Record statistics and random walks in financial time series

A thesis submitted towards partial fulfilment of
BS-MS Dual Degree Programme

by

Behlool Sabir

under the guidance of

Dr. M. S. Santhanam

Department of Physical Sciences
Indian Institute of Science Education and Research,
Pune

Certificate

This is to certify that this thesis entitled *'Record statistics and random walks in financial time series'* submitted towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research Pune represents original research carried out by *Behlool Sabir* at IISER, Pune, under the supervision of *Dr. M.S. Santhanam* during the academic year 2012-2013.

Student
BEHLOOL SABIR

Supervisor
DR. M.S. SANTHANAM

Dedicated to, people I love.

Acknowledgements

The success of this study required the help of various individuals. Foremost, I would like to express my sincere gratitude to my supervisor Dr. Santhanam for his continuous support in my thesis research work, for his ***patience***, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I cannot imagine having a better advisor and mentor for my project.

Besides my advisor, I would like to thank the rest of my thesis committee, for their encouragement and insightful comments.

My sincere thanks also goes to Dr. Arijit Bhattacharya, Dr. Saikrishanan Kayarat and to Dr. Ranjith Padinhateeri, for offering me project opportunities in their groups and leading me to work on diverse exciting projects.

I thank my room-mate Shashawat Antony, for the stimulating discussions, for the sleepless nights we were working together, and for all the fun we have had in the last couple of years. I would also like to express my heartfelt gratitude to my friends in IISER Pune: Ayesha Fatima, Ashutosh Agnihotri and all the others who were here to offer assistance and for all the healthy discussions we had. I would like to thank all the people from '06 and '07 batch for their continuous remote support and guidance and '08 batch for being gentle to me. I am grateful to Shweta for being with me through all the phases. I would also like to thank Surabhi Jirapure for assisting me in writing this piece with her literary skills.

I would also like to thank Taresh for his service in providing beverages and snacks which would turn siesta time into a fiesta and increase our working efficiency.

I would sincerely thank DST for the monetary support as it would had been a little difficult without them.

Linux, C, GCC, AWK and BASH: this piece of research would not have been possible without them!

Last but not the least; I would like to thank my family: my parents, brother and sister, for supporting me spiritually throughout my life.

Abstract

This project aimed at analysing the records statistics in stock price movements and mathematically model it. Probability distribution of the record gap for the stock price movements were determined. Power-law is observed in these probability distribution. Stochastic models namely, random walk without drift, random walk with finite drift and geometric random walk were simulated to generate time series which reproduces the signature properties of the stock price movements. These time series were then statistically analysed and probability distribution for record gaps was determined. Similar statistics were done for empirical stock market indices.

What?
"What" ain't no country I ever heard of. They speak English in "What"?
-Pulp Fiction

Contents

Chapter 1

Introduction

Record is the highest or lowest value which has not occurred earlier. The time of occurrence of the record is the record time. Let $X(t)$, $t \in 1, 2, 3, ...$ is a time series. For $X(T)$ to be a record (highest), its value should be such that $X(T) > X(T-1), X(T-2), ...X(1)$. For this event $X(t = T)$ will be the record value and T will be the time at which record was created. Record value is a captivating property to observe, like highest number of centuries scored by any batsman in cricket, maximum number of copies sold-out of any book, lowest value of dollar-rupee exchange in the past decade etc. Beyond such entertainment value, records can be studied to determine or predict the important future aspects, for instance occurrence of natural calamities [1], environmental changes [2, 3, 4], financial prices [5, 6, 7] etc.

There are processes and time series for which record time is as important as the record value. The emphasis in this thesis is laid on financial time series values for which the time at which records were observed is of considerable importance. There are several price movements related to financial data viz. currencies, stocks, funds, equities etc, but majority of the work here will be concentrated on individual stock price movements and on the stock market indices. In stock price movement, time series of *daily closing value* is studied, wherein daily closing value is the price of the stock or the value of the index at which market closes down. Large amount of stock market data was required to analyse records statistics. We have accessed them from publicly available sources, namely, *Yahoo Finance* [8] and *Google Finance* [9]. The data available in these web sources contain details like date, opening value, highest value, lowest value, closing value, volume and adjusted closing value for each trading day. Few rows from the actual data obtained from *Yahoo Finance* is shown in the table 1.1.

In majority of the current work, time series of the closing values of stocks are observed and the record time instances are filtered and analysed to deter-

Table 1.1: Few rows of the empirical data for *IBM* obtained from *Yahoo Finance*.

Date	Open	High	Low	Close	Volume	Adj Close
2012-08-17	19.52	19.53	19.26	19.52	14626800	19.52
2012-08-16	19.43	19.60	19.22	19.52	17835400	19.52
2012-08-15	19.29	19.40	19.18	19.29	10988200	19.29
2012-08-14	19.76	19.86	19.27	19.36	18077800	19.36
2012-08-13	19.69	20.07	19.48	19.62	13863900	19.62
2012-08-10	19.30	19.73	19.28	19.70	18170700	19.70
2012-08-09	19.40	19.56	19.06	19.41	20192600	19.41
2012-08-08	19.48	19.75	19.24	19.41	44990300	19.41
2012-08-07	18.56	19.05	18.51	18.96	19670100	18.96
2012-08-06	18.29	18.82	18.23	18.69	15318000	18.69
2012-08-03	17.83	18.33	17.72	18.26	18989200	18.26

mine the probability for the occurrence of the records in the future. Although, not much stress is given on the record values of the stocks, focus was laid on record gap distributions of the stock price movements. On determining this statistic on individual stocks and stock market indices, consistent power-law of the form $f(x) = x^{-\gamma}$ with $1 < \gamma < 2$ was observed. A longer time series was required to determine the exponent and to generalise and strengthen the argument of a consistent power law throughout the stock market data.

Occurrence of records in a typical financial time series are very limited. Financial data available in the public domain for even the old individual stocks for example IBM, HPQ (refer appendix A for the company and listing names) etc. has data for $\sim 50-60$ years. As stock market is operational for about 252 days every year there are ~ 13000 data points. Since the amount of data of this order is not sufficient for the purpose of analysis of records and statistics, therefore, a part of the dissertation was dedicated in finding the appropriate model to generate synthetic time series for stock prices. One of the first reported work, to model the financial data was done by Louis Bachelier [5] in his PhD thesis work. Louis Bachelier modelled asset return prices, the core assumption as a *random walk (RW)*.

The basic form of a random walk can be defined as:

$$X_{n+1} = X_n + \xi_n \,, \tag{1.1}$$

where ξ is independent and identically distributed (iid) random variable (RV). There are several previously studied financial data based on RW [10, 11, 12, 13, 14]. In my current work, RW time series were generated using iid RVs with uniform and normal distributions. Various statistics were observed, such as return values ($R = X_{n+1} - X_n$), record values (r), return records, record gaps (r_g), log return ($R_{\log}$) etc. To improve the understanding of the record statistics further, mean number of records were verified which was discussed by *Majumdar and Ziff(2008)* [15]. The main purpose of this dissertation is to analyse the record gap distribution employing time series and an ensemble of time series wherever necessary. Empirical stock data has some signature properties which are termed as *stylised facts* [16]. These facts are observed in almost all the stock price movement time series. These were complied over decades of observations and analysis of stock market data. Random walk model was tested against these stylized facts. The analysis of the model and further considerations yielded a need for a better model.

In the standard random walk, the mean position and hence for the drift is zero. Thus random walk does not show any preferred direction in the absence of this drift. In order the take into account drifts in the mean of the financial series, we consider random walk with a drift. Random walk when added with a constant drift (c) term is called as random walk with a drift or biased random walk.

$$X_{n+1} = X_n + \xi_n + c$$

Gregor Wergen et al. in 2011 [17] published a work that takes record analysis for random walk further to target the financial data. The work was focused primarily on the mean number of records. Using this as the model, work here was further developed to determine the record statistics, similar to what was determined for random walk model without drift. Further study of the model found a few crucial properties, such as variable return profile and desired record count of the stock price movements, missing.

Geometric random walk(GRW) is considered to be an appropriate model to generate stock price movement time series [18]. *LeRoy and Parke* in a study of volatility of market used GRW as the model to generate the synthetic stock price movements. GRW is defined as:

$$X_{n+1} = X_n \xi_n$$

GRW has varying steps (non uniform jumps) [19]. In this thesis, GRW model is used to generate time series. Stylized facts of stock price movements are tested upon them. To generate time series using GRW, data of distribution of log return is required. *Clark* in 1973 [20] proposed a model which claims log returns to be Gaussian distributed. Here in this thesis, log return distribution is assumed to be Gaussian. Though, this model is under scrutiny [21] and some even called it as a dangerous assumption to make [22]. Time series were generated using Gaussian log returns, and its ensemble was used to determine record gaps and its probability distribution. Record statistics were also done for the stock market indices.The aforementioned mentioned models namely, RW, biased RW and GRW were analysed from the perspective of stock market indices to determine the probability distribution of record gaps.

In the subsequent chapters, using some of the available models for stock price movements, we analysed record statistics and obtained numerical values of power-law exponents for the probability distributions of record gaps, i.e, the time intervals between the occurrence of subsequent records.

Chapter 2

Financial data analysis: empirical results

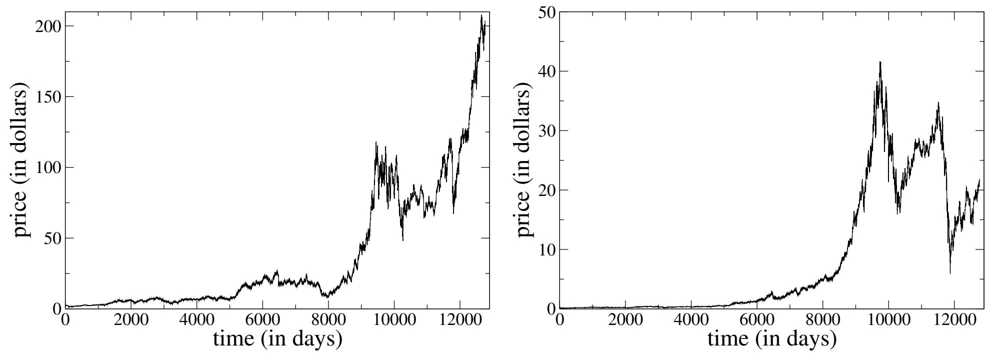
2.1 Historical data mining

To analyse the empirical data, the prime requirement is to gather the desired data which is open to public. Numerous long historical financial time series were required for statistical analysis and modelling. *Yahoo finance* and *Google finance* are one of the few free publicly available websites to fetch such financial data. These websites contain only daily values for long historical time series. Details like *opening value, highest value, lowest value, closing value, adjusted closing value* and *volume transacted* are available for each day.

2.1.1 Closing value

Desired values from the fetched data were the *adjusted daily closing* values. Adjusted closing value is the closing value which is *splits* and *dividends* corrected. Splits are the events when company revise the price of their respective stock thereby changes the number of stock owned by any individual. Most common split is when one share is replaced by two thereby reducing price of each stock to half of earlier. Historical data of closing value which is not corrected, is closing price without taking care of *dividends* and *splits*, if any. Figure 2.2 show raw closing value and adjusted closing value for *IBM* for year 1962-2012 plotted on top of each other and few of the splits are also marked inside the plot.

Figure 2.1(a) and 2.1(b) shows the historical *adjusted daily closing* values for two of the *New York Stock Exchange (NYSE)* listed stock prices. The increasing trend observed here in the long time lapse is because of the



(a) IBM (b) GE

Figure 2.1: Daily closing values from year 1962 to 2012

market inflation. Henceforth in this thesis, *closing value* should always be treated as *adjusted closing value*

2.2 Stylized facts

Stylized empirical facts are signature properties of a financial time series [16], and are based on a vast amount of financial data studied in the last 2-3 decades. Some of the important stylized facts are stated below.

1. Absence of autocorrelation in the daily return values : Let $X(t)$ represent the daily closing price of a stock at time t. Then, daily return $R(t)$ of the stock is defined as $R(t) = X(t+1) - X(t)$. One of the fundamental principles of the stock markets is that the stock price returns are memory-less. For the traders in the market, this implies that short-term profits cannot be made by relying upon the past performance. This idea is mathematically captured by the behaviour of the autocorrelation function $C(\tau)$ defined as,

$$C(\tau) = \lim_{T\to\infty} \frac{1}{T} \frac{\sum_{t=1}^{T} (R(t) - \bar{R})(R(t+\tau) - \bar{R})}{\sigma_R^2}, \ \tau \in [0, 1, 2, ..., T] \quad (2.1)$$

   where T is the total length of the time series, τ is the time lag, $\bar{R}$ is the mean of the return time series and σ_R is its standard deviation. This is illustrated for returns of IBM stock values in figure 2.3.

2. Distribution of daily return values shows heavy-tailed trends:

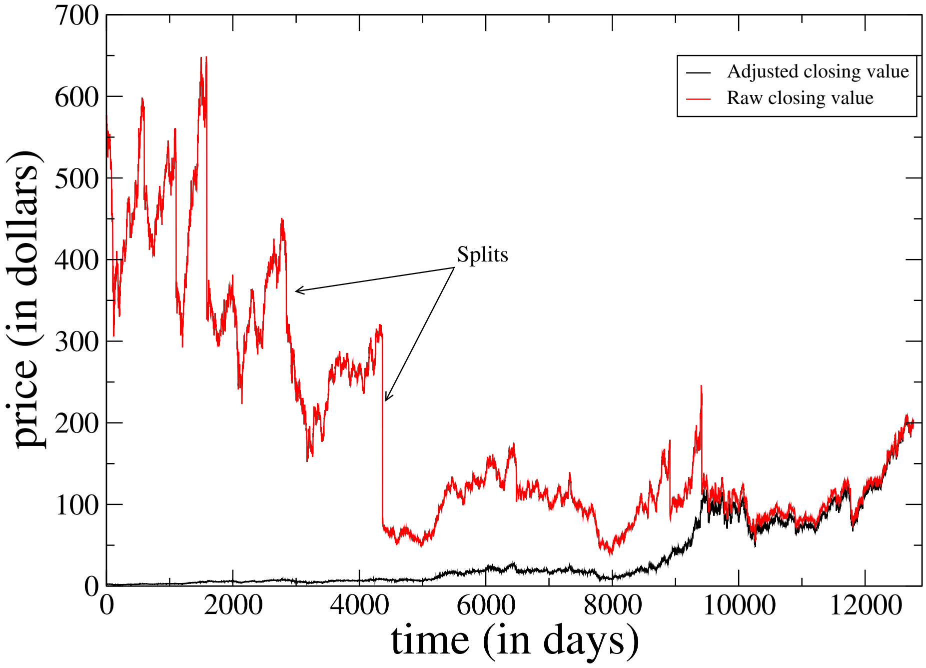


Figure 2.2: IBM adjusted and non-adjusted closing values

If the behaviour of the distribution, $f(R)$ as $R \to \infty$, is slower than exponential decay, then it is called a heavy-tailed distribution. It is generally observed that return distribution of the daily prices of the stock shows heavy-tailed trends. This is illustrated in figure 2.4 which shows the log-log plot of the distribution of the returns for daily stock prices of *IBM*, where $f(R) \sim R^{-1.7}$. *Power-law* in $f(R)$ as $R \to \infty$ is a signature of heavy-tail in return distribution.

3. Volatility of stock are clustered:

   Volatility is the measure of fluctuation in the time series. Volatility in financial time series tend to cluster and shows positive correlation. There are various ways of measuring volatility. One of them can be absolute returns $|R(t)|$, where returns can be of several types viz. daily, weekly, fortnightly, monthly etc, where clustering can be observed.

4. Slow decay of autocorrelation in the absolute returns:

   Absolute return values of the financial time series shows long range dependence, which implies the decay is slower than exponential, typically power-law decay of autocorrelation. Long range dependence can

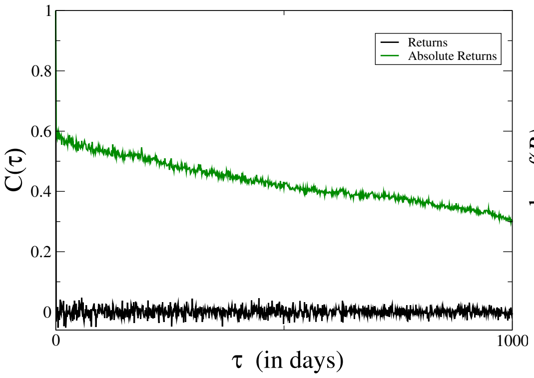


Figure 2.3: Autocorrelation of return values of IBM stock data

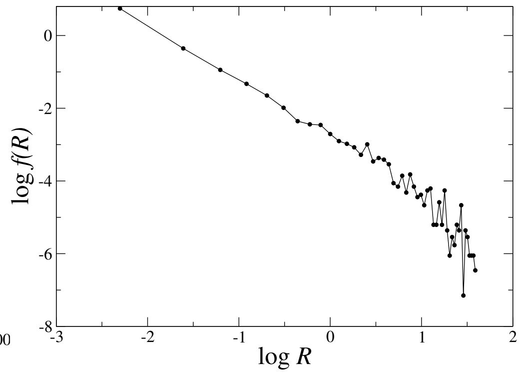


Figure 2.4: Log-log distribution of returns of IBM stock data

be defined as

$$C(\tau) \sim \tau^{-\alpha}, \quad (0 \leq \alpha \leq 1), \tag{2.2}$$

where γ is the autocorrelation exponent. In most of the stock market data, autocorrelation of absolute returns follows power-law of the form,

$$C(\tau) \sim \tau^{-\alpha}, \quad (0.2 \leq \alpha \leq 0.4). \tag{2.3}$$

The value of α indicates that it falls under the regime of long range dependence.

These statistical properties are observed in most of the stock-market data. Stylized facts were verified on several empirical data and the plots for *IBM* are shown in figure 2.3 and 2.4.

2.3 Record statistics

As used in common parlance, records are created when extreme values are reached at a given instant in time. Record position in a time series is a position, where the corresponding value till that time is the highest. Consider a time series x_t, $t = 1, 2, 3, ...$ At time $t = \tau$, x_τ will be called a record if $x_1, x_2, \cdots, x_{\tau-1} < x_\tau$ In figure 2.5, this is shown for the closing values of the IBM stock during 1962-2012. The enlarged version of same figure shows few of the record positions (indicated by the arrows) in this time series data. Records in the time series can be ranked as an integer sequence. By construction, the first occurrence of the record or r_1 corresponds to x_1. The rank r_2 corresponds to the second occurrence of record and so on. Consider r_i

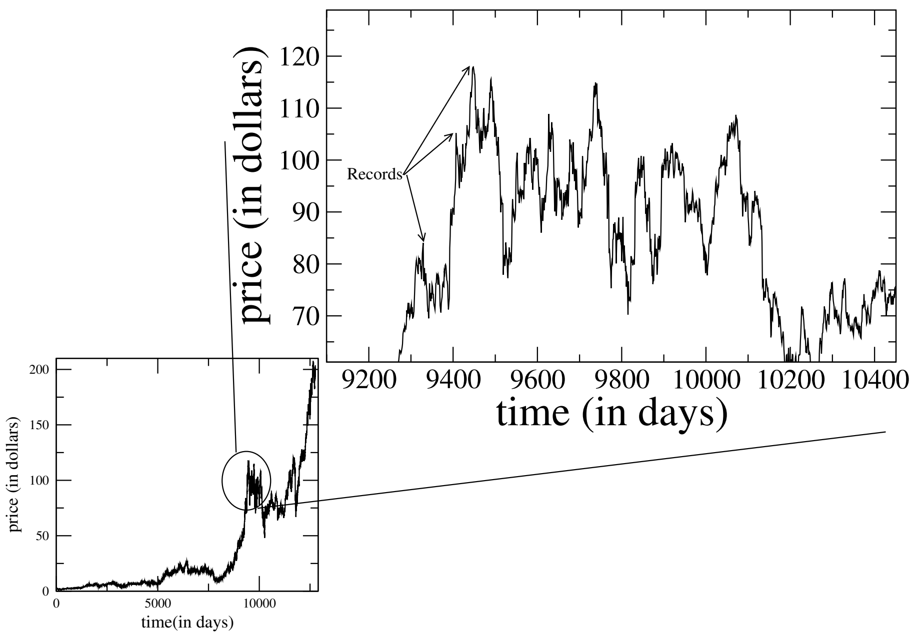


Figure 2.5: Figure shows the closing time series of IBM stock values for year 1962-2012 in the inset and few record positions are marked in the blown up region

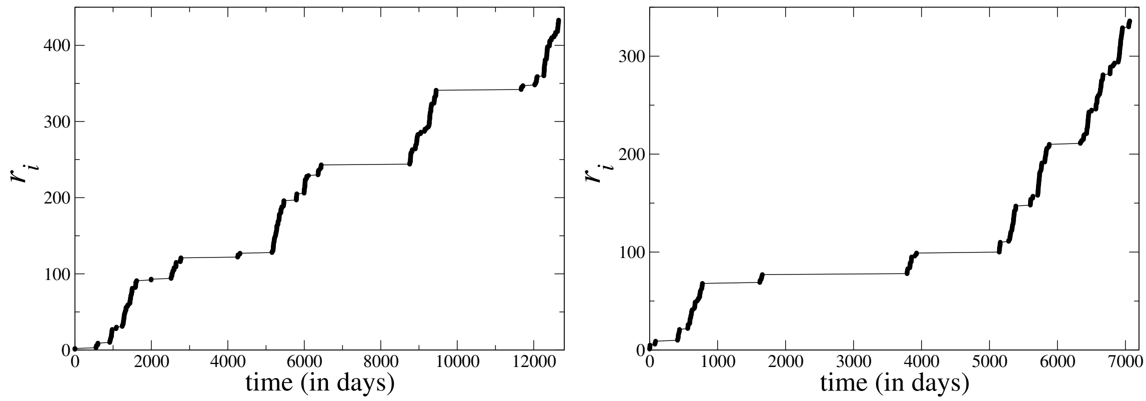
as the record position of the i^{th} record or i^{th} rank. Plots for record positions for stock data from *IBM* (Figure 2.6(a)) and *Apple* (Figure 2.6(b)) are shown here.

2.3.1 Mean number of records

In statistical analysis of records, mean number of records is an important characteristic. Mean number of records can be defined as the total number of records in a group of realisations till that time. Let $m(x(T))$ represent the total number of records in a given realisation of time series $x(t)$ until time T. Then the mean number of records is

$$\langle m(T) \rangle = \lim_{N\to\infty} \frac{\sum_{i=1}^{N} m(x_i(T))}{N} \tag{2.4}$$

This will be further discussed later in the thesis. Several other type of means can also be defined. For instance, mean number of records at any given time,



(a) IBM

(b) Apple

Figure 2.6: Rank of records against time

can be illustrated mathematically as

$$r'(t) = \frac{m(x(t))}{t} \tag{2.5}$$

Figure 2.7 shows mean number of records against time for stock data of *IBM* (Figure 2.7(a)) and *Exxon* (Figure 2.7(b)).

2.3.2 Record gap

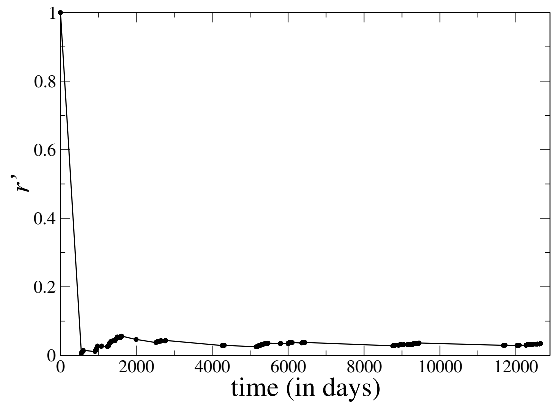
Record gaps are the intervals between adjacent record positions. As mentioned above in section 2.3, r_i is the record position for the i^{th} record. Then record gap, r_g, can be defined as

$$r_g(i) = t(r_{i+1}) - t(r_i),\ i \in [1, 2, 3, ...], \tag{2.6}$$

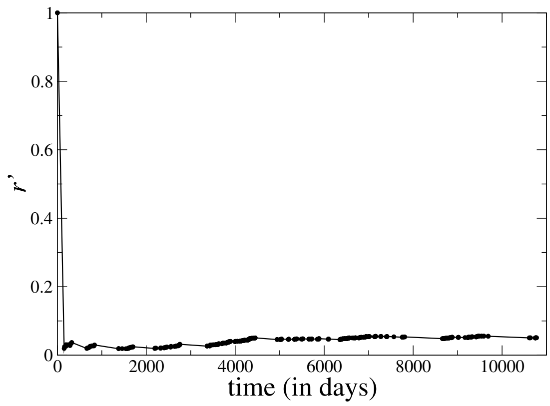
Total number of record gaps will always be one less than the total number of records.

Record gaps for closing values of empirical data of *IBM* are shown in figure 2.8. For the daily closing values of *IBM* for approximately 50 years $\sim$ 12000 days, 432 record positions were observed. This gives 431 values of record gaps. In this realisation, gaps ranging from as low as 1 day to highest of 2313 days are present, whereas mean of record gaps came out to be

$$\frac{\sum_{i=1}^{n} r_g(i)}{n} \sim 30,$$



(a) IBM



(b) Exxon

Figure 2.7: Mean number of records against time

which indicates that most of the records are clustering at the lower ends of the values. To quantitatively speculate the size of gaps which are more in number and the others which are rare, distribution of record gaps can be determined.

Record gap distribution

Record gap distribution, $\phi(r_g)$, is defined as number of values of r_g lying between r_g and $r_g + dr_g$. When area under the distribution curve is made unity, the plot can be treated as probability distribution of the record gap size. These statistics were performed on several empirical stock, individual as well as on index data. Plots for *IBM* and *GIS* in individual stocks and *DJT* and *DJU* in stock indices are shown here. From the distribution ($\phi(r_g)$) plots (figure 2.9) of the stock market data, it is evident that record gaps with low values are abundant compared to bigger record gaps. As these plots can also be treated as probability plots, it can be concluded that probability of gap of 1 between records for *IBM* is ~ 0.5 and for *GIS* ~ 0.4 in individual stocks case. For stock average indices *DJT* and *DJA* it is ~ 0.55. On further analysis and curve fitting, all of the above empirical data depicted a power-law trend. Following are the captured trends for each:

$$\phi(r_g)_{IBM} \propto r_g^{-1.64}$$

$$\phi(r_g)_{GIS} \propto r_g^{-1.49}$$

$$\phi(r_g)_{DJT} \propto r_g^{-1.54}$$

$$\phi(r_g)_{DJA} \propto r_g^{-1.66}$$

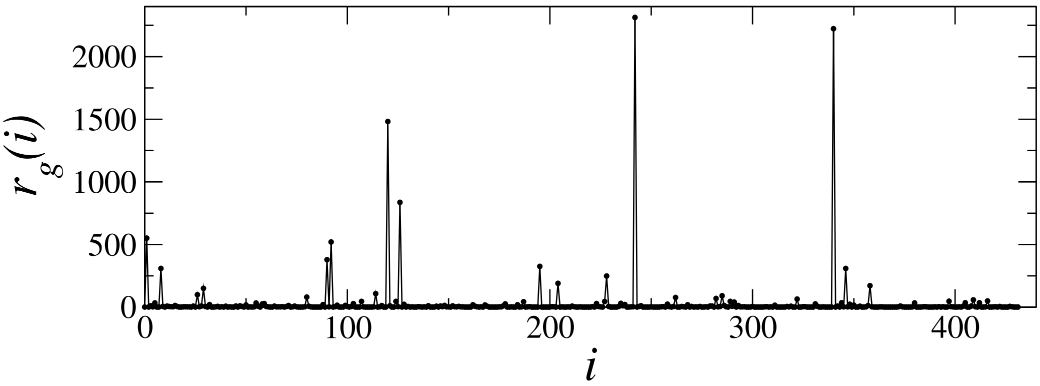
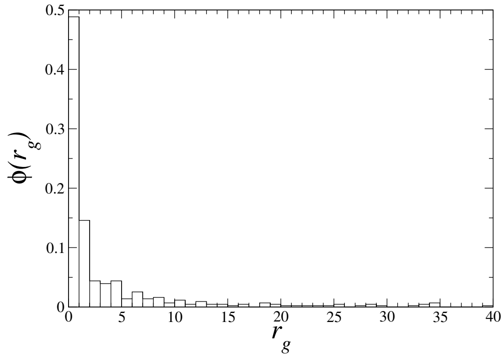
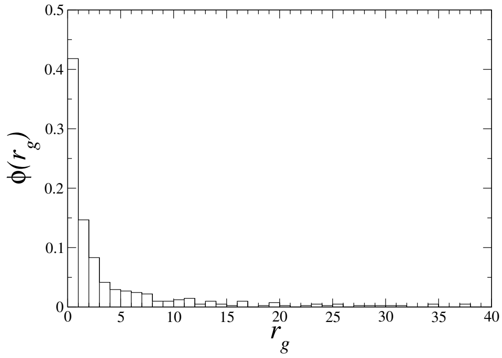


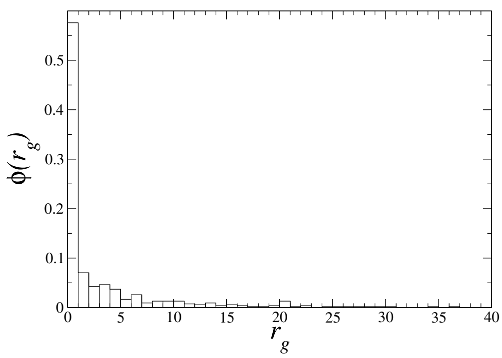
Figure 2.8: Record gap $r_g(i)$ plotted against i



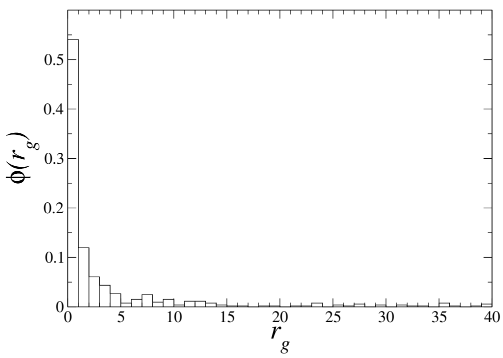
(a) IBM



(b) GIS



(c) DJT



(d) DJA

Figure 2.9: Record gap distribution or probability distribution of record gaps for empirical data of individual stocks (a & b) and market indices (c & d)

Power-law behaviour of the record gap distribution signifies strong clustering of record events in time scale of the time series. Detailed analysis of this will be covered in the upcoming chapters.

Chapter 3

Record statistics and financial time series

3.1 Financial time series and random walks

Random walk is considered as a corner stone for stochastic processes and statistical physics. At the first glance, financial time series also looks random. Financial time series are randomly evolved [5, 23, 6]. The earliest recorded study on the modelling of *stochastic process* [5] was done by Louis Bachelier in his PhD thesis work, which attempted to study the financial market using statistical tools and gave birth to *financial mathematics*. Several other works are done on stochasticity of the financial data [24] [25].

Another theory efficient-market hypothesis [24] which was proposed in the early 1960s and later published in an article, says that the financial prices and trades ruminates the public available information, which relates to the random-walk model.
Thomas Hellström in *A random walk through the stock market* [25], showed the links between the random-walk hypothesis and financial time series.

3.2 Records in random walks

Studying random walks and its statistics will help in providing a background to study the financial time series. Random walk is defined as:

$$X_n = X_{n-1} + \xi_n \tag{3.1}$$

where the seed, X_0, can be given any arbitrary value and ξ_n is an *independent and identically distributed (iid)* random variable.

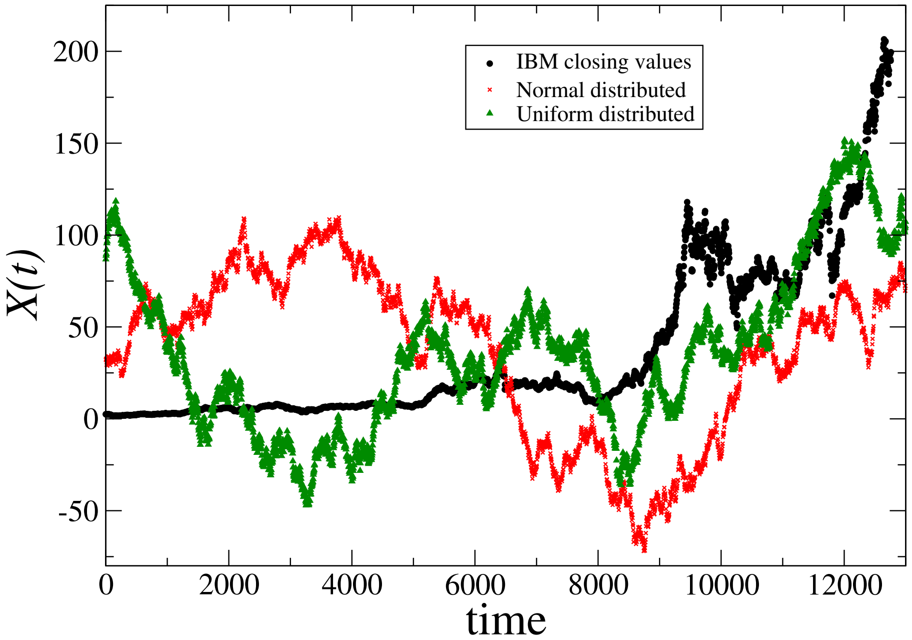


Figure 3.1: Comparative plot of random walk with normal and uniform distributed random variables and empirical *IBM* closing time series from year 1962-2012

3.2.1 Closing values

To gauge the relatedness of financial time series and random walk, a comparative study is done. Equating some of the signature statistics of financial time series with random walk's statistics are done in this section. Plot of random walk and empirical *IBM* time series of *closing value* is shown in the figure 3.1. Both the random walk series taken here are an ensemble average over 10^4 random walk realisations. It is ensured that mean and standard deviation of the random walk in both the cases is kept same as that of the empirical data. For both the time series, random walk of almost 13000 data points which is nearly 50 years of data, can show an overall declining trend and might even hit negative values unlike the most empirical data which has an increasing trend in long time lags and will never go negative.

3.2.2 Records

It is observed that occurrence of records in random walk is less as compared to empirical stock time series. In a time series of length 13000 data points for random walk, only 107 and 75 records points were observed for normally and uniformly distributed random variable respectively. Whereas 433 record points were observed in stock data of *IBM* for nearly the same length of time series. For further analysis and statistics of the random walk, longer time series (2520000 data points equivalent to 10000 years of time) will be considered so that enough data points come up for better statistics and approximations.

3.2.3 Stylized facts

For the extended time series of random walk realisations, autocorrelation for returns is observed. Returns in random walk will be same as the random variable (ξ_n). It is found that correlation is absent in both, uniformly distributed and normally distributed random walk time series return values, which is similar to what was observed in the empirical data. Return distribution of random walk realisations will be same as that of the random variables (ξ_n) (from Eq. 3.1), which in this case are uniformly and normally distributed.

3.2.4 Record gap distribution

Record gap distribution will be an interesting property to observe for random walks, which also is an emphasis of this dissertation. Figure 3.2 shows record gap distribution of random walk with uniformly distributed (figure 3.2(a)) and normally distributed (figure 3.2(b)) random variables. Descending pattern as observed for the empirical data (figure 2.9) is seen here as well. On observing the plot on the log-log scale, it shows an approximately straight line with a negative slope, this shows the trend observed is a power-law. Further curve fitting and determination of the exponent distribution of record gaps can be mathematically expressed as:

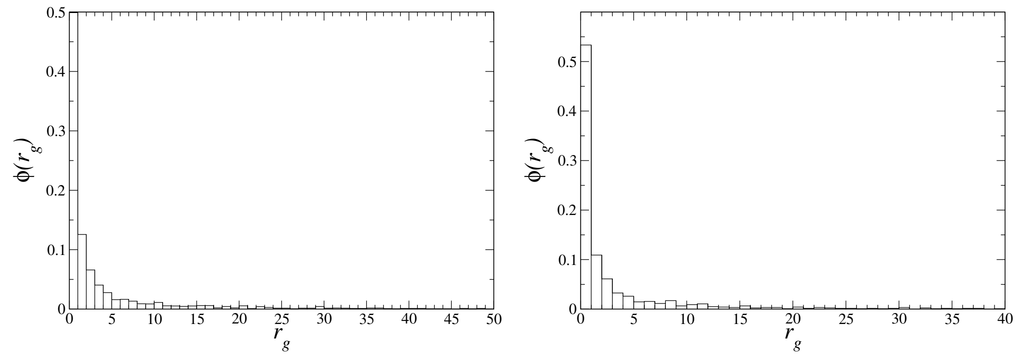
$$\phi(r_g)_{uniform} \propto r_g^{-1.63}$$

$$\phi(r_g)_{normal} \propto r_g^{-1.56}$$

3.2.5 Mean number of records

Probability of records in random walks [15] $P(M,t)$ of M records in t time steps where ($M \leq t+1$) is given by

$$P(M,t) = \begin{pmatrix} 2t - M + 1 \\ t \end{pmatrix} 2^{-2t+M-1} \tag{3.2}$$



(a) Uniformly distributed

(b) Normally distributed

Figure 3.2: Distribution of record gaps for random walk time series

above equation when approximated for large t gives

$$\langle M \rangle \sim \frac{2}{\sqrt{\pi}}\sqrt{t} \tag{3.3}$$

To verify the above equations (3.2 and 3.3), a simulation was performed which calculates mean number of records for a random walk with the desired random variable (uniform and normal). Figure 3.3 shows the mean number of records $\langle M \rangle$ plotted against the time step t for random walk with uniformly and normally distributed random numbers. Simulation was performed over 10^7 time steps for each realisation and then an ensemble average over 10^3 such realisations was done for both random walk, one as uniformly distributed and the other as normally distributed random variable.

3.2.6 How apposite is the model?

Random walk can be considered as a good model for closing value of stock market data, as some of the stylized facts like autocorrelation of return values have no memory, which is expected for stock data. Record gap distribution also matches with the empirical data's record gap distribution. On the contrary, facts like distribution of returns, number of records in a time series do not match with the empirical data of stock.

3.3 Random walks and finance data

Random walk model with *iid* random variables is not apt for record study of the stock market data. Records produced for random walk realisations were

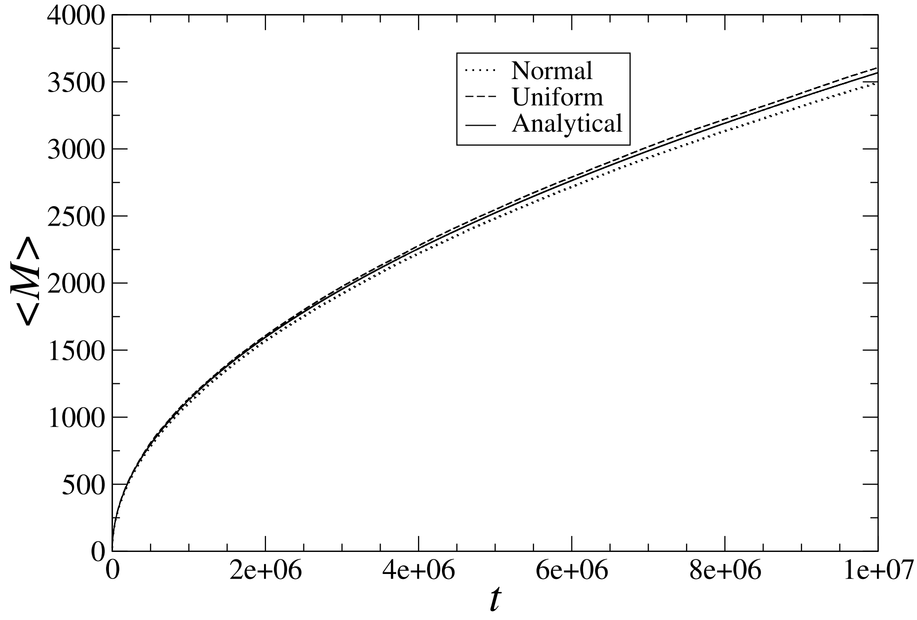


Figure 3.3: Mean number of record for RW with uniform and normal distributed RVs and the analytical result shown in Eq. 3.3 plotted against the time steps

very less as compared to what is expected in the empirical time series. So, a modified version of the random walk, *biased random walk*, is used[6, 23].

3.3.1 Biased random walk: The model

Random walks when given a constant drift/bias, give a broad trend to the time series. Random walk of the form

$$X_n = X_{n-1} + \xi_n + c \,, \tag{3.4}$$

where c is the constant bias/drift given to the system and ξ is the *iid* random variable, are biased random walks.

3.3.2 Closing value

Using this as the model, closing values are generated. Uniformly and normally distributed random variables are considered here. Simulations of random walk with a drift,were performed for time length approximately equal to the length of the empirical time series of the *IBM* closing values (~ 13000 days) and an ensemble average is taken over 10^4 such realisations. Drift is

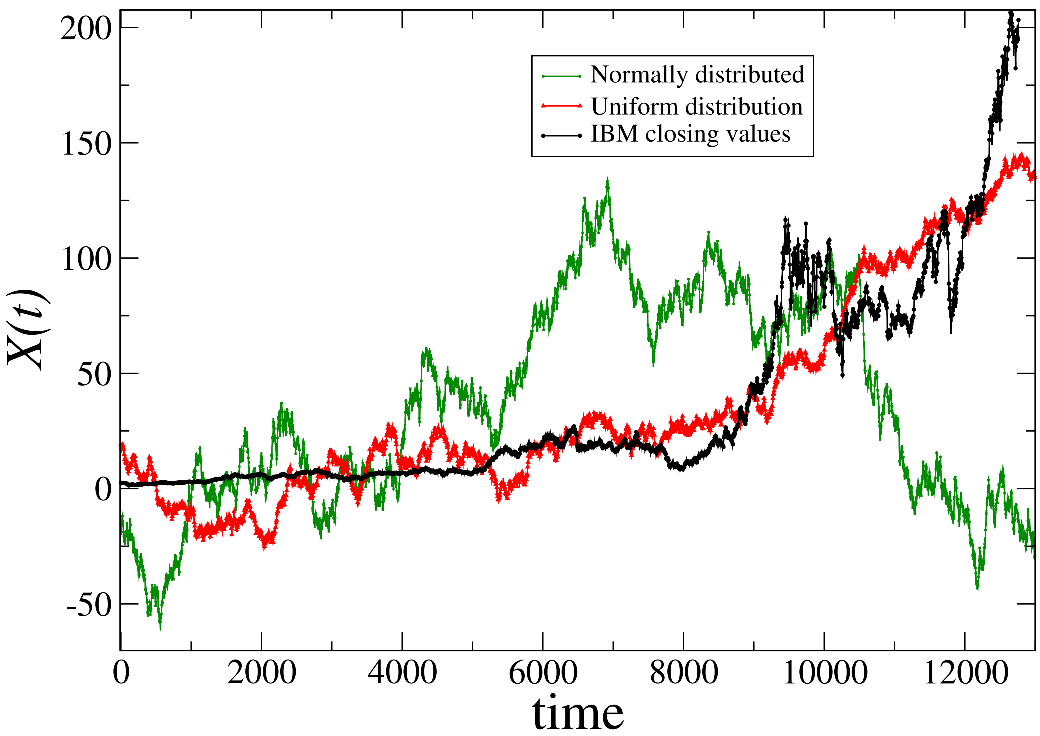


Figure 3.4: Comparative plot of random walk with a drift, with normal and uniform distributed random variables and empirical *IBM* closing time series from year 1962-2012

adjusted in such a way, that the final time series comes out to be very close to the empirical series statistically; in this case drift is taken as $c = 0.001$. Drift in the random walk is quantified [6] as c/σ, where σ is the standard deviation of the random variable. Plot of the ensemble average of random walk with a drift with random variables of both the types (normally and uniformly distributed) are shown in the figure 3.4. It was ascertained that both the random walks have the same mean and standard deviation.

3.3.3 Limitations of the model

Models such as biased random walk have a few problems and limitations to be used as stock market data. It is observed in the figure 3.4, like unbiased random walk, biased random walk can also hit negative values depending on the drift(c) given. Another major problem with both random walk and biased random walk is that, the return values distribution is irrespective of the part of the time series. Say for instance, if the stock starts at a price of \$2 and reaches a value of \$50 after some years. Then the return value

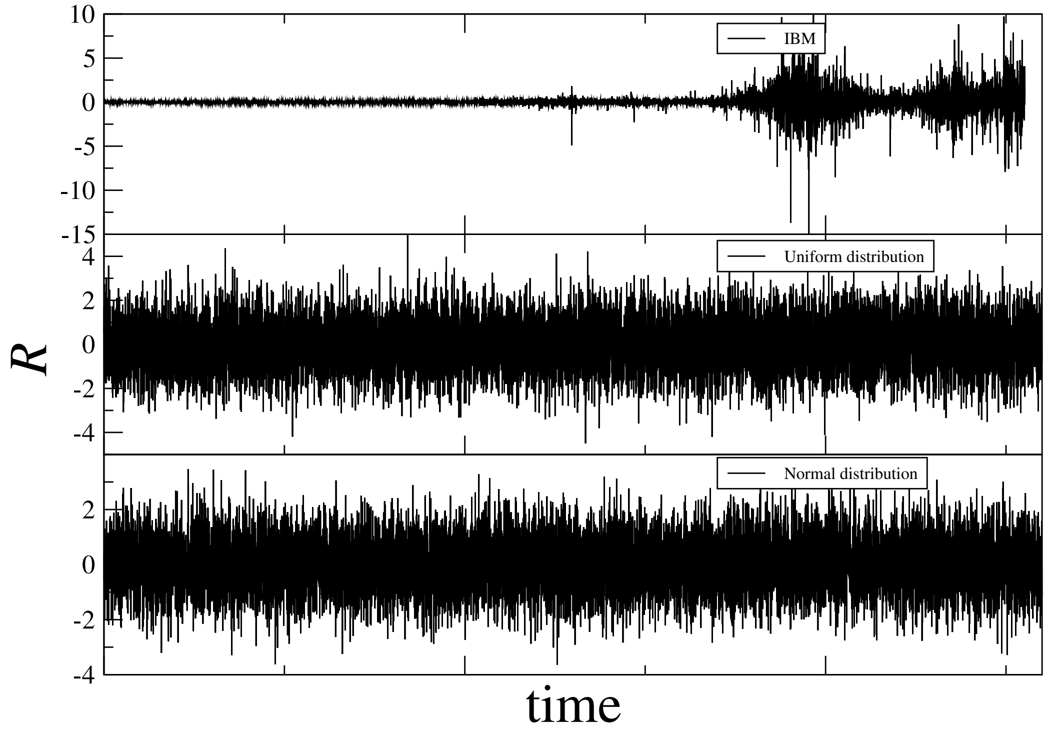
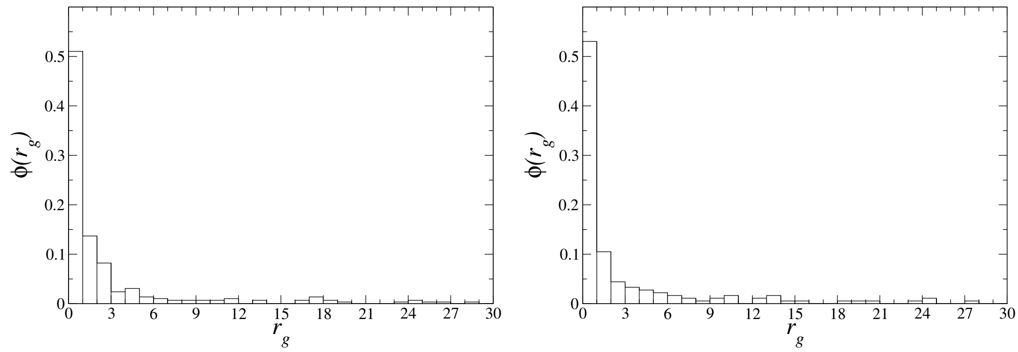


Figure 3.5: Return values for *IBM* (year 1962-2012) and RW with uniformly and normally distributed RVs plotted against time.

distribution will be the same in both the regions with closing values as \$2 and \$50, which is not the case with the empirical data. For empirical data, volatility varies with the increase/decrease in the closing value. Figure 3.5 depicts the profile change with respect to time in the empirical data of *IBM* which is compared with nearly constant profile for returns in random walk with uniformly and normally distributed random variables.

3.3.4 Record gap distribution

In biased random walk model simulation, significant number of records were observed. 293 and 182 record data points were observed for uniformly distributed and normally distributed random variable biased random walk respectively. Record gap values ranging from 1-3596 and 1-1765 were observed for uniformly and normally distributed random variable biased random walk respectively. Record gap distributions of the aforementioned time series are shown in the figure 3.6, which appears similar to the descending pattern observed for the empirical (figure 2.9) as well as the other random walk model without drift (figure 3.2). On taking ln on both the axes of the record dis-



(a) Uniformly distributed (b) Normally distributed

Figure 3.6: Distribution of record gaps for time series of random walk with a drift

tribution curve, data points showed an approximately straight line which suggests that record gap distribution of the simulated data for the model is a power-law. This can be mathematically expressed as:

$$\phi(r_g)_{uniform} \propto r_g^{-1.33}$$

$$\phi(r_g)_{normal} \propto r_g^{-1.21}$$

This when compared to the random walk model without drift model, is deviated from the empirical result of the record gap distribution.

Chapter 4

Geometric Random Walk

In the case of standard random walk time series given in Eq. 1.1, the change represented by $X_{n+1} - X_n$ is taken to be a random variable from a stationary distribution. In contrast, for most stock market data the change is not stationary 3.5 [26]. A more suitable model should also capture the non-uniform changes in stock market prices. In this context, geometric random walk is a commonly used model to generate synthetic stock market time series [18]. In contrast to the standard random walk, it has a peculiar behaviour of non-uniform jumps which implies that the change depends on X_n.

4.1 Geometric random walk: The model

GRW, like other random walk models, needs a seed X_0. It is defined by

$$X_{n+1} = X_n \xi_n \tag{4.1}$$

where ξ_n is a random variable from a suitable distribution. It is observed here that, ξ_n is the factor multiplied to the n^{th} term to get the $(n+1)^{th}$ term. Then, the precentage change in X_n is given by,

$$\left(\frac{X_{n+1}}{X_n} - 1 \right) 100 = (\xi_n - 1)100 \tag{4.2}$$

Thus, the percentage change depends on the random variable ξ_n (responsible for the irregularity in the series) and on X_n. This feature of geometric random walk is utilised to generate the synthetic financial data.

4.2 Geometric random walk results

4.2.1 Closing value

First, we discuss the simulated closing value obtained from GRW. It has been observed for the empirical stock market data that, the distribution of the *log return* defined as $R_{\log} = \log \frac{X_{n+1}}{X_n}$, is approximately Gaussian [20]. In Figure 4.2, the log-returns from IBM stock prices and GRW is shown for comparison. For the time series from GRW, we assume the changes to be a random variable from Gaussian distribution. Using the definition of geometric random walk (Eq. 4.1) and by taking logarithm, we have

$$\ln X_{n+1} = \ln X_n + \ln \xi_n \Rightarrow \ln \frac{X_{n+1}}{X_n} = \ln \xi_n$$

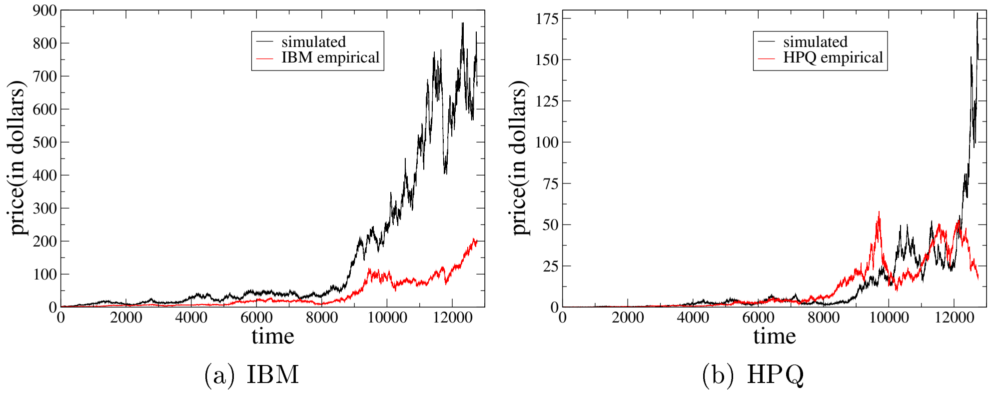
$$\Rightarrow X_{n+1} = X_n e^{\ln \xi_n} \tag{4.3}$$

Here, $\ln \xi_n$ is a normally distributed random variable. Time series is constructed by inserting arbitrary value to X_0 in Eq. 4.3 and generating normally distributed random numbers with the specified mean and standard deviation. We denote ξ to be a normally distributed random variable with zero mean and 1 as standard deviation. Now, we can generate normally distributed random variable of desired mean μ and standard deviation σ using,

$$\xi(\sigma, \mu) = \xi(1, 0)\sigma + \mu$$

Normally distributed random numbers are obtained numerically by using Box-Muller algorithm [27]. Inputs to be generated for the normally distributed random numbers for the synthetic stock time series, can be determined from the empirical data of the log return. In this section empirical data of *IBM* and *HPQ* are used for determining the mean and the standard deviation of the distribution of log returns. Mean and standard deviation for *IBM* are 0.00034 and 0.01617 respectively. Using these empirical parameters, time series of the closing value using geometric random walk is generated for *IBM* and *HPQ*. In figure 4.1 we compare the simulated and the empirical time series for both *IBM* and *HPQ*.

From the geometric random walk simulation, the distribution of log returns is computed and displayed in Fig. 4.2. It must be observed that the empirical log returns deviate from the assumed normality of log returns distribution.



(a) IBM (b) HPQ

Figure 4.1: Closing value of the empirical data plotted over the geometric random walk simulated curve

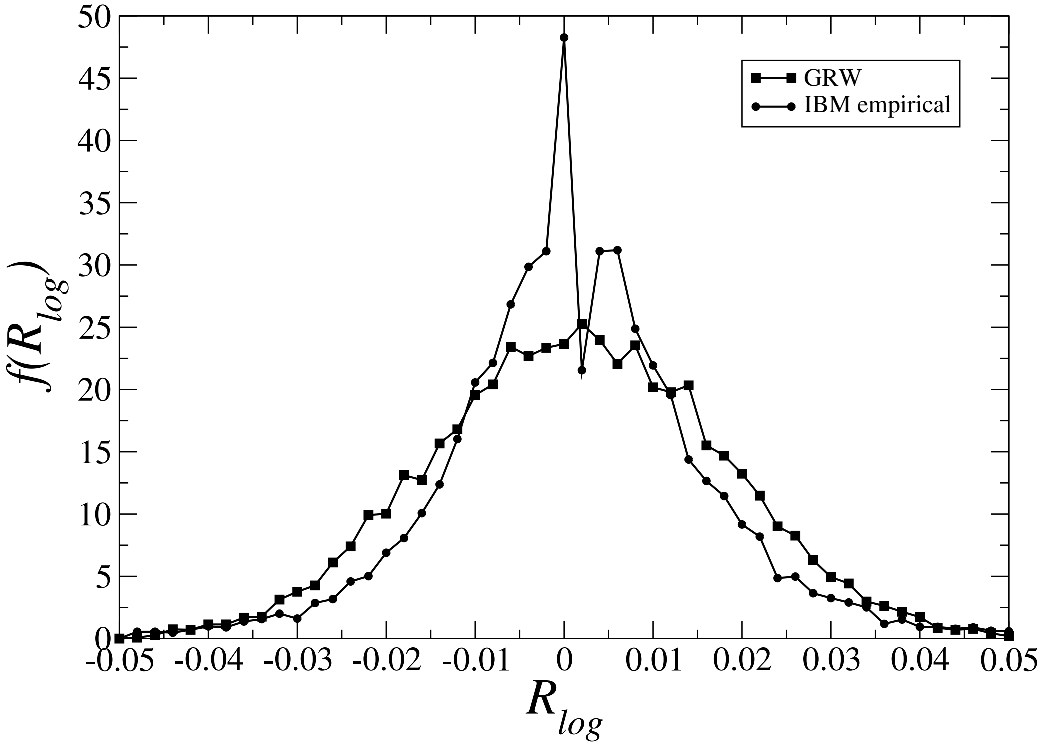


Figure 4.2: Distribution of the log return for *IBM* empirical stock data and geometric random walk.

4.2.2 Stylized facts

First, we look at the stylized facts for the time series obtained from geometric random walk. It is observed that the autocorrelation is absent in all the realisations of geometric random walk.

Distribution of the return values ($R = X_{n+1} - X_n$) of the time series generated by geometric random walk were calculated. On plotting the curve in the *log* scale gives a straight line which signifies the power-law behaviour of the return distribution. Figure 4.3 shows distribution of returns for the simulated time series of geometric random walk. In this, the mean and standard deviation of the empirical log return (R_{log}), needed for the geometric random walk were derived from the empirical data of *IBM* and *HPQ*. On linear regression of the data points, we have

$$f(R)^{grw}_{IBM} \propto R^{-1.82}$$

$$f(R)^{grw}_{HPQ} \propto R^{-1.61}$$

In addition, the autocorrelation of absolute return $|R|$ of the geometric random walk simulated time series indicates at least short term memory in the process.

4.2.3 Records

Total number of records present in a time series is one of the major drawback with the previous two models, namely random walk and biased random walk. Geometric random walk time series was generated and averaged over 500 realisations. Parameter inputs of the mean and standard deviation for the normally distributed random variable were derived from the log return of empirical data from IBM, HPQ, Apple, Exxon and GE. Table 4.1 shows the comparison between the number of records generated by GRW simulations and the empirical data. These numbers shows a clear improvement in the count of records over other models used such as RW and biased RW. This further increases the reliability in the statistics based on GRW model we are computing.

4.2.4 Record gaps distribution from GRW

Empirical data of individual stocks were used to determine the series of log return (R_{log}). Mean and standard deviation for these log return series were determined and used to generate synthetic time series using Eq. 4.3. The distribution of record gap is computed and after averaging over 500 realisations. Record gap distribution of the time series for various stocks are shown

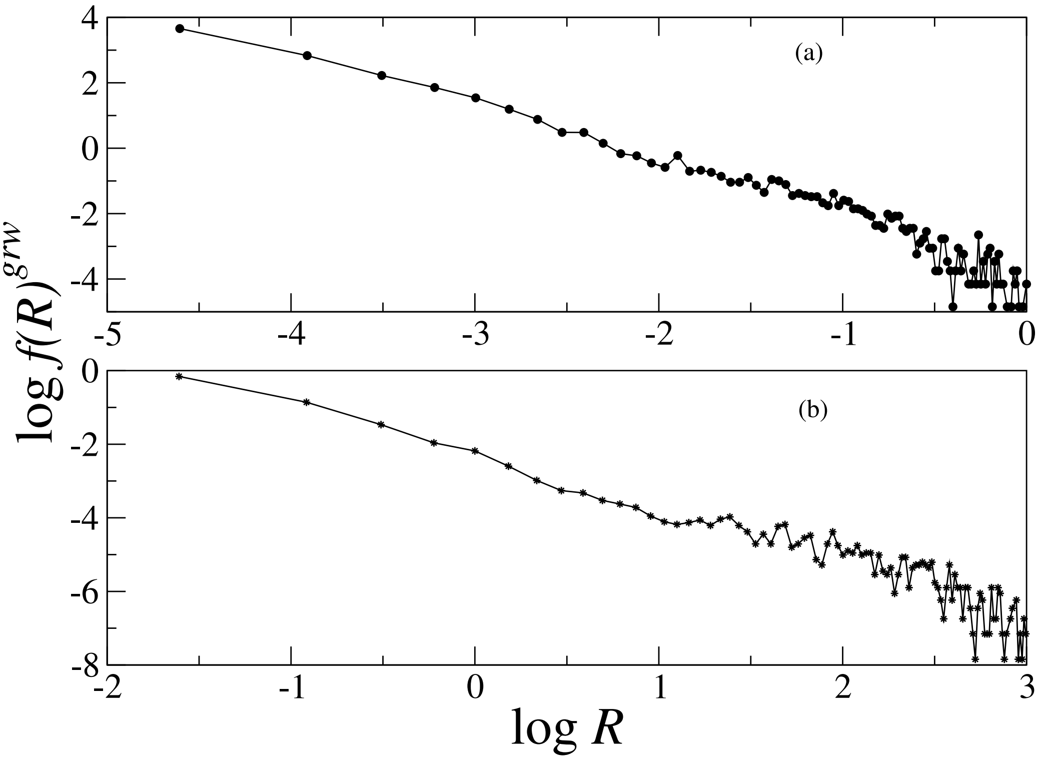


Figure 4.3: Return distribution for the simulated series of geometric random walk. Mean and standard deviation of random variable here are derived from (a) *IBM* and (b) *HPQ* empirical data.

Table 4.1: Number of records observed in the time series of various models and empirical data

Stocks/Models	Empirical	RW	Biased RW	GRW
IBM	433	92	92	461
Apple	336	67	65	260
Exxon	549	81	82	438
GE	414	87	96	215
HPQ	296	88	94	275

Table 4.2: Comparison of power-law exponent results for different *individual stocks* for the empirical and GRW data.

Stocks/Models	$\gamma_{empirical}$	γ_{grw}
IBM	1.64	1.56
Exxon	1.60	1.57
GE	1.63	1.55
HPQ	1.58	1.58

in the figure 4.4 as log-log plot. The approximate straight line in log-log plot indicates a power-law behaviour of the form $\phi(r_g)^{grw} \propto r_g^{-\gamma}$. In Table 4.2, the values of power-law exponent γ is listed along with their corresponding empirical result. The appearance of a power-law in the distribution of record gaps implies clustering of records for short time intervals, i.e, records tend to occur quickly in succession over time scales that are short in comparison with the observation time of the data. In all the data shown here, the observation time is of the order of few years and short time scales will correspond to few days. However, on longer time intervals of months to years, probability of occurrence of records is vanishingly small.

Results obtained above have some deviation from the empirical results. Distribution of the record gaps is analysed for the simulated data obtained from time series of 252000 data points, which is equivalent to 1000 years of data, and ensemble averaged over 500 realisations. Figure 4.5 is the log-log plot of the distribution of record gaps for the extended and averaged time series obtained by using the geometric random walk model. The distribution turns out to be a power-law and can be mathematically written as,

$$\phi(r_g)^{grw} \propto r_g^{-1.54} \tag{4.4}$$

Plot (figure 4.6) is shown to compare the power-law exponent of the distribution of the record gaps obtained from geometric random walk simulated data and the various empirical data of individual stocks.

Indices

In the analysis shown above, GRW was used as a model for individual stocks. In this section, we will treat GRW time-series as a stock market index and compare the model results with the empirical stock market indices. As discussed before, stock indices are indicators for a portfolio of stocks representative of the market. Record gap distribution for empirical index data is determined for several markets. Some of them are in the figure 4.7. These distributions also show power-law trends in record gap distribution.

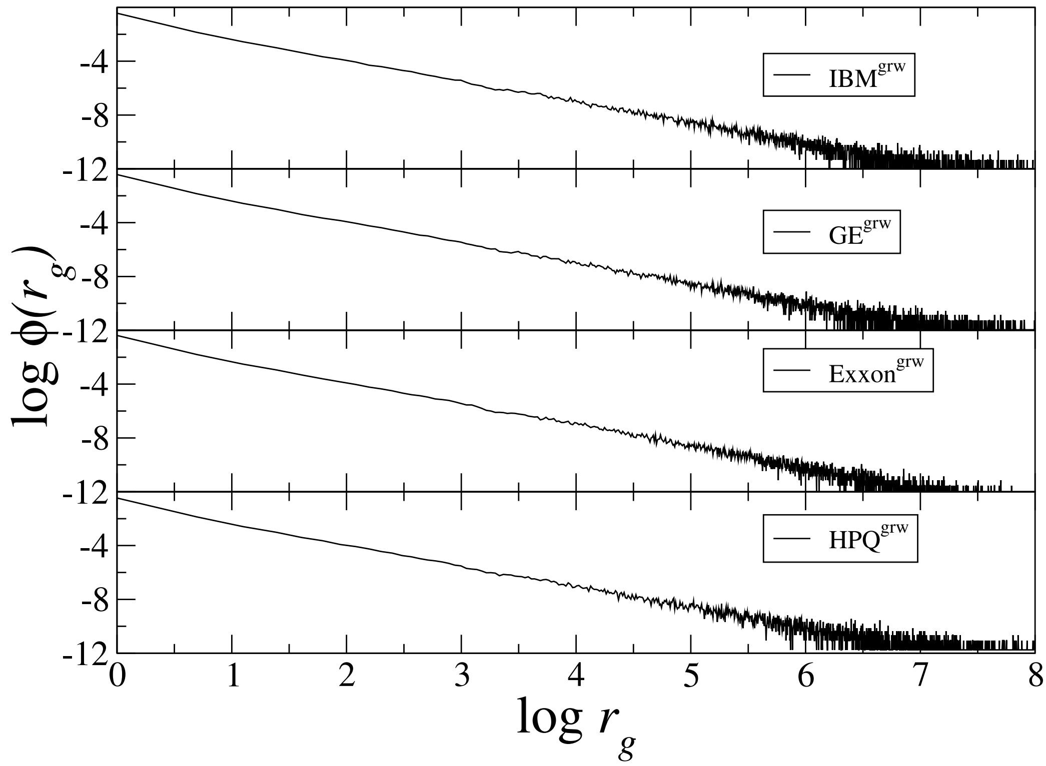


Figure 4.4: Record gap distribution for geometric random walk ensemble averaged time series plotted against record gaps on log-log scale using random variable features from the specified *individual stock* data.

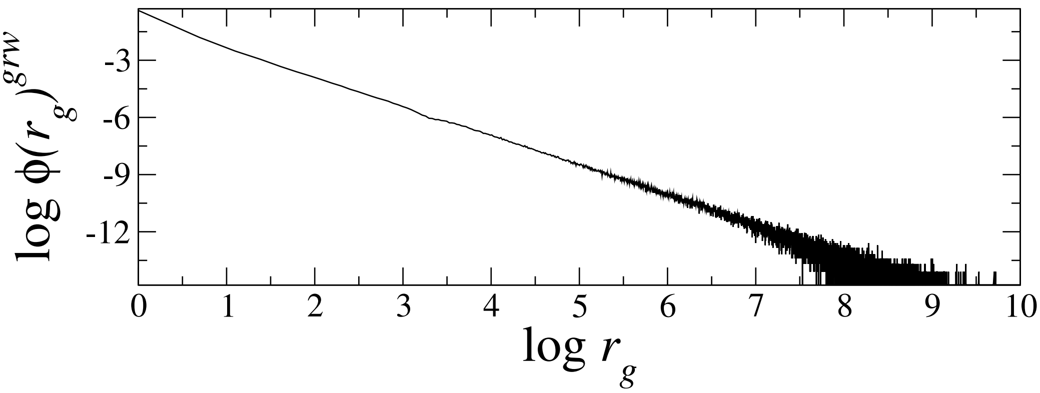


Figure 4.5: Record gap distribution for geometric random walk ensemble averaged time series, plotted against record gaps on log-log scale

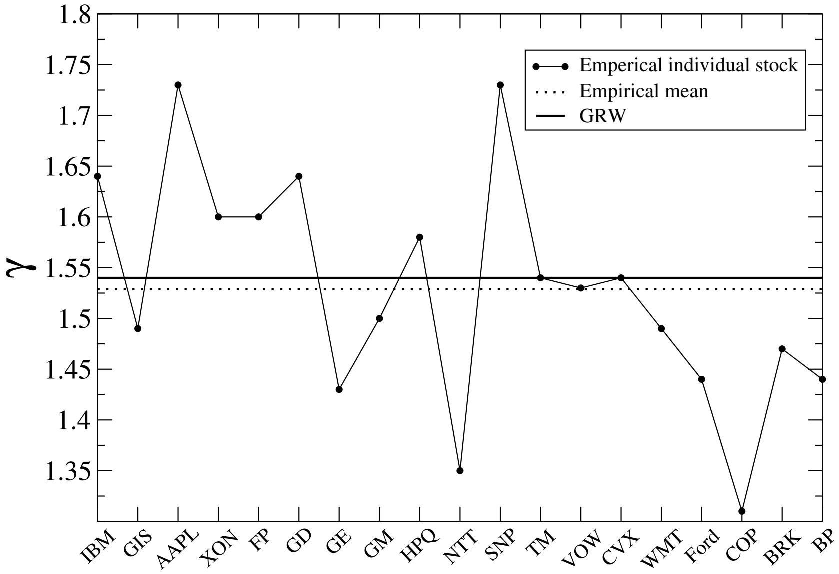
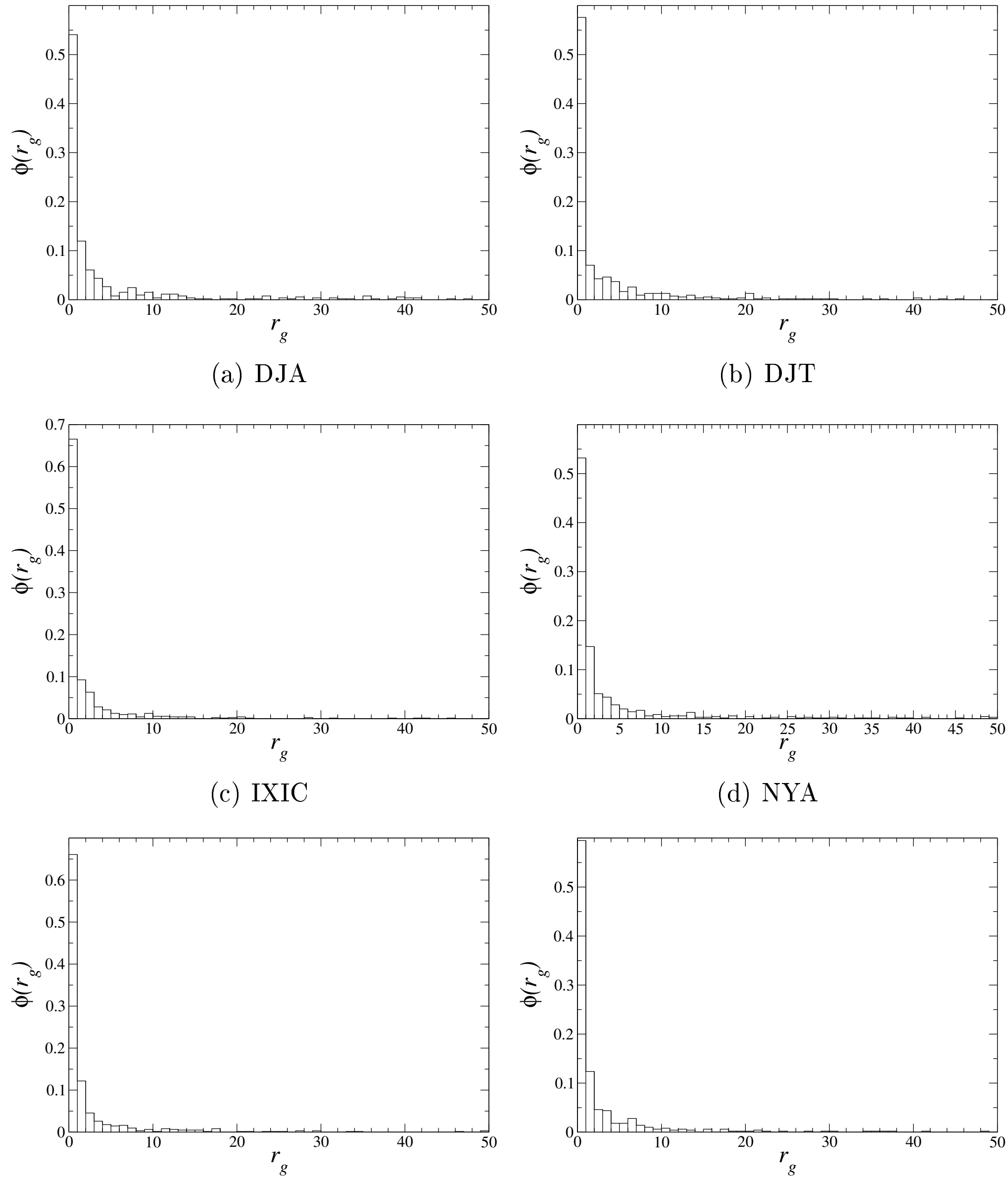


Figure 4.6: Power-law exponent for empirical data $log\,\phi(r_g)$ plotted against the corresponding listing names of the *individual stocks* and the arithmetic mean of the listed exponents. Same data is plotted for the GRW simulated result.

Table 4.3: Power-law exponents (γ) for distribution of record gap for empirical *stock market indices* and GRW generated time series.

Indices/Models	Empirical	GRW
NYA	1.70	1.55
DJA	1.67	1.55
DJT	1.60	1.58
IXBK	1.94	1.56
IXIC	1.50	1.57
IXIS	1.69	1.58



(a) DJA (b) DJT

(c) IXIC (d) NYA

(e) IXBK (f) IXIS

Figure 4.7: Record gap distribution for the empirical time series data of stock *market indices* plotted against record gaps.

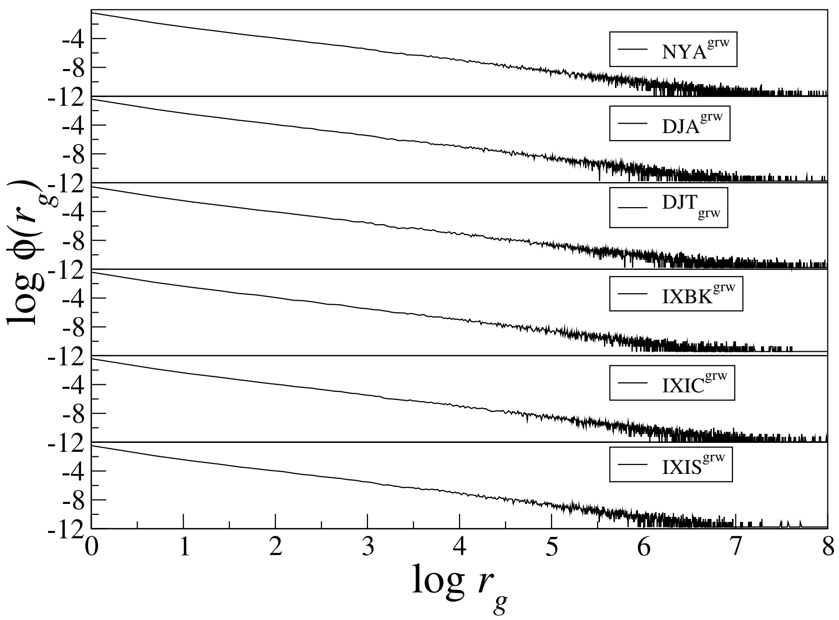


Figure 4.8: Record gap distribution for geometric random walk ensemble averaged time series are plotted against record gaps on log-log scale for *market indices.*

Further, for these stock market indices, simulated time series were generated assuming the log return distribution feature as normally distributed and adopting the mean and standard deviation from the empirical data. These time series were generated using the same length of data as that of the empirical data and were ensemble averaged over 500 realisations. Record gap distribution of the series were determined and is shown in log-log plot in the figure 4.8. The distributions indicate a power law similar to the case for individual stocks. The results can be represented as,

$$_{indices}\phi(r_g) \propto r_g^{-1.68} \tag{4.5}$$

$$_{indices}\phi(r_g)^{grw} \propto r_g^{-1.57} \tag{4.6}$$

Power-law exponents (γ) for the empirical stock indices are shown along with the synthetically generated record gap distributions using geometric random walk in the table 4.3. Figure 4.9 compares the power-law exponent of the distribution of the record gaps obtained from geometric random walk simulated data and the various empirical data of stock market indices.

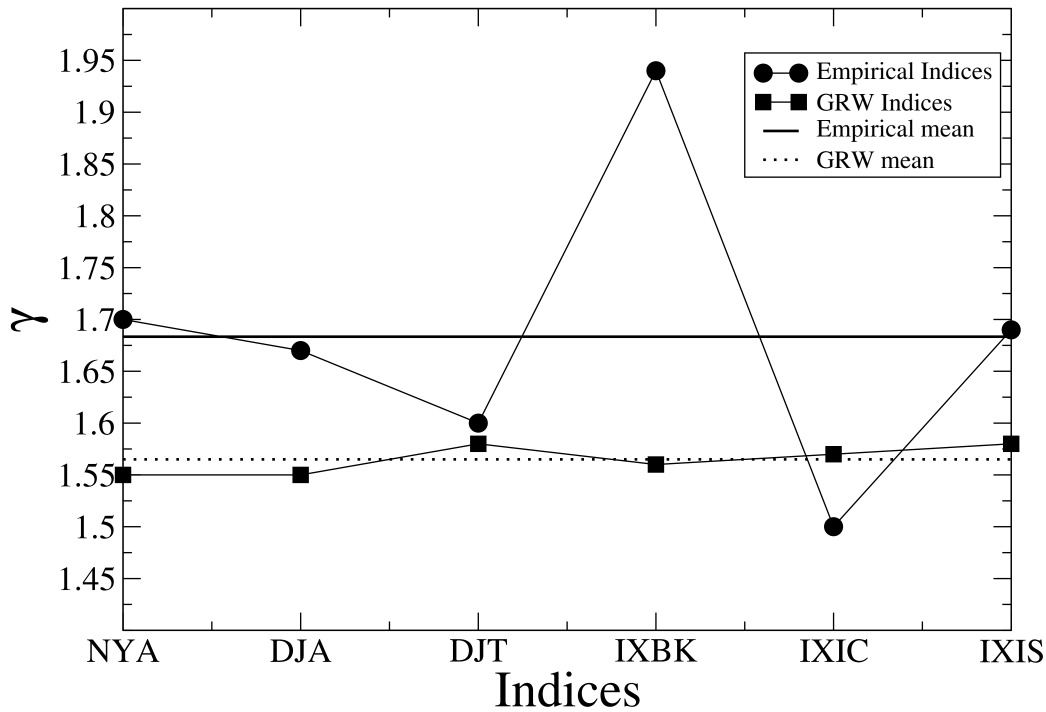


Figure 4.9: Power-law exponent for empirical data $log\,\phi(r_g)$ plotted against the corresponding listing names of the *market indices* and the arithmetic mean of the listed exponents. Data for the same is plotted for the GRW simulated result.

Chapter 5

Results and discussion

Probability distribution of the record gaps for empirical stock price movements were found to depict power-law with the exponent $1 < \gamma < 2$. This shows record gap clustering.

Probability distribution of the record gaps for stochastic processes like random walk, biased random walk and geometric random walk also depicts power-law.

- Random walk model:

  γ for probability distribution of record gap is:

$$\gamma_{uniform} \approx 1.63 \tag{5.1}$$

$$\gamma_{normal} \approx 1.56 \tag{5.2}$$

- Biased random walk model:

  γ for probability distribution of record gap is:

$$\gamma_{uniform} \approx 1.33 \tag{5.3}$$

$$\gamma_{normal} \approx 1.21 \tag{5.4}$$

Power-law for probability distribution for the record gaps was consistently observed in all the individual stock market price movements and was also observed in the stock market indices. Where the γ for the analysed individual stocks were ranging from minimum of 1.31 for stock prices of ConocoPhillips(NYSE listing name: COP) and maximum of 1.73 for stock prices of Sinopec Group(NYSE listing name: SNP). Arithmetic mean of the exponents for the analysed individual stock data is:

$$\gamma^{empirical} \approx 1.53 \tag{5.5}$$

Standard deviation of the exponent of the observed lot is:

$$\sigma_{\gamma}^{empirical} \approx 0.11 \tag{5.6}$$

Arithmetic mean for the analysed stock market indices is:

$$_{indices}\gamma^{empirical} \approx 1.68 \tag{5.7}$$

Standard deviation of the exponent of the observed lot is:

$$_{indices}\sigma_{\gamma}^{empirical} \approx 0.15 \tag{5.8}$$

Probability distribution of the record gaps observed for time series of geometric random walk modelled for individual stocks. Ensemble average of probability distribution for this shows power-law where exponent is given as:

$$\gamma^{grw} \approx 1.54 \tag{5.9}$$

GRW when modelled for stock market indices shows:

$$_{indices}\gamma^{grw} \approx 1.57 \tag{5.10}$$

Scope for future research

To generate the geometric random walk using the distribution of log return of the empirical data, better assumption than taking a normally distributed random variables is demanded. I would like to carry this work further and make a firm analytical model of probability distribution of record gap to determine the power-law exponent for the stochastic processes such as geometric random walk. This work can be applied to stock price movements with more expectation and can further be expanded for other commodity price movements.

References

[1] Q. Zhang, C. Zhu, and J. Cheng, "Preliminary study on the flooding and drought calamity during past 1500 years in the hai'an region, jiangsu province," *Chinese Geographical Science*, vol. 12, no. 2, pp. 146–151, 2002.

[2] G. Wergen and J. Krug, "Record-breaking temperatures reveal a warming climate," *EPL (Europhysics Letters)*, vol. 92, no. 3, p. 30008, 2010.

[3] S. Rahmstorf and D. Coumou, "Increase of extreme events in a warming world," *Proceedings of the National Academy of Sciences*, vol. 108, no. 44, pp. 17905–17909, 2011.

[4] G. Wergen, A. Hense, and J. Krug, "Record occurrence and record values in daily and monthly temperatures," *Climate Dynamics*, pp. 1–15, 2013.

[5] L. Bachelier, *Théorie de la spéculation.* Gauthier-Villars, 1900.

[6] G. Wergen, M. Bogner, and J. Krug, "Record statistics for biased random walks, with an application to financial data," *Physical Review E*, vol. 83, no. 5, p. 051109, 2011.

[7] H. Nagaraja and G. Barlevy, "Characterizations using record moments in a random record model and applications," *Journal of Applied Probability*, vol. 40, no. 3, pp. 826–833, 2003.

[8] http://in.finance.yahoo.com. Website.

[9] https://www.google.com/finance. Website.

[10] B. G. Malkiel, *A random walk down Wall Street: Including a life-cycle guide to personal investing.* WW Norton & Company, 1999.

[11] E. J. Chang, E. J. A. Lima, and B. M. Tabak, "Testing for predictability in emerging equity markets," *Emerging Markets Review*, vol. 5, no. 3, pp. 295–316, 2004.

[12] J. Belaire-Franch and K. K. Opong, "Some evidence of random walk behavior of euro exchange rates using ranks and signs," *Journal of banking & finance*, vol. 29, no. 7, pp. 1631–1643, 2005.

[13] T. Nakamura and M. Small, "Tests of the random walk hypothesis for financial data," *Physica A: Statistical Mechanics and its Applications*, vol. 377, no. 2, pp. 599–615, 2007.

[14] A. W. Lo and A. C. MacKinlay, *A non-random walk down Wall Street*. Princeton University Press, 2011.

[15] S. N. Majumdar and R. M. Ziff, "Universal record statistics of random walks and lévy flights," *Physical review letters*, vol. 101, no. 5, p. 050601, 2008.

[16] R. Cont, "Empirical properties of asset returns: stylized facts and statistical issues," 2001.

[17] G. Wergen, M. Bogner, and J. Krug, "Record statistics for biased random walks, with an application to financial data," *Physical Review E*, vol. 83, no. 5, p. 051109, 2011.

[18] S. F. LeRoy and W. R. Parke, "Stock price volatility: Tests based on the geometric random walk," *The American Economic Review*, vol. 82, no. 4, pp. 981–992, 1992.

[19] P. Krapivsky and S. Redner, "Random walk with shrinking steps," *American Journal of Physics*, vol. 72, p. 591, 2004.

[20] P. K. Clark, "A subordinated stochastic process model with finite variance for speculative prices," *Econometrica: Journal of the Econometric Society*, pp. 135–155, 1973.

[21] S. R. Hurst and E. Platen, "The marginal distributions of returns and volatility," *Lecture Notes-Monograph Series*, pp. 301–314, 1997.

[22] K. Fergusson and E. Platen, "On the distributional characterization of daily log-returns of a world stock index," *Applied Mathematical Finance*, vol. 13, no. 01, pp. 19–38, 2006.

[23] G. S. Satya N. Majumdar and G. Wergen, "Record statistics and persistence for a random walk with a drift," *Journal of Physics A: Mathematical and Theoretical*, vol. 45, no. 35, p. 355002, 2012.

[24] B. G. Malkiel and E. F. Fama, “Efficient capital markets: A review of theory and empirical work*,” *The journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.

[25] T. Hellström, “A random walk through the stock market,” *Licentiate Thesis, Department of Computing Science, Umeå University, Sweden*, 1998.

[26] R. N. Mantegna and H. E. Stanley, *Introduction to econophysics: correlations and complexity in finance.* Cambridge University Press, 1999.

[27] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in FORTRAN 77: Volume 1, Volume 1 of Fortran Numerical Recipes: The Art of Scientific Computing*, vol. 1. Cambridge university press, 1992.

Appendix A

Listing names

<u>Individual Stocks</u>

IBM-International Business Machines
GIS-General Mills, Inc.
APPL-Apple
XON-Exxon Mobil
FP-Total SA
GD-General Dynamics Corporation
GE-General Electric
GM-General Motors Company
HPQ-Hewlett-Packard Company
NTT-Nippon Telegraph and Telephone
SNP-China Petroleum & Chemical Corporation
TM-Toyota Motor Corporation Common
VOW-Volkswagen AG
CVX-Chevron Corporation
WMT-Wal-Mart Stores Inc.
FORD-Ford MOtors
COP-ConocoPhillips
BRK-Berkshire Hathaway Inc.
BP-BP p.l.c.

<u>Market Indices</u>

NYA-NYSE Composite Indes Percent OP
DJA-Dow Jones Composite Average
DJT-Dow Jones Transportation Average
DJU-Dow Jones Utility Average
IXBK-NASDAQ Bank
IXIC-NASDAQ Composite
IXIS-NASDAQ Insurance