikolaevich Kolmogorov, a 20th-century Soviet mathematician, in the early part of the last century [15]. It was later added to by another Russian Nikolai Vasilyevich Smirnov [26]. The KS test is a nonparametric test to decide if a given sample comes from a population with a specified continuous distribution or to check if two samples have the same continuous distribution. It operates under the hypothesis testing; the null hypothesis is the sample(s) is drawn from a specified or same distribution and the alternative hypothesis otherwise.

The test can also serve as a goodness-of-fit test with certain modifications. If we set the mean and variance of our sample distribution to those of the reference distribution we can then test a goodness-of-fit to the standard normal distribution. This is a special case for this test and is the one we will use for testing our data for normality.

The test works by comparing the cumulative distribution (CDF) of our reference or known distribution $P(x)$ with that of our sample $P_n(x)$. It measures the distances between the CDFs at each point and taking the maximum distance as the test statistic

$$D_n = max_x |P_n(x) - P(x)|. \tag{9}$$

If the distance is too large, for a given significance level, then $P_n(x)$ does not follow our specified distribution $P(x)$ and we reject the null hypothesis.

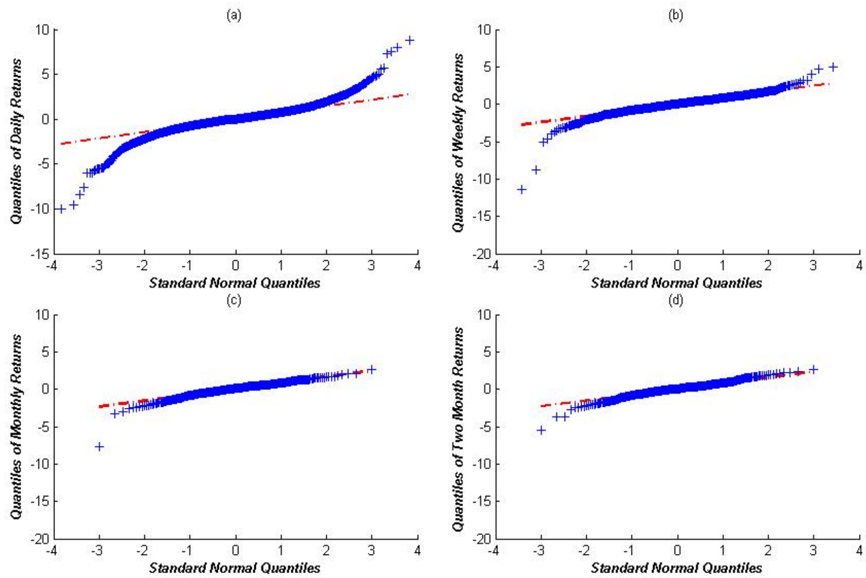
3.1.2 Q-Q Plot

The quantile-quantile (Q-Q) plot is an exploratory graphical tool used to compare two probability distributions by plotting there quantiles against each other. If a distribution X and a distribution Y are identically distributed then the plot of the X-quantiles versus Y-quantiles will produce a straight line configuration of slope 1 pointed towards the origin [28]. The plot can be modified to test if a distribution follows that of a normal distribution. By comparing quantiles of a sample against a known normal distribution, the Q-Q plot will show the sample quantiles linearly following the quantiles of the normal distribution if the sample is itself normal.

3.1.3 Results

To apply these tests we used the built in MATLAB commands *kstest()* and *qqplot()*. The *kstest()* command tests your data against the standard normal distribution returning an output of 1, rejecting the null hypothesis H_0 that the sample is normally distrusted or 0 otherwise. The command can also be adjusted to output a p-value. The p-value of a statistical test is the probability of obtaining a result equal to that or more severe then the one observed given that the H_0 is true. In other words the p-value is a measure of the strength of evidence against H_0. Before the test is performed, a threshold value is chosen, called the significance level of the test, traditionally 5%. A p-value less than this level means there is strong evidence against H_0. In our case if we get a p-value less than 0.05 we reject H_0 and say that our data is not normal. To apply the test we first need to standardize our data sets. We set the mean equal to zero (mean of the standard normal distribution) and the standard deviation equal to one (standard deviation of the standard normal distribution) of each data set. We can now use the test to check our data against the standard normal distribution. The result of these tests can be seen in table (**??**) below.

We can see that the p-values for the daily returns on both indices are very small, a lot smaller then the critical value of 0.05. This means that there is very strong evidence against H_0 in both cases and therefore it should be rejected. This is confirmed in the H_0 column with entries of 1 on both counts. This result means daily returns for both indices are not normally distributed. The same results are achieved for weekly and monthly returns for both indices and returns over two months on the FTSE 100. Notice, however, that the p-value increases as the length of the time interval increases. This would suggest that as the time scale becomes larger the behaviour of the distribution of returns tends to normal behaviour which is consistent with the findings from Gopikrishnan et al. [13]. In fact if we look at returns over two months on the S&P 500 the p-value is 0.0503, which is greater than the critical value of 0.05 and therefore we fail to reject H_0, that the distribution is



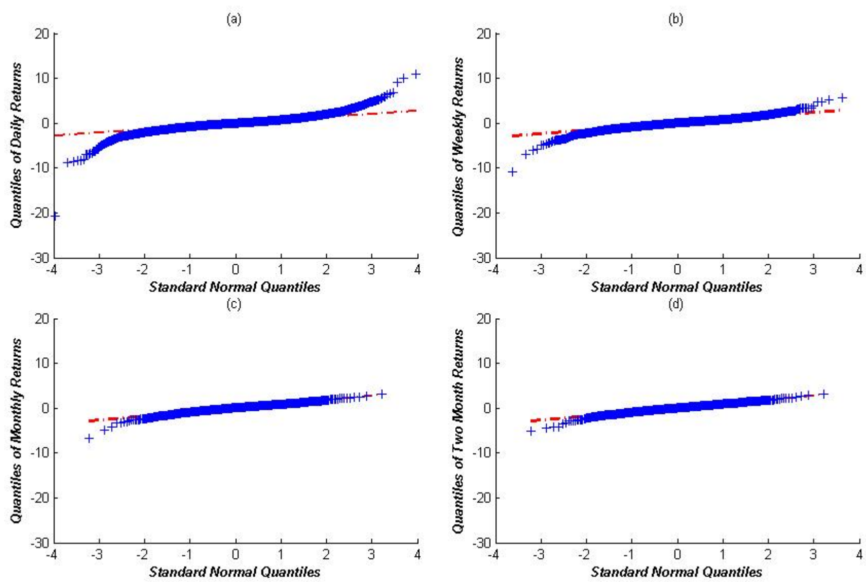


Figure 4: Q-Q plots of the daily (a), weekly (b), monthly (c) and two month (d) returns on the FTSE 100 index.

normal.

A graphical representation of the test is displayed in figure (5) for demonstrative purposes. Here the CDF of daily returns on the FTSE 100 (panel (a) on the left) and daily returns on the S&P 500 (panel (b) on the right) are plotted against the CDF of the stan-

Quantity			KS Test	
			H_0	$p - value$
FTSE 100	Returns	5-Min		
		15-Min		
		30-Min		
		Daily	1	9.9e-040
		Weekly	1	2.1e-005
		Monthly	1	0.0270
		Two Month	1	0.0317
	Volumes	Daily		
		Weekly		
		Monthly		
S&P 500	Returns	5-Min		
		15-Min		
		30-Min		
		Daily	1	2.4e-060
		Weekly	1	8.3e-010
		Monthly	1	0.0100
		Two Month	0	0.0503
	Volumes	Daily		
		Weekly		
		Monthly		
Synthetic Standard Normal			0	0.9611
Synthetic Power-Law				
Word Frequency in Moby Dick				

Table 1: Results of the K-S test applied to returns on the FTSE 100 and the S&P 500 indices. The test was also applied to a random sample of numbers that follow a normal distribution (this was achieved using the *randn()* command.)

dard normal distribution. It is clear that there is some distance between the CDF's of the returns and the CDF of the standard normal distribution meaning that the returns do not folllow normal behaviour. This is highlighted by the vertical black lines in both plots (these black lines were included manually for illustrative purposes only and are not the correct representations of the test statistics).

Next we consider plotting the our data using the *qqplot()* command to back up our findings from the K-S test. The *qqplot()* command plots our data against the standard normal distribution for comparison. If our sample is normal what you will find is that the qunatiles for our data linearly follow the same straight line as the standard normal. The plots of returns on the FTSE 100 and returns on the S&P 500 indices for the various times scales are shown in figure (4) and figure (**??**) respectively. In figure (4)we have plotted the

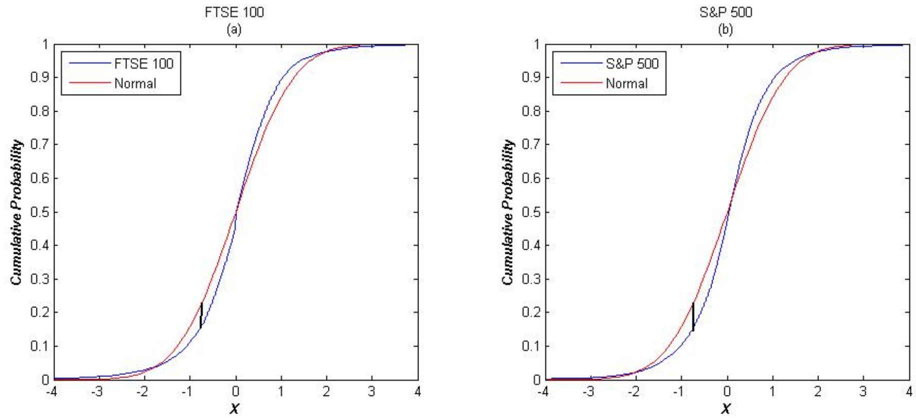


Figure 5: Illustration of the K-S test performed on daily returns for both the FTSE 100 (a) and the S&P 500 (b). Each plot shows the CDF of the normal distribution (red line), the empirical data (blue line) and the respective test statistics (black line). These black lines were included manually for illustrative purposes only and are not the correct representations of the test statistics.

daily (panel (a)), weekly (panel (b)), monthly (panel (c)) and two month (panel (d))

returns on the FTSE 100 index. In panel (a), the lower left extreme and the upper right extreme qunatiles (in blue) veer off from the standard normal quantiles (in red) resulting in the conclusion that the data is not normal. In panel (b), (c) and (d) the extremes appear to converge to the standard normal line which is consistent with our K-S test findings. Similar results can be seen for returns on the S&P 500 index in figure (**??**)

3.2 Leptokurtic Distribution

Recent studies of the price changes on financial market assets have shown that they follow power-law behaviour as opposed to Gaussian. The distribution of returns have

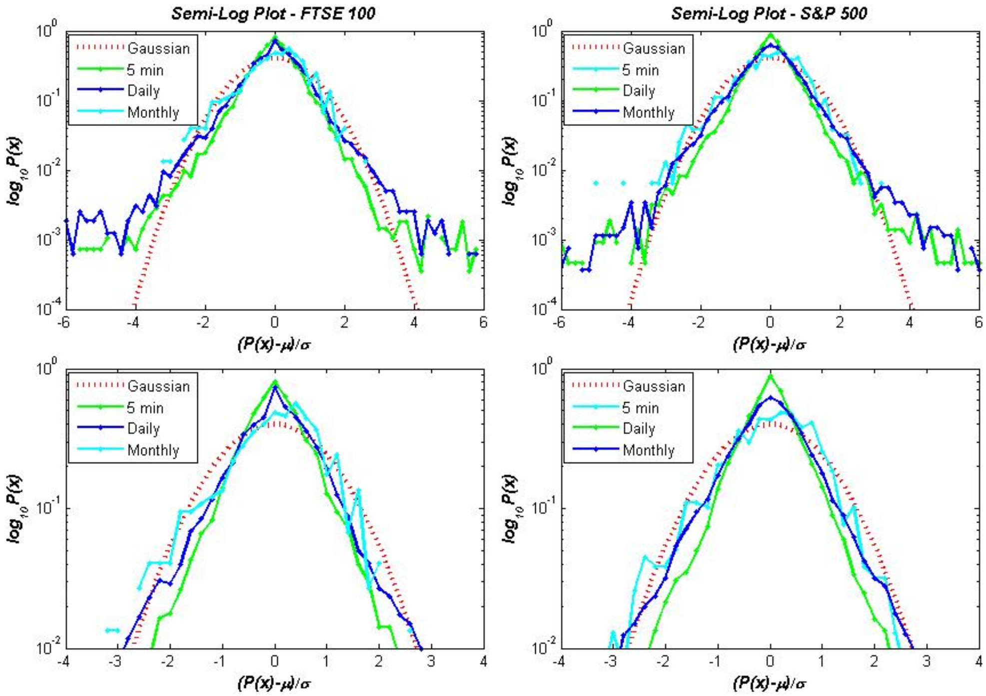


Figure 6: Semi-Log plots of returns on the FTSE 100 and S&P 500

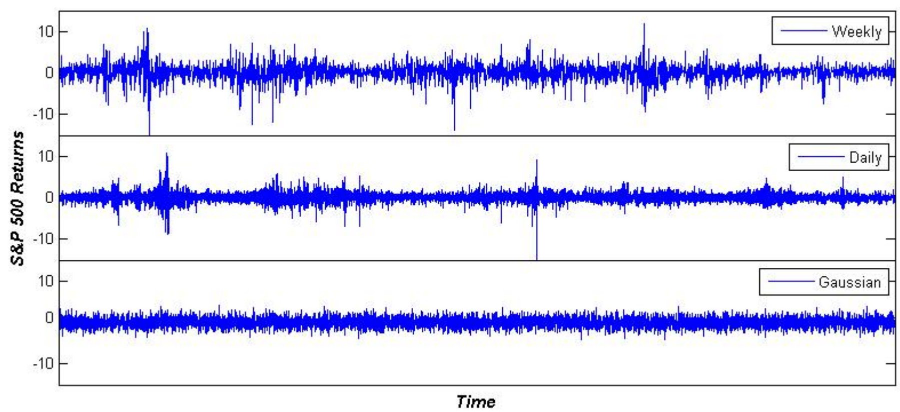


Figure 7: Noise

4 The Power-Law Distribution

In this section we follow the discussion in [22] to introduce some of the characteristics of the power-law distribution. A quantity follows a power-law distribution if the probability

$p(x)$ of taking on a value between x and $x + \mathrm{d}x$ is given by

$$p(x) = Pr(X = x) = Cx^{-\alpha}, \tag{10}$$

where X is a non-negative random variable, C is some constant and the exponent $\alpha > 0$. It follows that for positive values of α, $p(x)$ diverges as $x \to 0$. This would suggest that power-law distributions only hold over a certain range i.e. for values less than some point the distribution deviates. We denote this point x_{min}. This is the case for distributions we find in the real world. In such cases we say the *tail* of the distribution follows a power-law.

We can find the constant C by integrating over the region of the distribution and setting it equal to 1. The standard normalization requirement in probability theory is that the sum of all probabilities is 1. In this case the random variable x is continuous, therefore to sum up the probabilities we integrate,

$$\int_{-\infty}^{\infty} p(x)\,\mathrm{d}x = \int_{x_{min}}^{\infty} p(x)\,\mathrm{d}x = C\int_{x_{min}}^{\infty} x^{-\alpha}\,\mathrm{d}x = 1. \tag{11}$$

We are only concerned about the values for which the distribution exists, for $x \geq x_{min}$. We set the lower limit of the integral above to x_{min}, and essentially ignore values lower than x_{min}. Integrating we find

$$\frac{C}{1-\alpha}\left[x^{-\alpha+1}\right]_{x_{min}}^{\infty} = 1, \tag{12}$$

$$\Rightarrow C = (\alpha - 1)x_{min}^{\alpha-1}. \tag{13}$$

Notice that in (12) our equation only makes sense for $\alpha > 1$, otherwise it would diverge to infinity and we would not be able to normalize it. Re-arranging this we are able to solve for C. We can now get the required normalized quantity by subbing our value for C derived in equation (13) into equation (10) to get

$$p(x) = \frac{\alpha - 1}{x_{min}}\left(\frac{x}{x_{min}}\right)^{-\alpha}. \tag{14}$$

The main method for measuring power-law behaviour in a distribution mentioned in [22], takes advantage of the *complementary cumulative distribution function* or CCDF. If the cumulative distribution function $P(x)$ measures the probability that a random variable X with a given probability distribution $p(x)$ has a value less than or equal to x then the complementary cumulative distribution function is $1-P(x)$. In other words the probability that X takes on a value greater than or equal to x. In the case of a continuous distribution, it gives the area under the probability density curve from x to infinity. The CCDF is

$$P(x) = Pr(X \geq x) = \int_{x}^{\infty} p(x')\,\mathrm{d}x', \tag{15}$$

where we replace the variable x with a dummy variable x' to avoid confusion when integrating because one of our limits is also x. Using equation (10) and (13) we get the complementary cumulative distribution function

$$P(x) = \left(\frac{x}{x_{min}}\right)^{-\alpha+1}. \tag{16}$$

So far we have dealt with a random variable that is continuous, however some power-law distributions are discrete in nature for example, the number of citations a scientific paper receives. Therefore it is practical to explore the mathematics of a discrete power-law. Where a random variable X is discrete, it follows a power-law if the probability $p(x)$ of measuring the value x is given by

$$p(x) = Pr(X = x) = Cx^{-\alpha}. \tag{17}$$

Again the distribution is not defined at 0 for positive exponent values, so there must be a lower bound $x_{min} > 0$. Following the same procedure as before by summing up the probabilities and setting them equal to 1 we can find the constant C using equation (17)

$$\sum_{x=x_{min}}^{\infty} p(x) = C \sum_{x=x_{min}}^{\infty} x^{-\alpha} = 1, \tag{18}$$

and replacing $\sum_{x=x_{min}}^{\infty} x^{-\alpha}$ with $\zeta(\alpha, x_{min})$, where $\zeta(\alpha, x_{min})$ is the generalized form of the Reimann ζ-function we get

$$C\zeta(\alpha, x_{min}) = 1. \tag{19}$$

Rearranging equation (19) we finally get a value for the normalization constant C

$$C = \frac{1}{\zeta(\alpha, x_{min})}. \tag{20}$$

Using equation (20) we find the normalized expression for $p(x)$

$$p(x) = \frac{x^{-\alpha}}{\zeta(\alpha, x_{min})}. \tag{21}$$

Again it is valuable to consider the distribution's complementary CDF .The complementary CDF for a discrete random variable X is

$$P(x) = Pr(X \geq x) = \sum_{x \geq x_i} Pr(X = x_i) = \sum_{x \geq x_i} p(x_i), \tag{22}$$

and using equation (21) we find

$$P(x) = \frac{\zeta(\alpha, x)}{\zeta(\alpha, x_{min})}. \tag{23}$$

The functions for the discrete case are not as simple as those of a continuous random variable. In [3], Clauset et al. mention this fact and suggest ways in which one can conveniently approximate discrete power-law distributions from continuous ones. One way is to treat an integer power-law as if its values come from a continuous distribution but are rounded to the nearest whole number. They claim this is superior to the other approximation methods such as truncating or rounding down to the nearest integer value from continuous ones.

4.1 Scale Invariance

One curious property of a power-law is that its distributions are *scale free* [22]. This means that no matter what scale we examine our data at it retains its functional form. Given a relation like that in equation (10) we can multiply the variable x by a constant a

$$p(ax) = C(ax)^{-\alpha} = Ca^{-\alpha}x^{-\alpha} = a^{-\alpha}p(x) \propto p(x), \tag{24}$$

here scaling by the constant a results in simply multiplying the original power-law function by $a^{-\alpha}$. What does this mean? It demonstrates that the relative likelihood between small and large observations or events are the same, regardless of the choice of scale we decide to measure our observations on. Graphically the shape of the distribution is unchanged except for a multiplicative constant when plotted on logarithmic axes. This has the effect of essentially shifting the distribution up or down, depending on the value of the constant. This property is also what produces the signature straight-line when the distribution is plotted on log-log scales [22].

4.2 Mean and Variance

When one analyses the moments of a power-law distribution one discovers some other interesting attributes. If we first attempt to calculate the mean

$$\langle x \rangle = \int_{-\infty}^{\infty} x p(x)\,\mathrm{d}x = C \int_{x_{min}}^{\infty} x^{-\alpha+1}\,\mathrm{d}x = \frac{C}{2-\alpha}\left[x^{-\alpha+2}\right]_{x_{min}}^{\infty}, \tag{25}$$

we find that for $\alpha \leq 2$ the first moment of the distribution would diverge to infinity because of the upper limit of integration. We saw in the first part of this section the requirement that $\alpha > 1$ to avoid divergence, well in the range $1 < \alpha < 2$ the mean of the distribution is infinite and does not exist, along with all higher moments. For $\alpha > 2$, however, we find a closed form expression for the mean using equation (13)

$$\langle x \rangle = \frac{\alpha-1}{\alpha-2} x_{min}. \tag{26}$$

Looking at the second moment

$$\langle x^2 \rangle = C \int_{x_{min}}^{\infty} x^{-\alpha+2}\,\mathrm{d}x = \frac{C}{3-\alpha}\left[x^{-\alpha+3}\right]_{x_{min}}^{\infty}, \tag{27}$$

we find that for $\alpha \leq 3$ the second moment diverges, meaning that distributions with $\alpha \leq 3$ do not have finite variance either. In the range $2 < \alpha < 3$ the mean is finite but the second moment and higher moments are infinite. In fact we can extend this further to say that only the first $k < \alpha - 1$ moments exist for the power-law distribution by calculating the k^{th} moment

$$\langle x^k \rangle = C \int_{x_{min}}^{\infty} x^{-\alpha+k}\,\mathrm{d}x = \frac{\alpha-1}{\alpha-1-k} x_{min}^k. \tag{28}$$

For all moments up to $\alpha - 1$ the values are finite and are well defined. For moments greater than or equal to $\alpha - 1$, divergence occurs.

In reality however, it is possible to calculate the moments of a distribution following a power-law. If we observe data in the real world it is going to be of finite size and not infinite. The data set will be bounded by some upper limit, the maximum observed value and not infinity like we have in equation (25), (27) or (28). If we take a finite set of observations that follow a power-law like say, daily returns on the FTSE 100 in 2014 for example. It will have a well defined maximum value and we would therefore be able to calculate the average daily return for this period. If we then take a larger sample set, say, daily returns on the FTSE 100 for 2014 and 2013 there is a possibility of getting a larger maximum and therefore a larger average as a result. If we continue to increase the sample set the estimate of the mean will increase without bound and this is what the divergence of the moments of the power-law distribution is telling us [22].

5 Power-Law Behaviour

About 6% of the words we say or read or write is "the". The word "the" accounts for almost one in every sixteen words we encounter on a daily basis [29]. In fact if one was to rank the words used in a text or article a pattern emerges: the second most used word appears almost half as often as the first; the third most used word one third as often as the first; the fourth most used, one fourth as often; the fifth, one fifth as often and so on. It appears that word frequency is proportional to one over its rank. Bizarrely this occurs not only just in texts or articles, but in the entire English language and in any language [6, 23]. This pattern was first discovered by George Kingsley Zipf, an American linguist in 1935 [31] and since then has become known as Zipf's law (see figure (8)). This phenomenon, when plotted on a log-log graph, produces a straight line. This linear relationship on logarithmic scales is a power-law and has become one of the main identifiers for power-law behaviour. In this section we discuss methods of graphically recognising power-law distributions.

In his paper on power-law distributions Newman examines means of measuring power-law behaviour [22]. The standard strategy of plotting the histogram with equal bin length of a distribution on log-log scales Newman states is a "poor way to proceed". The graphs produced by this method have a lot of noise towards the tail end as a result of sampling errors. In panel (a) and (c) of figure (9) we apply this approach to synthetically generated power-law data (for further details see Appendix) and to the frequency of words used in the novel Moby Dick. We can see the straight line in the left of the diagram and the tail of the plot is very noisy. Here the power-law distribution begins to dwindle as the number of

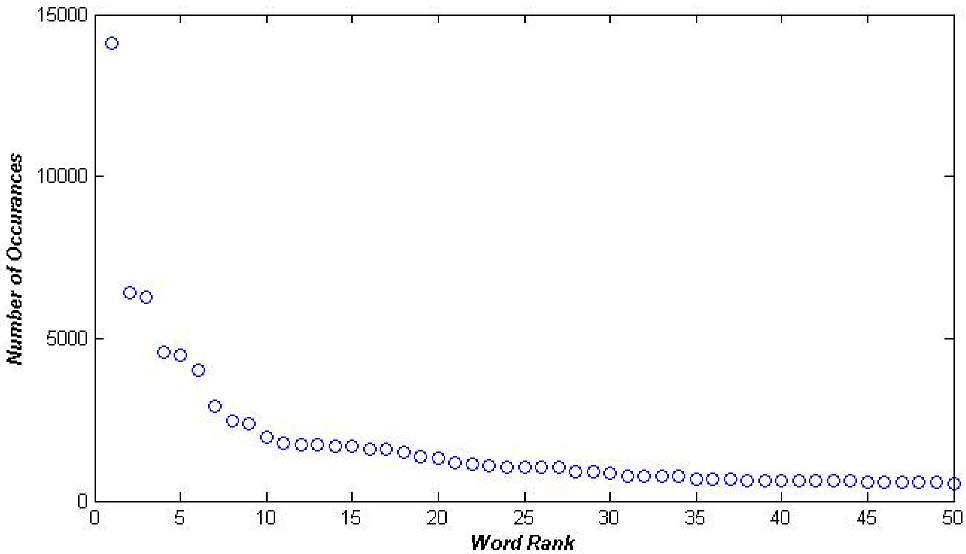


Figure 8: Word rank/frequencies of the fifty most used words in Hermann Melville's Moby Dick. This exhibits the pattern first discovered by George Zipf in 1935. The data was obtained from Aaron Clauset's (co-author of [3]) website at the Sante Fe Institute http://tuvalu.santafe.edu/ aaronc/powerlaws/data.htm.

samples per bin is very few. Hence the fractional change in the bin counts are very large which produces the noise. To avoid this one could simply discard these counts but then one runs the risk of losing valuable information about the distribution. Newman mentions an alternative solution whereby varying the length of each bin width such that each is a fixed multiple wider of the one previous. This is known as *logarithmic binning*. This method certainly improves our chances of getting the desired effect although some noise in the tail can still be experienced.

Newman states that this method is flawed as binning the data in such a way clumps all the counts together within a certain range in the same bin and thus we lose any valuable information contained in the individual values. To overcome this loss of information Newman suggest the best possible way to plot a power-law distribution is to plot the CCDF of the distribution on log-log scales. In panel (b) and (d) of figure 9 such a method is applied to the same data sets of (a) and (c). We can clearly see the straight line one would associate with a power-law and a significantly reduced level of noise (the noise at the end of the tail in panel (b) and (d) of figure 9 is a result of a small sample size. If the data size was a lot bigger this would not occur. This method is still superior to those mentioned before as each point is perfectly defined under the CCDF and can be plotted without binning).

To achieve this plot we used the *histc()* function in MATLAB which counts the samples in each bin of the histogram of our data set. We then took the cumulative sum of these

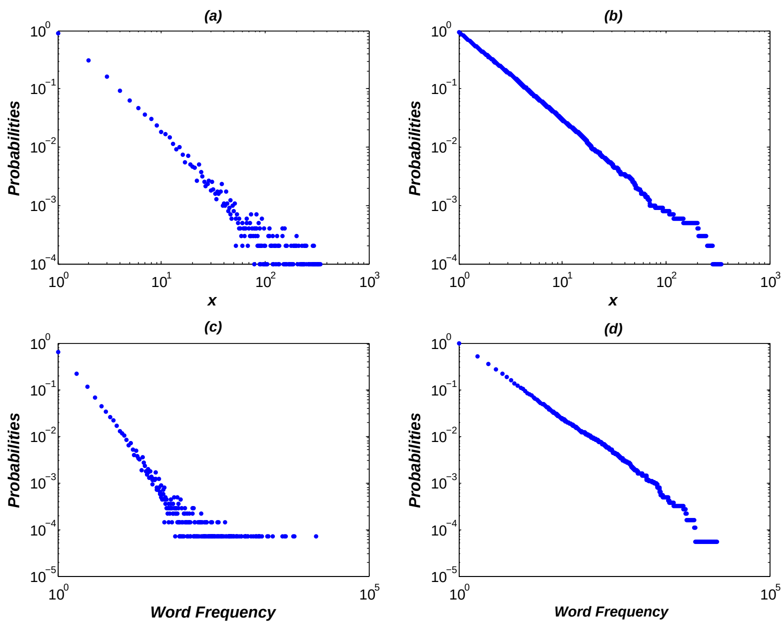


Figure 9: Panel (a) shows the standard method of plotting the histogram on log-log scales with equal bin length for 10^5 synthetic power-law distributed data points. Panel (c) shows the same method for word frequencies in Moby Dick. Both plots produced quite a lot of noise in the tail end of the graph. Panel (b) and (d) show the CCDF method for the synthetic and word frequency data sets respectively. The CCDF method is clearly the better of the two for achieving the linear relationship on log-log scales.

counts, essentially adding the first element to the second, that to the third, that to the fourth and so on. We took the complement of this result to produce the required CCDF and then plotted this on log-log scales.

We can now plot our data on log-log scales to detect power-law behaviour. Simply plotting the price changes straight up can be an issue however as the opening price may be the same as the closing resulting in a 0% return. Also the closing price may be less then the opening resulting in a negative return. Since we are plotting on log-log scales to detect a power-law relation some data manipulation is required. We discard the zero returns for each data set and then split each set into two, positive and negative tails. See (10) for both the positive and negative tails of the FTSE 100 and the S&P 500 for each time interval.

We have the combined all the intervals for each tail on the same scale for comparison. We see that probability of getting a smaller percentage returns are larger than that of large returns, suggestive of power-law behaviour. We can also see the signature straight line of a power law for returns on a daily timescale which may allow us to conclude that returns follow power-law behaviour. The straight power-law line is seen for the larger time scales as well but they are not as distinguishable. This may be a result of the data sets getting smaller and smaller but also, as Gopikrishnan et al. found, because returns tend toward normality for larger timescales. We can also see that the straight line does not cover the entire data set. This reiterates the fact that there is some value x_{min}, as we have seen in section 4.

In figure (11) we have plotted the volume of trade amounts on log-log scales. Similar to that of the returns, the power-law tail seems to fall off sharper as the time intervals become smaller. One distinguishable feature is that the volumes of trades have a deeper slope then that of the returns.

Plotting the CCDF of a distribution is useful for identifying power-law behaviour but it is not the most statistically sound method for claiming a distribution is power-law distributed. Nor does it gives any indication as the to values of the its parameters. In the next sections we look at methods for achieving these goals.

6 Detecting a Power-Law

In the previous section we used methods for graphically detecting power-law behaviour but we did not mention how to acquire the scaling parameter or the lower bound of the distribution. In this section we will use some of the mathematical techniques tested in

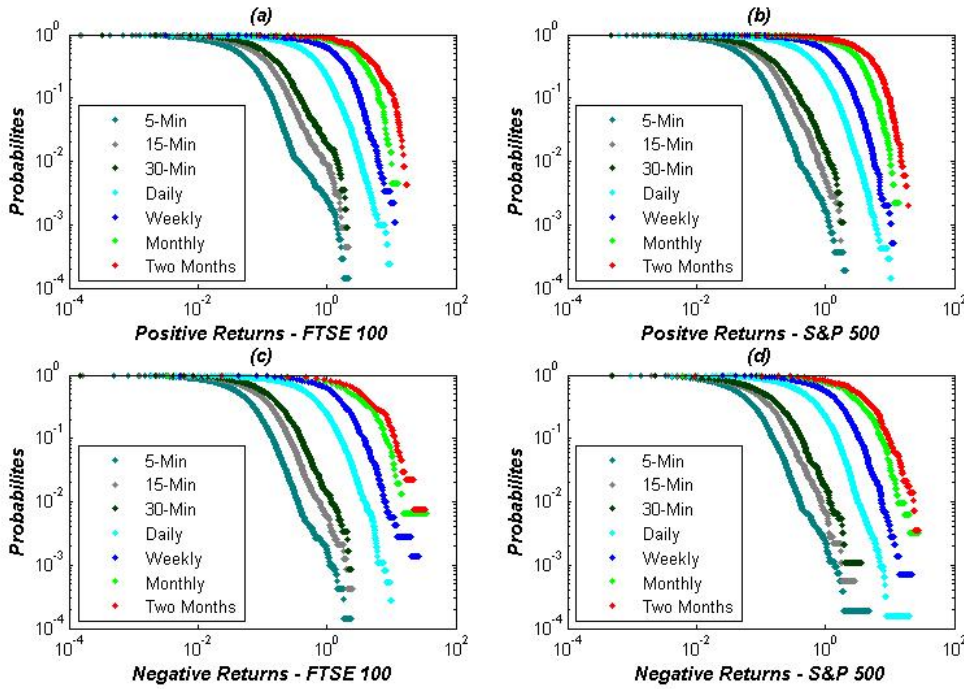


Figure 10: CCDF plots of positive (a) and negative (c) returns on the FTSE 100 index (left). Also the CCDF plots of positive (b) and negative (d) returns on the S&P 500 index. The y-axes reads as the probability of finding a return greater then or equal to some value x.

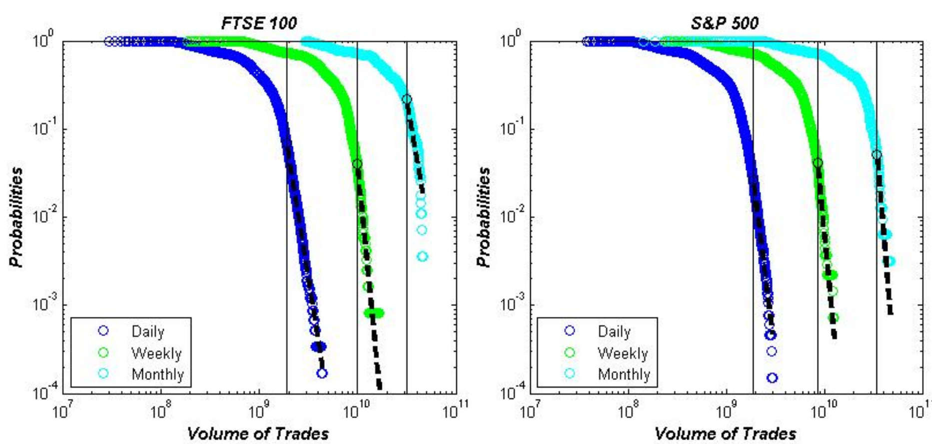


Figure 11: Volume of trades on the (a) FTSE 100 and (b) S&P 500

[3] for accurately achieving these values. First we will explain the approaches and derive the methods before applying them to our data sets. It is worth keeping in mind that the methods in this section allow us to fit a power-law distribution to a data set but bare no indication as to whether a power-law is a good fit or not. This is something we address in section 6.

6.1 Estimating the Lower Bound

As we have already established in section 3, the power-law distribution only exists for values of $x \geq x_{min}$. Thus for empirical data that exhibits power-law behaviour, it does so for values greater than some lower bound, which is why we conclude that the tail of the distribution is a power-law. Therefore if we are to accurately identify power-law behaviour, correctly estimating the value for $\hat{x}_{min}$ becomes very important (the '^' symbol above the x indicates that it is an estimate and not the true value. We use this symbol to denote an estimate going forward). As we will see in the next section, estimating the α parameter requires that a value of x_{min} be known. If we choose a value for $\hat{x}_{min}$ that is too low we end up attempting to fit a power-law to non power-law data which results in a biased estimate for our scaling parameter. Similarly, if we choose $\hat{x}_{min}$ too high we are discarding valid power-law data and again run the risk of arriving at a biased estimate for α due to the dwindling numbers in the tail of the distribution.

The most common approach to choosing $\hat{x}_{min}$ is to estimate visually the point where the CCDF of the distribution becomes straight on log-log scales. A conservative guess can be sufficient when the number of data points in the tail are quite large but as for most empirical cases, ours included, this is not appropriate for reasons mentioned above. This method is clearly subjective and therefore a more exact technique is required. Instead a method proposed in [4] can be applied to determine an accurate $\hat{x}_{min}$ value. The method compares the empirical data to a power-law model and selects an $\hat{x}_{min}$ value that best fits a power-law distribution to the empirical data. The method makes use of the Kolmogorov-Smirnov test or KS test, a statistical tool used for hypothesis testing that measures the distance between the CDF's of two distributions (see section 7.1.1 for more detail). The $\hat{x}_{min}$ value chosen is the value that minimizes the KS test statistic D_n, which is just simply the maximum distance between the CDF of the empirical data and the CDF of the power-law data

$$D_n = \min \left[\max_{x \geq x_{min}} |P_n(x) - P(x)| \right], \tag{29}$$

where $P_n(x)$ is the CDF of our empirical distribution and $P(x)$ is the CDF of the power-law model that best fits the empirical data. This method is run numerically for each proposed $\hat{x}_{min}$ and a scaling parameter $\hat{\alpha}$ is calculated for $x \geq \hat{x}_{min}$ (details of how to calculate $\hat{\alpha}$ are in the next section). Then the test statistic is computed comparing the empirical data with a power-law model with the calculated parameters $\hat{\alpha}$ and $\hat{x}_{min}$. This process is repeated until the $\hat{x}_{min}$ that produces the best fit is chosen.

This approach is recommended by Clauset et al. as the best method for estimating x_{min}. It can be used on both discrete and continuous distributions producing results as least as good if not better than other methods such as *Bayesian information criterion* or Hill plot. It also has the added advantage of being easy to implement and computationally more efficient and is the foundation of the method we use in section 6 to carry out a goodness-of-fit test on our empirical data.

6.2 Estimating Alpha

Traditionally the scaling parameter of a power-law is determined by first plotting the histogram on logarithmic scales which produces a straight line and then extracting the slope of this straight line by means of linear least-squares regression. The least-squares approach to regression analysis minimizes the sum of squared residuals, a residual being the difference between an observed value and the fitted value provided by a model. However this approach generates varying and inaccurate results depending on the choice of graphical method, i.e. logarithmic binning or cumulative distribution [1]. These arise due to breaches in the fundamental assumptions of linear regression as discussed in Clauset et al. such as independence and multivariate normality. Instead the more accurate procedure of estimating the scaling parameter α is using the method of *maximum likelihood estimation* [3]. This is the method that we will use to estimate α going forward.

Maximum likelihood estimation is a statistical procedure used to find estimates for a parameter of a model given observed data. The method was popularized by Ronald Fisher in the 1920's and works as follows: suppose we have observations of independent random variables $X_1, X_2, ..., X_n$ whose assumed probability distribution depends on an unknown parameter, in our case α. The aim is to acquire an estimate of the unknown parameter $\hat{\alpha}$ that is closest to the true value α given the observed values $x_1, x_2, ..., x_n$. In other words we hope to achieve an estimate that 'maximises' the probability or 'likelihood' of getting the data we observed - hence the name [21]. Following this approach we can derive an analytical formula for achieving a value of the scaling parameter α.

In this section derive an expression for both the continuous and discrete cases. Firstly we consider the normalized power-law distribution for a continuous random variable derived in section 3

$$p(x) = \frac{\alpha - 1}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\alpha}, \tag{30}$$

where x_{min} is the predetermined lower bound of the distribution and α is the scaling parameter. Since the random variables are independent the joint probability density function is simply the product of the marginal probability density functions. We call this the likelihood function, denoted $L(\alpha)$, and it is a function of the parameter we are trying to estimate, α

$$L(\alpha) = Pr(X_1 = x_1, X_2 = x_2, , X_n = x_n \mid \alpha) = p(x_1 \mid \alpha) \cdot p(x_2 \mid \alpha) \cdots f(x_n \mid \alpha), \tag{31}$$

which can be more neatly written as

$$L(\alpha) = \prod_{i=1}^{n} \frac{\alpha - 1}{x_{min}} \left(\frac{x_i}{x_{min}} \right)^{-\alpha}. \tag{32}$$

In order for us to find the maximum of this likelihood function, i.e. the value of α at which this function is at its maximum, we must differentiate with respect to α. To make this process easier we take the natural logarithm of both sides. The logarithm function is an increasing function, that is, if $x_1 < x_2$, then $f(x_1) < f(x_2)$. That means that the value of α that maximizes the natural logarithm of the likelihood function $ln(L(\alpha))$ is also the value of α that maximizes the likelihood function $L(\alpha)$. Moving forward in this manner and taking advantage of some of the characteristics of the logarithm function (i.e. $log \prod x_i = \sum log x_i$ and $loga/b = loga - logb$) we find

$$lnL(\alpha) = ln \prod_{i=1}^{n} \frac{\alpha - 1}{x_{min}} \left(\frac{x_i}{x_{min}} \right)^{-\alpha} \tag{33}$$

$$= \sum_{i=1}^{n} \left(ln(\alpha - 1) - lnx_{min} - \alpha ln \frac{x_i}{x_{min}} \right) \tag{34}$$

$$= nln(\alpha - 1) - nlnx_m in - \alpha \sum_{i=1}^{n} ln \frac{x_i}{x_{min}}. \tag{35}$$

This result gives a more manageable expression to differentiate than if we were to differentiate straight out. Differentiating equation (35) with respect to α and setting it equal to 0 yields the desired expression for for the maximum likelihood estimator $\hat{\alpha}$

$$\hat{\alpha} = 1 + n\left(\sum_{i=1}^{n} ln\frac{x_i}{x_{min}}\right)^{-1}. \tag{36}$$

There are some useful formal results that support the use of maximum likelihood estimation, some of which are referred to and proved in [3]. The most notable being, that as the number of data points from a distribution with parameter α get large i.e. as $n \to \infty$, the estimate tends to the true value, $\hat{\alpha} \to \alpha$. This result is also extended to the power-law case where the MLE $\hat{\alpha}$ converges to the true alpha.

Now we move on to the discrete case. If we consider the expression for an integer power-law distribution (20) we derived in section 3 and following the same protocol as the continuous case we get the likelihood function

$$L(\alpha) = \prod_{i=1}^{n} \frac{x_i^{-\alpha}}{\zeta(\alpha, x_{min})}, \tag{37}$$

where $\zeta(\alpha, x_{min})$ is the generalized zeta function. Taking the natural log of both sides gives

$$lnL(\alpha) = ln\prod_{i=1}^{n} \frac{x_i^{-\alpha}}{\zeta(\alpha, x_{min})} = \sum_{i=1}^{n}(-\alpha ln x_i - ln\zeta(\alpha, x_{min})) = -\alpha\sum_{i=1}^{n} ln x_i - nln\zeta(\alpha, x_{min}). \tag{38}$$

Differentiating with respect to α and setting it equal to 0 we obtain

$$\frac{-n}{\zeta(\alpha, x_{min})}\frac{\partial}{\partial\alpha}\zeta(\alpha, x_{min}) = \sum_{i=1}^{n} ln x_i \tag{39}$$

$$\frac{\zeta'(\hat{\alpha}, x_{min})}{\zeta(\hat{\alpha}, x_{min})} = -\frac{1}{n}\sum_{i=1}^{n} ln x_i, \tag{40}$$

where $\zeta'(\hat{\alpha}, x_{min})$ denotes differentiation with respect to α. This expression does not have an exact closed-form and therefore needs to be solved numerically to find a value for $\hat{\alpha}$ or one can numerically maximise the log-likelihood itself.

Alternatively, we can obtain an approximate expression for the discrete case from the continuous by rounding to the nearest integer. If we can find some expressional relation between these two cases we will be able to approximate $\hat{\alpha}$ for discrete values. Consider a differentiable function $f(x)$, with an indefinite integral $F(x)$ such that $F'(x) = f(x)$,

$$\int_{x-\frac{1}{2}}^{x+\frac{1}{2}} f(t)dt = F\left(x+\frac{1}{2}\right) - F\left(x-\frac{1}{2}\right). \tag{41}$$

Using the Taylor series expansion of $F(x)$ about x we find that the integral can be written as a series of differentiable functions

$$\begin{aligned} F\left(x+\frac{1}{2}\right) - F\left(x-\frac{1}{2}\right) &= [F(x) + \frac{1}{2}F'(x) + \frac{1}{8}F''(x) + \frac{1}{48}F'''(x) + \cdots] \\ &\quad - [F(x) - \frac{1}{2}F'(x) + \frac{1}{8}F''(x) - \frac{1}{48}F'''(x) + \cdots] \\ &= F'(x) + \frac{1}{24}F''(x) + \cdots \\ &= f(x) + \frac{1}{24}f'(x) + \cdots. \end{aligned} \tag{42}$$

If we replace x with x_{min} in the lower limit and replace the upper limit with infinity in the integral above and consider integrating the equivalent to summing over the domain we find the following relation

$$\int_{x_{min}-\frac{1}{2}}^{\infty} f(x')dx' = \sum_{x=x_{min}}^{\infty} f(x) + \frac{1}{24}\sum_{x=x_{min}}^{\infty} f''(x) + \cdots. \tag{43}$$

However, if we replace $f(x)$ for a power-law relation for some constant α and integrate we find

$$\int_{x_{min}-\frac{1}{2}}^{\infty} t^{-\alpha}dt = \frac{(x_{min}-\frac{1}{2})^{-\alpha+1}}{\alpha-1}, \tag{44}$$

but also

$$\int_{x_{min}-\frac{1}{2}}^{\infty} t^{-\alpha}dt = \sum_{x=x_{min}}^{\infty} x^{-\alpha} + \frac{\alpha(\alpha+1)}{24}\sum_{x=x_{min}}^{\infty} x^{-\alpha-2} + \cdots, \tag{45}$$

$$= \zeta(\alpha, x_{min})[1 + \mathcal{O}(x_{min}^{-2})], \tag{46}$$

where we have retained only the significant terms, making use of the fact that $x^{-2} \leq x_{min}^{-2}$. Equating both (44) and (46) we find an expression for $\zeta(\alpha, x_{min})$

$$\zeta(\alpha, x_{min}) = \frac{(x_{min}-\frac{1}{2})^{-\alpha+1}}{\alpha-1}[1 + \mathcal{O}(x_{min}^{-2})]. \tag{47}$$

Differentiating this with respect to α by using the chain rule and taking advantage of the fact that where $f(x) = a^x$, $f'(x) = f(x)ln(a)$ we find

$$\zeta'(\alpha, x_{min}) = \left(\frac{-(x_{min} - \frac{1}{2})^{-\alpha+1}}{(\alpha-1)^2} - \frac{(x_{min} - \frac{1}{2})^{-\alpha+1} ln(x_{min} - \frac{1}{2})}{(\alpha - 1)} \right) [1 + \mathcal{O}(x_{min}^{-2})] \tag{48}$$

$$= -\frac{(x_{min} - \frac{1}{2})^{-\alpha+1}}{(\alpha-1)} \left(\frac{1}{\alpha - 1} + ln\left(x_{min} - \frac{1}{2} \right) \right) [1 + \mathcal{O}(x_{min}^{-2})]. \tag{49}$$

We now have an expression for $\zeta(\alpha, x_{min})$ and $\zeta'(\alpha, x_{min})$. If we substitute these expressions into the ratio in (40), the maximum likelihood expression we found for the discrete case, we find

$$\frac{\zeta'(\hat{\alpha}, x_{min})}{\zeta(\hat{\alpha}, x_{min})} = -\frac{\frac{(x_{min} - \frac{1}{2})^{-\hat{\alpha}+1}}{(\hat{\alpha}-1)}}{\frac{(x_{min} - \frac{1}{2})^{-\hat{\alpha}+1}}{(\hat{\alpha}-1)}} \left(\frac{1}{\hat{\alpha} - 1} + ln\left(x_{min} - \frac{1}{2} \right) \right) [1 + \mathcal{O}(x_{min}^{-2})] \tag{50}$$

$$= -\left(\frac{1}{\hat{\alpha} - 1} + ln\left(x_{min} - \frac{1}{2} \right) \right) [1 + \mathcal{O}(x_{min}^{-2})]. \tag{51}$$

if we then equate (51) with what we have found in (40) we can find an expression for $\hat{\alpha}$ by neglecting the higher order terms

$$-\left(\frac{1}{\hat{\alpha} - 1} + ln\left(x_{min} - \frac{1}{2} \right) \right) = -\frac{1}{n} \sum_{i=1}^{n} ln x_i \tag{52}$$

$$\hat{\alpha} \approx 1 + n \left(\sum_{i=1}^{n} \frac{x_i}{x_{min} - \frac{1}{2}} \right)^{-1}. \tag{53}$$

The expression above is similar to the MLE we found for the continuous case, the difference being the $-\frac{1}{2}$ in the denominator. Clauset et al. states that this estimator behaves well for large x_{min} in particular $x_{min} \gtrsim 6$. In fact this is the estimator we use in our analysis for data sets where x_{min} is large, details of which will be discussed in the next section.

6.3 Implementation

To implement the methods discussed above we use the *plfit()* command (the command was written by Aaron Clauset and can be downloaded from his website at the Sante Fe Institute http://tuvalu.santafe.edu/ãaronc/powerlaws/data.htm). The command automatically detects whether the data is composed of real or integer values, and applies the appropriate method. For the discrete case the command generally uses the numerical maximization of the log likelihood approach we mentioned in the last section. However if the minimum value of a data set with discrete values is greater then 1000, the command uses the continuous approximation method. The fits are summarized in (Table 2), where the $\hat{x}_{min}$ and $\hat{\alpha}$ values are quoted for both tails of each time interval for the FTSE 100 and the S&P 500. The values for α for the financial data sets all seem to be in the range $2 \lesssim \alpha \lesssim 4$ for the most part which is consistent with the findings of [13]. All of these with the exception of two have an $\alpha > 3$ which means these distributions have a finite variance and by the central limit theorem would converge to a Gaussian. Shown also, are power-law fits to the synthetic power-law set and the word frequency set discussed in section 4 as well as a data set of 10^5 standard normal distributed random numbers. The synthetic power-law data is a set of continuous numbers drawn from a power-law distribution with a scaling parameter $\alpha = 2.5$ and $x_{min} = 1$. The methods are accurate to within roughly 1% of the true value for the synthetic data set which is a good indicator of their ability to produce reliable results.

7 Power-Law Hypothesis

We move on now to discuss how to determine the goodness-of-fit of a power-law distribution to observed data. In the last section we discussed how to fit a power-law distribution to a data set using maximum likelihood and the KS test but did not establish if the fit was a good fit or not. In this section we aim to do just that.

As we have mentioned already the conclusion that observed data follows a power-law from merely observing a straight line when plotting the data on log-log scales is not entirely justifiable. This feature, although necessary, is not sufficient as other distributions such as the exponential and the log-normal distributions also produce roughly straight lines on log-log scales. Therefore we require a protocol that will allow us to test our power-law hypothesis like a goodness-of-fit test. Goodness-of-fit tests are useful when it comes to hypothesis testing as they allow us to measure the differences between observed values and those expected under a particular model.

Clauset et al. propose a method for establishing the goodness-of-fit of a power-law model using a modified version of the KS test. The approach focuses on trying to distinguish statistical fluctuations from random sampling of a data set from those that arise because

Quantity			$\hat{x}_{min}$		$\hat{\alpha}$	
			Positive	Negative	Positive	Negative
FTSE 100	Returns	5-Min	0.1169	0.1447	3.3071	3.4719
		15-Min	0.2076	0.2142	3.0822	3.2521
		30-Min	0.2634	0.3333	2.8926	3.2449
		Daily	2.7380	1.3203	4.4195	3.3541
		Weekly	3.0596	2.2991	4.5049	3.2032
		Monthly	7.4844	2.0156	8.4706	2.1741
		Two Month	12.0333	8.4242	8.5092	4.0992
	Volumes	Daily	1.9120e+009		7.9901	
		Weekly	9.9098e+009		12.3328	
		Monthly	3.1434e+010		7.3665	
S&P 500	Returns	5-Min	0.1561	0.1056	3.3503	3.0772
		15-Min	0.1634	0.2307	2.8971	3.1200
		30-Min	0.2116	0.3233	2.7214	3.1464
		Daily	1.9333	1.4804	4.1907	3.8674
		Weekly	3.4825	2.2638	4.6499	3.2813
		Monthly	3.7109	8.2804	3.7440	4.8470
		Two Month	9.8177	6.3262	6.3051	3.4332
	Volumes	Daily	1.8970e+009		10.0203	
		Weekly	8.5503e+009		12.7005	
		Monthly	3.4619e+010		14.6238	
Synthetic Standard Normal			1.9045		6.7474	
Synthetic Power-Law			1.0008		2.5132	
Word Frequency in Moby Dick			7		1.9500	

Table 2: Summaries of power-law fits to the data sets analysed in this paper. The methods employed here were also conducted on the the computer generated power-law data we used in section 4, see figure (9). This data set was generated with $\alpha = 2.5$ and $x_{min} = 1$. These methods are relatively accurate at predicting these parameters. However the discrepancies reported here are probably as result of a small sample set.

the observed data is not power-law distributed. It is rare that when fitting a power-law distribution or any other distribution to observed data that the model or models fit exactly. Discrepancies arise between observed values and the model even if the observed values are derived from the theoretical model itself because of the random nature of sampling [3]. For example if you were to run a goodness-of-fit test on a data set of say 100 sample observations from a group of 1,000, the results you would get would vary if you were to run the test again on a different sample of 100. It is these statistical fluctuations that the procedure attempts to distinguish. The size of the data set also plays a huge factor in hypothesis testing, larger data sets lead to increased accuracy in determining the goodness-of-fit of a distribution but as data sets become smaller and smaller it becomes increasingly difficult to identify the observed distribution. That is why we prefer large sample sizes when hypothesis testing.

To differentiate between these fluctuations Clauset et al. propose an approach that compares the distance between an observed data set and the power-law model with the distances of a large number of synthetically generated power-law data sets drawn from the same model. If the observational distance is smaller then the typical synthetic ones then we conclude the power-law is a good fit and the distance between the observed data and the model are merely statistical fluctuations.

7.1 Implementation

We used the *plpva()* command in MATLAB (written by Aaron Clauset) to employ the method discussed above. The command takes an input of empirical observations and fits a power-law model to the data using the fitting procedures discussed in section 5 and then calculates the KS test statistic using (9). It then generates 1000 synthetic power-law data sets individually with the same parameters of the best fit to our empirical observations. Each synthetic data set is fitted to its own power-law model and a KS statistic is calculated respectively. Essentially we apply the first step to each of the 1000 generated power-law data sets. Insuring we are carrying out the same calculation across the board. Finally the command out puts a $p-value$ or plausibility value of the power-law hypothesis, which is just a count of the fraction of times the KS statistics is larger then the KS statistic computed for our empirical data. Hence a large $p-value$ is desirable here as we are measuring the plausibility of the hypothesis we are attempting to verify. This is contrasting of the usual treatment where one quotes a low $p-value$ to be statistically significant indicating that the 'null' hypothesis is unlikely to be true.

The synthetic data sets created using this command are generated in such a way that they follow a similar distribution to that of the empirical observations for values of $x < x_{min}$ and power-law distributed with the same parameters of the best fit to the empirical data set for $x \geq x_{min}$. To achieve this consider that for a distribution with n observations that follows a power-law, it has n_{tail} values in the tail of the distribution above x_{min} and $n - n_{tail}$ values for $x < x_{min}$. The command selects, with probability $1/n_{tail}$, a random number generated using equation (61) with power-law parameters equal to the best fit to the empirical distribution. With probability $(n - n_{tail})/n$, a number is picked at random from the observed data set in the range below the x_{min}. This value of this number is then used to create the part of the synthetic data set below x_{min}. This process is then repeated until the synthetic data set is created.

Clauset et al. recommend that a $p-value > 0.1$ be used as a decision rule when interpreting these results. They acknowledge that this is more conservative over the traditional $p - value = 0.05$ cut off point, and that the decision should depend on the context of the study. They also warn that very large $p - values$ should not be trusted when associated with small data sets. As we have mentioned above the power of a goodness-of-fit test will decrease for smaller and smaller sample sizes as it becomes increasingly difficult to distinguish between models. It is quite possible that a small sample empirical distribution could fit a power-law form even if it were not derived from the power-law model. Therefore we treat such instances with caution.

The results of this test applied to our data sets are formulated in (Table 3). The table displays a $p - value$, a 'gof' value or 'goodness-of-fit' value which is the KS test statistic measuring the maximum distance between the CDF of the empirical distribution and the power-law model. We provide the ratio x_{max}/x_{min}, encapsulating the span of the power-law tail for each data set. For a distribution to truly follow power-law behaviour one would expect to see the tail span several decades. Stumpf and Porter prescribe that a power-law should display linear behaviour on log-log tails for at least two orders of magnitude [27]. We will use this metric along with the gof value to aid in our analysis. The test was also applied to the synthetic data and word frequency data sets we have seen already. Both produce results that would allows to conclude that these were power-law distributed.

In comparison to this the data set containing 10^5 standard normal generated random numbers has a an extremely low $p - value = 0.0060$ and a small tail ratio. Also its gof value of 0.0522 is quite large in comparison to the synthetic power-law and word frequency results, almost a decade larger. These three results can be used as a bench mark to enable us to reach a firmer conclusion about our hypotheses. For example if we look at positive tail for two month returns on the FTSE 100. The quoted $p - value = 0.5500$ is very large indicating that a power-law is a good fit to the data however if we consider its gof value of 0.1029, this twice as large as the standard normal value, signifying a large distance away from the power-law model. Similarly the tail ratio is small then the standard normal value, which leads us to conclude that the p-value for the positive tail for two month returns on the FTSE 100 can not be trusted. On the other hand if values for the tail ratio and the gof are closer to those of the synthetic power-law and word frequency values, coupled with a $p - value \geq 1$, we can say, in terms of this analysis that a power-law is a good fit.

Quantity			gof		$p-value$		x_{max}/x_{min}	
			Positive	Negative	Positive	Negative	Positive	Negative
FTSE 100	Returns	5-Min	0.0369	0.0143	0	0.9580		
		15-Min	0.0219	0.0268	0.8740	0.5290		
		30-Min	0.0290	0.0218	0.7100	0.9980		
		Daily	0.0277	0.0301	0.9980	0.8250		
		Weekly	0.0488	0.0592	0.3330	0.0250		
		Monthly	0.0932	0.0698	0.7810	0.9580		
		Two Month	0.1029	0.0693	0.5500	0.8400	1.4859	
	Volumes	Daily	0.0317		0.2260			
		Weekly	0.0595		0.6790			
		Monthly	0.1145		0.0050			
S&P 500	Returns	5-Min	0.0182		0.8220	0.0410		
		15-Min	0.0330		0.1520	0.760		
		30-Min	0.0504	0.0377	0.0220	0.6730		
		Daily	0.0355	0.0290	0.1320	0.0650		
		Weekly	0.0509	0.0581	0.2610	1.00e-003		
		Monthly	0.0780	0.0806	1.00e-003	0.5920		
		Two Month	0.0691	0.0567	0.5350	0.5820		
	Volumes	Daily	0.0365		0.5220			
		Weekly	0.0490		0.8890			
		Monthly	0.0759		0.9230			
Synthetic Standard Normal			0.0522		0.0060		2.2142	
Synthetic Power-Law			0.0060		0.7300		279.97	
Word Frequency in Moby Dick			0.0093		0.5070		2012.29	

Table 3: Results of the goodness-of-fit test proposed in [3] to returns and volumes on the FTSE 100 and the S&P 500 for timescales ranging from 5 minutes to 2 months. Also tested were data sets containing 10^5 randomly generated numbers from a power-law and standard normal distribution and a data set of word frequency in the novel Moby Dick. Shown are the gof - KS test statitic, $p-value$ from the hypothesis test and tail ratio.

8 Conclusion

The study of power-law distributions is an active area of research today. In the field of finance and economics the distribution of returns have been shown to exhibit power-law behaviour. In this paper we discussed how to detect power-law behaviour in empirical financial data. We examined the various properties of mathematics behind the power-law distribution, namely its scale invariance, its existence for values greater then some lower bound x_{min} and its moments. We also saw that graphically detecting its existence in empirical data, although necessary, is not the best approach to claiming power-law behaviour. We dealt with techniques for fitting distributions with a power-law model before testing the ggodness-of-fit of these power-law forms. We applied these test to our data and found similar results to that in [13, 12] although we could not conclusively replicate there findings. This could most probably be down to the size of the data sets analysed in the this paper. The largest set analysed here was roughly 7000 data points compared to 10^6 studied by Gopikrishnan et al.Ẇe found that conclusive results on the distribution of returns and volumes of trades are difficult to obtain, and require a large amount of data to study the rare events that give rise to the tails(i.e. rare events are located in the tails of distributions, that is why the are rare, very little probability of happening). This was the main limitation of this study.

9 Appendix

9.1 Synthetic Power-Law Data

To generate power-law random numbers we use the *inversion method* as mentioned in [3] and [22]. It is a basic and simple method of pseudo-random number generation from any probability distribution given its cumulative distribution function [7]. The method takes independently identically distributed variables, u, that have a uniform probability density function

$$p(u) = \begin{cases} 1 & \text{for } u \in [0,1] \\ 0 & \textit{otherwise} \end{cases} \tag{54}$$

and computes the number x such that $P(x) = u$, where $P(x)$ is the cumulative distribution function of x. The generated number x is then taken to be a number drawn from the distribution of $P(x)$. In other words given an invertible cumulative distribution function we can generate the random variable $x = P^{-1}(u)$.

We can derive an expression using the above approach for generating random numbers distributed as a power-law. If we consider the continuous probability density $p(x)$ for $x \geq x_{min}$ and $p(u)$, the uniform probability density given in (54) we can relate the two as follows

$$p(x)dx = p(u)du \tag{55}$$

$$p(x) = p(u)\frac{du}{dx}. \tag{56}$$

Integrating both sides with respect to x over there respective domains gives us a relation between both cumulative distribution functions

$$\int_x^\infty p(x')dx' = \int_u^1 p(u')\frac{du'}{dx}dx. \tag{57}$$

Taking the right-hand-side we find

$$\int_u^1 p(u')\frac{du'}{dx}dx = \int_u^1 p(u')du' = \int_u^1 du' = [u']_u^1 = 1 - u, \tag{58}$$

where $p(u) = 1$ in $[u, 1]$. This is also equal to

$$1 - u = \int_x^\infty p(x')dx' = P(x) \tag{59}$$

or as desired

$$x = P^{-1}(1 - u). \tag{60}$$

For our case, we consider the cumulative distribution function (14) we derived in section 3. This gives us the following expression from which we can compute synthetic power-law observations

$$x = x_{min}(1 - u)^{-1/(\alpha-1)}. \tag{61}$$

MATLAB has a built in random number generator for uniform distributed numbers, *rand()*, which I utilised with the above equation to generate the synthetic power-law data seen in figure (**??**).

10 Bibliography

References

[1] Bauke, H. (2007) 'Parameter estimation for power-law distributions by maximum likelihood methods', The European Physical Journal B, 58(2), pp. 167-173

[2] Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997) *The Econometrics of Financial Markets*, Princeton, NJ: Princeton Univertsity Press.

[3] Clauset, A., Shalizi, C. R. and Newman, M. E. J. (2009) 'Power-Law Distributions in Empirical Data', SIAM Rev., 51(4), 661-703.

[4] Clauset, A., Young, M. and Gleditsch, K. S. (2007) 'On the Frequency of Severe Terrorist Events', Journal of Conflict Resolution, 51(1), 5888.

[5] Cryer, J. D. and Chan, K-S. (2008) *TIme Series: With Applications in R*, 2nd ed., Springer Science & Business Media.

[6] Dahl, H. (1979) *Word Frequencies of Spoken American English*, Essex, CT: Gale/Cengage Learning.

[7] Devroye, L. (1986). 'General Principles in Random Variate Generation' in Non-Uniform Random Variate Generation, New York: Springer-Verlag. 27-40.

[8] Fama, E. F. (1970) 'Efficient Capital Markets: A Review of Theory and Empirical Work', The Journal of Finance, 45(2), 383-417.

[9] Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H. E. (2003) 'A Theory of Large Fluctuations in Stock Market Activity', Cambridge, MA: Massachusetts Institute of Technology, Dept. of Economics.

[10] Ghasemi, A. and Zahediasl, S. (2012) 'Normality Test for Statistical Analysis: A Guide for Non-Statisticians', International Journal of Endocrinology Metabolism, 10(2), 486-489.

[11] Gopikrishnan, P., Meyer, M., Amaral, L. A. N. and Stanley, H. E. (1998) 'Inverse cubic law for the distribution of stock price variations', The European Physical Journal B, 3(2), 139-140.

[12] Gopikrishnan, P., Plerou, V., Gabaix, X. and Stanley, H. E. (2000) 'Statistical Properties of Share Volume Traded in Financial Markets', Physical Review E, 62(4), 4494-4496.

[13] Gopikrishnan, P., Plerou, V., Amaral, L. A. N., Meyer, M. and Stanley, H. E. (1999) 'Scaling of the Distribution of Fluctuations of Financial Market Indices', Physical Review E, 60(5), 5305-5316.

[14] Hull, J. (2006) 'The Black-Scholes-Merton Model' in *Options, futures, and other derivatives*, Upper Saddle River, N.J., Pearson/Prentice Hall.

[15] Kolmogorov, K. (1933) 'Sulla determinazione empirica di una legge di distribuzione', G. Ist. Ital. Attuari, 4, 8391.

[16] Liu, Y., Cizeau, P., Gopikrishnan, P., Meyer, M., Stanley, H. E., (1998) 'Volatility Studies of the S&P 500 Index', in J. Kertesz and I. Kondor eds., Econophysics, an Emerging Science, Dordrecht: Kluwer.

[17] Liu, Y., Gopikrishnan, P., Cizeau, P., Meyer, M., Peng, C. K. and Stanley, H. E., (1999) 'Statistical properties of the volatility of price fluctuations', Physical Rev., 60(2), 1390-1400.

[18] Mandelbrot, B. (1963) 'The Variation of Certain Speculative Prices', The Journal of Business, 36(4), 394-419.

[19] Milojevi, S. (2010) 'Power-law Distributions in Information Science Making the Case for Logarithmic Binning', JASIST, 61(12), 2417-2425.

[20] Mitzenmacher, M. (2004) 'A Brief History of Generative Models for Power Law and Lognormal Distributions', Internet Mathematics, 1(2), 226-251.

[21] Myung, I. J. (2003) 'Tutorial on maximum likelihood estimation', Journal of Mathematical Psychology, 47, 90100

[22] Newman, M. E. J. (2005) 'Power laws, Pareto distributions and Zipf's law',Contemporary Phy., 46(5), 323-351.

[23] Piantadosi, S. T. (2014) 'Zipf's word frequency law in natural language:a critical review and future directions', Psychonomic Bulletin & Review, 21(5), 1112-1130.

[24] Razali, N. M. and Wah, Y. B. (2011) 'Power comparisons of Shapiro-Wilk, Kolmogorov-SMirnov, Lilliefors and Anderson-Darling tests', Journal of Statistical Modeling and Analytics, 2(1), 21-33.

[25] Simon, H. A., (1955) 'On a Class of Skew Distribution Functions', Biometrika, 42(3), 425-440.

[26] Smirnov, N. (1948) 'Table for estimating the goodness of fit of empirical distributions', Annals of Mathematical Statistics, 19, 279-281.

[27] Stumpf, M. P. H. and Porter, M. A. (2012) ‘Critical truths about power laws’, Science, 335(6069), 665-666.

[28] Wilk, M. B. and Gnanaadesikan, R. (1968) ‘Probability Plotting Methods for the Analysis of Data’, 55(1), 1-17.

[29] Yang, C. (2013) ‘Who’s afraid of George Kingsley Zipf? Or: Do children and chimps have language?’, The Royal Statistical Society, 29-34.

[30] Yule, G. U. (1925) ‘A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S.’, Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character, 213, 21-87.

[31] Zipf, G. K. (1935) *The Psycho-Biology of Language*, Cambridge: Mass.