An Empirical Investigation of the Maslov Limit Order Market Model

Rasika M. Withanawasam
Peter A. Whigham
Timothy Crack
I. M. Premachandra

The Information Science Discussion Paper Series

Number 2010/04
March 2010
ISSN 1177-455X

University of Otago

Department of Information Science

The Department of Information Science is one of seven departments that make up the School of Business at the University of Otago. The department offers courses of study leading to a major in Information Science within the BCom, BA and BSc degrees. In addition to undergraduate teaching, the department is also strongly involved in postgraduate research programmes leading to MCom, MA, MSc and PhD degrees. Research projects in spatial information processing, connectionist-based information systems, software engineering and software development, information engineering and database, software metrics, distributed information systems, multimedia information systems and information systems security are particularly well supported.

The views expressed in this paper are not necessarily those of the department as a whole. The accuracy of the information presented in this paper is the sole responsibility of the authors.

Copyright

Correspondence

This paper represents work to date and may not necessarily form the basis for the authors' final conclusions relating to this topic. It is likely, however, that the paper will appear in some form in a journal or in conference proceedings in the near future. The authors would be pleased to receive correspondence in connection with any of the issues raised in this paper, or for subsequent publication details. Please write directly to the authors at the address provided below. (Details of final journal/conference publication venues for these papers are also provided on the Department's publications web pages: http://www.otago.ac.nz/informationscience/pubs/). Any other correspondence concerning the Series should be sent to the DPS Coordinator.

Department of Information Science
University of Otago
P O Box 56
Dunedin
NEW ZEALAND

Fax: +64 3 479 8311
email: dps@infoscience.otago.ac.nz
www: http://www.otago.ac.nz/informationscience/

Contents

List of Figures

List of Tables

Abstract

Modeling of financial market data for detecting important market characteristics as well as their abnormalities plays a key role in identifying their behavior. Researchers have proposed different types of techniques to model market data [1]. One such model proposed by Sergie Maslov, models the behavior of a limit order book. Being a very simple and interesting model, it has several drawbacks and limitations.

This paper analyses the behavior of the Maslov model and proposes several variants of it to make the original Maslov model more realistic. The price signals generated from these models are analyzed by comparing with real life stock data and it was shown that the proposed variants of the Maslov model are more realistic than the original Maslov model.

Chapter 1

Introduction

This paper is based on the research done by Sergie Maslov [2]. We propose two modifications (M_1,M_2) to the original Maslov model to make the model more realistic and their behavior in relation to the original Maslov model and range of real financial data is analyzed.

In Chapter 2, we discuss the Maslov original model along with the proposed modifications. Chapter 3 discusses our contribution in this research and Chapter 4 deals with the results we have obtained. The conclusion chapter summarizes the research findings and highlights the implication of the work and future extensions. In the Appendix, we discuss some important time series and financial data comparison techniques and their relevance to our study.

The following data sets are used in the analysis

Original Maslov simulation data, and modified Maslov data: A comprehensive description of these data is included in Chapter 2 along with the relevant information regarding the generation of these datasets.

DowJons index data (1998-2009): This is the second oldest market index in the US. This index gives an indication of how the prices of 30 large publicly owned companies traded in United States behave.

S&P 500 Index data (1998-2009): This is a price index which consists of 500 large-cap common stocks actively traded in the United States.

General Electric (1962-2009): General Electric Company is a multinational American technology and services company. In 2009 it was named as the world's largest company. We used daily, weekly and monthly returns of GE stocks traded in 2009 for our analysis.

MRO (2009): Marathon Oil is a leading integrated energy company with exploration and production activities based on countries including United States, Angola, Indonesia, and Norway.

DELL - Nasdaq trades in February 2007: Dell Inc is a multinational technology corporation residing in the US. This company designs, develops, manufactures, sells, and supports personal computers and other computer-related products. We used price returns of DELL stocks traded in the month of February 2007 for our analysis.

Chapter 2

Maslov Model and its Variants

2.1 Maslov Limit Order Market Model(M_0)

Maslov model (denoted by M_0) [2] simulates the behavior of a limit order book using a single stock with one trader submitting limit and market orders based on random logic. This model can be considered as a very basic layer of a complex financial market. As explained below, the author has used a very simple approach to model the market behavior of a limit order book. We define the behavior of Maslov's model using the following general notation:

P_0 **:** Starting Last Traded Price

LTP : Last Traded Price

MO : Market Order

LO : Limit Order

Δ **:** Price Off-set

BBP : Best Bid Price

BAP : Best Ask Price

θ **:** Order Size

p_0 **:** Probability of a buy order, with (1 - p_0) the probability of a sell order.

p_1 **:** Probability of a limit order given that it is a buy order and (1 - p_1) is the probability of a market order given that it is a buy order.

p_2 **:** Probability of a limit order given that it is a sell order and (1 - p_2) is the probability of a market order given that it is a sell order.

The Maslov model simulates buy orders with probability p_0 (p_0 = 0.5), and sell orders with probability 1 - p_0. Depending on whether the order type is buy or sell, a limit order takes place with probabilities p_1 and p_2 respectively (where p_1 = p_2 = 0.5). The price of a limit buy order is assumed to be LTP - Δ, where the price off-set (Δ) is assumed to be a uniformly distributed discrete random variable in the range $1 \leq \Delta \leq 4$. The price of a sell limit order is assumed to be LTP + Δ i.e., the Maslov model assumes that limit order buyers determine the buying price at a price slightly (Δ) lower than the last traded price & limit order sellers at a price slightly (Δ) higher than the last traded price.

The Maslov model basically considers three transactions namely buy orders, sell orders, and matching of a buy and a sell order.

An execution of a buy/sell matching transaction takes place in the following two cases.

- When a sell market order (MO) is submitted and is matched with the best price (highest) of the buy side of the book(i.e., BBP).
- When a buy market order (MO) is submitted and is matched with the best price (lowest) of the sell side of the book(i.e., BAP).

The order book of the Maslov model is maintained according to the following rules.

- Once a buy/sell matching transaction is taken place, the two orders are removed from the order book immediately.
- Traders can trade a fixed number of shares at a time i.e., θ is constant.
- If a buy market order is submitted when there are no sell orders in the order book, the buy market order is converted to a buy limit order & vice-versa.
- All the limit orders are assumed to be "good till canceled".

See Figure 2.1 and Algorithm 1 for additional details.

LTP generated from the Maslov model is recorded with time and it's behavior is analyzed. The Maslov model enforces that $BBP \leq LTP \leq BAP$ which which is not always observed in the behavior of a real stock market. Moreover, both limit buy and sell orders narrow the spread (the difference between the best ask price and best bid price) while market orders widen it. This results in a repetition of "cone shapes" in the price signal(See the Figure 4.2(a) on page 4.2(a)). When the spread is high, some sudden large drops and jumps in the price signal can be seen.

The original Maslov paper [2] analyzes the behavior of a stock market limit order book by relating its behavior to some real financial market movements. Some of the features the author analyzes are fat tail property of the probability distribution of price fluctuations, crossovers between two power law regions in the same distribution, long range correlations of the volatility, and the Hurst exponent of the price.

Maslov model addresses the following empirical theories and produces evidence to prove them using the numerical results produced by the model.

- The histogram of the short time lag increments of market price has a different Gaussian shape with sharp maximum and broad wings.

  So, according to the current consensus on the shape of the distribution, it shows the characteristics of a Pareto-Levy distribution [3, 4] up to a certain value, with a power low exponent of $1 + \alpha \sim 2.4 - 2.7$, and then it crosses over either to a steeper power law with an exponent of $1 + \alpha \sim$ 3.7 - 4.3 or to an exponential decay. In both these cases, this crossover ensures a finite variance (second moment) of the distribution.

- When computed with time scales less than several trading days, the graph of price vs. time have a Hurst exponent ([2, 5]) $H \sim 0.6 - 0.8$, which is different to the value corresponding to the ordinary random walk which is $H \sim 0.5$.

- The volatility of price(second moment of price fluctuations) should exhibit a correlated behavior. It should show some clustering of volatility i.e., having regions of high amplitude data separated by relatively low amplitude regions visible in time vs. price increments plot. Volatility clustering affects the shapes of the autocorrelation function of the volatility (price increments) as a function of time. The autocorrelation function of price increments should decay according to the power law with a very small exponent in the range 0.3 - 0.4 and with no apparent cut-off.

The Maslov paper [2] presents the following empirical evidence to support the above empirical theories.

- It compares the price vs. time graph and price increments vs. time graph with ordinary random walk graphs with similar attributes and shows that both graphs are drastically different from a random walk. Unlike a random walk, the author has observed some price increment clustering where the regions with high volatility that are separated by some quiet regions. He computes the Hurst exponent of the price graph using the Fourier transform of the price signal by taking the average over many runs of the model. He also claims that the relationship of the Fourier transform of the autocorrelation function of the price signal and the value of the Hurst exponent is of the form $S(f) \sim f^{-(1+2H)}$. Maslov results show that the log-log plot of S(f) of a price signal of length 2^{18}, averaged over multiple realizations resulted in a value of the Hurst estimate approximately equal to 0.25. This corresponds to the decay of $S(f) \sim f^{-3/2}$. But the value he obtained differs from the short term Hurst exponent $H \sim 0.6 - 0.7$, corresponding to real stock prices.

- Maslov argues that the amplitude of price fluctuations generated from his model has long range correlations while signs of price fluctuations having short range correlations. He has used the autocorrelation function of the absolute values of price increments $S(t)_{abs}$ to show this correlation behavior. According to him, the autocorrelation function of the absolute values of price increments behaves according to the power law tail with an exponent of $S(t)_{abs} \sim t^{-1/2}$. He also derives the Fourier transform of $S(t)_{abs}$ which has a clear form of $f^{-1/2}$. The exponent he has got for his simulation was not very different from 0.3 which is the corresponding value for real data such as S&P 500 stock index. Then he analyses correlations of signs of price fluctuations (changes) using Fourier transform of auto correlation function and the results shows that the behavior is much closer to frequency independent forms such as white noise characteristics. He compares this with real stock prices to show that real data also has similar long range (lag is less than 30 minutes) correlations of signs of price increments.

- The histogram of price increments measured over time lags 1, 10, and 100 provides strong support for non-Gaussian distribution which is very close to the shape of real stock prices. As the lag increases the peak of the histogram gradually softens (gets closer to Gaussian), while the wings remain strongly non-Gaussian. Also his analysis on the log-log plot of the histogram of lag 1 for data collected during 3.5×10^7 time steps shows that log-log plot has two distinguishable power low regions separated by a large crossover approximately around 1. According to Maslov the reason behind this is unknown. Exponents of these two regions are measured to be $1 + \alpha = 0.6 \pm 0.1$ and 3 ± 0.2. A similar crossover of two power low regions was reported in real stock price fluctuations in NYSE with the exponents ranging between $1.4 - 1.7$ and $4 - 4.5$. Power low exponent of the far tail, $1 + \alpha = 3$, stays right at the borderline, separating the Pareto-Levy region with power low exponent $1 + \alpha < 3$, where the distribution has an infinite second moment(i.e., variance). According to the author, although his model shows very long range correlations in price fluctuations, one should not expect convergence of a price fluctuation distribution to a Pareto-Levy or Gaussian as lag is increased.

We propose the following modifications to the original Maslov model(M_0) in order to make the original Maslov model more realistic.

2.1.1 Modified Maslov Model 1(M_1)

In M_1, the probabilities (p_0,p_1,p_2) and the price determination logic remain as per Maslov original model(M_0) but Δ is assumed to be a uniformly distributed discrete random variable in the range $-1 \leq \Delta \leq 2$ (Refer the Algorithm 2 and Table 2.1) where as M_0 assumes Δ to be in the range $1 \leq \Delta \leq 4$.

Maslov model does not allow the order price to overlap with the contra side of the order book; which prevents probable trades of aggressive limit orders. In real markets when a trader wants to buy at the market price, he/she simply submits a market order with the awareness of the top of the book price, but before his order arrives to the market, there can be some other orders that hit the book which would remove several top most price points. As a result, our trader might get an unexpected price. Hence market orders involve some risk. Therefore some traders submit LOs with the price of top of the book in contra side instead of submitting a market order in order to minimize the risk. This ensures him that he would not receive the worst price even though he would miss the best price that he wanted to get. In order to incoperate this behavior we select the values for Δ in the range $-1 \leq \Delta \leq 2$, which would allow the model to go into either side of the LTP when determining the next LOP. This strategy allows the model to have aggressive limit order matchings. This type of trading cannot takes place in the original Maslov model M_0.

2.1.2 Modified Maslov Model 2(M_2)

In M_2, Δ is assumed to be a uniformly distributed discrete random variable in the range $1 \leq \Delta \leq 4$ and the probabilities are assumed to be p_0,p_1,p_2 = 0.5 as per the original Maslov model M_0. The difference in M_2 is that for buy orders it is assumed that LOP = BAP $-\Delta$ and for sell orders LOP = BBP + Δ. When the contra side is empty, limit buy and sell order prices are defined to be LTP $-\Delta$ and LTP $+\Delta$ respectively (See the Algorithm 3 and Table 2.1).

This modification is introduced taking in to account of the behavior of a rational market and it's traders. This is also based on observed patterns of traders who are trading in a market. The logic behind this modification is that the top of the book price of the contra side being used to determine the limit order price of a side.

```
Input  : Δ = 1, 2, 3, 4
Output: LOP
/* Buy ?                                   */
if (Buy) then
    LOP = LTP - Δ;
end
/* Sell                                    */
else
    LOP = LTP + Δ;
end
```

Algorithm 1: Algorithm for determining Limit Order Prices of M_0.

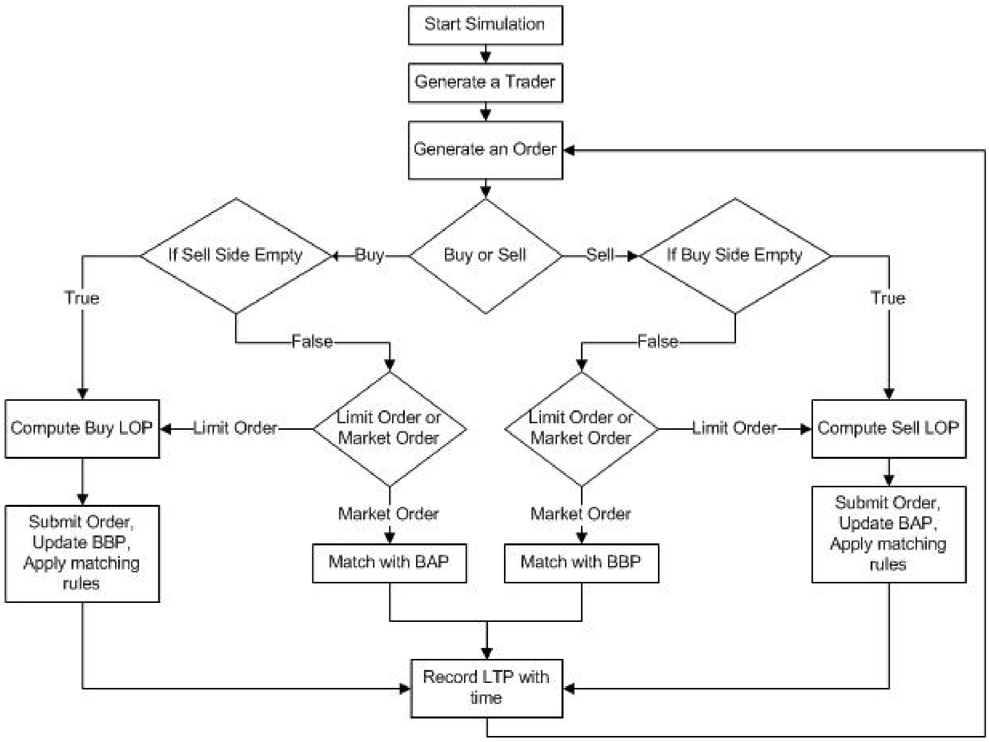


Figure 2.1: Flow Diagram of M_0.

Input : $\Delta = -1, 0, 1, 1$
Output: LOP

```
/* Buy ?                                                    */
```
if *(Buy)* **then**
    LOP = LTP - Δ;
end
```
/* Sell                                                     */
```
else
    LOP = LTP + Δ;
end

Algorithm 2: Algorithm for determining Limit Order Prices of M_1

Input : $\Delta = 1, 2, 3, 4$
Output: LOP

```
/* Buy ?                                              */
if (Buy) then
   /* Sell Side Empty ?                               */
   if (Sell Side Empty) then
      LOP = LTP - Δ;
   end
   else
      LOP = BAP - Δ;
   end
end
/* Sell                                               */
else
   /* Buy Side Empty ?                                */
   if (Buy Side Empty) then
      LOP = LTP + Δ;
   end
   else
      LOP = BBP + Δ;
   end
end
```

Algorithm 3: Algorithm for determining Limit Order Prices of M_2

Model	Lmit Buy Price				Limit Sell Price			
	No Bids Asks	No Asks Bids	No Bids No Asks	Bids Asks	No Bids Asks	No Asks Bids	No Bids No Asks	Bids Asks
M_0 Δ=1,2,3,4	(LTP-Δ)	(LTP-Δ)	(LTP-Δ)	(LTP-Δ)	(LTP+Δ)	(LTP+Δ)	(LTP+Δ)	(LTP+Δ)
M_1 Δ=-1,0,1,2	(LTP-Δ)	(LTP-Δ)	(LTP-Δ)	(LTP-Δ)	(LTP+Δ)	(LTP+Δ)	(LTP+Δ)	(LTP+Δ)
M_2 Δ=1,2,3,4	(BAP-Δ)	(LTP-Δ)	(LTP-Δ)	(BAP-Δ)	(LTP+Δ)	(BBP+Δ)	(LTP+Δ)	(BBP+Δ)

Table 2.1: Price Computation Logics of M_0, M_1, and M_2

Chapter 3

Methodology

The contribution of our research is twofold.

- Analyze the behavior of time series and financial measures for the M_0 and its variants M_1 and M_2.
- Compare these behaviors with real data.

In order to add some realness to the original Maslov model M_0, we have been experimenting with various forms of simple variations. These experiments were based on the price determination logic of the limit orders. In order to make the models simple, we refrained from introducing any complex patterns or special behaviors. Here the starting price (P_0) is selected as 1000. Price and time data were recorded over 10000 time steps after initial time steps of 1000 and all the results were computed and averaged over five hundred simulations.

We used two measures to analyze the behavior of the datasets that were used in our research: the Hurst exponent and fitting a probability distribution to logarithmic price returns.

The following sections (3.1 and 3.2)) describe the two main measures used to explain the behavior of the selected datasets.

3.1 Hurst Exponent

Hurst exponent (normally denoted by H) is used in areas such as applied mathematics, fractals and chaos theory, long memory processes, and spectral analysis. It has different but related meanings in different contexts.

Hurst exponent is a measure of whether the data is a pure random walk or has underlying trends and hence, it is considered as a measure of predictability of a series. Random Gaussian process with an underlying trend should have some degree of autocorrelation. If this autocorrelation has a very long (infinite) decay or long range correlations, it is referred to as a long memory process with a Hurst exponent value $0.5 < H < 1.0$.

This long memory behavior could be due to a sudden impact that affects a process. In such process, although the impact is sudden, the underlying process takes some time to come back to its normal behavior. This is due to the memory which is carried through with the process itself. For example, although a large buy or sell order can cause a sudden change in stock price, stock price behavior takes some time to come back to its normal operation. Hurst estimate can be used as an indication of this type of behavior of processes. It can always be used to compare behaviors of memory-less processes like random walks [6].

It is a measure of “dependence index” in fractal geometry. It is a measure of relative tendency of how close it is to the mean or cluster in a direction. So it is a measure of persistence (i.e., the characteristic or tendency of underlying series to continue in its current direction). If the Hurst exponent value is between 0.5 and 1, the process can be considered as a persistence series (which has positive autocorrelation) meaning that if the process has an increment between times t-1 and t, then there is a high possibility of having an increment between times t and t+1 as well. If H is between 0 and 0.5, it is an anti-persistence series (which has negative auto-correlation). In other words, if the process shows an increase between times t-1 and t, there is a high possibility of having a decrease in between t and t+1. If it is equal or closer to 0.5, this implies that it is a random and unpredictable series. This behavior is called "mean reversion" [6, 7].

Hurst exponent is related to fractal dimension as well. Fractal shapes can be identified as shapes which can be made with a large number of similar basic shapes. Some examples of this type of fractal shapes are fern leaf and Sierpinski pyramid. Fractal dimension D is a statistical quantity which gives an indication on how completely a fractal shape appears to fill the space. This fractal dimension is used to measure the roughness of the coast line. Hurst estimation directly relates to the fractal dimension such that D = 2 - H and lies in between 0 and 1 with higher values indicating a smoother trend, less volatility, and less roughness. In fact, the Hurst exponent was developed in the field of hydrology as a result of an attempt to obtain the optimum dam size for the Nile River by analyzing changing rain and drought conditions over a long period of time. It has been showed that the height of the Nile River measured annually over many years gave a value of H = 0.77 [6].

For brown noise (Sometimes referred to as random walk or Gaussian noise), the estimated Hurst Exponent is around 0.5. For white noise, the value of the Hurst Exponent is around 0 and for the popular Levy stable processes and truncated Levy processes with parameter α and γ Hurst Exponent = γ/α for $\gamma < \alpha$ and 1 for $\gamma = \alpha$ (See Appendix A.2 for Stable Pareto Distributions).

In financial world, many economists and statisticians have been trying to model the stock market behavior using various models. One basic model used for this is the random walk model. In order to model stock market behavior using this model, they have assumed that the distribution of stock returns follow a normal distribution. Under that assumption, methods like Value at Risk (risk of loss measured on a specific portfolio of financial asset) have been developed. But later some argued that the distribution of price returns does not follow a normal distribution and because of this, the rescaled range analysis or Hurst Exponent analysis was introduced. For example, daily return on stocks behave according to the Gaussian distribution. In terms of correlation, return of yesterday may not have any relationship with the return today. However, when the return period increases, the distribution gets closer to a log-normal distribution. Here the extreme values or tails of the distribution follow a power law. These longer return time series shows some amount of autocorrelation and a non-random Hurst exponent. It is observed that the return period increases as the value of the Hurst exponent increases and gets closer to 1 (correlation increases and shows long memory behavior). Many researchers (mainly Peters, 1991) [8] have proved that stock returns have characteristic $H > 0.5$, so that their behavior is distinct from random walk and is not generated by a stochastic process generating non-correlated values. This was referred to as long term memory behavior in stock returns [5, 7].

There are various methods in practice to estimate the Hurst exponent value. Widely used methods are rescaled-ranged computation, wavelet based method, and graphical methods.

The rescaled range is a statistical measure of the variability of a time series. In other words, it is a statistical technique used to detect the presence or absence of trend in time series by finding the Hurst exponent. It is computed by dividing the range of the values by the standard deviation over the same portion of time series. If the maximum, minimum, and standard deviation values of time series of size n are x, y, and S respectively, which have a range, R = x - y, the rescaled range of the series is defined as R/S. When we increase the number of observations n, the rescaled range value also increases. The slope of the doubly logarithmic plot of rescaled range (R/S) vs. sample size (n) gives the Hurst exponent H [6].

Wavelet based method is also used to estimate the Hurst exponent in many cases. In graphical method, price auto-correlation function and Fourier transform of price auto-correlation function is used. Fourier transform of price auto-correlation function of the price signal S has a relationship with the Hurst exponent H of the form $S \sim f^{(1+2H)}$. The slope of the doubly logarithmic plot of this is related to the Hurst exponent as $2H + 1$[2, 5]. In a long memory process, the decay of the autocorrelation function follows a power law. The Hurst exponent relates to the power exponent as 1 - $\alpha/2$ [2].

The Hurst exponent is non deterministic so it is only an estimation based on observed data. The only way to test the estimated value is by comparing it with a dataset with known Hurst exponent.

In our research, the Hurst exponent values are estimated using the re-scaled range computation method. Re-scaled range method was chosen after experimenting for consistency and accuracy with the other two main methods. Here, the logarithmic price returns were used in order to reduce the overall market movement and all the aforementioned measures were computed for all the selected datasets and compared the values with each type of data available. All the results were computed and averaged over five hundred simulations.

3.2 Fitting Pareto Levy Distribution to Logarithmic Price Returns

Researchers have shown that extreme high values of stock return values follow a simple power law distribution (Pareto-Levy distribution) and they have proposed various ways to estimate the parameters of this distribution and to find the exact range of order statistics which follows this distribution, by estimating cut-off value γ (See Appendix A.2). In order to get accurate estimates of parameters. There should be a minimum number of order statistics values, which is at least 50 [9].

Both arithmetic and logarithmic stock returns (daily, weekly, and monthly) can be modeled using the Pareto Levy distribution (See Appendix A.2). In addition to stock prices, indexes variations, volumes and volatility decay distributions may also be modeled using the same distribution [9, 10, 11, 12, 13, 14].

Since we cannot deduce a closed form of the Stable-Pareto distribution (See Appendix A.2), we cannot fit it directly to a financial data series. Therefore, we use some closed form versions of that distribution to fit for the data set. The most commonly used distribution to fit extreme variations of financial data is the simple power law distribution and some of its variations. Power-law distributions include continuous distributions such as simple power law, power law with cut-off, exponential distribution, stretched exponential distribution, log-normal and discrete distributions such as power law, Yule distribution, exponential distribution, Poisson distribution used for financial data analysis [9].

The following sections (3.2, 3.2, and 3.2) describe various methods that could be used to fit the Pareto-Levy distribution [9, 15, 16, 17, 18, 19, 20].

We assumed that extreme values of logarithmic price returns follow the given power-law distribution (Pareto-Levy distribution) and estimated the parameters.

$f(x) = \alpha\gamma^{\alpha}x^{-(1+\alpha)}, (x > \gamma)$

Where α, γ are parameters of the distribution.

Graphical Methods

The most commonly used method to fit this distribution is simple histogram analysis. The probability density function of the Pareto-Levy distribution follows $ln(f(x)) = -(1 + \alpha) * ln(x) + constant$ form. This is actually a straight line in log-log plot of the histogram of data between ln(f(x)) and ln(x). So slope S of the straight line is related to the exponent of the power law distribution as S = (1+α). The starting value that straitens the log-log plot is taken as the estimated value for γ.

The following steps were used to find α and γ values.

- Consider only positive logarithmic price returns.

- Plot a histogram with a suitable class size.
- Divide the frequencies of each class by n to get probability values, where n is the sample size.
- Plot log(probabilities) against the log(mid values) of each class.
- Least squares linear regression is used to estimate the slope of the doubly logarithmic plot.
- Compute α using the slope and estimate γ using the graphical method.
- Repeat the same steps for negative price returns.

Some other alternative methods for estimating the cut-off value are

- Using the value which straightens the log-log plot of probability density or cumulative density functions
- Identifying the point beyond which the plot of estimated power exponent and the minimum value (cut-off) is stable (This is basically known as the Hill plot) [9].

Maximum Likelihood Method with Goodness of Fit Tests

When fitting a probability distribution to the price increments and logarithmic return values, we used the Pareto Levy stable distribution (A simple power law distribution, See the Algorithm 6).

We have used maximum likelihood estimation [21] combined with goodness of fit test [12] for parameter estimation. Maximum Likelihood method is a statistical method used to fit a statistical model to the dataset in order to provide estimates for models parameter. Maximum likelihood Hill estimator method is used to estimate the power exponent value α of the distribution and selected goodness of fit tests [12] were carried out to get the best fitted parameter (α) along with the cut-off value (γ). Normally power-law exponent (α) is assumed to be greater than one because exponent less than one is not normal and cannot exist in nature [9].

Maximum likelihood method is considered as the most accurate and robust method in practice. Maximum Likelihood Method is more accurate than the Least Square regression method when fitting these distributions. This is because Least Square regression method (or graphical method in other words) is considered as a subjective method.

Parameters were estimated for each run and were averaged over five hundred runs.

In terms of goodness of fit methods, distance between empirical distribution and theoretical distribution is used to find the best fitted values of a distribution. Values obtained by Goodness of Fit tests or in other words the above mentioned distances are known as test statistic values. There are number of ways of calculating these test statistics. For non normal data, commonly used method is Kologorov-Smirnov (KS) method. There exists some other goodness of fit methods which compute the best fitting parameters for a dataset such as Kuiper, Cramer-Von-Mises/Watson and Anderson-Darling tests. Sometimes modified goodness of fit statistics gives more accurate results than normal statistics. Modification to goodness of fit statistics is done by weighting data to avoid some insensitiveness around extreme limits. In these extreme limits of data, distribution tends to get close to zero or one. So re-weighting assures uniform sensitivity across the whole range of test statistic values.

Cramer-Von-Mises test is minimum distance estimation method used to find goodness of fit by comparing probability distribution (theoretical) with a given empirical distribution function. Here the test statistics values are used in hypothesis testing to find the best fitted parameters. This test can be performed in between two empirical distributions, and it is called the Cramer-Von-Mises two sample test [9]. According to many researchers, re-weighted KS and Kuiper methods are not much different from standard KS statistics [9, 12].

In our research, Anderson-Darling minimum value test [10, 17] and hypothesis testing with Cramer-Von-Mises/Watson statistics were used (See Algorithms 4) [16, 12].

```
/* Test Statistics Formulas of Goodness of Fit Tests */
```

$$Anderson - Darling(A^2) = -n - [\Sigma_i(2i-1)(log z_i + log(1 - z_{n+1-i}))]/n \tag{3.1}$$

$$Cramer - Von - Mises(W^2) = 1/12n + \Sigma_i(z_i - (2i - 1/2n))^2 \tag{3.2}$$

$$Watson(U^2) = W^2 + n(z - 1/2)^2 \tag{3.3}$$

$$Where, z_i = F(x_i) \tag{3.4}$$

$$n = samplesize \tag{3.5}$$

(3.6)

Algorithm 4: Test Statistics Formulas of Goodness of Fit Tests

In Anderson-Darling minimum value test the minimum distance between empirical distribution and theoretical distribution or in other words minimum test statics value is used to find the best fitting parameters of the distribution (See the Figure 3.1 on page 14).

In Cramer-Von-Mises/Watson test, P-Value associated with the test statistics is used for hypothesis testing (See the Figure 3.2 on page 15) [11]. Here the test statistics are often compared with tabulated critical values (corresponding to significance levels) to take the decision of ruling out a hypothesis. Also, hypothesis testing is involved with comparing probability value (p-value) associated with the critical value to take that decision. The NULL hypothesis (i.e., the empirical and theoretical distributions are identical) is rejected if the calculated test statistics is greater than the critical value obtained from a critical values table [12] for a given significance level or, if the calculated p-value is lower than the significance level.

P-value is defined as the probability of test statistic values which is larger than the critical level. P-Value can be obtained by analyzing the empirical distribution and a larger number of synthetic distributions which have been derived from the power law distribution we used to fit, with the estimated exponent and cut-off. This means compare the empirical test statistic value with test statistic values of each synthetic distribution and its own distribution, and selecting the fraction of synthetic statistics which exceeds the empirical test statistics as the P-Value of the empirical distribution. If the P-Value is large and close to one, the difference between empirical and theoretical distributions can be only due to statistical fluctuations. If the value is small the theoretical distribution cannot be fit to the dataset [11, 12].

KS method is more accurate for small number of observations; roughly around 1000. A major draw-back in Anderson-Darling method is that it estimates large numbers as the cut-off value γ. Even though it is considered as a better method compared with KS statistics in terms of the sample size, it is more suitable for distributions with larger number of samples in the tail of the distribution. When the cut-off (γ) value is large, the number of samples taken in to consideration for fitting is less, so it leads to an increase in the statistical error of the estimated values and badly affects the ability of validating the most suitable distribution for dataset [9].

Apart from those methods Anderson-Darling statistics were used in a similar way to the Cramer-Von-Mises test for hypothesis testing and found that the results are similar to the Cramer test. In order to increase the accuracy of the testing method, Anderson-Darling minimum value test was performed after removing the outliers (values greater than $Q_1 + 3 * IQR$) (See Appendix A.1) from the minimum value selection range. In Hypothesis testing, if the method does not reject the hypothesis, we assumed all data follows the given distribution and estimated the parameters from the whole dataset. Estimated values were analyzed in all theses methods and their accuracy is investigated in each case.

After analyzing dataset parameters such as sample size and data distribution, and considering their advantages and disadvantages, we chose the set of fitted parameter values using the Cramer-Von-Mises/Watson goodness of fit test combined with maximum likelihood Hill estimator method [11] to analyze extreme values of logarithmic price returns in our datasets.

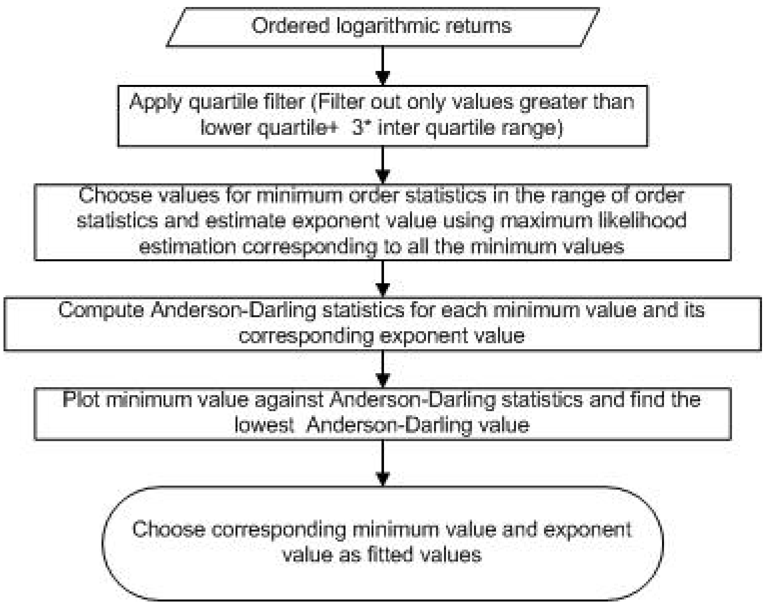


Figure 3.1: Anderson Darling Method

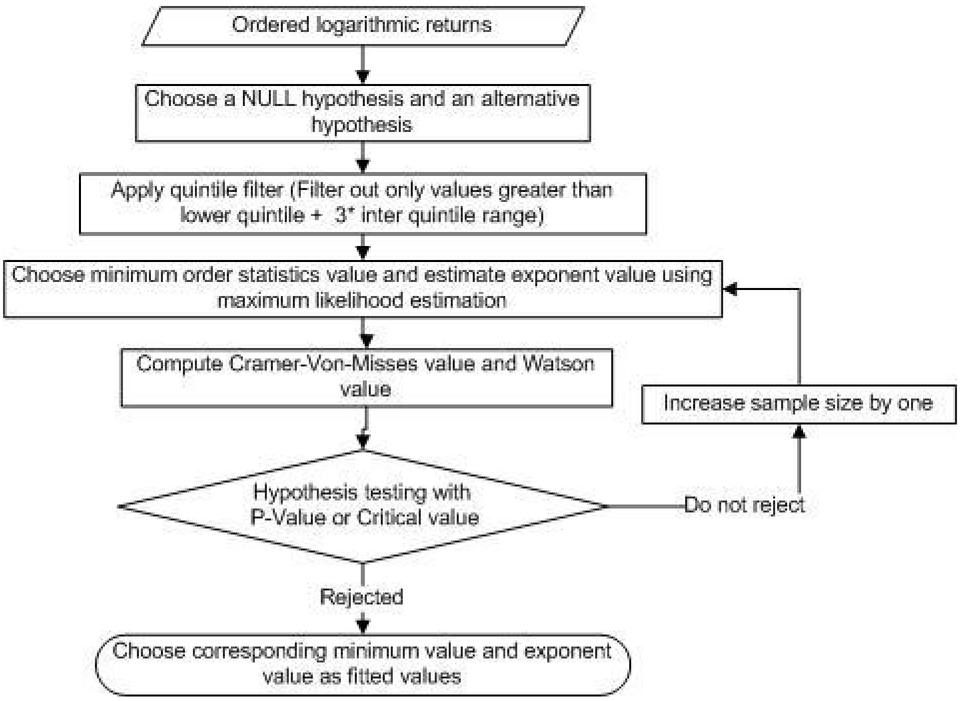


Figure 3.2: Cramer-Von-Mises test

Chapter 4

A Comparison of Maslov, Variants and Real Data

4.1 Numerical Results

When analyzing price time plots of M_0 and its variants, it can be clearly observed that in M_0 there is an abnormal cone shape behavior of prices, which is not visible in real stock prices(See Figures 4.2 and 4.3). Apart from that the daily price deviation (i.e., the difference of daily high price and the daily low price) is very high in M_0 with compared to M_1 and M_2 (See the Figure 4.2). M_2 showed the lowest deviation compared to the other two models (See Figure 4.9). The cone shapes that can be observed in M_0 are a resultant of the behavior of limit and market orders with empty book states. Limit orders build the book while market orders are consuming the book. Once it gets empty, formation of the cone shape stops and the next order price is determined with respect to the LTP and again it starts to build the cone with limit and market orders.

In Figures 4.4 and 4.5, we have given the histograms of price increments and logarithmic price returns for combined prices of one hundred runs of Maslov model and its variants. We can clearly observe the sharp peak and fat tail characteristics in these graphs which can be seen in financial data graphs.

Also when analyzing the price density plots (See Figures 4.6 and 4.7) of all three models we can clearly see that M_0 has a multi modal price density plot which is a resultant of layered price values in the price time graph. But when we analyzed this among one hundred simulations, all three models showed similar characteristics. Real price graph of DELL data in NASDAQ (Figure 4.8) shows multi modal characteristics but the range of price deviations is not as high as per Maslov models. The bi-modal characteristic of M_0 and M_1 price density plots leads to a much flatter distributed density plot than M_2, which has high price concentration around the starting value. In the spread histogram of M_2, it is observed that the spread values are low when compared to M_0 and M_1.

When analyzing the spread histogram plot of three Maslov variants in Figure 4.9, we can clearly observe that the M_2 model shows a clear difference compared to M_0 and M_1. In M_2, the number of small size spread values is high compared to other two.

The Hurst exponent estimate, Pareto exponent estimate for negative and positive returns (including their standard errors) and auto correlation decay values have been computed using the aforementioned techniques for M_0 and its variants are listed in Tables 4.1, 4.2 and 4.3 and the corresponding values for real data samples are listed in Table 4.5 and 4.7.

When comparing the behavior of the Hurst exponent, we can clearly see that in all the Maslov models, the Hurst exponent value is lower than the value obtained for real data samples. But in M_2, we can see this value is much lower than that of M_0 and M_1. We can observe that in both real data and Maslov data when the return period increases the value of the Pareto exponent reduces.

In the auto correlation decay exponent, we can see that all indexes show a value slightly less than than 0.4 while real prices show values closer to 0.3. In all Maslov models, there is no clear distinction in values but they are much closer to price returns than index return values.

In terms of the Pareto exponent α, we can see that the values of the Pareto exponent are slightly higher in all the Maslov Models when compared with real data. Specially in modification two, this value is much higher than the other Maslov versions (See the Tables 4.3 and 4.5).

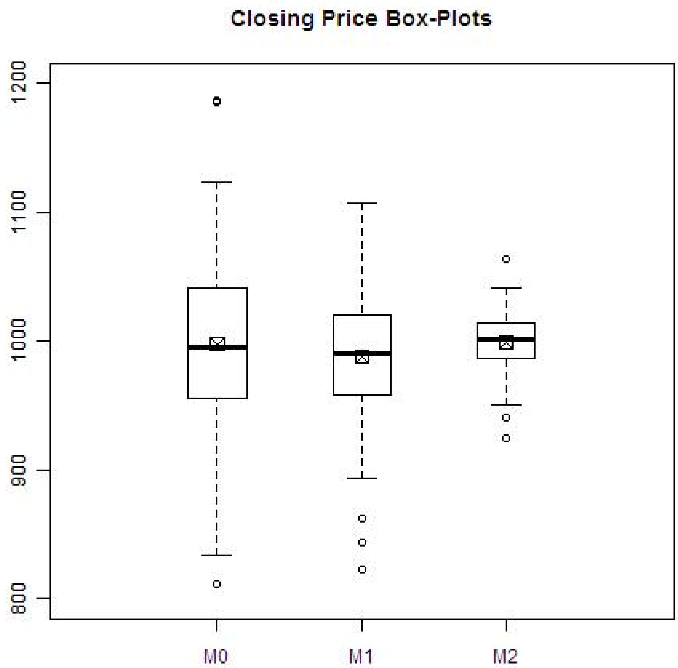
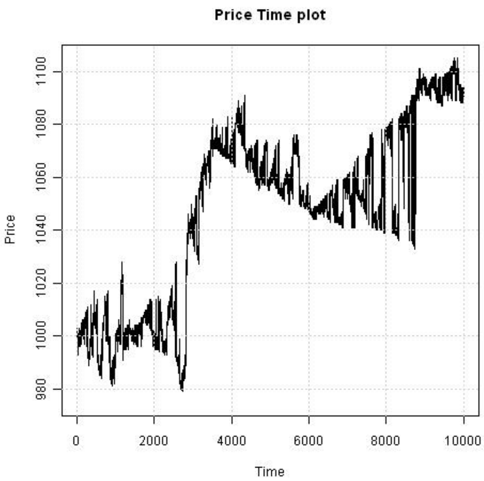


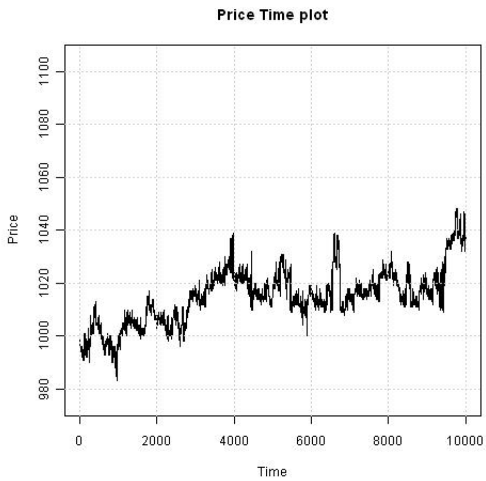
Figure 4.1: Closing price box-plots of Maslov model and it's variants.

Model	Prices	Price Increments	Logarithmic Price Returns
M_0	0.19	0.31	0.31
Std Err	0.00	0.00	0.00
M_1	0.10	0.36	0.36
Std Err	0.00	0.00	0.00
M_2	0.08	0.28	0.28
Std Err	0.00	0.00	0.00

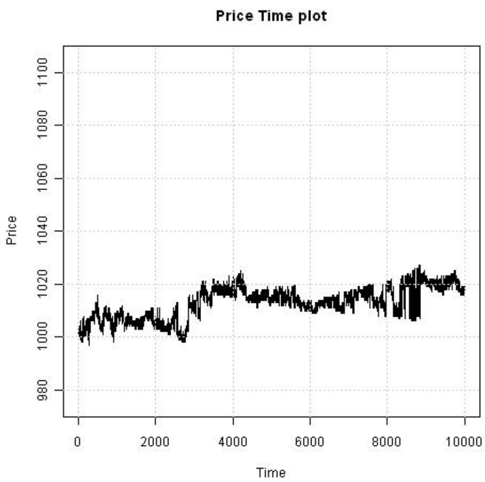
Table 4.1: Hurst Exponent Estimates: This table summarizes values obtained for the Hurst exponent for prices, price increments, and logarithmic price returns of three Maslov variants for 500 samples



(a) M_0



(b) M_1



(c) M_2

Figure 4.2: Price Vs. Time graphs over three Maslov variants using the same random seeds, with $p_0 = 1000$.

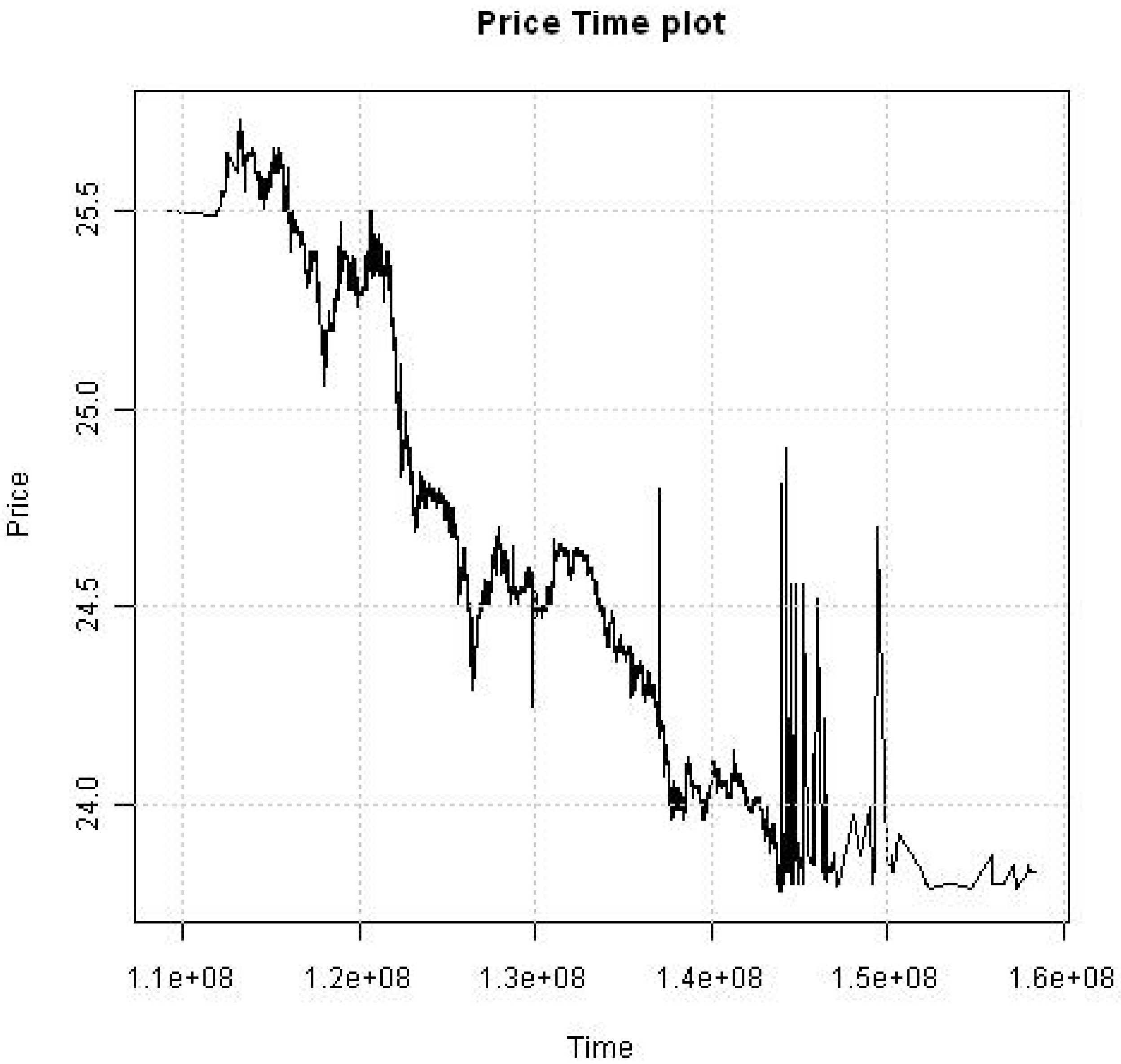
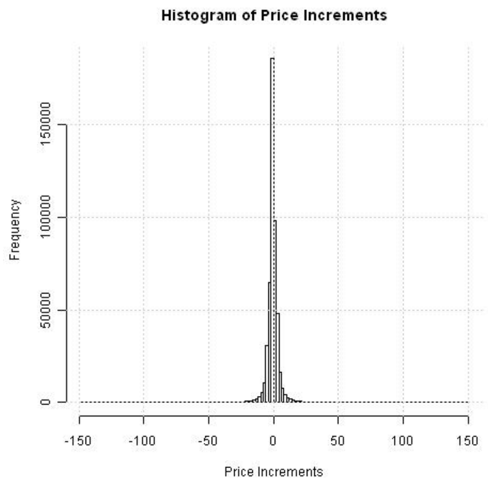
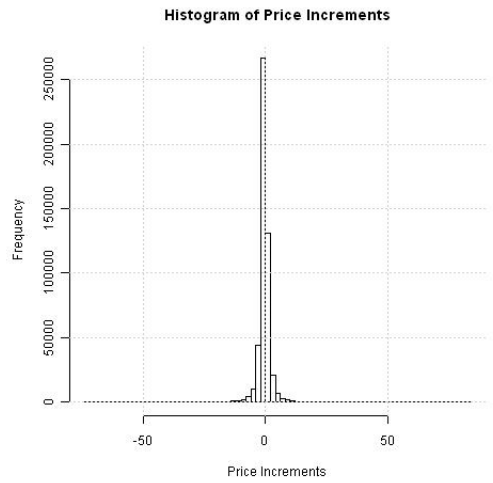


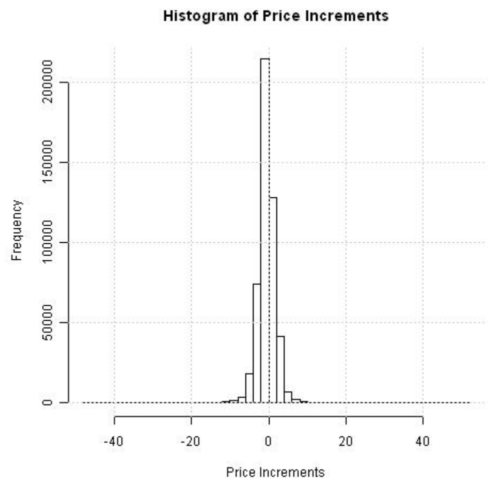
Figure 4.3: Price Vs. Time graph :DELL NASDAQ(1^{st} of February 2007)



(a) M_0

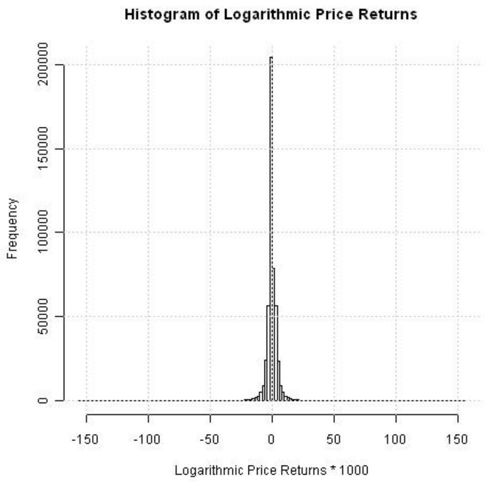


(b) M_1

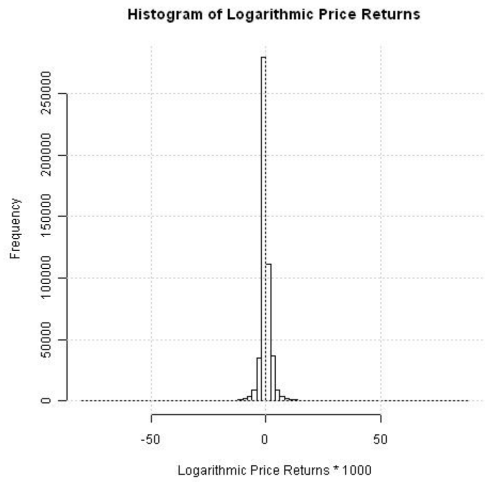


(c) M_2

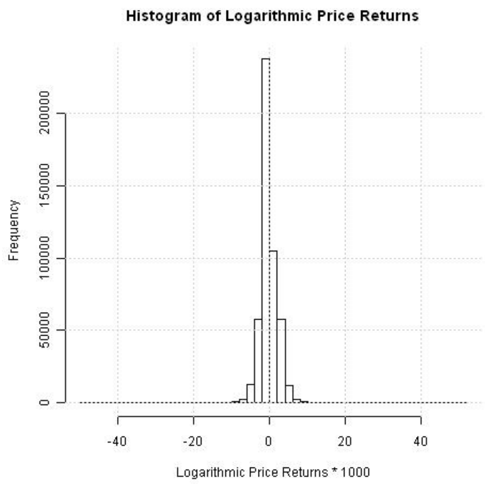
Figure 4.4: Histograms of price increments over three Maslov variants when data belonging to 100 time series samples are combined. Price increment is $P_t - Pt - 1$. We assume P_0 = 1000 and simulate a time series of 10000 observations



(a) M_0



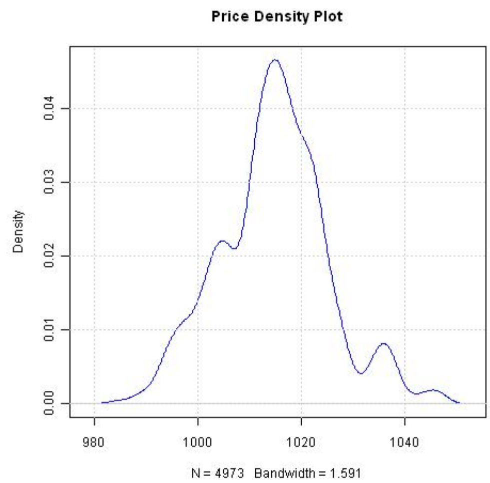
(b) M_1



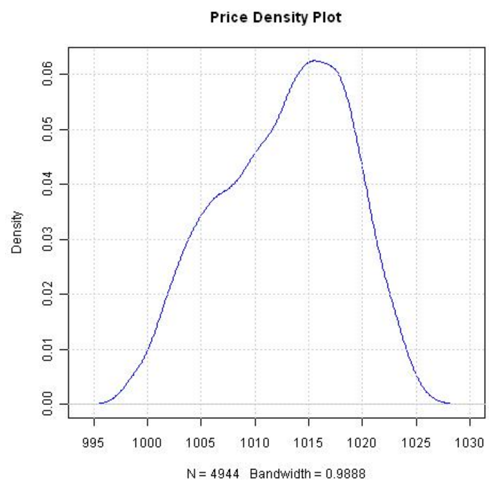
(c) M_2

Figure 4.5: Histograms of logarithmic price returns over three Maslov variants (All the logarithmic price return values are multiplied by 1000). Logarithmic Price Return is $log(P_t) - log(Pt - 1)$. We assume P_0 = 1000 and simulate a time series of 10000 observations.

(a) M_0

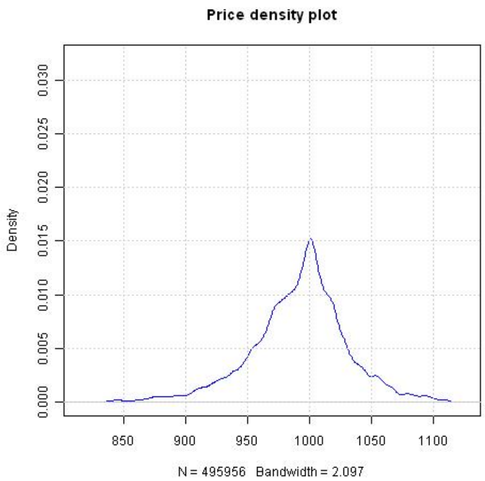


(b) M_1

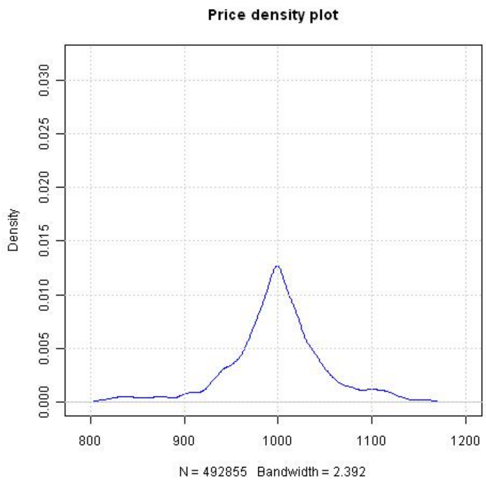


(c) M_2

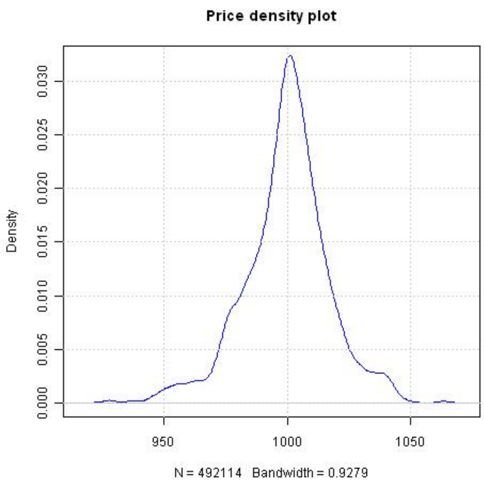
Figure 4.6: Price density plots over three Maslov variants. For single time series shown in Figure 4.2.



(a) M_0

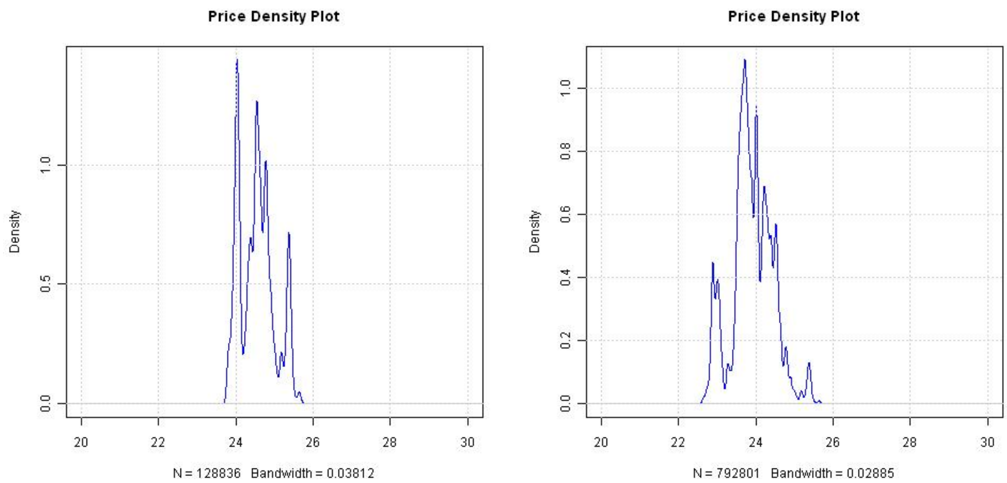


(b) M_1



(c) M_2

Figure 4.7: Price density plots Over three Maslov variants when data belonging to 100 time series samples each with 10000 observations are combined



(a) Price density plot for DELL prices on 1^{st} of February 2007

(b) price density plot for DELL prices in February 2007

Figure 4.8: First graphs shows the price density plot for DELL on 1^{st} of February 2007 and the second graph shows the same plot for the whole month of February

Model	Samples	Pareto Negative	Pareto Positive	Cut-off Negative	Cut-off Positive	ACF decay
M_0	500	2.34	2.34	12.15	12.11	0.37
Std Err		0.03	0.05	0.05	0.06	0.00
M_1	500	1.75	1.76	4.90	5.04	0.35
Std Err		0.01	0.01	0.03	0.01	0.00
M_2	500	4.78	4.84	8.12	8.15	0.34
Std Err		0.16	0.19	0.05	0.05	0.00

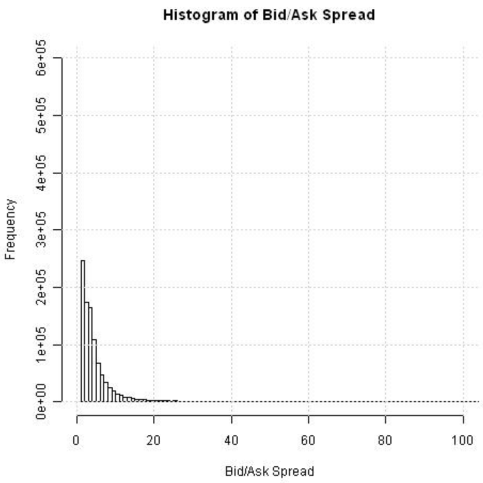
Table 4.2: Maslov Results for Price Increments: This table summarizes values obtained for Pareto exponent for negative increments, Pareto exponent for positive increments, cut-off value for negative increments, cut-off value for positive increments and exponent of Auto-correlation decay of absolute values of price increments of three Maslov models for 500 samples.

Model	Samples	Pareto Negative	Pareto Positive	Cut-off Negative	Cut-off Positive
M_0	500	2.20	2.20	0.001	0.001
Std Err		0.04	0.04	0.00	0.00
M_1	500	2.00	1.94	0.004	0.005
Std Err		0.01	0.01	0.00	0.00
M_2	500	4.29	4.18	0.007	0.007
Std Err		0.16	0.15	0.00	0.00

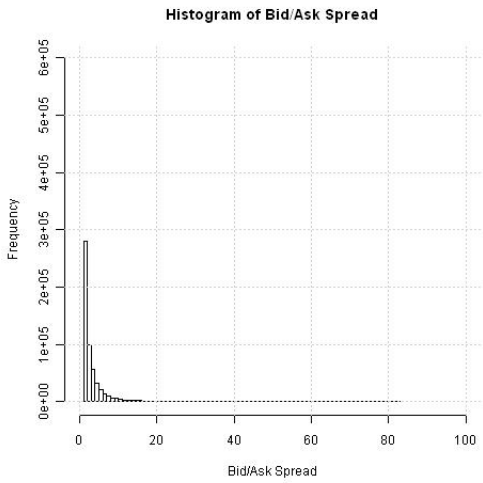
Table 4.3: Maslov Results for Logarithmic Price Returns: This table summarizes values obtained for Pareto exponent for negative returns, Pareto exponent for positive returns, cut-off value for negative returns, cut-off value for positive returns of three Maslov models for 500 samples.

Sample	Negative Pareto Distribution			Positive Pareto Distribution		
	Mean	Mode	Variance	Mean	Mode	Variance
M_0	0.0018	0.001	7.6388*10^{-6}	0.0018	0.001	7.6388*10^{-6}
M_1	0.008	0.004	N/A*	0.0103	0.005	N/A*
M_2	0.0091	0.007	8.4805*10^{-6}	0.0092	0.007	9.2909*10^{-6}

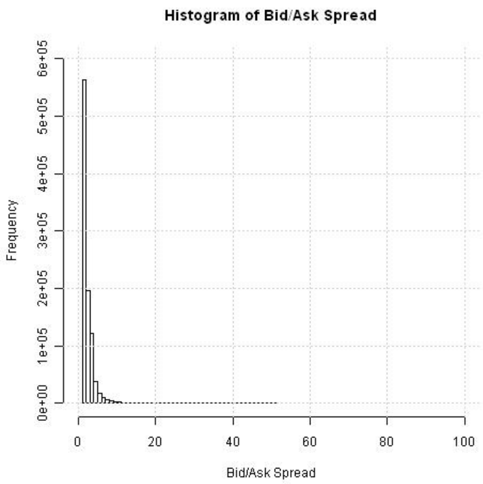
Table 4.4: Probability Distribution Analysis: This table summarizes mean, mode, and variance values of the fitted Pareto distributions of negative and positive logarithmic returns over three Maslov variants. (*: Variance does not exist when $\alpha < 2$.)



(a) M_0



(b) M_1



(c) M_2

Figure 4.9: Spread histograms over three Maslov variants when data belonging to 100 time series samples each with 10000 observations are combined

Sample	Return Frequency	Hurst Exponent	Pareto Negative	Pareto Positive	Cut-off Negative	Cut-off Positive
Dow-Jons	Daily	0.53	2.45	2.45	0.02	0.01
	Weekly	0.53	3.37	2.15	0.07	0.02
	Monthly	0.57	1.10	2.33	0.02	0.04
S&P 500	Daily	0.56	2.64	2.92	0.01	0.02
	Weekly	0.60	2.39	2.77	0.02	0.02
	Monthly	0.73	2.20	2.70	0.05	0.04
General Electrics	Daily	0.48	2.25	3.04	0.04	0.04
	Weekly	0.47	2.04	2.70	0.06	0.06
	Monthly	0.50	1.70	2.76	0.11	0.08
MRO	Daily	0.48	1.93	1.30	0.02	0.02
DELL		0.53	1.72	1.79	0.39	0.75

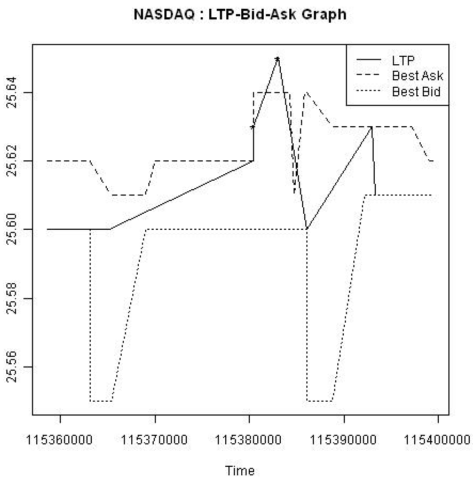
Table 4.5: Real Data Results for Logarithmic Price Returns: This table summarizes values obtained for the Hurst exponent, Pareto exponent for negative returns, Pareto exponent for positive returns, cut-off value for negative returns, cut-off value for positive returns for real data samples

Sample	Negative Pareto Distribution			Positive Pareto Distribution		
	Mean	Mode	Variance	Mean	Mode	Variance
Dow-Jons						
Daily	0.033	0.02	0.001	0.016	0.01	0.0002
Weekly	0.099	0.07	0.002	0.037	0.02	0.004
Monthly	0.22	0.02	N/A*	0.070	0.04	0.006
S&P 500						
Daily	0.016	0.01	0.0001	0.030	0.02	0.0003
Weekly	0.034	0.02	0.001	0.031	0.02	0.0004
Monthly	0.0916	0.05	0.019	0.063	0.04	0.002
General Electrics						
Daily	0.072	0.04	0.009	0.059	0.04	0.001
Weekly	0.117	0.06	0.169	0.095	0.06	0.004
Monthly	0.267	0.11	N/A*	0.125	0.08	0.007
MRO						
Daily	0.041	0.02	N/A*	0.086	0.02	N/A*
DELL						
	0.931	0.39	N/A*	1.699	0.75	N/A*

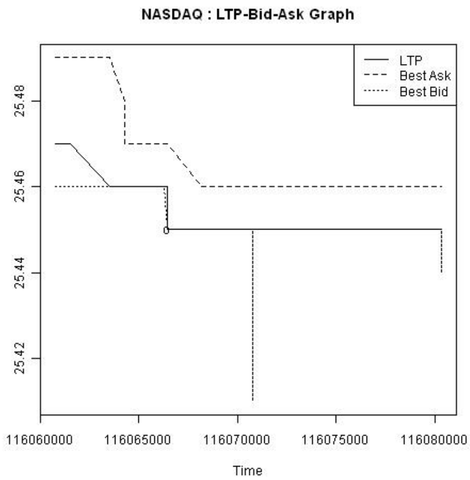
Table 4.6: Probability Distribution Analysis: This table summarizes mean, mode, and variance values of the fitted Pareto distributions of negative and positive logarithmic returns of selected real data samples. (*: Variance does not exist when $\alpha < 2$.)

Sample	Return Frequency	ACF decay
Dow-Jons	Daily	0.38
	Weekly	0.36
	Monthly	0.32
S&P 500	Daily	0.37
	Weekly	0.36
	Monthly	0.36
General Electrics	Daily	0.24
	Weekly	0.27
	Monthly	0.32
MRO	Daily	0.28
DELL		0.32

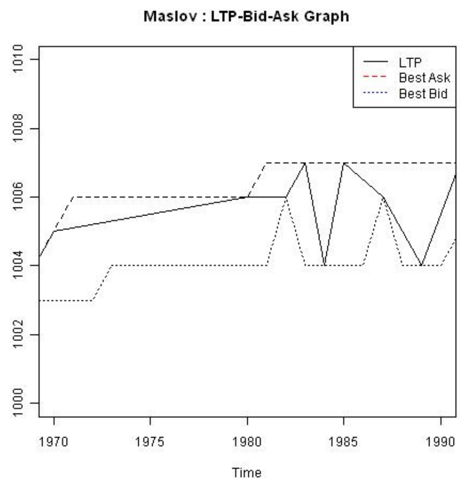
Table 4.7: Real Data Results for Price Increments: This table summarizes values obtained for the exponent of Auto-correlation decay for selected real data samples



(a) DELL Nasdaq, LTP Above Best Ask



(b) DELL Nasdaq, LTP Below Best Bid



(c) Maslov

Figure 4.10: Last Traded Price, Best Bid and Best Ask behavior of real data (DELL-Nasdaq) and Maslov variants

Chapter 5

Conclusion

Maslov [2] simulates the behavior of a limit order book using a single stock with one trader submitting limit and market orders based on random logic. In our research we analyzed the properties of this model and compared its behavior to real data. In order to produce more realistic behavior the Maslov model, we propose two variants of it and discuss their behavior in relation to the original Maslov model and some real data samples.

Our numerical analysis reveals that the behavior of the Maslov model deviates from real financial data and the modified versions (specially the second modification M_2) showed a much closer relationship to the same. We could observe it mainly from the behavior of price vs time graph (See the Figure 4.2 on page 18). In analyzing the behavior of the Maslov model & it's variants time series techniques such as the Hurst exponent analysis and Histogram analysis were used.

The next step in our research involves introducing some complex behaviors to the model such as considering the market conditions before computing the limit order price and introducing evolutionary strategies to the order generation logic.

Appendix A

Appendix

A.1 Methods for Time Series Analysis

In order to analyze the model data and financial data, we used various time series analysis and comparison techniques [22].

A time series is a collection of data items observed through repeated measurements over a certain period of time. There are two main types of time series available in practice, namely stock series and flow series. Stock series is a measure of certain attributes at a point in time. For example, the monthly labor force survey is a stock measure because it indicates whether a person is employed in that particular month or not. Flow series is a measure of activity over a given period. For example, manufacturing is a flow series measure, because daily manufactured amounts are summed to give a total value for production for that particular period of time [22, 23].

Systematic effects and calendar related effects which can occur in any kind of time series are called seasonal effects. For example, a sharp increase in stock trading can occur around December in response to the Christmas period. Natural Conditions (unexpected weather patterns such as snow in summer), Business and Administrative decisions (Start and end of the school term) and Social and Cultural aspects (Christmas) can cause seasonal variations in time series data. Seasonal effects can be identified by regularly spaced peaks/troughs which have a consistent direction and approximately the same magnitude every year, relative to the trend. Another form of seasonal effect is the trading day effect, that is the number of trading days in a given month depends on holidays (the exact date of holidays such as Easter changes) which leads to different effects for the same reason in different periods.

Seasonal adjustment is removing seasonal effects from a time series. But when a time series is dominated by the trend or irregular(random) components, the seasonality adjustments for that particular series may be inappropriate, and it is very difficult to identify and remove seasonality from a series. Hence most often introduction of an artificial seasonal element to the series is recommended [22, 23].

The trend of a time series is defined as the long term movement without any calendar related and irregular effects. The irregular component of a time series or the residual part is what remains after the seasonal and trend components are removed. Random variations of a time series are fluctuations which are not systematic or predictable. In a highly irregular series, this behavior could mask the trend and seasonality behaviors and dominates the series.

It is very hard to compare two time series with periodic data, because of the inaccuracy and time delays in the identification of turning points and structural shifts. Turning points are the points of time where the direction of the underlying trend of the series changes. For example, it is very difficult to locate the time point where a consistently decreasing series begins to rise steadily. When comparing two time series values, we may miss these turning points and leads to an inaccurate results [22, 23].

We can use decomposition methods to separate out these parts from a time series. The main decomposition models are additive or multiplicative, but there are some other forms in practice such as pseudo additive. An additive model is suitable if the amplitudes of both the seasonal and irregular components do not vary as the level of the trend varies. A multiplicative model is suitable if the amplitudes of both the seasonal and irregular variations increase as the level of the trend increases. A multiplicative model cannot be used when the original time series contains very small or zero values for each of its components. In this case, a pseudo additive model (a combination of additive and multiplicative models) is used. Pseudo additive model assumes that seasonal and irregular variations are both dependent on the trend but independent of each other. The pseudo-additive model continues the convention of the multiplicative model to have both the seasonal factor and the irregular factor centered around one [22, 23].

Extreme values or outliers are the values of a time series which are unusually different compared to other data. These values could distort the overall underlying movement of a time series by affecting the trend. It is necessary to detect and correct for outliers in order to improve modeling of the three time series components (trend, seasonal and irregular).

Quantile filtering is a common way of extracting outliers from a time series. The set of values beyond the limit $Q_1 + 3 * IQR$ are considered to be outliers of a series where $Q_1, Q_2, and Q_3$ are the first, second, and third quartiles of the dataset.The inter quartile range $(IQR = Q_3 - Q_1)$ is defined as the spread of the middle 50% of the data and is often used as a measure of spread. This is also known as the mid spread, and is a measure of statistical dispersion of data, calculated by difference between the third and first quartiles [22, 23].

Trend breaks can be possible due to economic policy decisions, changes in population behavior and changes in the way an attribute is measured. Seasonal breaks are changes in the seasonality of a series, which do not affect the level or the trend of the series. They may be affected by social traditions, administrative practices or technological innovations [22, 23].

In terms of detection and correction of these effects, forward factors and concurrent analysis are two main approaches to derive seasonal and trading day factors. The forward factors method is basically an annual analysis of the latest available data to predict the seasonal and trading day factors for the next year. Concurrent analysis is re-estimating seasonal factors as each new data point becomes available. This method is more computationally intensive than the forward factor method, but the seasonal factors will be more responsive to dynamic changes. Methods of adjustment can be divided in to two main methods: indirect or aggregate method of adjustment and direct or disaggregate method of adjustment. The indirect method seasonally adjusts each of the lower component series individually, then sums all the values to obtain the seasonally adjusted series for the total. The direct method of adjustment involves summing up of all the original series to form a total series and then seasonally adjusting the total series directly. If the component series has very different seasonal patterns, then the indirect seasonal adjustment is appropriate. However if seasonality is low and difficult to identify in the individual series, then using direct seasonal adjustment can remove any residual seasonality from the aggregate series [22, 23].

Stationarity of a time series is a main characteristic that is analyzed in time series analysis. Most techniques used in time series forecasting expect stationarity condition to be satisfied. i.e, a time series must follow a first and second order stationary process. First Order Stationarity implies that its expected value remains the same at any time. For example, a financial time series becomes first order stationary when its trend component is removed by some mechanism such as differencing. A series is second order stationary, if it is first order stationary and the covariance between two time seres values is a function of time difference only. In case of financial time series, they can be made second order stationary if we remove its variance by applying some kind of mechanism such as taking the square root [22, 23].

Filtering techniques are used to extract useful information such as the cyclic component from a time series. These filters are a direct implementations of input-output relationships. Differencing and filtering are used as data pre-processing techniques before applying effective and efficient time series modeling methods [22, 23].

A.1.1 Smoothing Techniques

Smoothing is a technique used to reduce the variability of a dataset. Smoothing reduces variance by averaging over the periodogram of neighboring frequencies and introduces bias because the expectation of neighboring periodogram values is not identical to the selected frequency. Over smoothing is a serious issue. Tapering corrects the bias introduced from the finiteness of the data. The expected value of the periodogram at a certain frequency is not quite equal to the spectral density. It can be affected by the spectral density at neighboring frequencies. For the spectral density which is more dynamic, more tapering is required. Smoothing introduces bias, but reduces variance. Tapering decreases bias and introduces variance and also attempts to remove the influence of side lobes that are introduced by the spectral window [22, 23].

Exponential smoothing: Exponential smoothing is a technique that can be applied to time series data to prepare smoothed data and to make forecasts. In exponential smoothing, the weighted average of the time series are calculated by assigning exponentially decreasing weights with time. i.e., the exponentially smoothed value for time period t is $S_t = \alpha x_{t-1} + (1-\alpha)s_{t-1}$, where α is smoothing factor and $1 < \alpha < 2$.

Exponential smoothing is a commonly used technique in financial market and economic data.

Simple moving average: The simplest way to smooth a time series is to calculate a simple (unweighted) moving average. The smoothed value is just the mean of the last k observations of the series. One main disadvantage of this technique is that it cannot be used to smooth the first k -1 terms of the time series. When calculating the simple moving average for period k, unlike in exponential smoothing, is this equal weight is given for each observation.

Weighted moving average: This calculated as a weighted moving average using a set of weighting factors. Sum of all the weight factors should be equal to one. Weight factors are chosen in such a way that more weight is given for most recent time series values and less weight is given for old time series values. This technique has the same disadvantage as the simple moving average technique.

Exponential moving average: In this case, the current smoothed value (y'_t) is computed as the simple weighted average of the current observation at t (y_t) and the previous smoothed value at t-1 (y'_{t-1}), where α is the smoothing factor, and $0 < \alpha < 1$ gives the formula y'_t = $(1-\alpha)y'_{t-1}$ + αy_t.

If the value of α is close to one, it gives less smoothing effect and gives greater weight to recent changes in the data. If α closer to zero, it gives a greater smoothing effect and is less responsive to recent changes. There is no formalized and correct procedure for choosing α. Sometimes expert knowledge is used to choose an appropriate factor and least squares method is used to optimize the selected value.

A.1.2 Time Series Models

Time domain models and frequency domain models are used for time series analysis. One way of analyzing financial time series is to model the process using some statistical methods. An accepted model for stock price series is the famous random walk or Brownian motion model proposed by Osborne in

1959 [24]. Osborne suggests that stock return is a random variable that follows zero-mean Gaussian distribution. Other models, such as the linear correlative model for stock returns have also been used in the research literature.

A.1.3 Autocorrelation Analysis of a Time Series

Autocorrelation and partial auto correlation are used in modeling a time series model for analysis. Autocorrelation measures the similarity between time separated observations as a function. Simply, it is the cross correlation of a signal with itself. Autocorrelation is a mathematical tool measures the repeated patterns in a time series. It can also be used to identify the missing fundamental frequencies in a signal implied by its harmonic frequencies. Autocorrelation of a random time series process describes the correlation between values of the series at different points in time and provides a strong scale free measure of the strength of statistical dependence, as a function of the two times or of the time difference. Its value must lie in the range [-1, 1]. When the autocorrelation function is normalized by mean and variance of that particular series, it is referred to as the autocorrelation coefficient [22, 23].

Partial autocorrelation function is very important when identifying autoregressive and autoregressive moving average models for time aeries using the Box-Jenkins approach [22]. Partial autocorrelation of lag k is the autocorrelation between t and t + k with the linear dependence of t + 1 through to t + k - 1 removed. This is useful in identifying the order of an autoregressive model. The partial autocorrelation of an AR(p) process for lags greater than p is zero. If the sample autocorrelation plot indicates that an AR model may be appropriate, then the sample partial autocorrelation plot is examined to identify the order of the AR process [22, 23].

Model Selection Criteria [22, 23]:

- If none of the simple autocorrelation coefficients are significantly different from zero, the series can be identified as a random number of white noise series. This kind of series cannot be modeled by an autoregressive model as there is no information involve for modelling.

- If the simple autocorrelation coefficients decrease linearly, pass through zero and become negative, or if the simple autocorrelations show a wave like cyclic pattern while cutting the zero line several times, this series can be identified as not stationary; it should be differenced once or more times using an appropriate transformation to convert the series in to a stationary one before it is modeled with an autoregressive model.

- If the simple autocorrelation coefficients indicate seasonal patterns, (i.e., if there are almost equally spaced cyclic autocorrelation peaks) the series is not stationary and it should be differenced with a gap (approximately equal to the seasonal interval) before applying a model.

- If the simple autocorrelation coefficients are decreasing exponentially but approaching zero gradually while the partial autocorrelation coefficients are significantly non zero for some small number of lags and also if they are not significantly different from zero, this series could be modeled with an autoregressive process(AR(p) process).

- If the partial autocorrelation coefficients are decreasing exponentially but approaching zero gradually while the simple autocorrelations are significantly non-zero for some small number of lags and they are not significantly different from zero, this series could be modeled with a moving average process(MA(q) process).

- If the partial and simple autocorrelations both converge to zero for longer lags, but neither actually reaches zero after any particular lag, this series may be modeled by a combination of autoregressive and moving average processes(ARMA(p,q) process.

A.2 Pareto-Levy Stable Distributions

Many empirical quantities of financial data cluster around extreme values [25]. As a result of this, the Pareto-Levy family of distributions always go hand in hand with financial data. We have used a simple power-law distribution (a member of Pareto-Levy distribution family) to fit extreme variations of logarithmic returns of various datasets used in our research (See the Algorithm 6).

Prior to Paul Levy's mathematical analysis, some analysts investigated histograms of some variables which generally looked like normal distributions but deviated from the actual shape of the normal distributions. They identified some sharp peak and fat-tailed characteristics in these distributions and named them leptokurtic. In 1915, economist Wesley Claire Mitchell [26] showed that the distribution of the percentage changes in stock prices deviate from the normal distribution. This means that the probability of having extremely large fluctuations or extremely small fluctuations is high compared to having moderate fluctuations (higher proportion of probability is in the tails of the distribution compared to normal distribution).

Distribution of a random variable can be considered as stable if the linear combination of two independent copies of that particular random variable has the same distribution. So in general, if x and y are random variables of two stable distributions, x+y also has a stable distribution. These stable distributions are called Levy alpha-stable distributions [3, 4]. *Ex: If x_1 and x_2 are two independent copies of random variable x, then the distribution of ax_1+bx_2 has the same distribution as $cx + d$. It becomes strictly stable only if $d = 0$.*

The normal distribution is one variation of stable distribution. According to the central limit theorem, a properly normed sum of a set of random variables with finite variance converges into a normal distribution as the number of variables increases. Stable distributions which are not normal are called stable Paretian distributions after Vilfredo Pareto. All stable distributions are infinitely divisible (See A.2.1). Gnedenko and Kolmogorov state that the sum of random variables drawn from a power law tail distribution ($x^{-(1+\alpha)}$) with exponent $1 + \alpha$ will converge to a stable distribution as the number of variables increases [3, 4].

It is not possible to analytically derive the probability density function for general stable distributions. However Paul Levy discovered a generic characteristic formula (See Algorithm 5) for all stable distributions [3, 4].

There are four main parameters in stable Pareto distributions:

α : Stability parameter α, also known as characteristic exponent or peakedness of the distribution determines the type of the distribution where $0 < \alpha \leq 2$. For normal distribution $\alpha = 2$. The second (variance) or higher moments exist only when $\alpha = 2$.

β : Skewness parameter. This is a measure of asymmetry, identified as the third central moment of the distribution (When $\alpha < 2$ the second or higher moments does not exist for the distribution). β can be any real value in the range [-1,1]. For normal distribution or any other symmetric distributions $\beta = 0$. When $\beta < 0$ or $\beta > 0$, the distribution skews left and right, respectively.

c : Scale or dispersion parameter (which is a measure of the width of the distribution). This can refer to any positive real value and this value is related to the standard deviation of normal distributions.

In non-normal distributions this is not related to the standard deviation but in non normal stable distributions it is infinite.

μ : This represents the shift of the distribution, Also known as the mean or a measure of centrality. This can refer to any real value. Distribution mean exists and equal to the value of μ only when $\alpha > 1$.

In general, the parameters β and α are considered to be shape parameters while μ and c are location and scale parameters respectively.

A Log-log plot of probability density functions of symmetric centered stable distributions show power law behavior for large x, with the slope or power law exponent equal to $-(\alpha + 1)$. So the parameter α increases the peakedness of the distribution goes down while the slope of the log-log plot becomes steeper. Log-log plot of skewed centered stable distributions probability density functions show the power law behavior for large x. The slope of the linear portions is equal to $-(\alpha + 1)$.

```
/* Characteristic Function (Pareto-Levy Distribution)                */
```

$$CharacteristicFunction = \phi(\omega) \tag{A.1}$$

$$log(\phi(\omega)) = \imath\mu\omega - \mid c\omega \mid^{\alpha} (1 - \imath\beta F(\omega, \alpha, c)) \tag{A.2}$$

$$F(\omega, \alpha, c) = sgn(\omega)tan(\pi\alpha/2), if(\alpha \neq 1) \tag{A.3}$$

$$= -(2/\pi)log(\mid c\omega \mid), if(\alpha = 1) \tag{A.4}$$

$$sgn(\omega) = 1, \omega > 0 \tag{A.5}$$

$$= 0, \omega = 0 \tag{A.6}$$

$$= -1, \omega < 0 \tag{A.7}$$

Algorithm 5: Characteristic Function (Pareto-Levy Distribution)

A.2.1 Flavors of stable distributions

Following distributions can be defined as special cases of Stable Pareto Distribution.

Some of the derived distributions of stable Pareto distributions [3, 4].

Normal Distribution : For the normal distribution $\alpha = 2$, $\beta = 0$, variance $s^2 = 2c^2$, and $\mu = mean$.

Cauchy Distribution : For Cauchy distribution, $\alpha = 1$, $\beta = 0$. Cauchy distribution does not have a mean value, so the central moments are not defined.

Levy Distribution : For Levy distributions, $\alpha = 1/2$ and $\beta = 1$, where c is the scale parameter of the distribution.

Except for the normal distribution with $\alpha = 2$, all other stable distributions are leptokurtotic and heavy-tailed in shape. The normal distribution, Cauchy distribution, and Levy distribution are considered as special cases of stable distribution because all of them posses the aforementioned characteristics.

When relating this behavior with financial data, it is observed that the linear behavior with slope $(\alpha+1)$ can be observed after some value of $x = \gamma$ and in the range $[\gamma, \infty]$. Determination of the exponent $(\alpha + 1)$ and the cut-off parameter (or threshold) γ is done by performing a simple graphical method (obtaining the value which separate two power law regions) to the log-log scale. But this procedure is considered as subjective. Hence methods such as goodness of fit tests 3.2 can be used to estimate the cut-off parameter of the distribution.

Problems Associated with Fitting Data to a Probability Distribution [9, 10, 11]:

The main problem of fitting financial data to a distribution is to choose the best distribution to fit. Commonly used method to address this issue is by means of a hypothesis which states that the given data set has been drawn from a particular distribution and to rule out the other competing hypothesis while proving the selected one.

When selecting an appropriate data range to fit to a given distribution, we often use a cut-off value. Choosing this cut-off value is also problematic because if we choose a very low value we may be selecting data which have not come from the selected distribution. On the other hand, if we chose a very large cut-off value we might be omitting legitimate values which actually follow the selected distribution.

When we fit a data set to a distribution, maximum likelihood or any other method gives us only the best fit of the given distribution to the given data set. But it doesnt give any warnings or errors if the given dataset does not follow the given distribution. In fact, there can be some other distributions which would be best fits to the given dataset. So our fitting method does not imply that our dataset actually follows the given distribution.

There can be deviations when we try to fit a known distribution to a data set which was drawn from that particular distribution, because of the random nature of the sampling. Addressing the issue of finding the best distribution type to fit a given data set is a big problem in financial world [9, 10, 11].

Methods that can be Used to Validate the Fitted Values :

Likelihood ratio test is used to compare distributions with one another. It simply calculates the likelihood of two distributions and chooses the one with the higher likelihood. Also sign of the logarithmic ratio between two likelihood values can be used to find the best fit. Non parametric bootstrap method can be used overcome the uncertainty of estimated data. It is done by randomly selecting a large number of sequences (1000) from the original dataset and estimating cut-off and exponent values for each of those datasets to get the average of the estimated values. We have used this method when fitting real data to the distribution.

Monte-Carlo power test is used to find the best goodness of fit method. It is simply analyzing the test statistic values and gives evidence on the speed of convergence of the method and effect on values of the parameters for convergence. According to this test, modified Cramer-Von-Mises test gives more accurate and robust values above all the other methods [9]. We can use the p-value approach to find out the best fitting distribution. We can simply calculate the p-value of competing distributions and compare with the main (assumed) distribution to get the best fit. If the P-value of our assumed distribution is large, then the assumption is not ruled out [11].

The following sections (A.2.1 and A.2.1)) describe the terms used to describe a probability distribution.

Infinite divisibility If x is any random variable with cumulative distribution function F, and F is infinitely divisible for every positive integer n, then there exist n different independent identically distributed random variables x_1+x_2+....+x_n (with a cumulative distribution function F_n) whose sum is equal to x. Normal distribution, Cauchy distribution and all other members of the stable distribution family, Poisson distribution, negative binomial distribution, exponential distribution, geometric distribution, Gamma distribution and degenerate distribution are examples of infinitely divisible distributions. The uniform distribution and binomial distribution are not infinitely divisible. Also if a given characteristic function F can be represented as the n^{th} power of some other characteristic function for every integer n, both F and the corresponding density function are called infinitely divisible [3, 4].

First few central moments and their interpretations

- The zeroth central moment is one.
- The first central moment is zero.
- The second central moment is variance; the square root of it represents the standard deviation.
- The third central moment represents skewness
- The fourth central moment represents kurtosis

Pareto Levy Relationship with Maslov paper [2]

- Histogram of short time lag increments of market price, generated by the Maslov model has a Gaussian shape with sharp maximum and broad wings (i.e., high data concentration in extreme values). So according to the current consensus of this peculiar distribution, up to a certain level it shows the characteristics of Pareto-Levy distribution, with a power law exponent of $1 + \alpha_1 \sim 2.4 - 2.7$, and then it crosses over either to a steeper power law with an exponent $1 + \alpha_2 \sim 3.7 - 4.3$ or to an exponential decay. In both cases this crossover ensures a finite variance (second moment) of the distribution [2].
- Maslov analyzed the histogram of price increments measured within time lags of 1, 10, and 100. The overall shape of these histograms is strongly non Gaussian and it is very close to the shape of real stock prices. As the lag increases the sharp maximum peak of the histogram gradually softens (close to Gaussian), while the wings remain strongly non Gaussian. Also his analysis on the log-log plot of histogram with lag 1 for data collected during 3.5 * 10^7 time stamps shows the log-log plot has two distinguishable power law regions separated by a large crossover around the increment approximately equal to 1 due to some unknown reason. The exponents of these two regions are measured to be $1 + \alpha_1 \sim 0.6 \pm 0.1$ and $1 + \alpha_2 \sim 3 \pm 0.2$. A similar crossover of two power law regions was reported in real stock price fluctuations in NYSE with the exponents $1+\alpha_1 \sim 1.4-1.7$ and $1 + \alpha_2 \sim 4 - 4.5$. The power law exponent of far tail $1 + \alpha = 3$ stays right at the borderline, separating the Pareto-Levy region with power law exponent 1 + α <3, where the distribution has infinite second moment(variance). And also he does not expect a convergence of a price fluctuation distribution to a universal Pareto-Levy or Gaussian as lag is increased [2].

```
/* Probability Distribution Function of Simple Power-Law
   Distribution:                                                    */
```

$$F(x) = 1 - (\gamma/x)^{\alpha} \tag{A.8}$$

Where α is power exponent and γ is cut-off (x ¿ γ).

```
/* Probability Density Function of Simple Power-Law Distribution:
   */
```

$$f(x) = (\alpha\gamma^{\alpha})/x^{\alpha+1}, \alpha > 0, \gamma > x \tag{A.9}$$

$$Mean = \alpha\gamma/(\alpha - 1), \alpha > 1 \tag{A.10}$$

$$Mode = \gamma \tag{A.11}$$

$$Variance = \gamma^2\alpha/(\alpha - 1)^2(\alpha - 2), \alpha > 2 \tag{A.12}$$

Algorithm 6: Pareto-Levy Distribution

Bibliography

[1] F. Slaninaa. Critical comparison of several order-book models for stock-market fluctuations. *THE EUROPEAN PHYSICAL JOURNAL B*, 61:225240, 2008.

[2] S. Maslov. Simple model of a limit order-driven market. *Physica A*, 278:571–578, 2000.

[3] S. Alan and O.J. Keith. *Kendalls Advanced Theory of Statistics*. C. Griffin, Pennsylvania State University, sixth edition edition, 1992. 151-152.

[4] D. Anirban. *Asymptotic Theory of Statistics and Probability*. Springer, Department Of Statistics, Purdue University, first edition edition, 2008. 71-72.

[5] R. Eduardo A.R. Jose, A. Jesus and F.A. Guillermo. Time-varying hurst exponent for us stock markets. *Physica A*, 387:6159–6169, 2008.

[6] C. Lento. A synthesis of technical analysis and fractal geometry - evidence from the dow jones industrial average components. *Working Paper*, pages 1–42, 2009.

[7] F. Lillo and J. D. Farmer. The long memory of the efficient market. *Quantitative Finance Papers*, 8(cond-mat/0311053):1–5, 2004.

[8] E. E. Peters. *Chaos and order in the capital markets: a new view of cycles, prices, and market volatility*. Wiley, New York, first edition edition, 1991.

[9] R.S. Cosma C. Aaron and M.E.J. Newman. Power-law distribution in empirical data. *SIAM*, Review 51:661–703, 2009.

[10] H.F. Coronel-Brizio and A.R. Hernandez-Montoya. On fitting the pareto-levy distribution to stock market index data: selecting a suitable cut-off value. *Physica A*, 354:437–449, 2005.

[11] H.S.M.N. Abu W.C. Chin and I. Zaidi. A simple power-law tail estimation of financial stock return. *Sains Malaysiana*, 38(5):745749, 2009.

[12] M.A. Stephens. Goodness of fit for the extreme value distribution. *Biometrika*, 64:583–588, 1977.

[13] M.P. Larry R. Mezbahur and C.H. Herbert. A modified anderson-darling test for uniformity. *Bull. Malays. Math. Sci. Soc*, 2(29):11–16, 2006.

[14] M.M. Mark B.A. Inmaculada and K.P. Anna. Parameter estimation for the truncated pareto distribution. *Journal of the American Statistical Assoc*, 101:6159–6169, 2006.

[15] J.C.W Rayner and D.J. Best. *Smooth Tests of Goodness of Fit*. OXFORD UNIVERSITY PRESS, New York Oxford, 1989.

[16] S. Hokky and S. Yohanes. Statistical facts of artificial stock market, comparison with indonesian empirical data. *Finance*, 1(0408004), 2004.

[17] H.F. Coronel-Brizio anb A.R. Hernandez-Montoya. On fitting the pareto-levy distribution to stock market index data: Selecting a suitable cut-off value. *Physica A*, 354:437–449, 2005.

[18] G. Sneh and S. Samuel. *Goodness-of-Fit Tests for Pareto Distribution*. Birkhuser Boston, statistical models and methods for biomedical and technical systems edition, 2008. Statistics for Industry and Technology.

[19] V. Choulakian and M.A. Stephens. Goodness-of-fit tests for the generalized pareto distribution. *Technometrics*, Fill.

[20] C.R. de la Cruz-Laso H.F. Coronel-Brizio and A.R. Hernandez-Montoya. Fitting the power-law distribution to the mexican stock market index data. *Quantitative Finance Papers*, 1(cond-mat/0303568), 2003.

[21] J. Aldrich. R. a. fisher and the making of maximum likelihood 1912-1922. *Statistical Science*, 12(3):162–176, 1997.

[22] E.P.B George and M.J Gwilym. *Time Series Analysis ,forecasting and control*. HOLDEN-DAY, 4432 Telegraph Avenue,Oakland,California, revised edition, 1976.

[23] L.B. Bruce and T.O. Richard. *Forcasting and Time Series :An Applied Approach*. Duxbury Press, Belmount California, third edition edition, 1993.

[24] M.F.M. Osborne. Brownian motion in the stock market. *Operations Research*, 10:345–379, 1959.

[25] R. Cont. Volatility clustering in financial markets: Empirical facts and agent-based models. *Working Paper*, 2005.

[26] L. Cabral. Black swans, fat tails, and movie revenues. *Technical Note*, pages 1–2, 2009.