DEGREE PROJECT IN COMPUTER ENGINEERING,
FIRST CYCLE, 15 CREDITS
STOCKHOLM, SWEDEN 2018

Monte Carlo Simulations of Stock Prices

Modelling the probability of future stock returns

TOBIAS BRODD

ADRIAN DJERF

KTH ROYAL INSTITUTE OF TECHNOLOGY
SCHOOL OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

Monte Carlo-simuleringar av aktiekurser

Sannolikhetsmodellering av framtida aktiekurser

Tobias Brodd, Adrian Djerf

Examensarbete inom Datateknik DD142X
Handledare: Arvind Kumar
Examinator: Örjan Ekeberg

CSC KTH 6 Juni 2018

Abstract

The financial market is a stochastic and complex system that is challenging to model. It is crucial for investors to be able to model the probability of possible outcomes of financial investments and financing decisions in order to produce fruitful and productive investments.

This study investigates how Monte Carlo simulations of random walks can be used to model the probability of future stock returns and how the simulations can be improved in order to provide better accuracy. The implemented method uses a mathematical model called Geometric Brownian Motion (GBM) in order to simulate stock prices. Ten Swedish large-cap stocks were used as a data set for the simulations, which in turn were conducted in time periods of 1 month, 3 months, 6 months, 9 months and 12 months.

The two main parameters which determine the outcome of the simulations are the mean return of a stock and the standard deviation of historical returns. When these parameters were calculated without weights the method proved to be of no statistical significance. The method improved and thereby proved to be statistically significant for predictions for a 1 month time period when the parameters instead were weighted.

By varying the assumptions regarding price distribution with respect to the size of the current time period and using other weights, the method could possibly prove to be more accurate than what this study suggests. Monte Carlo simulations seem to have the potential to become a powerful tool that can expand our abilities to predict and model stock prices.

Sammanfattning

Den finansiella marknaden är ett stokastiskt och komplext system som är svårt att modellera. Det är angeläget för investerare att kunna modellera sannolikheten för möjliga utfall av finansiella investeringar och beslut för att kunna producera fruktfulla och produktiva investeringar.

Den här studien undersöker hur Monte Carlo-simuleringar av så kallade random walks kan användas för att modellera sannolikheten för framtida aktieavkastningar, och hur simuleringarna kan förbättras för att ge bättre precision. Den implementerade metoden använder den matematiska modellen Geometric Brownian Motion (GBM) för att simulera aktiepriser. Tio svenska large-cap aktier valdes ut som data för simuleringarna, som sedan gjordes för tidsperioderna 1 månad, 3 månader, 6 månader, 9 månader och 12 månader.

Huvudparametrarna som bestämmer utfallet av simuleringarna är medelvärdet av avkastningarna för en aktie samt standardavvikelsen av de historiska avkastningarna. När dessa parametrar beräknades utan viktning gav metoden ingen statistisk signifikans. Metoden förbättrades och gav då statistisk signifikans på en 1 månadsperiod när parametrarna istället var viktade.

Metoden skulle kunna visa sig ha högre precision än vad den här studien föreslår. Det är möjligt att till exempel variera antagandena angående prisernas fördelning med avseende på storleken av den nuvarande tidsperioden, och genom att använda andra vikter. Monte Carlo-simuleringar har därför potentialen att utvecklas till ett kraftfullt verktyg som kan öka vår förmåga att modellera och förutse aktiekurser.

Contents

1 Introduction **1**
1.1 Relevance of the Study 1
1.2 Problem Statement 2
1.3 Previous Research 2
1.4 Scope of Study 2
1.5 Disposition of the Report 3

2 Background **4**
2.1 The Stock Market and Investment Strategies 4
2.2 Monte Carlo Simulations (MCS) 4
2.2.1 Calculation of π Using the Monte Carlo Method . . . 5
2.3 Simulations of stock prices 6
2.4 Geometric Brownian Motion (GBM) 8
2.5 Mean and Standard Deviation 8
2.5.1 Calculations using Non-Weighted Mean and Standard Deviation 9
2.5.2 Calculations using Weighted Mean and Standard Deviation 9

3 Method **10**
3.1 Data Selection 10
3.2 Testing Environment 12
3.3 Methodology 12
3.3.1 Simulating Stock Prices 12
3.3.2 Interpreting the Simulations 13
3.3.3 Analysing the Simulations 13

4 Results **14**
4.1 Simulations Using Non-Weighted Mean and Standard Deviation 14
4.2 Simulations Using Weighted Mean and Standard Deviation . 18
4.3 Significance Tests 20
4.3.1 Significance Tests for each Time Period 20
4.3.2 Significance Tests for each Stock 21

5 Discussion **21**
5.1 Analysis of the Distributions 21
5.2 Analysis of the Accuracy 22
5.3 Analysis of the Significance Tests 24
5.4 Future research 24

6 Conclusion **25**

1 Introduction

The financial market is a prime example of a complex stochastic system [8]. This means that the market is composed of several different components which all operate on a limited amount of information, or in other words a high level of uncertainty and randomness. These components are for example the companies on the stock market, the investors who trade shares of the stocks and the current state of the economic cycle. Other factors that also influence the market is the ongoing speculation conducted by ordinary citizens, the interest rate for loans (which generally determines how much people spend), natural resources (e.g. oil) and also major political events (such as elections) [16]. These numerous factors effect one another, which makes the financial market a very complex system.

1.1 Relevance of the Study

In the corporate world it is of great interest to construct models which predict and forecast future outcomes of for example stock prices. Fund managers, for example, often handle large amounts of money that belongs to both large corporations as well as to individuals. To handle the client's savings in a responsible and hopefully fruitful way, investors need to use models that simulate prices, options and derivatives, in order to make well-funded investment and financing decisions [15].

Since the research on stock market prediction and probability modelling can result in new knowledge which enables corporations as well as individual investors to make a sustainable profit, there is a natural incentive to conduct research in the field.

The stochastic and complex nature of the financial market is also appealing to more theoretically focused scientists. Before computers were available for commercial use, trading stocks was done in person at physical exchanges (such as the New York Stock exchange and the Chicago Board Option Exchange). The most reliable information to base trades upon was knowledge of the companies on the market [11]. Today's trading is completely different. Nowadays, mathematical methods and algorithms are developed to create a prediction of future market outcomes. These models are then developed into actual computer programs that can execute transactions at a very high speed, so that the information reaches the investor as soon as possible. Therefore, subjects such as numerical methods, algorithm design, big-data analysis and probability theory are all relevant in the field called computational finance [21].

1.2 Problem Statement

This report will investigate a method of probability modelling using a mathematical approach called Monte Carlo simulation. The aim of the report is to answer the following questions:

- How can Monte Carlo simulations of random walks be used to model the probability of future stock returns?
- What is the accuracy of this simulation method?
- How can the method for simulating stock prices be modified in order to improve the accuracy?

1.3 Previous Research

The Monte Carlo method is used when simulating systems that operate on a high level of uncertainty. Application areas of the method include physical sciences (e.g. molecular systems), computational biology (e.g. studying biological systems involving proteins) and artificial intelligence for games (e.g. determining the best move in a game) [13].

In the finance and business sector the method is mainly used for evaluating risks. Business risk analysts can factor in parameters such as levels of sale, interest rates and change of tax laws to evaluate possible future business scenarios. The method is also used for pricing options, a financial instrument with multiple sources of uncertainty [12].

The method does not seem to be used to any larger extent for modelling the probability of stock returns. There exists some more informal work on the topic (conducted by for example hobbyist investors), but not many published scientific papers. Therefore, the research in this report will build upon the knowledge obtained from other fields than stock market modelling (such as the ones mentioned above) and apply this knowledge to a relatively unexplored research field.

1.4 Scope of Study

There exists many ways of using Monte Carlo simulations to model stochastic variables. This study will focus on one of the more basic variations and will only use the expected level of return and the volatility (calculated based on historical data) as input parameters. The modifications in order to improve the simulation will therefore mainly focus on calculating these parameters in a way that better represents the data.

To analyse the performance of our simulation, we will run tests on ten Swedish large-cap stocks. Large-cap stocks refers to stocks with a high level of capitalisation [17]. It might be more optimal to run the simulations on a larger data set consisting of more diverse stocks, but this would take a considerable amount of time and resources, which we do not have when conducting our research.

The simulations depend quite heavily on processing power. Since we only have access to personal computers with limited processing power in this project, we will not run our simulations hundreds of millions of times. The limit will therefore be set to 100 000. This limitation will make our research more feasible, since the simulations can be run on our own computers, but might provide a slight inaccuracy (compared to a study conducted on supercomputers).

1.5 Disposition of the Report

Section 1 (Introduction) of the report introduces the field of stock market analysis, the report's problem statement and discusses the scope of our study.

Section 2 (Background) will present the stock market and investments concepts and introduce relevant theory behind Monte Carlo simulations (covering topic such as random walks and Brownian motion).

Section 3 (Method) will present how we have conducted our research of Monte Carlo simulations. This includes a discussion of how data was selected, how software was developed to run simulations, how we determined the accuracy of a simulation and how we made improvements to the most basic form of simulation.

Section 4 (Results) will present the probabilities for the stocksa and the accuracy of the probability models for both the basic form of Monte Carlo simulations and the modified implementation.

Section 5 (Discussion) will discuss the methods used in this report, as well as the obtained results. The discussion will also reflect upon what factors make a prediction/model of probabilities trustworthy and accurate.

Section 6 (Conclusion) will briefly present the main findings and conclusions of this study.

2 Background

2.1 The Stock Market and Investment Strategies

Most corporations start out as small businesses with a minimal capital. To be able to further expand the business, a corporation generally needs to raise capital. The most common way of raising capital is by selling shares of the business to investors, in order to use the acquired capital to finance growth in some way. These ”shares” of the company are referred to as stocks. The stock market is essentially a large auction where shares of different companies are being sold to investors [23].

Due to factors such as greed, fear and expectations, a share in a company can sell for far more, or for far less, than the company's actual intrinsic value [19]. Investors therefore strive to allocate money in a stock when the price is low, and to sell the stock when the price is high, in order to make a profit. However, these investments in the stock market are not risk-free. If an investor for example misjudges a company's ability to grow, then it is possible that the stock price has decreased since the initial allocation, and therefore results in a loss.

In order to increase chances of making profitable investments, it is necessary to thoroughly analyse potential stocks using some sort of method. There are mainly two different methods for investment analysis: fundamental analysis and technical analysis. Fundamental analysis is based on companies financial statements and technical analysis exclusively look at historic prices and numbers using mathematical models. However, they ultimately serve the same purpose, to indicate whether a stock is worth investing in or not [2]. Analysing random walks is a third method for predicting future stock returns. This third method is what is analysed in this study [9].

2.2 Monte Carlo Simulations (MCS)

The terms Monte Carlo simulation, Monte Carlo method and Monte Carlo experiment refers to a wide range of computational algorithms that utilise randomness in some way. Some problems, which contain a high degree of uncertainty (e.g. several stochastic factors or variables), might be impossible to solve accurately in a algebraic or numerical manner, but by using randomness, it is possible to obtain a good enough approximation, or to at least model the probabilities of the problems outcomes [18].

A Monte Carlo simulation usually follow these steps:

1. A domain of possible inputs is defined

2. Random values are generated from a probability distribution (which depends on the domain of possible inputs)
3. Computation or analysis is performed on the generated values
4. A conclusion can be drawn

A classic example to showcase the use of randomness is the approximation of the mathematical constant $\pi = 3.141592...$ using Monte Carlo simulations.

2.2.1 Calculation of π Using the Monte Carlo Method

We assume a circle with a diameter $d = 1$ (and a radius $r = \frac{1}{2}$), enclosed by a square with side length 1. We know that the area A of the circle is:

$$A = r^2\pi = \frac{1}{2} \cdot \frac{1}{2}\pi = \frac{\pi}{4} \tag{1}$$

We now sample random coordinates from the normal distribution $N(0, 1)$. The points can vary from (0, 0) to (1, 1). A number of points will fall inside of the circle, whilst other points will fall outside of the circle. We keep track of the number of points inside of the circle N_{inner}, and of the total number of points N_{total}.

If we divide the number of points that fall inside of the circle by the total number of points, we get an approximation of the circle's area:

$$\frac{N_{inner}}{N_{total}} \approx \frac{Area \quad of \quad Circle}{Area \quad of \quad Square} = \frac{A}{1} = A \tag{2}$$

Therefore, if we combine equation (1) and (2), we obtain an approximation of π:

$$\frac{\pi}{4} \approx \frac{N_{inner}}{N_{total}} \Leftrightarrow \pi \approx \frac{N_{inner}}{N_{total}} \cdot 4$$

For example, in Figure 1 we have 786 sampled points inside of the circle (red) out of 1000 sampled points (black + red) which results in:

$$\pi \approx \frac{N_{inner}}{N_{total}} \cdot 4 = \frac{786}{1000} \cdot 4 = 3.144$$

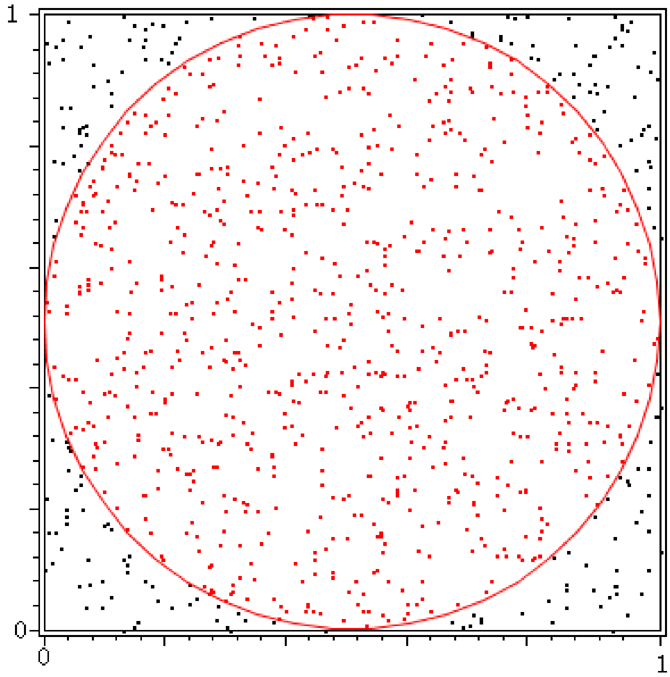

Figure 1: A total of 1000 sampled points (black + red). 786 fell inside of the circle (red), while the other 214 fell outside of the circle (black).

This example [1] shows the power of sampling a large number of random values within a certain domain, and using the result to make some sort of approximation. If we extend the number of generated points in the example above, to let's say 1 000 000, then the result would be significantly more accurate. The constant π can be calculated "infinitely" precise by the perfect circle's circumference and diameter. Other more complex problems on the other hand, such as stochastic processes, can't be calculated just as an expression of numbers. In those problems the power of randomness can be truly useful in order to obtain a good enough approximation of the problem at hand.

2.3 Simulations of stock prices

In our case we want to use Monte Carlo simulations to determine the probability that a stock increases or decreases in value. We will assume that a stock price gives daily returns that follow a normal distribution (a continuous probability distribution).

The first thing which needs to be done is to estimate the expected level of return, and the volatility of the stock. We assume that we have access to an array of historical prices. The expected level of return can then be estimated by calculating the mean of the historical returns. The volatility can be estimated by calculating the standard deviation of the historical returns. From the expected level of return and from the volatility of the stock we can set up a probability distribution which attempts to model the behaviour of the stock [10].

The simulation itself is the act of selecting random numbers out of this probability distribution function. By sampling say ten values from the distribution, we get a sense of how the stock could potentially behave within the next ten days. However, just one simulation won't really give us an insight in the probability of the stocks future returns. The likelihood of the stock following just one single random simulation is close to zero [10].

The real insight in the stock and the possible outcomes of the future is gained when running thousands of simulations, by creating thousands of random price curves, which all are different, but at the same time share some of the key characteristics of the historical price data. These are so called random walks. It is believed that stocks follow a random walk, which implies that simulating a stock using random variables may work [7].

As an example of what the simulations could look like, see Figure 2. It showcases 100 simulations of the Ericsson B (ERIC_B) stock on a period of 12 months.

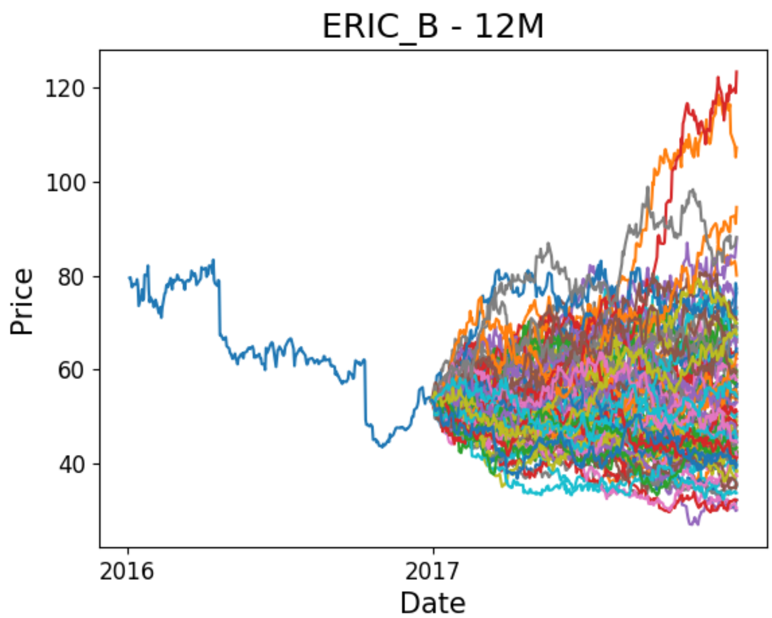

Figure 2: 100 simulations of a stock price 12 months into the future.

2.4 Geometric Brownian Motion (GBM)

Geometric Brownian Motion, or GBM for short, is a continuous-time stochastic process that satisfies the Stochastic Differential Equation (SDE):

$$dS_t = \mu S_t dt + \sigma S_t dB_t$$

where S_t is a stochastic process and B_t is a Brownian motion (a Wiener process) characterised by the following properties:

- $B_0 = 0$
- B has both stationary and independent increments
- B has Gaussian increments

What the SDE essentially means in this study is that each increment in time results in the price of the stock moving with a drift ($\mu S_t dt$) and a shock ($\sigma S_t dB_t$). The drift can be seen as the general direction of the stock's price whereas the shock is a random amount of volatility that acts on the stock's price. The shock is what will create the curve's noise or jaggedness.

To get the formula for the GBM we must find a solution to the SDE. It turns out that the SDE has an analytic solution under Itō's interpretation, unlike other SDE:s, defined as:

$$S_t = S_0 e^{(\mu - \frac{\sigma^2}{2})t + \sigma B_t}$$

This is what in this study is used to model the prices of a stock. We will however see the stochastic process S_t as the price of a stock S at time t and B_t as normal distributed variable with $\mu = 0$ and $\sigma = 1$ [22, 20].

2.5 Mean and Standard Deviation

The mean value and standard deviation are two common variables used in statistics. The mean value tells us how big the capacitance is "in average" while the standard deviation tells us how much the capacitance differ [4]. There exists two widely used methods for calculating mean and standard deviation, the first using non-weighted values and the second using weighted values. These two methods will be used in this study and are therefore described here in the following subsections.

2.5.1 Calculations using Non-Weighted Mean and Standard Deviation

Calculating the non-weighted mean (μ) and non-weighted standard deviation (σ) is very common and can be described by the following formulas:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$$

where x_i are samples that we want to calculate the mean/standard deviation for [4].

2.5.2 Calculations using Weighted Mean and Standard Deviation

Calculating the weighted mean ($\hat{\mu}$) and weighted standard deviation ($\hat{\sigma}$) is a bit more complicated, but can be described by the following formulas:

$$\hat{\mu} = \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} w_i}{(\sum_{i=1}^{N} w_i)^2 - \sum_{i=1}^{N} (w_i)^2} \sum_{i=1}^{N} w_i (x_i - \mu)^2$$

where x_i are samples that we want to calculate the mean/standard deviation for. The weights w_i can be calculated using the formula:

$$w_i = (1 - \alpha)^i$$

where α is a constant [14].

3 Method

3.1 Data Selection

The examined data consists of 10 Swedish stocks from the OMXS30 index listed on the NASDAQ Nordic exchange. The stocks were manually chosen based on their total volume and their trends. For this study, only high volume stocks with a variety of positive and negative trends were selected. Volume was chosen as preference because we only wanted to analyse large-cap stocks. A variety of trends was preferred because we wanted to know whether the simulations work on both positive and negative trending stocks or not. The chosen stocks are shown in Figure 3.

The data sets were downloaded directly from NASDAQ Nordic and have a span of 6 years, 2012-01-01 to 2017-12-31. This span was chosen to allow 5 years of historical prices to be used for calculating statistical values. The last year of historical prices was used as a reference for comparing the simulated and actual prices. The span includes daily data and since not all days are banking days, only 1505 days of data were downloaded. To account for company actions such as stock splits, dividends and distributions, only adjusted closing prices were used.

The charts for the ten analysed stocks are showcased on the next page.

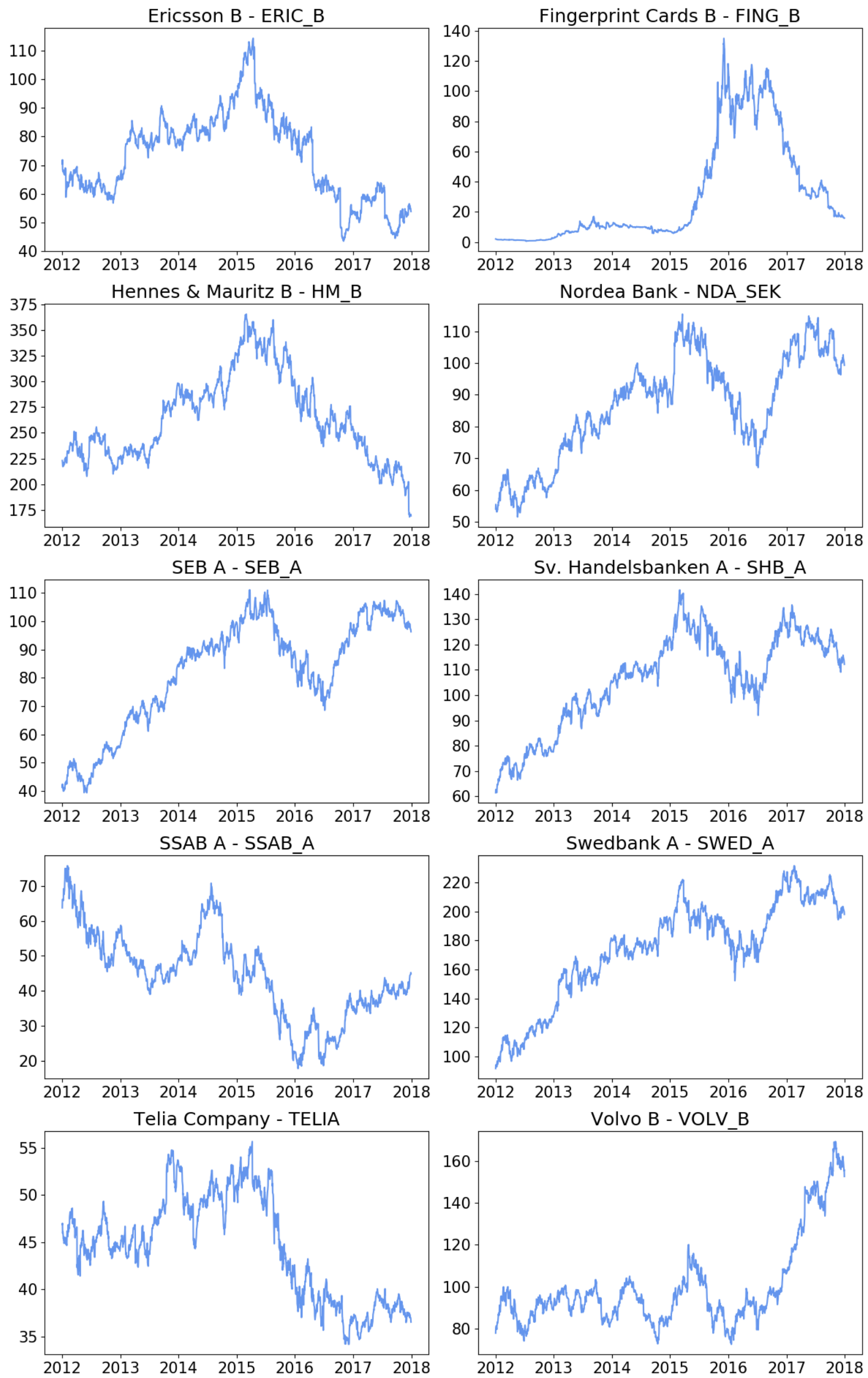

Figure 3: The analysed stocks. Prices are on the y-axes and years are on the x-axes.

3.2 Testing Environment

The simulations were run on a MacBook Pro with an Intel Core i7 processor (clock speed 3.3 GHz) and 16 GB (2133 MHz) RAM running macOS High Sierra version 10.13. The simulation code was written in C++ using the C++11 standard. Statistical functions were imported from GSL, GNU Statistical Library, version 2.4. The results were analysed and plotted using Python 3.6.

3.3 Methodology

3.3.1 Simulating Stock Prices

The idea behind simulating future stock prices and modelling the probability of stock prices is to generate thousands of random walks based on a stochastic stock price model. Each random walk follows a Geometric Brownian Motion (GBM) (Section 2.4) using $(\mu - \frac{\sigma^2}{2})dt$ as drift and $\sigma\sqrt{dt}N(0,1)$ as shock. μ is the mean historical return, σ is the standard deviation of historical returns and dt is the time step of the simulation. In this study, μ and σ were calculated using the two methods described in Section 2.5. An α value of $\frac{2}{30+1}$ was used for calculating the weights.

In each step of the Monte Carlo simulation, the next price in the random walk was calculated using the formula:

$$S_i = S_{i-1}e^{(\mu - \frac{\sigma^2}{2})dt + \sigma\sqrt{dt}N(0,1)}$$

where $i = 1, ..., \frac{t}{dt}$, S_0 is given by the latest historical price and $N(0,1)$ is a random value from the normal distribution with mean 0 and standard deviation 1. This step was then performed until $\frac{t}{dt}$ prices were generated. Here, t is the time period that should be simulated. For each simulation, the simulated end price was then added to a list for later use.

Using the described Monte Carlo simulation, 5 different time periods t (represented as days) were simulated. Each time period used a dt value of 1 (1 day), which means that each step in the simulation added a new *daily* price to the random walk. 100 000 simulations were run per time period, which results in 100 000 simulated end prices for each stock S. The time periods all had a start date of 2017-01-02 (first banking day of 2017), but differed in their end date. The following end dates where used:

- 2017-01-31 (1 month)
- 2017-03-31 (3 months)
- 2017-06-30 (6 months)
- 2017-09-30 (9 months)
- 2017-12-31 (12 months)

3.3.2 Interpreting the Simulations

In this study, the simulated prices were assumed to follow a log-normal distribution [3]. This meant that the resulting list of simulated end prices could be analysed using ordinary statistical functions. As one problem statement in this study concerned the accuracy of the used method based on whether a stock will go up or down in the future, it was crucial to calculate the probabilities of these cases. The functions gsl_cdf_lognormal_Q and gsl_cdf_lognormal_P from the GSL library were therefore used for calculating the probabilities. gsl_cdf_lognormal_Pinv was used for calculating the inverse of gsl_cdf_lognormal_P. In the end, the following results were saved for each stock S and time period t:

- S_t
- μ
- $P(X > S_0)$, the probability of drawing a stock price sample above the start price
- $P(X \leq S_0)$, the probability of drawing a stock price sample below the start price
- -1σ percentile, the price for which 15.87% of all prices are below
- $+1\sigma$ percentile, the price for which 84.13% of all prices are below

where S_t is the actual price of the stock for time period t and S_0 is the start price of the stock stimulation.

3.3.3 Analysing the Simulations

The saved results, consisting of results from all simulations, were then analysed using a Python program. The program would classify each simulated stock S at time period t as one of the following four outcomes:

- The resulting probability of the price going up was greater than 50% and the actual end price was larger than the start price. This meant that the positive trend was correctly predicted. This outcome is in this study called true positive (TP).
- The resulting probability of the price going up was less than 50% and the actual end price was greater than the start price. This meant that the positive trend was not correctly predicted. This outcome is in this study called false positive (FP).
- The resulting probability of the price going up was less than 50% and the actual end price was less than the start price. This meant that the

negative trend was correctly predicted. This outcome is in this study called true negative (TN).

- The resulting probability of the price going up was greater than 50% and the actual end price was less than the start price. This meant that the negative trend was not correctly predicted. This outcome is in this study called false negative (FN).

After the classifications were made, the results were grouped by classification and then plotted. The distribution for each stock and time period was also plotted. Three lines, showing the price for the -1σ percentile, $+1\sigma$ percentile and the actual end price, were also added to each distribution plot.

4 Results

Because two methods of calculating the mean value and standard deviation were used in this study, this section is divided into two subsections.

4.1 Simulations Using Non-Weighted Mean and Standard Deviation

For the following results, the method using non-weighted values for calculating mean and standard deviation was used.

The results are showcased on the next page.

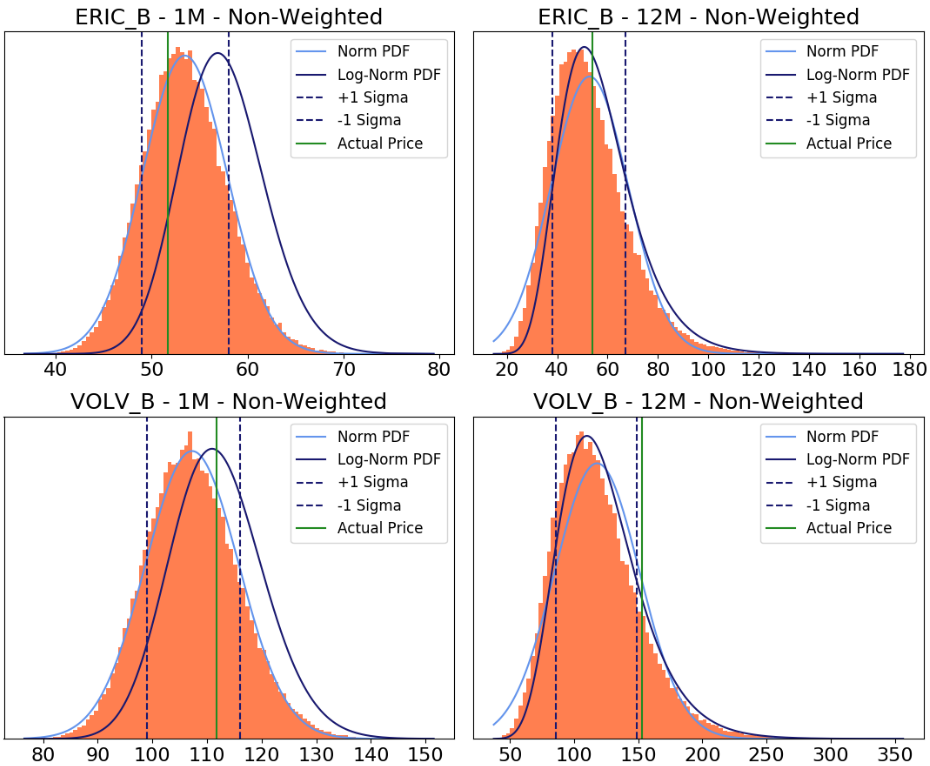

Figure 4: The distribution of the simulated end prices for Ericsson B and Volvo B using non-weighted parameters. End prices are on the x-axes.

Figure 4 shows the distributions of the stock prices for the time periods 1 month and 12 months. The plotted curves describe the normal and the log-normal probability density functions. Worth noting is that for short time periods, the curve for the log-normal distribution did not seem fit the actual data, whereas the normal distribution did. There was also a difference between the stocks themselves about how well the normal and log-normal distribution functions fitted the actual data. For example, the log-normal distribution fitted much better for Fingerprint Cards B than for Ericsson B. The spread between the -1σ and $+1\sigma$ percentiles also increased with time. This meant that the price interval which $\sim 70\%$ of the simulated end prices belongs, increased with time.

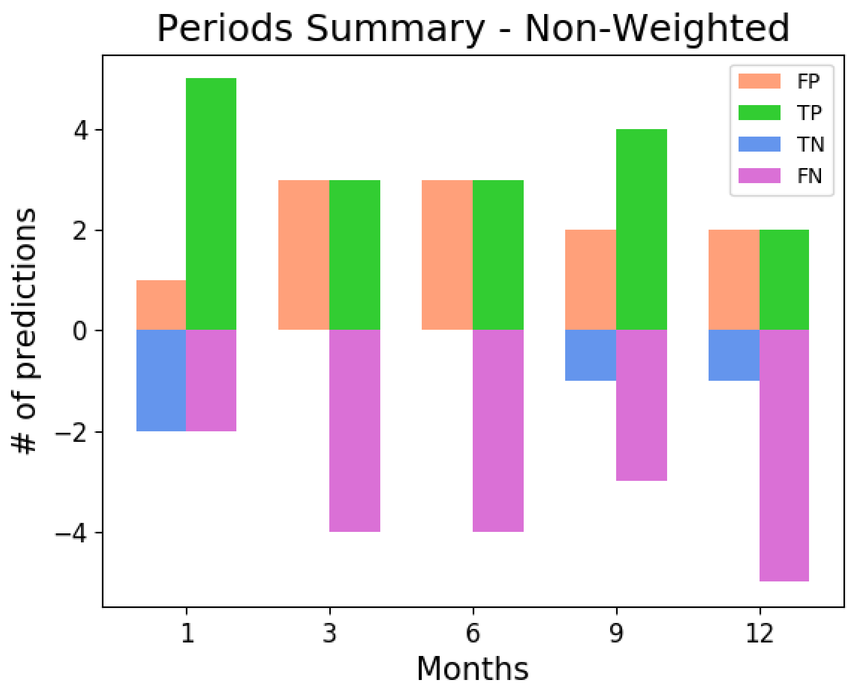

Figure 5: Predictions using non-weighted parameters for each analysed time period. TP and TN together form the set of correct predictions.

Figure 5 shows the number of true positives, false positives, true negatives and false negatives for each simulated time period. True positives and true negatives together form the set of correct predictions. This meant that for the 1 month forecast, our method got $5 + 2 = 7$ out of 10 correct, which is a 70% accuracy. For the following two time periods the accuracy of our method was 30%. The accuracy was slightly improved for the 9 months period, which got an accuracy of 50%. For the 12 months period, an accuracy of only 30% was achieved. Another thing worth noting is that the analysed stocks' actual trends were distributed 56% up and 44% down, which is almost a uniform distribution whereas there were substantially more true positive than true negatives in the results.

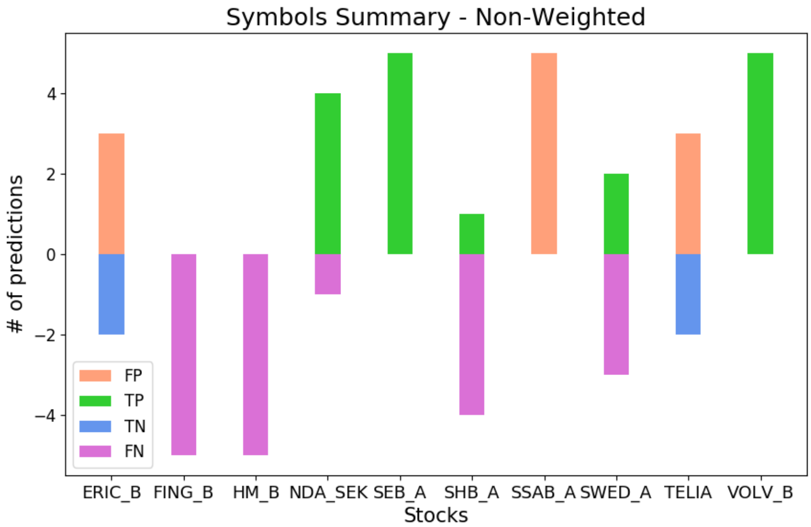

Figure 6: Predictions using non-weighted parameters for each analysed stock. TP and TN together form the set of correct predictions.

Figure 6 shows that while there were no time period for which the method had 100% correct predictions, there were two stocks that the method got 100% correct, namely SEB A and Volvo B. Both of these stocks had a positive trend for all time periods. There were also three stocks, Fingerprint Cards B, Hennes & Mauritz B and SSAB A, that the method did not have any correct predictions for. Fingerprint Cards B and Hennes & Mauritz B both had a negative trend for all time periods while SSAB A had a positive trend for all time periods. There was also one stock where the method predicted correct in 80% of all cases, namely Nordea Bank. The complete opposite were true for Handelsbanken A for which the method had 80% incorrect predictions.

4.2 Simulations Using Weighted Mean and Standard Deviation

For the following results, the method using weighted values for calculating mean and standard deviation was used.

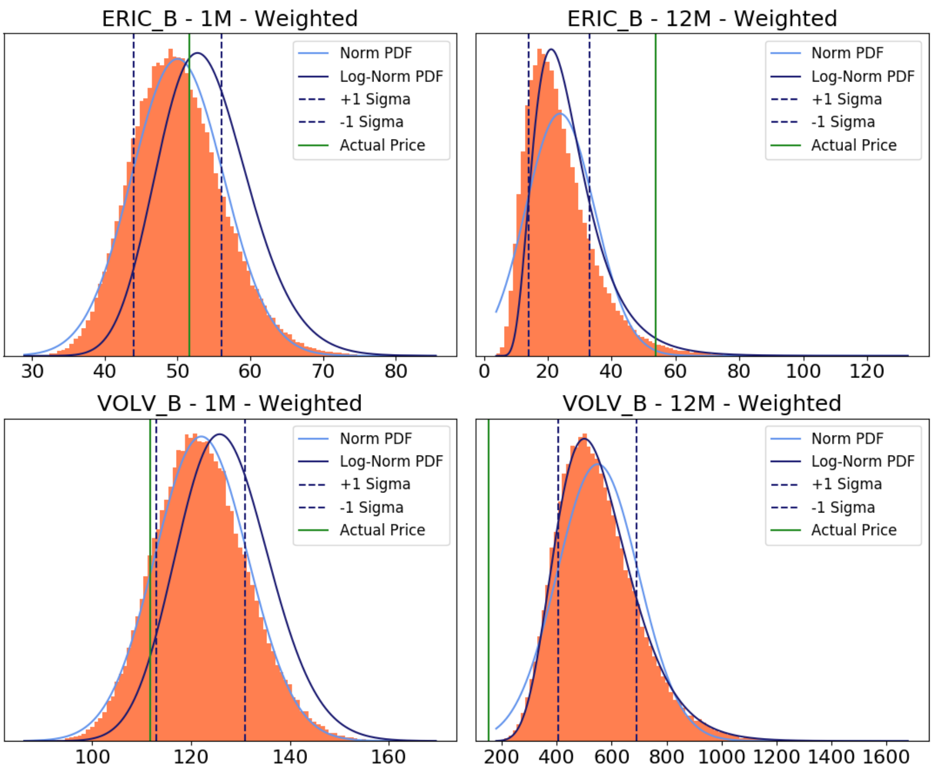

Figure 7: The distribution of the simulated end prices for Ericsson B and Volvo B using weighted parameters. End prices are on the x-axes.

As in Section 4.1, Figure 4 shows the distributions of the stock prices for the time periods 1 month and 12 months. The plotted curves describe the normal and the log-normal probability density functions. Here the log-normal distribution seemed to better fit the actual data than in Section 4.1. The spread also seemed better as most graphs showed a smaller or equally small interval to which $\sim 70\%$ of the end prices belonged.

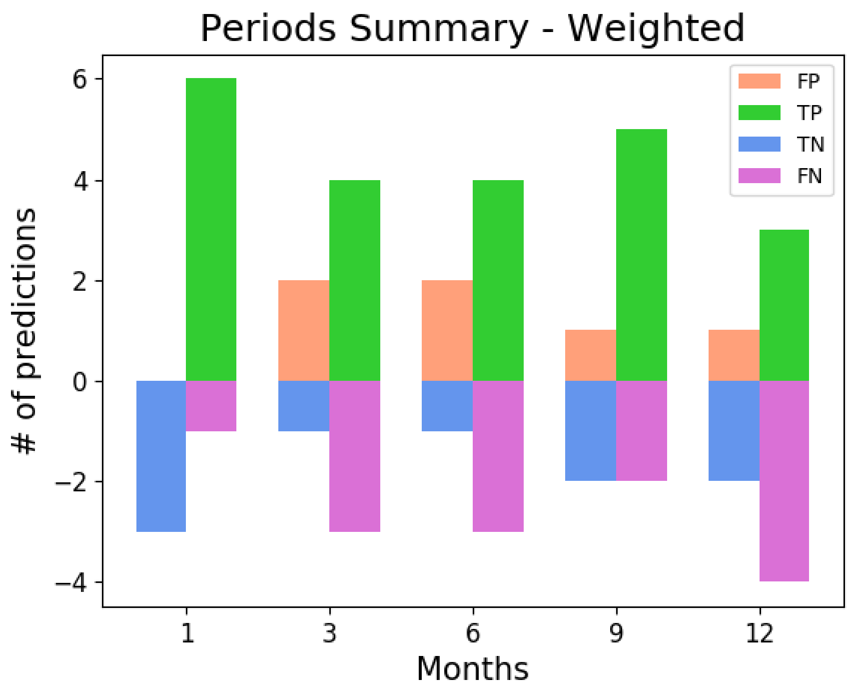

Figure 8: Predictions using weighted parameters for each analysed time period. TP and TN together form the set of correct predictions.

As in Section 4.1, Figure 8 shows the number of true positives, false positives, true negatives and false negatives for each simulated time period. True positives and true negatives together form the set of correct predictions. Compared to the previous section, the results were improved for all time periods. Forecasting 1 month now got a 90% accuracy, compared to the 70% before. For the 3 months and the 6 months periods, the results were 50%. The 9 months period saw an increase from 50% to 70% and the 12 months period increased from 30% to 50%. Using a weighted mean and standard deviation the lowest accuracy was increased from 30% to 50% and the highest accuracy from 70% to 90%.

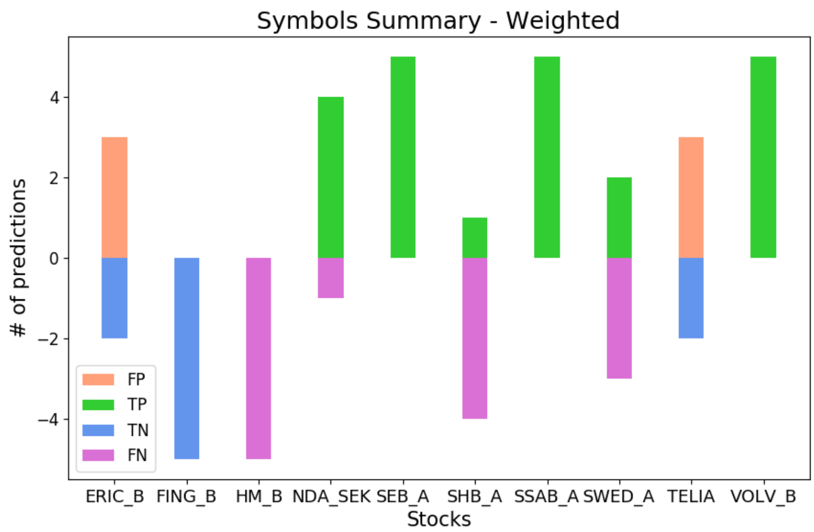

Figure 9: Predictions using weighted parameters for each analysed stock. TP and TN together form the set of correct predictions.

Figure 9 shows that there were four stocks that the method got 100% correct, Fingerprint Cards B, SEB A, SSAB A and Volvo B. Three out of these four stocks had a positive trend for all time periods while one had a negative trend for all time periods instead. There was also one stock, Hennes & Mauritz B, that the method did not have any correct predictions for. Hennes & Mauritz B had a negative trend for all time periods and the method did not correctly predict the negative trend for any of the time periods. The rest of the results were otherwise the same as in Figure 6.

4.3 Significance Tests

All significance tests were performed by following the examples in Blom [5].

4.3.1 Significance Tests for each Time Period

To show whether our results are significant or not, a significance test were performed for each time period. The hypothesis used for the tests was:

H_0: The predictions are random.
H_a: The predictions are not random.

To test the null hypothesis, the Cumulative Density Function (CDF) was calculated for a binomial distribution using $p = \frac{1}{2}$ and $n = 10$. p is the probability and n, the number of stocks, is 10. The significance level, α, was set to 5% which is a common level used in the scientific field [5]. What this essentially means is that we calculate the probability of our results being drawn from the binomial distribution. If this probability isn't likely to happen, which is determined by the α level, we reject our null hypothesis.

For the standard method, no results were significant. To be able to show significance, 80% correct predictions were needed. As our results show, the best case was 70%. For the weighted method, the predictions of the 1 month period were significant. For all other periods, no significance could be proven.

4.3.2 Significance Tests for each Stock

The same hypothesis and test was used in 4.3.1 to test whether there were any significant results for the individual stocks. n was for these tests changed from 10 (the number of stocks) to 5 (the number of time periods). The conclusions that could be made from the tests were that the results for the method that did not use any weighted parameters were significant for Nordea Bank, SEB A and Volvo B. The method using weighted parameters added two more stocks to the list, namely Fingerprint Cards B and SSAB A. In order for the results to prove significant for a stock, 80% correct predictions were needed.

5 Discussion

This report has shown how the Monte Carlo method can be used in a relatively simple way in order to generate possible future outcomes of stock prices. The following sections are going to analyse the distributions, the accuracy of the method and the significance tests.

5.1 Analysis of the Distributions

When analysing the distributions of the stock prices we found that for shorter time periods the curve for the log-normal distribution did not perfectly fit the price distribution, whilst the normal distribution seemed to fit quite well. This is interesting since we assumed that the prices should in fact follow a log-normal distribution. Antoniou et al. [3] says that stock prices often follow a log-normal distribution, but that it may not always be the case and our results seem to indicate that as well. The fact that prices on the shorter time periods do not follow the assumed distribution might contribute to a certain amount of accuracy loss, since our calculations are based upon assumptions which in this case does not seem to hold for all time periods. On the other hand, for the longer time periods the distribution of the prices did fit the assumed log-normal distribution curve. One possible explanation to this might be that prices spread more as time passes, and more closely fall into a log-normal distribution after a large number of days. Perhaps prices at the shorter time periods should be assumed to follow a normal distribution, whilst the prices at the long time periods should be assumed to follow a log-normal distribution to get more accurate results. Our empirical data shows that this assumption could be closer to the "truth" than assuming a log-normal distribution for all time periods.

What is really interesting when looking at for example Figure 4 is that the price spread between the -1σ and $+1\sigma$ percentiles is quite small, just a few SEK for the the 1 month period. We also see that the actual end price is inside the interval (or very near) for most graphs. This seems to indicate that it is possible model a stock's price using Monte Carlo simulations based on the GBM model. This might be true for at least shorter time periods since there is larger spread when forecasting long time periods. The longer time periods therefore results in wide predictions that may not be helpful for an investor. If we instead look at Figure 7 where weighted parameters were used we actually get smaller, but less accurate intervals when the actual end price is taken into account. Using weights seems to have an exaggerating effect on the distribution and therefore also on the price predictions. The same applies for almost all stocks that were analysed. We can therefore conclude that even though weighted values may predict trends better as Figures 8, 9 show, they may result in worse price predictions.

5.2 Analysis of the Accuracy

While the most basic simulation (using a non-weighted mean and standard deviation) was developed and tested, it could clearly be observed that the method was not significant for all stocks. The method did however work well on certain stocks such as SEB A and Volvo B which were correctly predicted for all time periods. After inspecting the individual predictions for these stocks, we could quickly spot a trend. Most of the correctly predicted stocks increased in price for all time periods and were correctly predicted that way too. We also noted that there were substantially more true positives than true negatives in our results as seen in Figure 8, which could indicate a flaw in our method. The method should be equally good or bad at predicting positive and negative trends since the data almost has a uniform distribution. Figure 8 also shows this flaw quite well because almost all stocks with a negative trend were incorrectly predicted.

It can be observed from Figure 8 that the approach using weights was substantially better than not using any weights. The prediction accuracy increased for all time periods which is quite impressive, but not surprising since we believe that more recent data should matter more than data from a long time ago. One interesting thing to note is that when the method did not use any weights, Fingerprint Cards B could not be predicted at all. When adding weights to the parameters we see that the accuracy increased from 0% to 100%. By taking a look at Figure 3 we also see that in order to predict the negative trend for 2017 one would only need to look at the previous few months. If a longer time period than that is taken as input, one would probably predict that Fingerprint Cards B would have a positive trend in 2017 which would be incorrect. We can therefore come to the conclusion that the method should use different weights that depends on the historical prices for the analysed time period to get an improved accuracy. In this study only one set of weights were used. Different weights could (and should) be tested in order to see if the results can be improved even more. In this study we can however conclude that the weights are of significant importance to the result of the simulations.

So what are the trends among the stocks that our method correctly predicted and what conclusions can be drawn from this information? Generally, the accurately predicted stocks had a positive trend throughout all time periods. Because our data was not exactly 50/50 in regards to how many stocks that had a positive/negative trend, an alternative strategy to ours would be to just predict that all stocks will go up for all time periods. As the market generally increases in value each year [6], this might actually be a really good strategy for long term investments. Even though this might work, our method has an edge over that strategy, it can predict negative trends. Stocks that for each period sometimes had a positive trend and sometimes had a negative trend were however much harder to predict. Both Ericsson B and Telia Company were of this type and could not accurately be predicted. More interestingly some stocks seem to have the same structure when looking at their plots. For example, compare SEB A and Handelsbanken A in Figure 3. Both seem to have the same plot, but the results for SEB A were significant and they were not significant for Handelsbanken A. What conclusion can be drawn from this? Well, if we look more closely to the plots, we can see that Handelsbanken A actually had a positive trend for the 1 month period, and then switched direction and had a negative trend for the rest of the periods. This "trend switch" seems to have confused our method and is not something it can handle. Dynamically set parameters could possibly solve this as the mean return calculated on a short time interval would change sign with the trend. It could also be that stocks are not as random as we initially thought. We believe that company actions or actions towards a company may have an impact on the company's stock. Since our method does not consider any fundamental analysis, a lawsuit would for example be nearly impossible to predict, but could have a significant impact on the stock. Our method may therefore prove to be more accurate on instruments that are less coupled to company actions or actions on a company. An index, such as OMXS30, would perhaps be a good instrument to try this theory on as indices consist of several stocks. We believe that for an instrument based on an collection of stocks, company actions could cancel each other out and therefore be more stable and less prone to quick trend switches.

5.3 Analysis of the Significance Tests

In this study we performed significance tests for both time periods and the individual stocks. What we can see from the significance tests for the time periods is that our method's results are not significant in most cases. This could mean that our method probably does not work for all types of stocks and that it is not a general method. If we however look at the individual stocks we see that the results were significant for several stocks. This tells us that for a certain type of stock, it could be possible to use our method to predict future stock prices and trends. It also seems to indicate that the results can change from being significant to not significant and vice verse depending on how the significance tests are calculated. Maybe the used significance test is too simple and should be replace with a more advanced test to better conclude whether the results actually are significant or not.

We can conclude that this study shows that Monte Carlo simulations certainly can be used in the finance field, but that modelling of stock prices might not be the optimal use of this method. The results themselves could however probably be improved by varying the assumed distribution as well as using another set of weights for calculating the mean and standard deviation parameters for the GBM equation.

5.4 Future research

The Monte Carlo method is not really used in today's applications regarding pure stock analysis. This may be due to the fact that the method may have been tried before, and not proven to be superior to for example fundamental analysis or technical analysis (Section 2.1). There is however a large number of further questions which have not yet been answered. How would the quality of the simulations change if the stocks are categorised and treated differently? How would the simulations behave if the parameters were not assumed to be constant, but variables instead? Would it be beneficial to add more parameters to the simulations, such as inflation and current rates? Is the method more accurate of other types of instruments such as indices?

We do believe that this report showcases that the method has potential to be improved and thereby becoming a powerful tool that will expand our abilities to predict and model stock prices.

6 Conclusion

From this study we can conclude that it is possible to model the probability of future stock returns using Monte Carlo simulations with Geometric Brownian Motion as underlying stock price model. The accuracy of such approach may however not be good enough for most investors who want to analyse trends, since no trend predictions for any of the five predicted time periods turned out to be significant using our approach. Even when analysing the stocks individually, only predictions for two out of the ten analysed stocks were significant. The accuracy of trend predictions could however be improved using weights when calculating the parameters for the Geometric Brownian Motion model. Using this approach trend predictions for one forecasting period and five stocks were shown to be significant. This improvement comes at a cost as the distributions for the stocks generally seemed to fit the actual end price less accurately than when non-weighted parameters were used. More research should be conducted to conclude whether the Monte Carlo method can be used for business purposes.

References

[1] Academo. *Estimating Pi using the Monte Carlo Method.* URL: https://academo.org/demos/estimating-pi-monte-carlo/.

[2] Steven B Achelis. *Technical Analysis from A to Z.* McGraw Hill New York, 2001.

[3] I Antoniou et al. "On the log-normal distribution of stock market data". In: *Physica A: Statistical Mechanics and its Applications* 331.3 (2004), pp. 617–638. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2003.09.034. URL: http://www.sciencedirect.com/science/article/pii/S0378437103008987.

[4] Gunnar Blom. *Sannolikhetsteori och statistikteori med tillämpningar.* Vol. 5:14. Studentlitteratur, 2005. Chap. 10.3. ISBN: 978-91-44-02442-4.

[5] Gunnar Blom. *Sannolikhetsteori och statistikteori med tillämpningar.* Vol. 5:14. Studentlitteratur, 2005. Chap. 13.2, 13.3. ISBN: 978-91-44-02442-4.

[6] Rich Habits (Summit of Coin). *WHY DOES THE STOCK MARKET GO UP OVER TIME?* URL: http://www.summitofcoin.com/blog/6/14/2017/why-does-the-stock-market-go-up-over-time.

[7] Godfrey Michael D., Granger Clive W. J., and Morgenstern Oskar. "THE RANDOM-WALK HYPOTHESIS OF STOCK MARKET BEHAVIOR". In: *Kyklos* 17.1 (), pp. 1–30. DOI: 10.1111/j.1467-6435.1964.tb02458.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-6435.1964.tb02458.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-6435.1964.tb02458.x.

[8] Bernard Lapeyre Damien Lamberton. *Introduction to Stochastic Calculus Applied to Finance, Second Edition.* Chapman and Hall/CRC, 2007.

[9] Eugene F Fama. "Random Walks in Stock Market Prices". In: 51 (Jan. 1995).

[10] Jamieson S. (Python For Finance). *Monte Carlo Simulation in Python – Simulating a Random Walk.* URL: http://www.pythonforfinance.net/2016/11/28/monte-carlo-simulation-in-python/.

[11] Oxford Futures. *THE ART OF HAND SIGNALS.* URL: http://www.oxfordfutures.com/docs/Handsignals.pdf.

[12] Paul Glasserman. *Monte Carlo methods in financial engineering.* Vol. 53. Springer Science & Business Media, 2013.

[13] Vassili Kolokoltsov. *Nonlinear Markov processes p. 375.* Cambridge Univ. Press, 2010.

[14] GNU Statistical Library. *Weighted Samples.* URL: https://www.gnu.org/software/gsl/doc/html/statistics.html%5C#weighted-samples.

[15] Cathedral Consulting Group LLC. *Forecasting: The Importance of a Financial Model*. 2013.

[16] Cory Mitchell. *4 Factors That Shape Market Trends*. URL: https://www.investopedia.com/articles/trading/09/what-factors-create-trends.asp.

[17] NASDAQ. *Large Cap*. URL: https://www.nasdaq.com/investing/glossary/l/large-cap.

[18] Department of Physics Indian Institute of Technology Guwahati S. B. Santra. *Monte Carlo Simulation technique*. URL: http://www.iitg.ac.in/nwast2015/slides/sbs.pdf.

[19] Hersh Shefrin. *Beyond greed and fear: Understanding behavioral finance and the psychology of investing*. Oxford University Press on Demand, 2002.

[20] Karl Sigman. "Simulating Brownian motion (BM) and geometric Brownian motion (GBM)". In: (2013). URL: http://www.columbia.edu/~ks20/4404-Sigman/4404-Notes-sim-BM.pdf.

[21] Edward P.K. Tsang and Serafin Martinez-Jaramillo. *Computational Finance*. University of Essex, 2004.

[22] Ward Whitt. "A Quick Introduction to Stochastic Calculus". In: (2007). URL: http://www.columbia.edu/~ww2040/4701Sum07/lec0813.pdf.

[23] Finance Zacks. *Purpose of the Stock Market*. URL: https://finance.zacks.com/purpose-stock-market-2365.htmlp.