Master Thesis
Computer Science
Thesis No: MSC-2009:38.
May 22nd 2009

Brownian Dynamic Simulation to Predict the Stock Market Price

Authors

Ramana Reddy Dappiti
Mohan Krishna Thalluri

Department of Interaction and System Design
Blekinge Institute of Technology
Box 520
SE – 372 25 Ronneby
Sweden

This thesis is submitted to the Department of Interaction and System Design, School of Engineering at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Computer Science. The thesis is equivalent to 20 weeks of full time studies.

Contact Information:
Author(s):

Ramana Reddy Dappiti
E-mail: ramanareddy113@yahoo.co.in

Mohan Krishna Thalluri
E-mail: kittu1412001@yahoo.co.in

University advisor: Stefan J. Johansson
Email: sja@bth.se
Address: Blekinge Institute of Technology, SE-371 79 Karlskrona, Sweden
Phone: +46 455 38 50 00 | Fax: +46 455 38 50 57

Department of
Interaction and System Design
Blekinge Institute of Technology
Box 520
SE - 372 25 Ronneby
Sweden

Internet : www.bth.se/tek
Phone : +46 457 38 50 00
Fax : + 46 457 102 45

ABSTRACT

Stock Prices have been modeled using a variety of techniques such as neural networks, simple regression based models and so on with limited accuracy. We attempt to use Random Walk method to model movements of stock prices with modifications to account for market sentiment. A simulator has been developed as part of the work to experiment with actual NASDAQ100 stock data and check how the actual stock values compare with the predictions. In cases of short and medium term prediction (1-3 months), the predicted prices are close to the actual values, while for longer term (1 year), the predictions begin to diverge. The Random Walk method has been compared with linear regression, average and last known value across four periods and has that the Random Walk method is no better that the conventional methods as at 95% confidence there is no significant difference between the conventional methods and Random Walk model.

Keywords: Stock Price, Simulation,
Random Walk, NASDAQ100.

PREFACE

We would like to thank the following individuals for their contribution in one way or another to our thesis. Thesis Advisor Mr. Stefan J.Johansson and Thesis Responsible Mr. Guohua Bai

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to one and all for helping me in working on this thesis work. I would also thank my family members whose blessings always have a profound influence over me. I would like to thank my university advisor Stefan Johansson for supervising me on this interesting research topic and for his keen guidance and support throughout the course of this research thesis and in providing a quality report. I convey my regards to you for giving me this opportunity to present my research work on challenging aspects of designing simulator for predicting values. Finally I thank everyone who has directly and indirectly encouraged me with this thesis work

CONTENTS

List Of Figures:

List of Tables:

INTRODUCTION

Brownian motion and Stock Market Prediction

Prediction of stock markets has been the research interest of many scientists around the world. Speculators who wish to make a “quick buck” as well as economists who wish to predict crashes, anyone in the financial industry has an interest in predicting what stock prices are likely to be. Clearly, there is no model which can accurately predict stock prices; else markets would be absolutely perfect! However, the problem is pertinent and any improvement in the accuracy of prediction improves the state of financial markets today. This forms the broad motivation of our study.

Our work focuses on development of a model which improves upon trivial methods such as regression techniques to predict stock prices. Concepts from physics have already been successfully applied to stocks and other derivatives, a prime example being the Black Scholes model [3].Similarly, we attempt to simulate stock price movements and check whether we are able to predict stock market movements over a month’s time. Even if partially successful the simulator thus developed can be handy tool for the financial industry which uses backward looking methods such as simple extrapolation of patterns to predict stock market prices.

We have specifically focussed on using with some modification a Random Walk Model. The concept of Brownian motion and random walk has been used for modelling a variety of physical microscopic processes [6]. As a mathematical model, it has been used to model various processes, including the modelling of markets. In the simplest terms, a "random walk" is essentially a Brownian motion where the previous change in the value of a variable is unrelated to future or past changes [6].

CHAPTER 1: Background

1.1 Modeling Work Described in Literature

Atsalakis and Valavanis (2009) have surveyed the stock market forecasting techniques and show that the main methods used today are artificial neural networks, auto regression and random walk as is shown by the table extracted from the author's work [1]. The table is reproduced in Appendix II. The different types of models in use are discussed below:

1.2 Random and Continuous Random Walk Models

Financial markets have been studied from the random walk point of view. The random walk formalism was the first model known in finance, having been suggested by Bachelier (1900) to describe stock market dynamics. Bachelier modeled the price evolution assuming that prices change one unit at each time step with a probability p of going up and $1 - p$ of going down. Thus, there are only two possible events.

This process is called the binomial model and is the simplest random walk. In 1965, Montroll and Weiss published a paper on continuous-time random walks (CTRWs) in which the waiting-time between two consecutive jumps of a diffusing particle is a real positive stochastic variable. This was the starting point for several developments on the physics of normal and anomalous self-diffusion.

A modified form of random walk is used in physics to model anomalous diffusion, by incorporating a random waiting time between particle jumps. The analogy in Finance is the particle jumps are log-returns and the waiting times measure delay between transactions. These two random variables (log-return and waiting time) are typically not independent [12]. For these coupled CTRW models the authors have proposed to compute the limiting stochastic process. The authors have presented applications with tick-by-tick stock and futures data. Scalas (2006) has also studied CTRW models and their application to Finance and Economics [12].

1.3 Finite Difference Stock Market Model

Melecky and Sergyeyev (2008) have recently proposed a finite-difference stock market model which involves usage of intrinsic values. The authors suggest that using a deterministic delay difference model for the time series of the closing stock price and the intrinsic value of the stock can better other contemporary models. The unique feature of this model is the equation describing the evolution of the intrinsic value. The authors show that in comparison with the real-world data, upon a suitable choice of parameters the model predictions are close to actual values of the real stock for short time horizons [13]. The results from the sample cases tested by the authors are shown below:

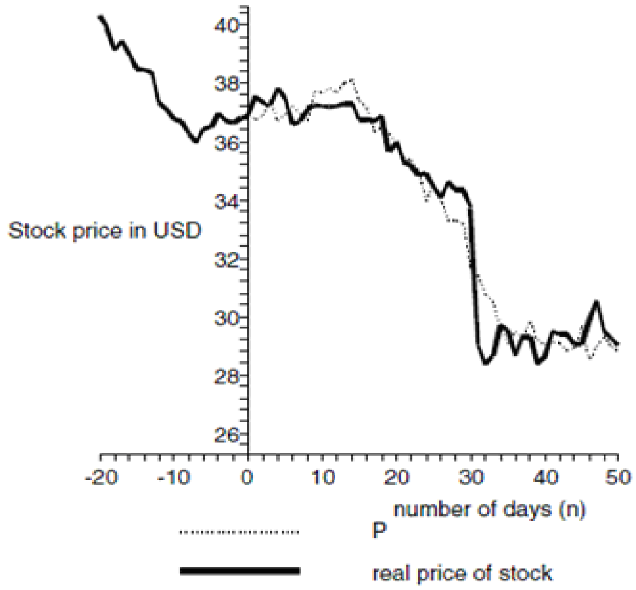


Fig 1: Evolution of GM Stock (Predicted and actual values are very similar)

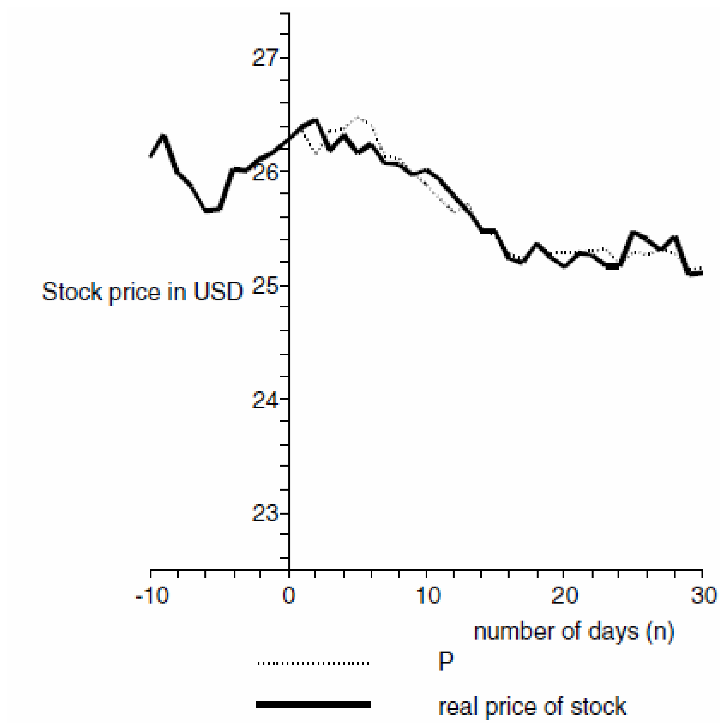


Fig 2: Evolution of Microsoft Stock (Predicted and actual values are very similar)

Upon looking at the graphs above it is clear that indeed a correct set of parameters can lead the model to produce results which are reasonably similar, at least on the qualitative level, to the price dynamics of the real stock market.

1.4 Genetic Algorithms

Cajueiro et al, have attempted to predict stock market crashes taking the case of Brazilian Stock Market [4]. The authors attempt to detect the development of bubbles and crashes in individual stocks using a genetic algorithm. Using the model the authors calibrate the parameters of the model. The most liquid stocks in Brazil have been tested. The results show that the empirical results are consistent with the prediction hypothesis and the method applied can be used to forecast the end of asset bubbles or large corrections in stock prices.

1.5 Markov–Fourier grey model

Hsu et al, have recently attempted to improve upon the forecasting techniques by using the Markov-Fourier grey model [7]. The model in an integration prediction method which includes the grey model (GM), Fourier series, and Markov state transition to predict the turning time of Taiwan weighted stock index (TAIEX) for increasing the forecasting accuracy [7]. The authors use two parts of the forecast a) Build an optimal grey model from a series of data and b) Use the Fourier series to refine the residuals produced by the mentioned model [7]. The authors demonstrate that this unique approach gets the better result performance than that of the other methods.

1.6 Neural Network Models

Neural network has been popular in time series prediction in financial areas because of their advantages in handling nonlinear systems. Lin et al have recently presented a study of using a novel recurrent neural network–echo state network (ESN) to predict the next closing price in stock markets [9]. The authors show how this method under the right set of parameters can prevent coarse prediction performance.

1.7 Exchange Rate Modeling

There are a variety of models for predicting foreign exchange rates which is one of the important financial instruments. Leu et al have developed a fuzzy time series model which has been used successfully to predict both stock prices as well as foreign exchange rates [8]. The authors use the word distance based fuzzy time series model (DBFTS) [8], and differentiate this model from regular fuzzy logic models by using the distance between fuzzy logic relationships (FLRs) to select prediction rules [8]. The two factors considered in the model are the exchange rate itself and the variable set which affect exchange rates. Exchange rate data released by the Central Bank of Taiwan is used for validating the model [8]. The authors have reported that the model actually out-performs the random walk method as measured by the Mean Square Error.

CHAPTER 2: Methodology

2.1 Approach:

Literature of work already done on random walk model shows that equations have been developed to study the dynamic behavior of particles, whose mass and size is much larger than the host medium (Langevin Equation). We can use the most important stocks, for example NASDAQ 100 stock, which are "heavier" in comparison to other stocks and model their movement over time. The broad steps in the model could be as follows:

1. Collect historical stock price data for the influential stocks in the market. Assume this is time, t = 0.

2. Establish the correlation between these stock prices

3. Use equations of Brownian Motion available in literature to describe the speed of movement of stock prices as a function of other stock prices.

4. Represent each stocks movement as a function of a random force exerted by other stocks in the market. The force one stock exerts on another will also depend on the correlation between these stocks, which have been measured in the second step.
5. Using the speed of stock movement, predict the stock's position in the next time interval, an incremental time from t=0.

6. Repeat the calculations from step 3 to again predict stock prices the next instant.

2.2 Challenge/problem focus

Stock price movements are difficult to predict using the available simplistic methods which are backward looking. Since investors and financial institutions bear a lot of risk in stocks, it is pertinent to understand how stock market can be predicted. This forms the basic premise of our thesis work.

Stock index movement prediction through a simple yet realistic model can be very beneficial to the financial industry. Concepts from physics have already been successfully applied to stocks and other derivatives, a prime example being the Black-Scholes model [3]. Similarly, we attempt to simulate stock price movements and check whether we are able to predict stock market movements over a month's time. Even if partially successful the simulator thus developed can be handy tool for the financial industry which uses backward looking methods such as simple extrapolation of patterns to predict stock market prices.

2.3 Research Questions

Following insight could potentially be had from this study which can be of immense value to the financial world at larger or to fund managers who are using outdated techniques to predict stock prices. The essential question we are trying the answer through this work is:

1. How can our random walk based model be used to predict stock market movements?
2. How good is the prediction as quantified through standard measures of error?
3. What correction or modification can improve the model further?
4. How does this compare with trivial methods, specifically linear regression, average and last known value?

2.4 Hypothesis

The basis of random walk models [15] in general and specifically our work is that stock prices have a long term steady growth rate as well as random movements in small time intervals which influence stock price movements. In addition in our model we wish to incorporate the important element of "market sentiment" which moves all shares in tandem up or down. The major hypothesis therefore is that:

- With the combination of the "market sentiment" factor as well as the random movements added to the steady growth rate we should be able to predict better than trivial methods such as linear regression or last known value.

2.5 Research Methodology

A quantitative research has been used for conducting this research work. This research helps the investors in predicting the real values than assuming the values Extensive literature, analytical thinking and experimental study are involved in this research project.

Quantitative research includes the experimental study and designing a simulator for predicting the values**.** Results from the experimental study can be seen in chapter 5.
The following figure 3 will illustrate these phases we have followed in the research.

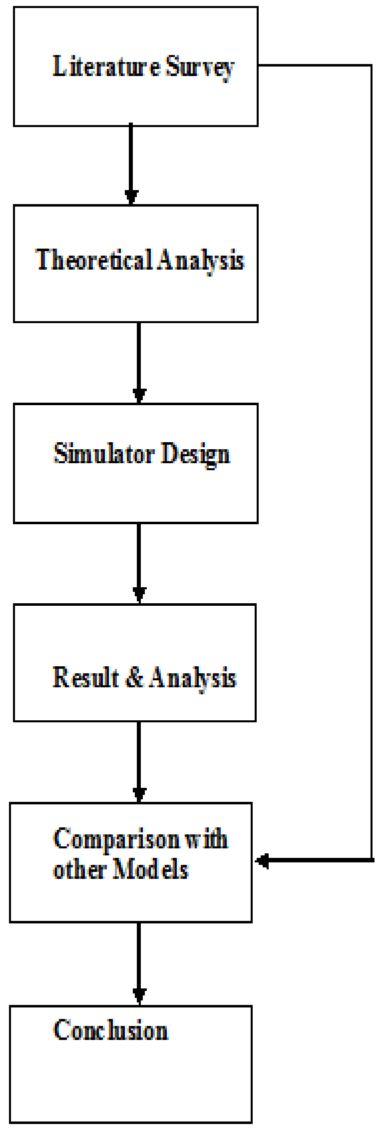


Fig 3. Overview of Research Methodology

2.6 Type of Thesis

This thesis work is an academic research based project as we apply different theories and models. In this work we developed a simulator to find out the prediction values of stock market by using the Brownian theory and Random Walk model.

2.7 Scope of Thesis

This thesis work addresses some of the most important challenges faced by the financial markets. We tried to analyze the historic values and using the scientific methods we calculated the predicted values for the stock prices.

CHAPTER 3: Theoretical work

3.1 Random Walk Model to Predict Stock Prices

We use the random walk model with some modifications to predict stock prices. Specifically we take sample stock files from NASDAQ100 which is a popular index and is representative of the mood of the market. The stock data has been downloaded from publicly available resources [17]. The data series runs from 3rd January 1995 until 15th February 2009 and hence is very comprehensive. For certain stocks such as Yahoo, Google etc. the data series is available post 2004, when these companies went public.

The methodology is described in the steps below:

Step 1: Accept User Inputs on the following parameters

- Stock Name which needs to be predicted, S
- Days into the future when the prediction needs to be made, D_{future}
- Number of data points from the past to be used, D_{past}
- Number of iterations, N_{iter}

Step 2: An array with the data is initialized as shown in the Fig 4 below:

Initialize the array with actual data before the baseline date, d and 0 for all dates after the baseline date.

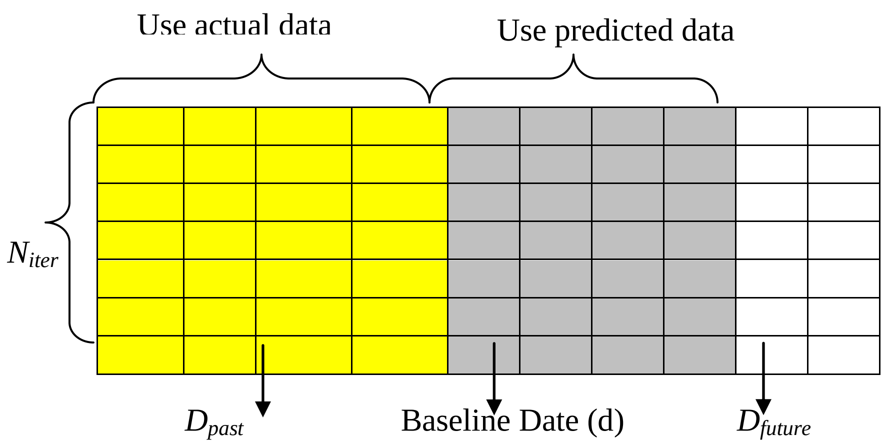


Fig 4a: The array used to implement random walk model

Step 3: As shown in the schematic above, use past data to calculate the following:

- Rate initial, $R_d = S_d/S_{d-1}$
- Rate for earlier days, $R_{di} = S_i/S_{i-1}$
- Standard Deviation of Returns = STDEV(R_{Dpast}, $R_{Dpast+1}$, ……………………R_{d-1})

Step 4: Calculate change in the stock price as follows: $\Delta S = S_{di+1} - S_{di} = S_{di}R_{di}\Delta t + S_{di}N(0,1)\sigma\sqrt{\Delta t}$

- Where : Sdi = Stock price at the beginning of the period i
    - ΔS = Stock price change during time Δt
    - $S_{di}R_{di}\Delta t$ = Expected value of stock price change during time Δt
    - N(0,1) = Normally distributed random variable with mean 0 and standard deviation = 1
    - σ is the standard deviation of the stock price

Step 5: Calculate the new stock price value by using the ΔS calculate and populate the array with the new stock price. Hence, $S_{di+1}=S_{di}+\Delta S$

Step 6: Print S_{di+1}

Step 7: i=i+1

Step 8: Repeat Steps 4 through 6 until ($i=D_{future}$)

Step 9: Iteration Count = Iteration Count + 1

Step 10: If Iteration Count = N_{iter} Step 12, Else Step 11

Step 11: Repeat Steps 2 through Step 10

Step 12: END

The modifications to the usual random walk algorithm is the usage of bounds to keep the predicted values practical, which in our case is within +/- 30% of the previous days value and the usage of an additional parameter called "Market State."

Market State defines the market's sentiment. Sometimes, even when the companies in the economy have not changed anything, the markets boom, while other times, without any company related reason, the stock prices crash. This is reflected in the Capital Asset Price Model byβ. In our case, we increase the average by 10% if the market sentiment is high, whereas if the market sentiment is low, the average rate calculated is reduced 10%. Therefore, we can take into account the obvious effect of the market sentiment as well.

The schematic in Fig 4b shows the steps for each iteration run by the simulator

3.2 Tracking Errors in Prediction

In order to understand whether the prediction is good or not, we use the Root Mean Square Error (RMSE) which is a standard measure used to quantify deviation from actual values. Since actual values are also available for the periods for which we can calculate RMSE using the following formula:

$$\mathrm{RMSD}(\theta_1,\theta_2) = \sqrt{\mathrm{MSE}(\theta_1,\theta_2)} = \sqrt{\mathrm{E}((\theta_1-\theta_2)^2)} = \sqrt{\frac{\sum_{i=1}^{n}(x_{1,i}-x_{2,i})^2}{n}}.$$

Here RMSD (Root Mean Square Deviation) is identical to RMSE and as shown the errors for each prediction are squared and the root of their mean is calculated. In the above formula, θ_1 and θ_2 are the actual and predicted values between the baseline date and D_{future}.. The advantage of using RMSE instead of absolute deviation on the final day of prediction is that RMSE is calculated over the entire prediction duration, rather than 1 day. Therefore it is a better reflection of the deviation of the model from reality.

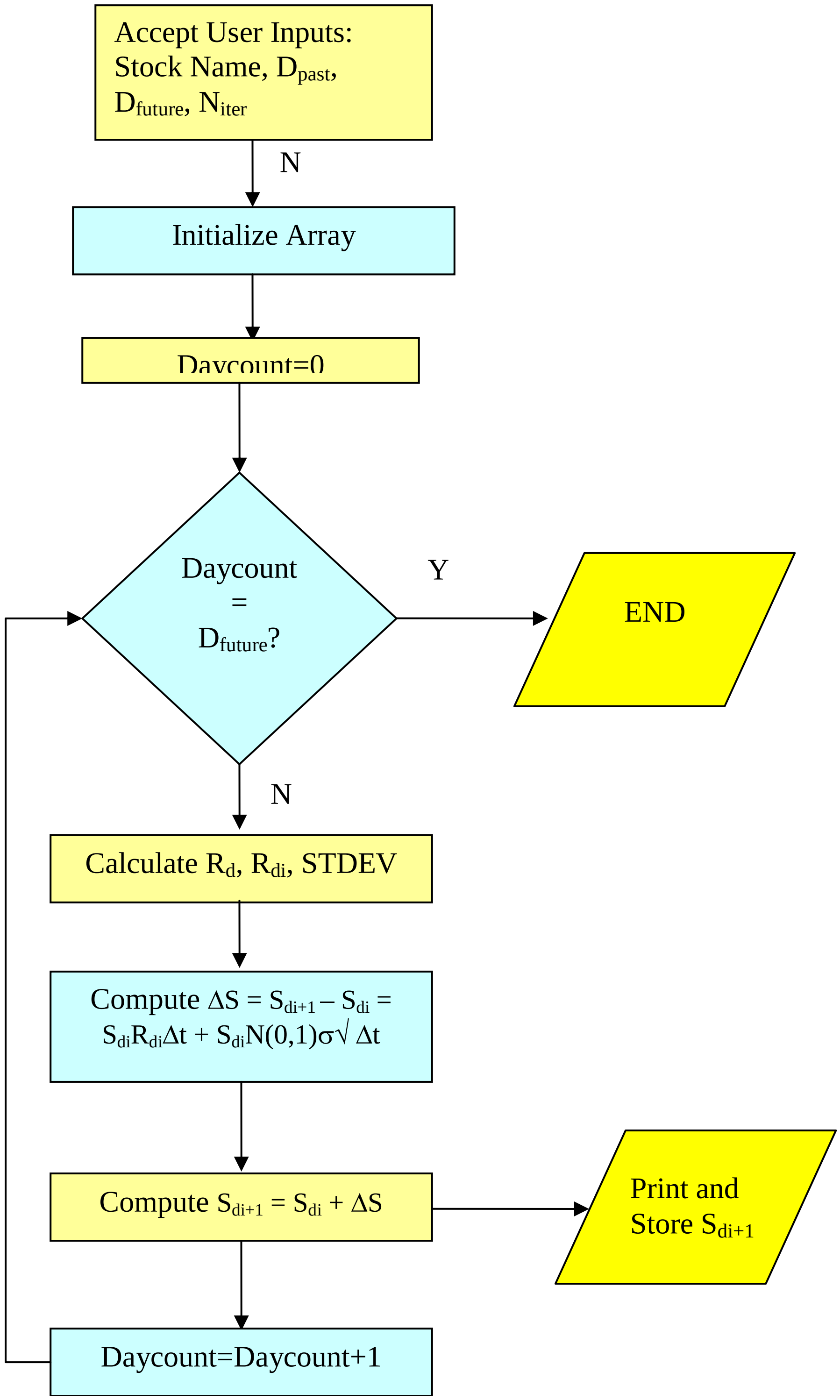


Fig 4b: Flowchart for each iterations steps

3.3 Comparison with conventional techniques

Conventional techniques such as regression to predict stock prices are used by amateur stock price analysts. Neural Networks are also oftentimes used to predict stock prices as shown in the literature survey. These are based on historical data of the stock's performance.

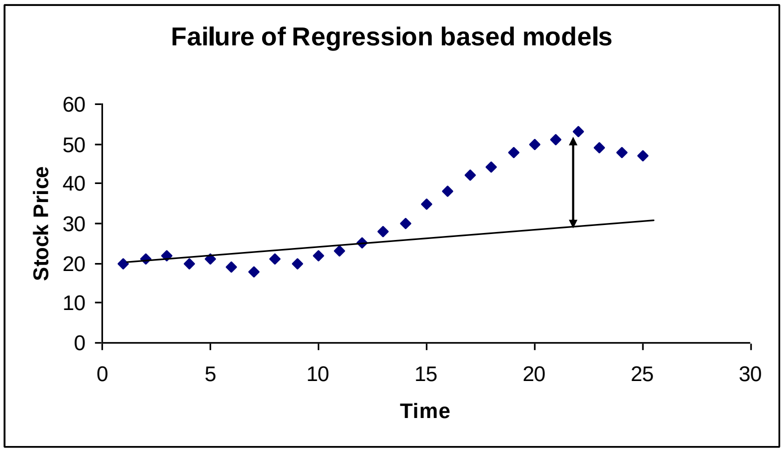


Fig 5: Regression based methods suffer from drawbacks

In the case of regression, the assumption is that past performance can be extrapolated to predict the future. However, this does not take into account effects such as market sentiment or the random noise which is very common in stocks. How simple regression can fail is shown in the Fig 5. The data is made up; however this could be imagined as the stock of an IT company during the dot com boom. The actual data is shown by dots while the regression line based on the first few data points. As shown, the linear regression falls behind the actual data as the stock unexpectedly booms. The difference shown is approximately 25 points, showing how these models can fail.

Similarly, neural networks learn from past data by taking into account patterns which have been seen in the future. However, these patterns might not be observed in the future, therefore, simple neural networks might not be able to predict stocks with much accuracy. The evolved neural network based prediction models discussed in the section on literature survey are definitely much more robust in predicting stock performance. An actual case study is described in the Results and Discussion section.

CHAPTER 4: Empirical Study

4.1 Simulator

A simulator has been developed based on the modified random walk model to demonstrate how the model can be used to predict stock prices in the future. We also show what the limitations of the model using various scenarios are. The simulator has been programmed using .NET framework with the input parameters as described in the model description above. A user friendly GUI has also been provided for convenience and aesthetics. A snap shot of the GUI of the simulator is shown in Fig 6 below.

The simulator takes the following inputs. The stock values are fed in the form of a text file (filename.txt) using the browse functionality in the software. The simulator has been coded in such a way that it takes the 2000^{th} row in the text file containing stock data as the baseline date. The simulator then requires the field on "Number of days used for prediction" to be filled. If we say 10 days, the simulator will use the stock data for 10 days before the baseline date. The next input which the simulator requires is the number of days one needs to predict. If this is 10, the simulator predicts stock prices for 10 days following the baseline date.

The number of iterations specifies how many runs of the simulation will be used to calculate the average predicted stock prices.

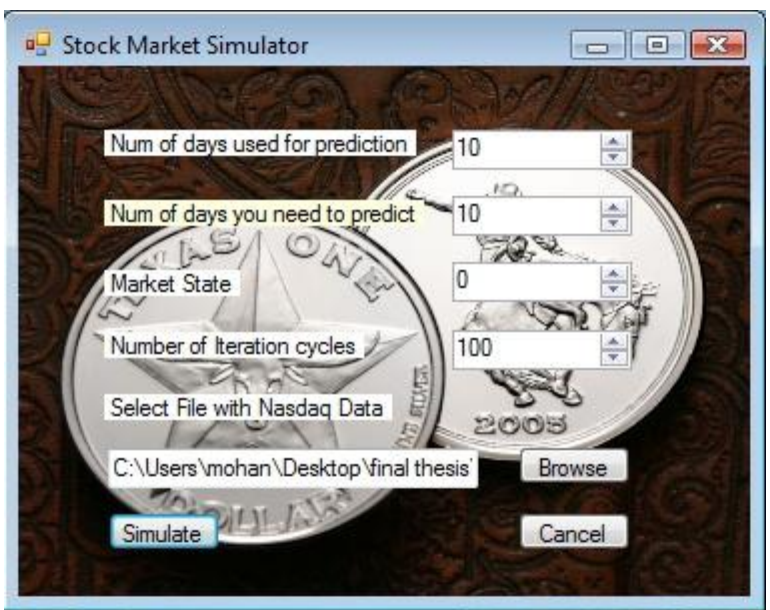


Fig 6: Stock Market Simulator Developed

CHAPTER 5: Results

The model developed incorporates various elements into the prediction. These are shown schematically below in the Fig 7. We discuss each one of them below:

Expected Value: Based on the performance and the pedigree of the company, the stock's returns have an expected value. This is a function of the company's strengths and competencies as well its consistency. Therefore, taking the past data into account the change in the stock price can be estimated. For example, if the stock has been falling in the past, the expectation is that if nothing significant has happened, the stock will continue to fall. This does not incorporate factors such as "market sentiment".

In our model, the term $S_{di}R_{di}\Delta t$ represents the expected change in stock price.

Market Sentiment: The sentiment of the market is another relevant factor. Macro economic conditions drive the markets upwards or downwards. For example, in recent times all stocks have tumbled, irrespective of whether the company has actually seen a decline in sales or not. This is because the market sentiment over powers the individual stock's expectation. Market leaders such as Google, IBM, P&G have seen a reduction in market capitalization recently. Hence, even if fundamentals are right, the market sentiment can drive down prices.

In our model, the usage of the parameter, "State of the Market" mimics this.

Variability: There is always some random noise in the stock return movements which cannot be explained. This is the main argument of the random walk model. Using σ of the stock returns and using random numbers drawn from a normal distribution with mean 0 and standard deviation of 1 (white noise), we can incorporate this noise into the model. The term $S_{di}N(0,1)\sigma\sqrt{\Delta t}$

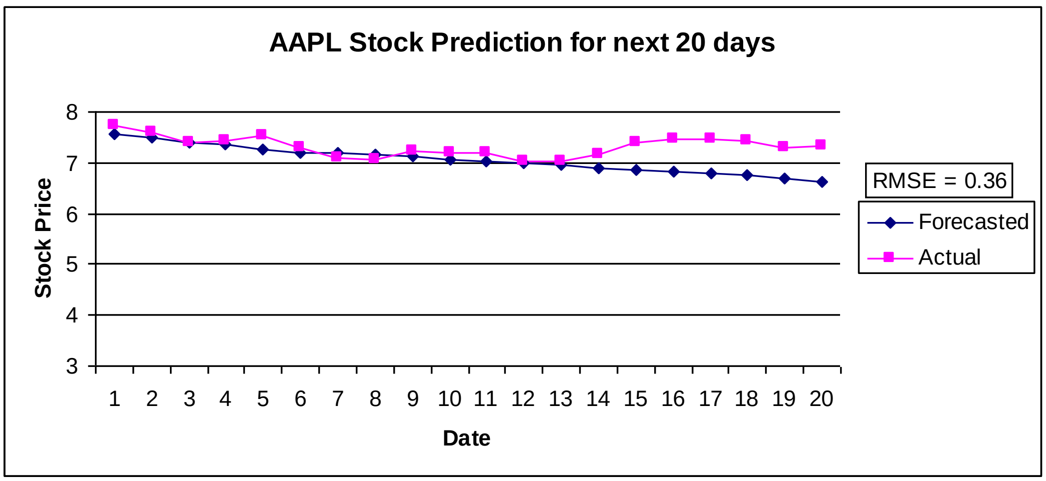
Fig 7: Factors which affect stock prices

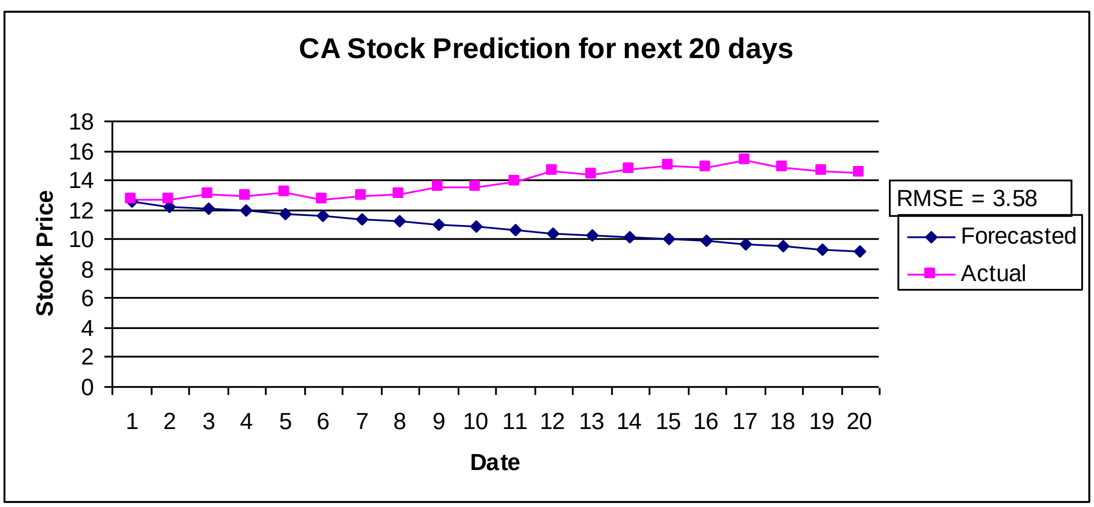
5.2 Impact of Various Parameters on Results

Several test cases from the NASDAQ100 stocks have been taken to test the software developed. This also includes NASDAQ100 index which is representative of the stock market overall. We run the simulator and plot the actual as well as predicted values in the graph. Alongside the RMSE has been calculated to quantify the deviation from the actual values. We have used the baseline date as Mid December (11^{th} – 16^{th} Dec., 2002) for the purpose of these experiments. This corresponds to Row 2000 in the stock data files used. The scenarios are described in the sub-sections below. All stock acronyms are explained in Appendix II:

Test Case 1 (Short term Prediction)

Prediction of Stock Prices 20 days into the future (short term) using the past 20 days data as basis. N_{iter}=100. Market State=0 (Normal).





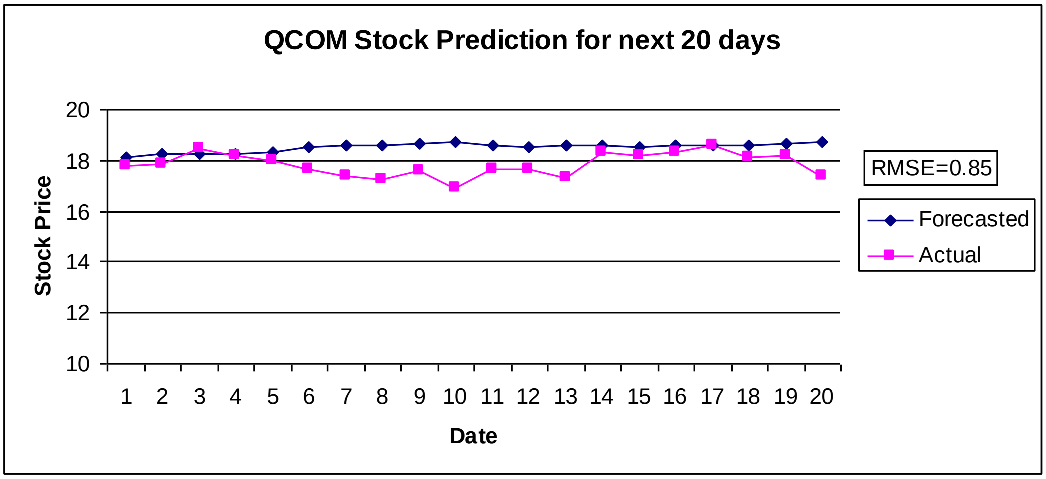


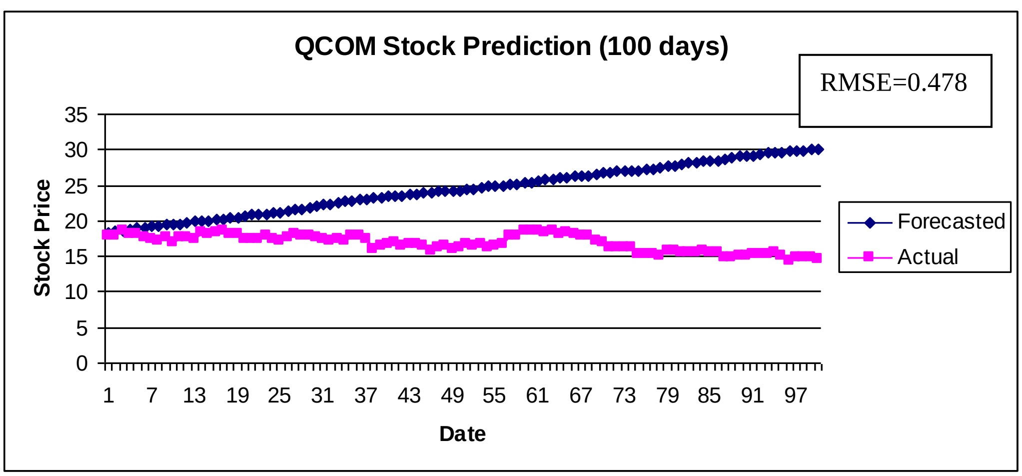
Fig 8: Actual Vs. Predicted Stock Prices (Short term)

From the test cases discussed above, it is clear that for short term (approximately a month's stock trading time), the model robustly predicts the stock prices with very low RMSE values. The error in prediction could be low because of two reasons:

1) The prediction horizon is small
2) The model is actually working

Test Case 2 (Medium term Prediction)

To understand which of the two is correct, we use the model to predict stock prices over a larger time horizon (100 days, or a quarter of trading time). N_{iter}=100. D_{past}=100



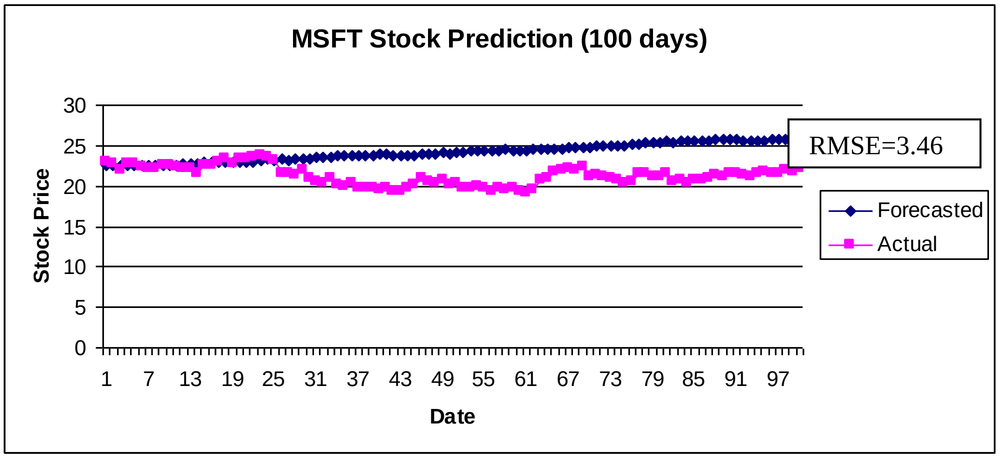
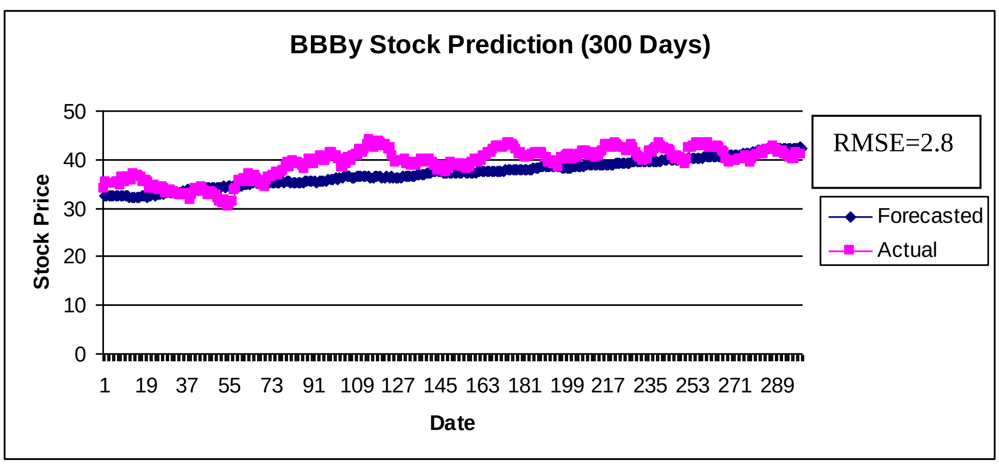


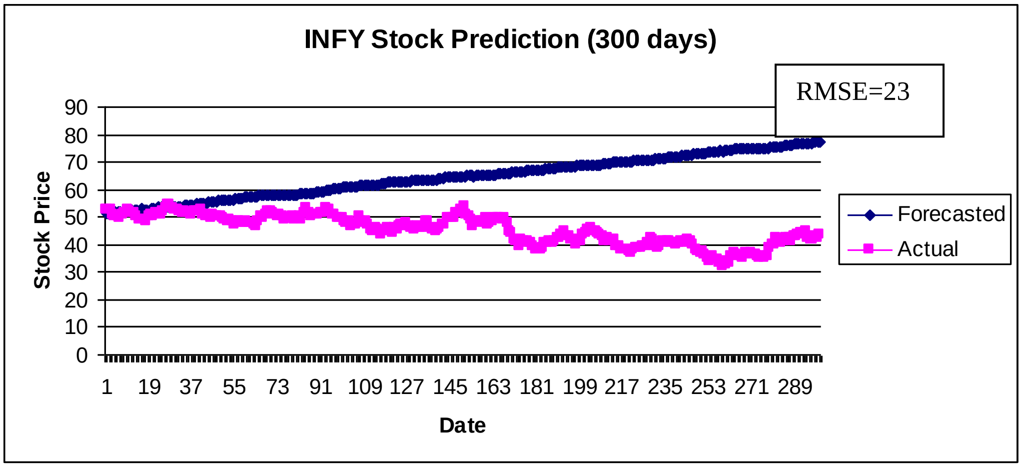
Fig 9: Actual Vs. Predicted Stock Prices (Medium term)

Using the above randomly drawn stocks we can see that the stock price predictor can actually predict stock price movements fairly accurately. However in the case of QCOM, the predictions are deviant from actual values and hence we need to test this in the "down state" of the simulator to understand whether results can be improved further or not. To further test the robustness of the tool, we use it to predict stock prices for a year's trading time (approximately 300 days).

Test Case 3 (Long term Prediction)

D_{future}=300 days, or a year of trading time. N_{iter}=100. D_{past}=300





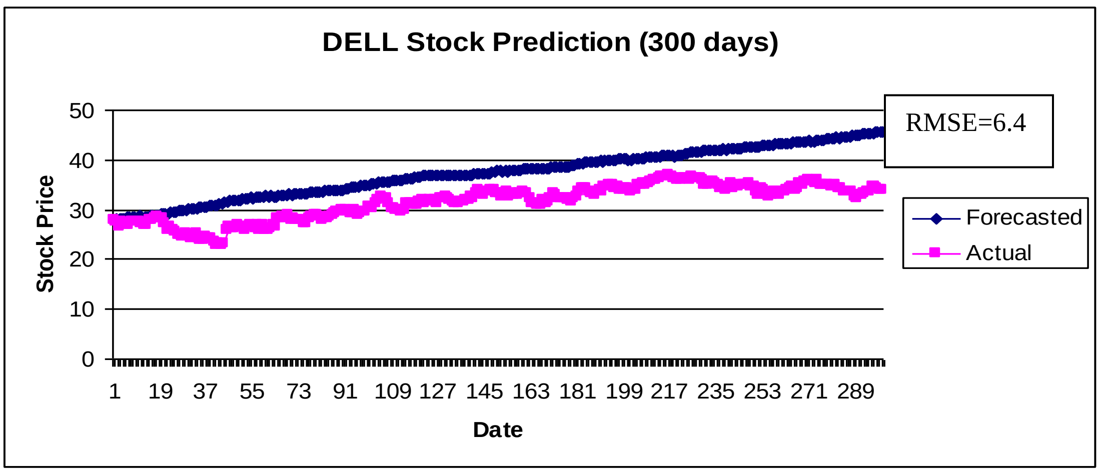


Fig 10: Actual Vs. Predicted Stock Prices (Long term)

As shown in the cases considered above the prediction is fairly robust with the model, however in the case of INFY the prediction deviates from the actual values and the prediction is not accurate.

As is clear from the above cases, the prediction model has its own limitations. However, largely, the results are accurate. This should be reflected in the predictions of NASDAQ100 index data. These are discussed in the results below:

Test Case 4 (Aggregate Index Prediction)

D_{future}=100 days, or a year of trading time. N_{iter}=100. D_{past}=100

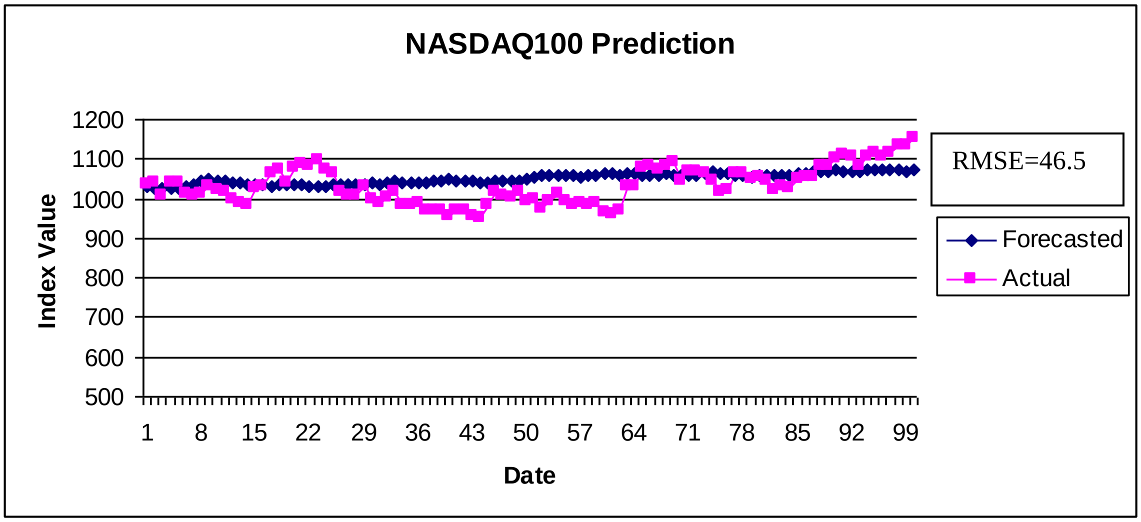
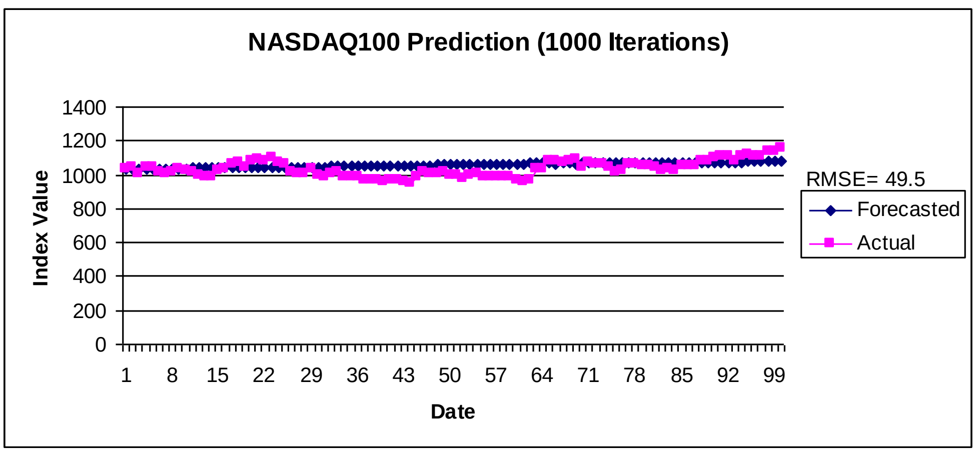


Fig 11: Actual Vs. Predicted Index Prices (Affect of aggregation)

As shown above the index values are predicted fairly robustly because aggregation leads to moderation in the variation.

Impact of number of Iterations



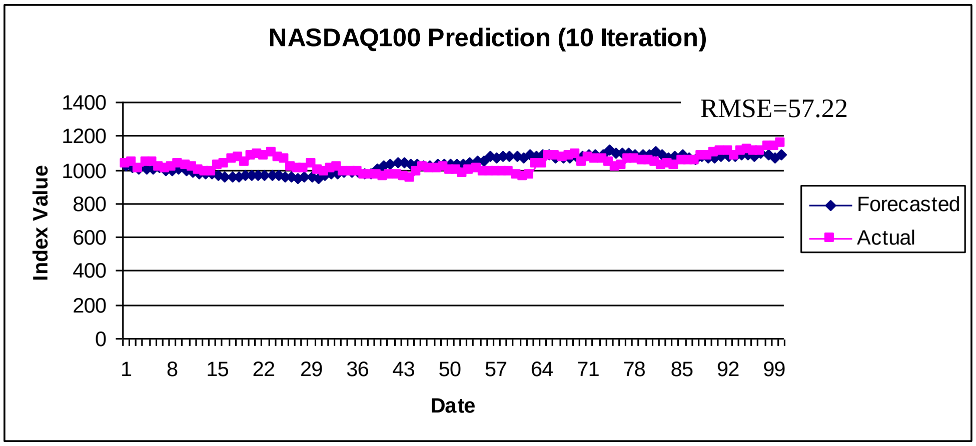


Fig 12: Actual Vs. Predicted Index Price (Affect of # of Iterations)

From this result it is clear that increasing the number of iterations can lead to better results.

5.3 Impact of State of Economy

We measure the impact of state of the system on stocks to understand how the predicted values change with the "boom" state, "bearish" state and "normal" state (Parameter values are 1, -1, 0 respectively). Since we have the prediction of QCOM where the stock price declined whereas our prediction was high (Case II), we see if using the state -1 improves the prediction:

As shown in the graph (Fig 13), the RMSE reduces, showing that using the state -1 does help. However the impact is low. Therefore there is a need to increase the magnitude of impact of the parameter in the algorithm.

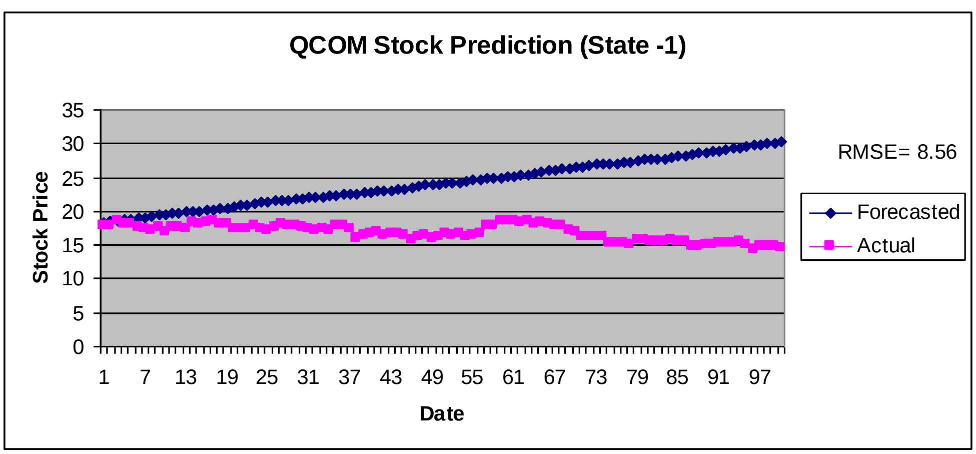


Fig 13: Actual Vs. Predicted Stock Prices (Impact of State)

5.4 Comparison Case Study

We compare the robustness of the model developed against trivial methods such as linear regression, last known value and average. Taking the NASDAQ100 index values as well as several stocks from the index, we compare the RMSE for different methods. All methods use the previous 100 day data (from the baseline date) and predict the next 100 day stock prices. The other methods are described below:

Linear Regression: In order to make an exact comparison, the baseline date for regression is taken as the one corresponding to the 2000^{th} row of the stock text file. The method here is as follows:

Step 1: Fit a linear model (Y = A + BX) on the past 100 days of stock. Here, X would be the Day number while A is the stock price.
Step 2: Using the linear model developed in Step 1, we predict the values for Day numbers 101, 102…… 200. This gives us the predicted values.
Step 3: Calculate the RMSE as has been described earlier.

We have used standard MS Office FORECAST function to implement the linear regression predictions. The FORECAST function forces a best fit line of the form Y=MX+ C on the data set (D_{past},…. $D_{baseline}$) and determines the values of the slope, M and the Y-intercept C, through minimization of the mean square error between actual values and predicted values. Thereafter for each day from $D_{baseline+1}$, …… D_{future}, the values are predicted by using the best fit line, Y = MX + C. This gives us the predicted stock price values through linear regression.

Average: Average simply implies the average of stock prices from Day 1 to Day 100.

Last Known Value: We assume that the value of the stock will remain the same as was observed on Day 100 and then compute RMSE by comparing this with the actual stock prices. As is evident this is most trivial of all methods and can only lead to better performance in case of very stable stock performance.

Time Periods: In order to compare the performance of the random walk models, we use four periods. Period 1 corresponds to a baseline date of Mid Dec. 1998, while period 2 corresponds to a baseline date of Mid Dec. 2002. Using these are the baseline date and D_{past} = 100 days, D_{future} = 100 days, N_{iter} = 100, the simulator predicts the stock price for 100 days into the future from the baseline date. By testing the simulator over two periods we can show whether the Random Walk methods or its variants performs better that the other methods irrespective of the time. Also, 1998-1999 was a period of rapid stock market appreciation while 2002-2003 was the period following the “dot-com burst”. The effectiveness of the market sentiment factor we have added to the Random Walk Model can be tested. Additionally, two more periods have been added in order to make the comparison comprehensive. The Period 3 corresponds to Mid. Dec 1996 (corresponding to row number 500 as the baseline date) and Period 4 correspond to Early Dec. 2006 (corresponding to row number 3000 as the baseline date).

Stocks Chosen: 22 Stocks have been chosen randomly from the set of NASDAQ100, along with the NASDAQ100 index itself. These have been used for the purpose of the comparison study.

Results:
The RMSEs for the predictions have been tabulated for Period 1 and Period 2 in Tables 1 and 2 below while the comparison of relative performance of the models has been illustrated in Tables 3 and 4. The lowest RMSEs have been highlighted in Tables 1, 2.

	Linear Regression	**Average**	**Last Known Value**	**Random Walk (0 State)**	**Random Walk (+1 State)**	**Random Walk (-1 State)**
NASDAQ100	329.57	645.54	644.58	139.46	150.15	136.40
AAPL	1.92	1.17	1.60	2.07	**1.83**	1.87
BBBY	1.51	4.72	5.71	1.37	1.38	1.06
CA	10.56	6.30	7.24	4.79	**5.49**	5.34
CMSCA	3.46	6.35	7.30	**1.74**	**2.12**	1.75
DELL	6.90	12.40	15.35	7.19	**6.21**	6.47
ERTS	2.26	**1.19**	1.99	2.84	2.81	2.78
FLEX	1.23	5.44	5.97	3.97	**4.00**	3.28
HOLX	0.27	0.92	2.07	1.18	1.04	1.10
IACI	0.96	1.19	0.91	1.09	**1.07**	0.93
INFY	9.67	14.60	15.27	2.77	2.57	2.71
INTC	3.25	7.57	8.94	7.12	7.54	6.97
MSFT	7.10	11.08	10.62	2.66	2.95	2.70
PCAR	0.99	0.96	0.95	2.38	**2.41**	2.20
PDCO	2.39	**1.05**	1.56	2.42	2.57	2.24
QCOM	4.35	4.40	4.09	5.05	**5.03**	5.00
ADBE	**0.78**	2.25	1.98	1.59	1.57	1.62
CTAS	6.08	10.93	11.65	4.18	4.64	**3.91**
FWLT	20.12	10.54	12.28	11.80	11.96	**9.97**
LLTC	5.53	9.53	8.97	**2.80**	2.84	2.94
MCHP	1.66	2.00	**1.26**	2.98	3.13	2.86
VRTX	4.12	**2.05**	2.70	3.97	4.81	4.05
XLNX	3.76	8.95	10.59	1.73	1.82	**1.65**

Table 1: Period 1 (Baseline Date: Mid Dec. 1998, Row 1000 in input file): RMSE for different stocks and NASDAQ100 index for Random Walk (States +1, 0, -1), Linear Regression, Average and Last Known Value

	Linear Regression	Average	Last Known Value	Random Walk (0 State)	Random Walk (+1 State)	Random Walk (-1 State)
NASDAQ100	89.54	70.24	144.95	46.57	54.01	**40.57**
AAPL	0.78	0.47	0.60	0.48	0.48	**0.35**
BBBY	2.58	3.26	5.30	3.94	3.04	**2.89**
CA	4.03	1.96	5.14	5.13	5.19	**4.57**
CMSCA	1.57	3.31	4.38	1.07	1.40	**0.80**
DELL	5.06	1.89	4.67	**3.00**	3.05	3.43
ERTS	6.22	5.29	3.38	2.18	2.43	**2.07**
FLEX	1.71	0.52	0.92	2.54	2.67	**2.14**
HOLX	1.82	0.84	0.95	0.94	0.98	1.13
IACI	0.94	0.63	1.47	0.75	0.68	**0.67**
INFY	9.54	4.22	1.70	**5.55**	6.28	6.24
INTC	1.49	0.87	1.14	**1.34**	1.54	1.72
MSFT	4.72	1.41	1.18	3.46	**2.61**	2.82
PCAR	0.86	2.63	4.39	2.02	2.29	**1.75**
PDCO	1.37	3.16	1.36	1.22	**1.11**	1.24
QCOM	6.91	1.56	2.11	8.75	9.12	**8.21**
ADBE	1.97	4.11	3.14	1.74	**1.50**	1.53
CTAS	15.69	8.35	**7.23**	13.32	13.15	11.71
FWLT	6.36	**4.60**	5.07	5.79	6.04	5.98
LLTC	2.25	4.57	3.73	**1.54**	1.97	1.93
MCHP	8.41	**1.93**	1.95	7.22	8.81	8.23
VRTX	3.11	6.29	6.10	1.94	**1.48**	2.00
XLNX	1.93	5.01	4.98	**1.89**	3.19	3.47

Table 2: Period 2 (Baseline Date: Mid Dec. 2002, Row 2000 in input file): RMSE for different stocks and NASDAQ100 index for Random Walk (States +1, 0, -1), Linear Regression, Average and Last Known Value

The relative performance of the Random Walk methods and its variants as compared to other methods has been illustrated through the tables below. If the random walk method or its variants performs better than the other methods, it is represented with a "+" symbol, while in case the performance is worse, the "-" symbol is used. The bottom most rows shows the % cases in which Random Walk performs better than the other methods.

	Random Walk '0' State			Random Walk '1 State'			Random Walk '-1 State'		
	Vs. Linear	Vs. Average	Vs. Last Known	Vs. Linear	Vs. Average	Vs. Last Known	Vs. Linear	Vs. Average	Vs. Last Known
NASDAQ	+	+	+	+	+	+	+	+	+
AAPL	-	-	-	+	-	-	+	-	-
BBBY	+	+	+	+	+	+	+	+	+
CA	+	+	+	+	+	+	+	+	+
CMSCA	+	+	+	+	+	+	+	+	+
DELL	-	+	+	+	+	+	+	+	+
ERTS	-	-	-	-	-	-	-	-	-
FLEX	-	+	+	-	+	+	-	+	+
HOLX	-	-	+	-	-	+	-	-	+
IACI	-	+	-	-	+	-	+	+	-
INFY	+	+	+	+	+	+	+	+	+
INTC	-	+	+	-	+	+	-	+	+
MSFT	+	+	+	+	+	+	+	+	+
PCAR	-	-	-	-	-	-	-	-	-
PDCO	-	-	-	-	-	-	+	-	-
QCOM	-	-	-	-	-	-	-	-	-
ADBE	-	+	+	-	+	+	-	+	+
CTAS	+	+	+	+	+	+	+	+	+
FWLT	+	-	+	+	-	+	+	+	+
LLTC	+	+	+	+	+	+	+	+	+
MCHP	-	-	-	-	-	-	-	-	-
VRTX	+	-	-	-	-	-	+	-	-
XLNX	+	+	+	+	+	+	+	+	+
	48%	**61%**	**65%**	**52%**	**61%**	**65%**	**65%**	**65%**	**65%**

Table 3: Period 1 (Baseline Date: Mid Dec. 1998, Row 1000 in input file)
Comparison of Linear Regression vs. Random Walk (+1, -1, 0)

	Random Walk 'State 0'			Random Walk 'State +1'			Random Walk 'State -1'		
	Vs. Linear Regression	Vs. Average	Vs. Last Known Value	Vs. Linear Regression	Vs. Average	Vs. Last Known Value	Vs. Linear Regression	Vs. Average	Vs. Last Known Value
NASDAQ	+	+	+	+	+	+	+	+	+
AAPL	+	-	+	+	-	+	+	+	+
BBBY	-	-	+	-	+	+	-	+	+
CA	-	-	+	-	-	-	-	-	+
CMSCA	+	+	+	+	+	+	+	+	+
DELL	+	-	+	+	-	+	+	-	+
ERTS	+	+	+	+	+	+	+	+	+
FLEX	-	-	-	-	-	-	-	-	-
HOLX	+	-	+	+	-	-	+	-	-
IACI	+	-	+	+	-	+	+	-	+
INFY	+	-	-	+	-	-	+	-	-
INTC	+	-	-	-	-	-	-	-	-
MSFT	+	-	-	+	-	-	+	-	-
PCAR	-	+	+	-	+	+	-	+	+
PDCO	+	+	+	+	+	+	+	+	+
QCOM	-	-	-	-	-	-	-	-	-
ADBE	+	+	+	+	+	+	+	+	+
CTAS	+	-	-	+	-	-	+	-	-
FWLT	+	-	-	+	-	-	+	-	-
LLTC	+	+	+	+	+	+	+	+	+
MCHP	+	-	-	-	-	-	+	-	-
VRTX	+	+	+	+	+	+	+	+	+
XLNX	+	+	+	-	+	+	-	+	+
	78%	**39%**	**65%**	**65%**	**43%**	**57%**	**70%**	**48%**	**61%**

Table 4: Period 2 (Baseline Date: Mid Dec. 2002, Row 2000 in input file)
Comparison of Linear Regression vs. Random Walk (+1, -1, 0)

The results for Period 3 and Period 4 are shown in the Appendix IV. We have chosen the Random Walk Model with State 0 for the analysis. The results along with the T-Test statistical analysis results are presented there.

Statistical Analysis using the Hypothesis Test

Whether the Random Walk Method and its variants is any better than the other methods has been studied using the Hypothesis Test which is formulated as follows:

H0: Random Walk Methods and its variants lead to RMSEs which are not different than the RMSEs obtained from other methods
H1: Random Walk Methods and its variants lead to RMSEs which are significantly different than the RMSEs obtained from other methods

This can also be reformulated in the following manners:

H0: RMSE (Other Methods) – RMSE (Random Walk, Variants) is not greater than 0
H1: RMSE (Other Methods) – RMSE (Random Walk, Variants) is significantly greater than 0

By comparing the difference between the RMSEs obtained for Random Walk Variants and Other Methods with 0, we can check for significance in difference. This can be done using the t-test. We have used the Data Analysis tool pack from Excel to conduct the t-test comparing the following with the Hypothesized difference of 0.

1. RMSE (Linear Regression) - RMSE (Random Walk State 0)
2. RMSE (Linear Regression) - RMSE (Random Walk State +1)
3. RMSE (Linear Regression) - RMSE (Random Walk State -1)
4. RMSE (Average) - RMSE (Random Walk State 0)
5. RMSE (Average) - RMSE (Random Walk State +1)
6. RMSE (Average) - RMSE (Random Walk State -1)
7. RMSE (Last Known Value) - RMSE (Random Walk State 0)
8. RMSE (Last Known Value) - RMSE (Random Walk State +1)
9. RMSE (Last Known Value) - RMSE (Random Walk State -1)

The results for these comparisons are shown below in the tables 5, 6 for the two periods, Period 1 and Period 2 at two different confidence intervals:

		RW			RW State '+1'			RW State '-1'		
Period	Confidence Level	LR	Avg	LKV	LR	Avg	LKV	LR	Avg	LKV
Row 1000	85%	Sig. Better	At Par	At Par	Sig. Better	At Par	At Par	Sig. Better	At Par	At Par
	95%	At Par	At Par	At Par	At Par	At Par	At Par	At Par	At Par	At Par

Table 5: T-Test: Statistical Analysis Results, Period 1
(Mid Dec. 1998, Baseline corresponding to Row 1000 in input file)

	Confidence Level	RW			RW State '+1'			RW State '-1'		
Period 2: Row 2000	**85%**	Sig. Better	At Par	At Par	Sig. Better	At Par	At Par	Sig. Better	At Par	At Par
	95%	At Par	At Par	At Par	At Par	At Par	At Par	At Par	At Par	At Par

Table 6: T-Test: Statistical Analysis Results, Period 2
(Mid Dec. 2002, Baseline corresponding to Row 1000 in input file)

The corresponding results from Period 3 and Period 4 are shown below:

	Confidence	**RW Vs. LR**	**RW Vs. Average**	**RW s. LKV**
Row 500 (Mid Dec. 1996)	95%	At Par	At Par	At Par
	85%	Sig. Poor	Sig. Poor	At Par
Row 3000 (1st Week Dec, 2006)	95%	Sig. Poor	At Par	At Par
	85%	Sig. Poor	Sig. Poor	Sig. Poor

Table 7: **T-Test: Statistical Analysis Results, Period 3 and Period 4**
(Mid Dec. 1996 and Early Dec. 2006 as Baseline dates)

The significantly poor results are highlighted in Red.

The detailed results of the T-tests are shown in the Appendix III.

Absolute Deviation

A comparison of absolute deviation from actual stock price has also been done for 4 different methods: Last known value, average stock price, linear regression and average. Here we follow the same method as described in the section above with the following changes:

1. The error is the deviation is the absolute value of the 200th day stock price vs. predicted price (Therefore it is not the overall RMSE, but the final day's prediction error)
2. Average stock price, takes a simple average of the last 100 day stock price
3. Last known value, takes the closing stock price on Day 1

The results have been show in the Figure 14, 15 below:

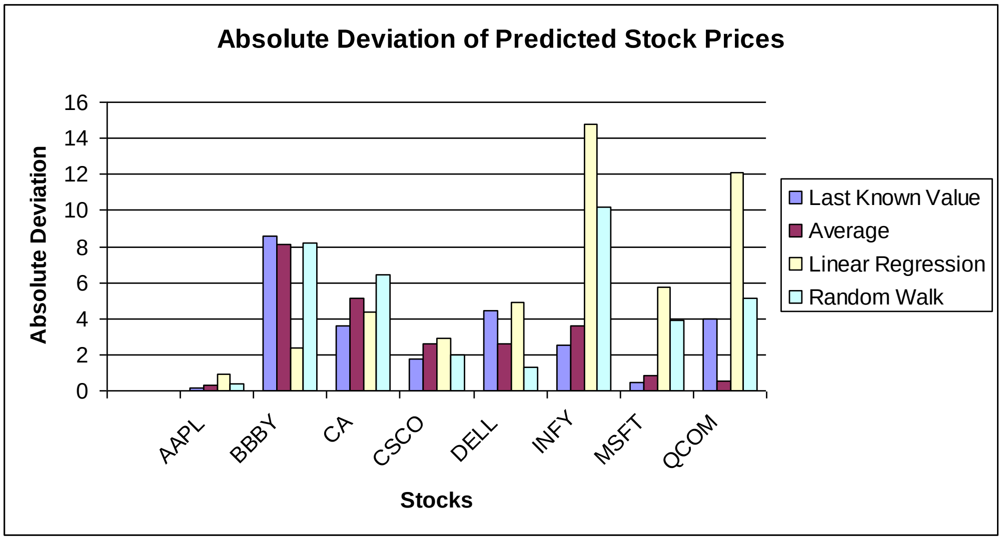


Fig 14: Absolute Deviation of predicted values from actual values for different stocks (Baseline date Mid Dec, 2002 corresponding to row 2000 in input file)

As shown, the prediction errors are similar across the methods with each method out-performing the other across stocks. For the aggregate index (NASDAQ100) the comparison of deviation across methods is shown in Fig 17. This also shows the random walk is better suited to predict indices only, where it has an some advantage.

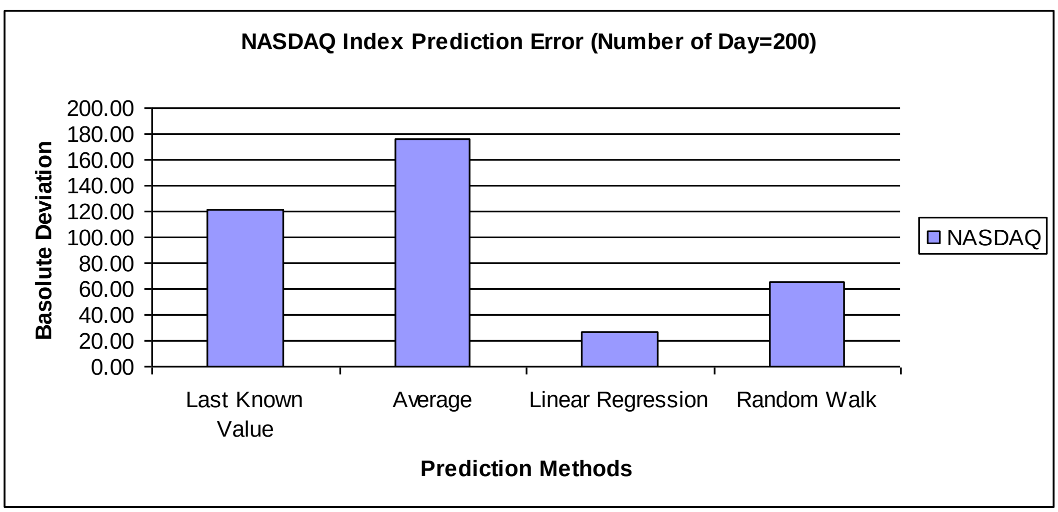


Fig 15: Absolute Deviation of predicted values from actual values for NASDAQ100 (Period 2, Baseline date Mid Dec. 2002, corresponding to row 2000 in input file)

CHAPTER 6: Discussion

The experiments show how for small time horizons the simulator produces excellent results; however limitations start showing up as the time horizon increases. Furthermore the impact of various parameters is clear. As variability reduces on aggregation and NASDAQ100 is an aggregated index, we expect it the results to be good.

The number of iterations leads to better results because the confidence interval reduces with the increase in number of samples and hence RMSE should reduce. The other important aspects are what is the impact of the state of the economy and how does the results compare with other methods. These are discussed in the sections below:

Comparison with other "trivial" methods

The comparison data shown in the result section of the report shows the RMSE values obtained for different methods for different stocks. The ones with the lowest and highest RMSE values are highlighted in yellow.

As we can see, across all the periods, there is no generalization that can be made as for some periods random walk has a directional to significant advantage whereas for others it has a directional to significant disadvantage as compared to the other models. Testing at different confidence levels show that across different periods, Random Walk's performance ranges from significantly superior to significantly inferior. This leaves us with no specific conclusion. **Overall, using the 95% confidence interval T-tests as the basis, Random Walk method offers no advantage over trivial methods such as Linear Regression, Average and Last Known Value.**

However, for index price prediction, Random walk based model seem to perform somewhat better as shown in the graph below. The results for NASDAQ100 index prediction are shown Fig 18 below:

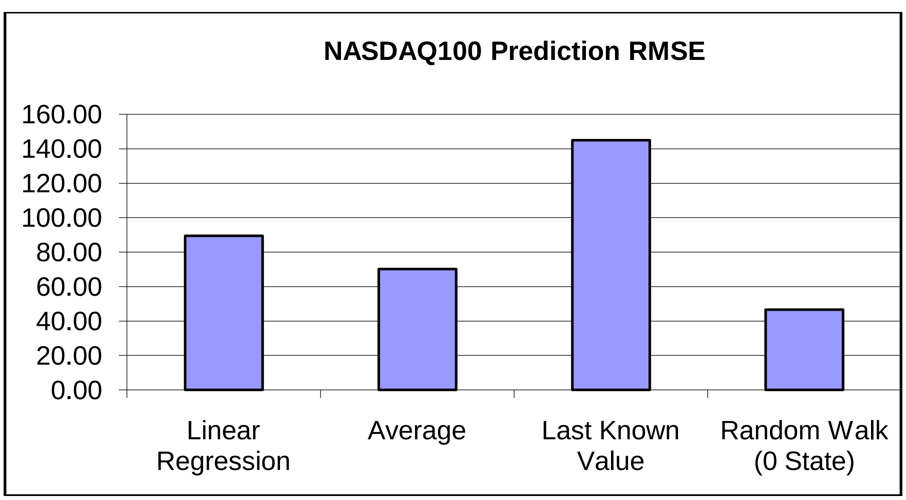


Fig 16: Random Walk model predicts the NASDAQ100 better (Lower RMSE)

Impact of State or Market Sentiment

From the results, it is clear that the market sentiment has no significant impact on the overall results. A comparison of the RMSE for NASDAQ100 prediction for the different methods is shown in the figure below:

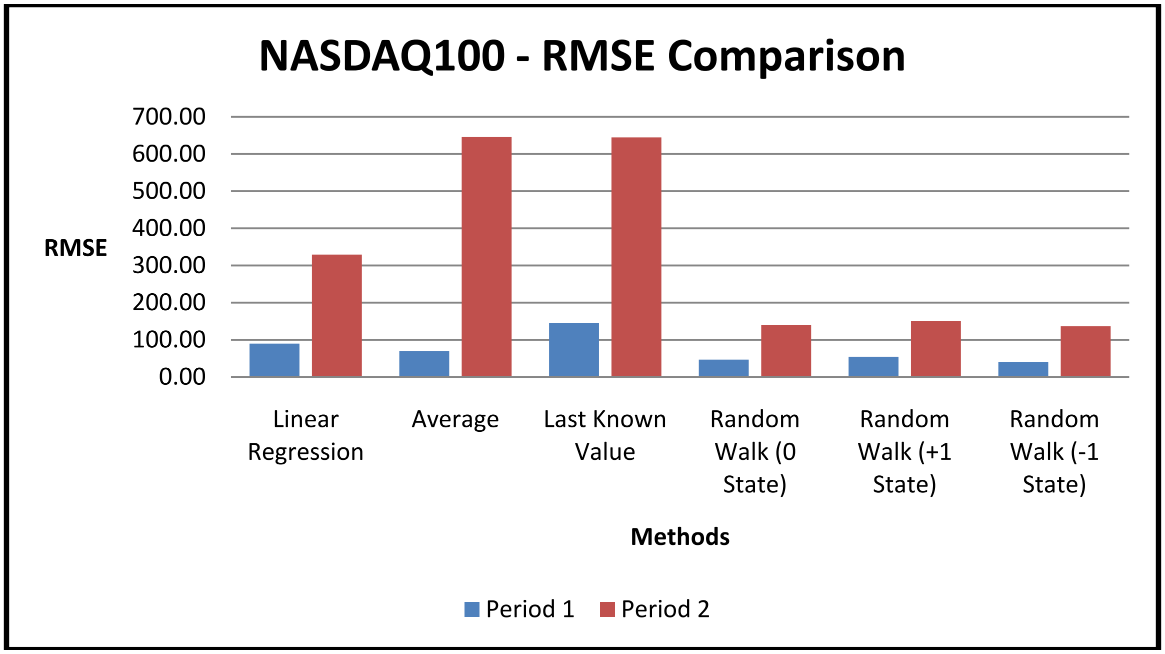


Figure 17: Random Walk model predicts the NASDAQ100 better (Lower RMSE)

The important observation from Figure 19 is the fact that in Period 2, corresponding to Mid Dec. 2002, the predictions of NASDAQ100 are much better for all random walk variants as compared to the other methods. **This shows that for index predictions in a period which does not have "normal" market sentiment, the random walk model performs better than the other methods.** However, this is only one data point and hence a general conclusion cannot be drawn.

CONCLUSIONS

Based on the results obtained from the simulator, we see that the model predicts the stock prices reasonably, but with certain limitations. Using a very long period for prediction does show higher deviation from the actual data. Similarly, for single stocks the deviations from actual values are similar to average based model. Overall, the random walk method falls short of our expectations and offers no advantage as compared to trivial methods when it comes to stock predictions. Random walk also seems to be directionally better than other methods for predicting indices such as NASDAQ100.

One known drawback of random walk models is that the model leads to "jumpy" changes in the predicted prices [5]. This the author argues is because during a small time interval, say 1 day, the standard deviation in stock returns can be much larger than the mean stock return rate. Since the model calculates stock price changes, ΔS, where a term has standard deviation being multiplied by time, this leads to "jumpy" changes in stock prices. However, when a large number of stocks are studied as an aggregate, the standard deviation values reduce. In fact, our model performs best for the aggregated NASDAQ100 index as aggregation reduces variability and the random walk model comes up with less choppy values. This is the benefit of using a random walk model instead of other methods.

Similarly, the states used in the model are effective but not to the extent required. However, the model is a starting point for future research. By modifying the algorithm appropriately, its robustness can be improved. A simple modification which could be tried could be to use moving average instead of simple average in the random walk algorithm to compute the stock value change. We could also consider developing a hybrid model which uses the random walk model along with genetic algorithm, which allows the model to self correct as it comes across deviations from actual data.

REFERENCES

[1] Atsalakis, G. S., Valavanis, K. P., (2009), surveying stock market forecasting techniques – Part II: Soft computing methods, *Expert Systems with Applications*, 36, 5932–5941

[2] Bachelier, L., 1900, Théorie de la Spéculation, Gauthier-Villars, Paris (as studied from the translation: Davis, M. and Etheridge, A. (2006), Louis Bachelier's Theory of Speculation: The Origins of Modern Finance, Princeton University Press)

[3] Black, F.; Scholes, M. (1973). "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy* 81 (3): 637–654

[4] Cajueiro, D., O, Tabak, B. M, Werneck, F. K., (2009), Can we predict crashes? The case of the Brazilian stock market, *Physica A*, 388 , 1603-1609

[5] Grinde, R., (2004), Random Walk Models, Retrieved online from http://pubpages.unh.edu/~rbg/TMS/SupportFiles/Random%20Walk%20Models.pdf on April 8, 2009.

[6] Hänggi, P. and Marchesoni, F., (2005), Introduction: 100 years of Brownian motion**,** *CHAOS*, 15, 026101

[7] Hsu, Y. T., Liu, M., C., Hung, H., F., (2009), Forecasting the turning time of stock market based on Markov–Fourier grey model, *Expert Systems with Applications*, 36, 8597–8603

[8] Leu, Y., Lee, C.P, Kou, Y. Z, (2009), A distance-based fuzzy time series model for exchange rates forecasting, *Expert Systems with Applications* 36, 8107–8114

[9] Lin, X., Yang, Z., Song, Y., (2009), Short-term stock price prediction based on echo state networks, *Expert Systems with Applications*, 36, 7313–7317

[10] Lo AW, MacKinlay AC. Stock prices do not follow random walks: evidence from a simple specification test. *Rev Financial Stud*, 1988;1:41–66.

[11] Masoliver, J., Montero, M., Perell, J., Weiss, G. H., (2006), The continuous time random walk formalism in financial markets, *Journal of Economic Behavior & Organization*, Vol. 61, 577–598

[12] Meerschaert, M. M., and Scalas, E., (2006), Coupled continuous time random walks in finance, *Physica A*, 370, 114–118

[13] Melecky, J. and Sergyeyev, A., (2008), A simple finite-difference stock market model involving intrinsic value, *Chaos, Solitons and Fractals* 38 (2008) 769–777

[14] Montroll, E. W., Weiss, G. H., (1965), Random walks on lattices, II, *J. Math. Phys*., 6, 167– 181.

[15] Pearson, K., 1905. The problem of the Random Walk, *Nature*, 72, 294

[16] Scalas, E., (2006), The application of continuous-time random walks in finance and economics, *Physica A*, 362, 225–239

[17] http://www.optiontradingtips.com/resources/historical-data/nasdaq100.html

Appendix I: Stock Market Models in Literature (Atsalakis and Valavanis, 2009)

Table 4
Comparative studies

Article	ANNs	LR, MLR	ARMA, ARIMA	GA	RW	B & H	Others
Andreou et al. (2000)	•						•
Armano et al. (2004)	•					•	
Atsalakis and Valavanis (2006a)						•	
Atsalakis and Valavanis (2006b)						•	
Baba and Kozaki (1992)							
Baek and Cho (2002)						•	
Barnes et al. (2000)		•	•				•
Bautista (2001)					•		
Brownstone (1996)		•					
Cao et al. (2005)		•					•
Casas (2001)						•	
Chandra and Reeb (1999)						•	
Chaturvedi and Chandra (2004)		•					
Chen et al. (2003)					•	•	
Chen et al. (2005a, 2005b)	•						
Chenoweth and Obradovic (1996)	•					•	
Chun and Park (2005)					•		
Doesken et al. (2005)						•	
Donaldson and Kamstra (1999)			•				•
Dong et al. (2003)		•					
Dourra and Siy (2002)							•
Egeli et al. (2003)	•						
Fernandez-Rodriguez et al. (2000)					•	•	
Harvey et al. (2000)		•				•	
Huang et al. (2005)	•						•
Hui et al. (2000)	•						
Kanas and Yannopoulos (2001)		•					
Kim (1998)	•						
Kim and Han (1998)						•	•
Kimoto et al. (1990)		•					
Kosaka et al. (1991)							•
Koulouriotis (2004)							•
Koulouriotis et al. (2001)						•	
Koulouriotis et al. (2002)		•					
Koulouriotis et al. (2005)	•	•					•
Lam (2001)						•	
Leigh et al. (2002)						•	
Malliaris and Salchenberger (1993)							
Mizuno et al. (1998)	•						•
Motiwalla and Wahab (2000)		•				•	
Nishina and Hagiwara (1997)	•						
Oh and Kim (2002)	•					•	
Olson and Mossman (2003)	•						
Pai and Lin (2005)			•				•
Pan et al. (2005)							
Pantazopoulos et al. (1998)							•
Phua et al. (2001)							
Qi (1999)		•					
Quah and Srinivasan (1999)							•
Raposo et al. (2002)	•						•
Rast (1999)	•						
Rech (2002)			•				•
Refenes et al. (1993)		•					
Schumann and Lohrbach (1993)			•				
Setnes and Van Drempt (1999)		•				•	
(Siekmann et al., 1999)		•			•	•	•
Steiner and Wittkemper (1997)							
Thammano (1999)	•						

Appendix II: Stock Acronyms

NASDAQ100	Index of top 100 influencing stock (price, volume) on NASDAQ
AAPL	Apple Inc.
ADBE	Abobe Systems Inc.
BBBY	Bed Bath and Beyond Inc.
CA	CA Inc.
CMSCA	Comcast Corporation
CTAS	Cintas Corporation
DELL	DELL Inc.
ERTS	Electronic Arts Inc.
FLEX	Flextronics International Ltd.
FWLT	Foster Wheeler AG.
HOLX	Hologic Inc.
IACI	Interactive Corp
INFY	Infosys Technologies Ltd.
INTC	Intel Corporation
LLTC	Linear Technology Corporation
MCHP	Microchip Technology Inc.
MSFT	Microsoft Corporation
PCAR	PACCAR Inc
PDCO	Patterson Companies Inc.
QCOM	Qualcomm Inc.
VRTX	Vertex Pharmaceuticals Inc.
XLNX	Xilinx Inc.

Appendix III: Results of t-Test

Period 1 Statistical Analysis Results (95% Confidence)

Period 1: T-Test for RANDOM WALK State 0 Vs. LINEAR REGRESSION

	Variable 1	*Variable 2*
Mean	18.63	9.44
Variance	4614.05	809.34
Observations	23.00	23.00
Pearson Correlation	1.00	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	1.11	
P(T<=t) one-tail	0.14	
t Critical one-tail	1.72	
P(T<=t) two-tail	0.28	
t Critical two-tail	2.07	

Period 1: T-Test for RANDOM WALK State 1 Vs. LINEAR REGRESSION

	Variable 1	*Variable 2*
Mean	18.63	10.00
Variance	4614.05	939.63
Observations	23.00	23.00
Pearson Correlation	1.00	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	1.11	
P(T<=t) one-tail	0.14	
t Critical one-tail	1.72	
P(T<=t) two-tail	0.28	
t Critical two-tail	2.07	

Period 1: T-Test for RANDOM WALK State -1 Vs. LINEAR REGRESSION

	Variable 1	*Variable 2*
Mean	18.63	9.12
Variance	4614.05	774.59

Observations	23.00	23.00
Pearson Correlation	1.00	
Hypothesized Mean Difference	0.00	
Df	22.00	
t Stat	1.14	
P(T<=t) one-tail	0.13	
t Critical one-tail	1.72	
P(T<=t) two-tail	0.27	
t Critical two-tail	2.07	

Period 1: T-Test for RANDOM WALK State 0 Vs. Average

	Variable 1	*Variable 2*
Mean	33.53	9.44
Variance	17817.46	809.34
Observations	23.00	23.00
Pearson Correlation	1.00	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	1.10	
P(T<=t) one-tail	0.14	
t Critical one-tail	1.72	
P(T<=t) two-tail	0.28	
t Critical two-tail	2.07	

Period 1: T-Test for RANDOM WALK State +1 Vs. Average

	Variable 1	*Variable 2*
Mean	33.53	10.00
Variance	17817.46	939.63
Observations	23.00	23.00
Pearson Correlation	1.00	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	1.10	
P(T<=t) one-tail	0.14	
t Critical one-tail	1.72	
P(T<=t) two-tail	0.28	
t Critical two-tail	2.07	

Period 1: T-Test for RANDOM WALK State -1 Vs. Average

	Variable 1	*Variable 2*
Mean	33.53	9.12
Variance	17817.46	774.59
Observations	23.00	23.00
Pearson Correlation	1.00	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	1.11	
P(T<=t) one-tail	0.14	
t Critical one-tail	1.72	
P(T<=t) two-tail	0.28	
t Critical two-tail	2.07	

Period 1: T-Test for RANDOM WALK State 0 Vs. LAST KNOWN VALUE

	Variable 1	*Variable 2*
Mean	34.07	9.44
Variance	17733.79	809.34
Observations	23.00	23.00
Pearson Correlation	1.00	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	1.13	
P(T<=t) one-tail	0.14	
t Critical one-tail	1.72	
P(T<=t) two-tail	0.27	
t Critical two-tail	2.07	

Period 1: T-Test for RANDOM WALK State +1 Vs. LAST KNOWN VALUE

	Variable 1	*Variable 2*
Mean	34.07	10.00
Variance	17733.79	939.63
Observations	23.00	23.00
Pearson Correlation	1.00	
Hypothesized Mean Difference	0.00	

Df	22.00
t Stat	1.13
P(T<=t) one-tail	0.14
t Critical one-tail	1.72
P(T<=t) two-tail	0.27
t Critical two-tail	2.07

Period 1: T-Test for RANDOM WALK State -1 Vs. LAST KNOWN VALUE

	Variable 1	Variable 2
Mean	34.07012	9.121053
Variance	17733.79	774.589
Observations	23	23
Pearson Correlation	0.997675	
Hypothesized Mean Difference	0	
Df	22	
t Stat	1.135013	
P(T<=t) one-tail	0.134293	
t Critical one-tail	1.717144	
P(T<=t) two-tail	0.268585	
t Critical two-tail	2.073873	

Period 1: Statistical Analysis Results (85% Confidence)

Period 1: T-Test for RANDOM WALK State 0 Vs. LINEAR REGRESSION

	Variable 1	*Variable 2*
Mean	18.63	9.44
Variance	4614.05	809.34
Observations	23.00	23.00
Pearson Correlation	1.00	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	1.11	
P(T<=t) one-tail	0.14	
t Critical one-tail	1.06	

P(T<=t) two-tail	0.28
t Critical two-tail	1.49

Period 1: T-Test for RANDOM WALK State +1 Vs. LINEAR REGRESSION

	Variable 1	*Variable 2*
Mean	18.63	10.00
Variance	4614.05	939.63
Observations	23.00	23.00
Pearson Correlation	1.00	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	1.11	
P(T<=t) one-tail	0.14	
t Critical one-tail	1.06	
P(T<=t) two-tail	0.28	
t Critical two-tail	1.49	

Period 1: T-Test for RANDOM WALK State -1 Vs. LINEAR REGRESSION

	Variable 1	*Variable 2*
Mean	18.63	9.12
Variance	4614.05	774.59
Observations	23.00	23.00
Pearson Correlation	1.00	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	1.14	
P(T<=t) one-tail	0.13	
t Critical one-tail	1.06	
P(T<=t) two-tail	0.27	
t Critical two-tail	1.49	

Period 1: T-Test for RANDOM WALK State 0 Vs. Average

	Variable 1	*Variable 2*
Mean	33.53	9.44
Variance	17817.46	809.34
Observations	23.00	23.00

Pearson Correlation	1.00
Hypothesized Mean Difference	0.00
df	22.00
t Stat	1.10
P(T<=t) one-tail	0.14
t Critical one-tail	1.06
P(T<=t) two-tail	0.28
t Critical two-tail	1.49

Period 1: T-Test for RANDOM WALK State +1 Vs. Average

	Variable 1	*Variable 2*
Mean	33.53	10.00
Variance	17817.46	939.63
Observations	23.00	23.00
Pearson Correlation	1.00	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	1.10	
P(T<=t) one-tail	0.14	
t Critical one-tail	1.06	
P(T<=t) two-tail	0.28	
t Critical two-tail	1.49	

Period 1: T-Test for RANDOM WALK State -1 Vs. Average

	Variable 1	*Variable 2*
Mean	33.53	9.12
Variance	17817.46	774.59
Observations	23.00	23.00
Pearson Correlation	1.00	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	1.11	
P(T<=t) one-tail	0.14	
t Critical one-tail	1.06	
P(T<=t) two-tail	0.28	
t Critical two-tail	1.49	

Period 1: T-Test for RANDOM WALK State 0 Vs. LAST KNOWN VALUE

	Variable 1	*Variable 2*
Mean	34.07	9.44
Variance	17733.79	809.34
Observations	23.00	23.00
Pearson Correlation	1.00	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	1.13	
P(T<=t) one-tail	0.14	
t Critical one-tail	1.06	
P(T<=t) two-tail	0.27	
t Critical two-tail	1.49	

Period 1: T-Test for RANDOM WALK State +1 Vs. LAST KNOWN VALUE

	Variable 1	*Variable 2*
Mean	34.07	10.00
Variance	17733.79	939.63
Observations	23.00	23.00
Pearson Correlation	1.00	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	1.13	
P(T<=t) one-tail	0.14	
t Critical one-tail	1.06	
P(T<=t) two-tail	0.27	
t Critical two-tail	1.49	

Period 1: T-Test for RANDOM WALK State -1 Vs. LAST KNOWN VALUE

	Variable 1	*Variable 2*
Mean	34.07	9.12
Variance	17733.79	774.59
Observations	23.00	23.00
Pearson Correlation	1.00	
Hypothesized Mean Difference	0.00	
df	22.00	

t Stat	1.14
P(T<=t) one-tail	0.13
t Critical one-tail	1.06
P(T<=t) two-tail	0.27
t Critical two-tail	1.49

Period 2 Statistical Analysis Results (95% Confidence)

Period 2: T-Test for RANDOM WALK State 0 Vs. LINEAR REGRESSION

	Variable 1	*Variable 2*
Mean	7.78	5.32
Variance	330.43	90.24
Observations	23.00	23.00
Pearson Correlation	0.99	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	1.31	
P(T<=t) one-tail	0.10	
t Critical one-tail	1.72	
P(T<=t) two-tail	0.20	
t Critical two-tail	2.07	

Period 2: T-Test for RANDOM WALK State +1 Vs. LINEAR REGRESSION

	Variable 1	*Variable 2*
Mean	7.78	5.78
Variance	330.43	120.62
Observations	23.00	23.00
Pearson Correlation	0.99	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	1.28	
P(T<=t) one-tail	0.11	
t Critical one-tail	1.72	
P(T<=t) two-tail	0.21	
t Critical two-tail	2.07	

Period 2: T-Test for RANDOM WALK State -1 Vs. LINEAR REGRESSION

	Variable 1	*Variable 2*
Mean	7.78	5.02
Variance	330.43	68.33
Observations	23.00	23.00

Pearson Correlation	0.98	
Hypothesized Mean Difference	0.00	
Df	22.00	
t Stat	1.30	
P(T<=t) one-tail	0.10	
t Critical one-tail	1.72	
P(T<=t) two-tail	0.21	
t Critical two-tail	2.07	

Period 2: T-Test for RANDOM WALK State 0 Vs. Average

	Variable 1	*Variable 2*
Mean	5.96	5.32
Variance	200.54	90.24
Observations	23.00	23.00
Pearson Correlation	0.96	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	0.53	
P(T<=t) one-tail	0.30	
t Critical one-tail	1.72	
P(T<=t) two-tail	0.60	
t Critical two-tail	2.07	

Period 2: T-Test for RANDOM WALK State +1 Vs. Average

	Variable 1	*Variable 2*
Mean	5.96	5.78
Variance	200.54	120.62
Observations	23.00	23.00
Pearson Correlation	0.96	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	0.18	
P(T<=t) one-tail	0.43	
t Critical one-tail	1.72	
P(T<=t) two-tail	0.86	
t Critical two-tail	2.07	

Period 2: T-Test for RANDOM WALK State -1 Vs. LINEAR REGRESSION

	Variable 1	*Variable 2*
Mean	5.96	5.02
Variance	200.54	68.33
Observations	23.00	23.00
Pearson Correlation	0.95	

Hypothesized Mean Difference	0.00
df	22.00
t Stat	0.66
P(T<=t) one-tail	0.26
t Critical one-tail	1.72
P(T<=t) two-tail	0.52
t Critical two-tail	2.07

Period 2: T-Test for RANDOM WALK State 0 Vs. LAST KNOWN VALUE

	Variable 1	*Variable 2*
Mean	9.38	5.32
Variance	877.04	90.24
Observations	23.00	23.00
Pearson Correlation	0.95	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	0.94	
P(T<=t) one-tail	0.18	
t Critical one-tail	1.72	
P(T<=t) two-tail	0.36	
t Critical two-tail	2.07	

Period 2: T-Test for RANDOM WALK State +1 Vs. LAST KNOWN VALUE

	Variable 1	*Variable 2*
Mean	9.38	5.78
Variance	877.04	120.62
Observations	23.00	23.00
Pearson Correlation	0.96	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	0.90	
P(T<=t) one-tail	0.19	
t Critical one-tail	1.72	
P(T<=t) two-tail	0.38	
t Critical two-tail	2.07	

Period 2: T-Test for RANDOM WALK State -1 Vs. LAST KNOWN VALUE

	Variable 1	*Variable 2*
Mean	9.38	5.02
Variance	877.04	68.33
Observations	23.00	23.00
Pearson Correlation	0.94	
Hypothesized Mean Difference	0.00	
Df	22.00	
t Stat	0.95	
P(T<=t) one-tail	0.18	
t Critical one-tail	1.72	

P(T<=t) two-tail	0.35
t Critical two-tail	2.07

Period 2 Statistical Analysis Results (85% Confidence)

Period 2: T-Test for RANDOM WALK State 0 Vs. LINEAR REGRESSION

	Variable 1	*Variable 2*
Mean	7.78	5.32
Variance	330.43	90.24
Observations	23.00	23.00
Pearson Correlation	0.99	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	1.31	
P(T<=t) one-tail	0.10	
t Critical one-tail	1.06	
P(T<=t) two-tail	0.20	
t Critical two-tail	1.49	

Period 2: T-Test for RANDOM WALK State +1 Vs. LINEAR REGRESSION

	Variable 1	*Variable 2*
Mean	7.78	5.78
Variance	330.43	120.62
Observations	23.00	23.00
Pearson Correlation	0.99	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	1.28	
P(T<=t) one-tail	0.11	
t Critical one-tail	1.06	
P(T<=t) two-tail	0.21	
t Critical two-tail	1.49	

Period 2: T-Test for RANDOM WALK State -1 Vs. LINEAR REGRESSION

	Variable 1	*Variable 2*
Mean	7.78	5.02
Variance	330.43	68.33
Observations	23.00	23.00

Pearson Correlation	0.98
Hypothesized Mean Difference	0.00
df	22.00
t Stat	1.30
P(T<=t) one-tail	0.10
t Critical one-tail	1.06
P(T<=t) two-tail	0.21
t Critical two-tail	1.49

Period 2: T-Test for RANDOM WALK State 0 Vs. Average

	Variable 1	*Variable 2*
Mean	5.96	5.32
Variance	200.54	90.24
Observations	23.00	23.00
Pearson Correlation	0.96	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	0.53	
P(T<=t) one-tail	0.30	
t Critical one-tail	1.06	
P(T<=t) two-tail	0.60	
t Critical two-tail	1.49	

Period 2: T-Test for RANDOM WALK State +1 Vs. Average

	Variable 1	*Variable 2*
Mean	5.96	5.78
Variance	200.54	120.62
Observations	23.00	23.00
Pearson Correlation	0.96	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	0.18	
P(T<=t) one-tail	0.43	
t Critical one-tail	1.06	
P(T<=t) two-tail	0.86	
t Critical two-tail	1.49	

Period 2: T-Test for RANDOM WALK State -1 Vs. Average

	Variable 1	*Variable 2*
Mean	5.96	5.02
Variance	200.54	68.33
Observations	23.00	23.00
Pearson Correlation	0.95	

Hypothesized Mean Difference	0.00
Df	22.00
t Stat	0.66
P(T<=t) one-tail	0.26
t Critical one-tail	1.06
P(T<=t) two-tail	0.52
t Critical two-tail	1.49

Period 2: T-Test for RANDOM WALK State 0 Vs. LAST KNOWN VALUE

	Variable 1	*Variable 2*
Mean	9.38	5.32
Variance	877.04	90.24
Observations	23.00	23.00
Pearson Correlation	0.95	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	0.94	
P(T<=t) one-tail	0.18	
t Critical one-tail	1.06	
P(T<=t) two-tail	0.36	
t Critical two-tail	1.49	

Period 2: T-Test for RANDOM WALK State +1 Vs. LAST KNOWN VALUE

	Variable 1	*Variable 2*
Mean	9.38	5.78
Variance	877.04	120.62
Observations	23.00	23.00
Pearson Correlation	0.96	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	0.90	
P(T<=t) one-tail	0.19	
t Critical one-tail	1.06	
P(T<=t) two-tail	0.38	
t Critical two-tail	1.49	

Period 2: T-Test for RANDOM WALK State -1 Vs. LAST KNOWN VALUE

	Variable 1	*Variable 2*
Mean	9.38	5.02
Variance	877.04	68.33
Observations	23.00	23.00
Pearson Correlation	0.94	
Hypothesized Mean Difference	0.00	

df	22.00
t Stat	0.95
P(T<=t) one-tail	0.18
t Critical one-tail	1.06
P(T<=t) two-tail	0.35
t Critical two-tail	1.49

Appendix IV: Comparison Case Study: Period 3 and Period 4

Comparison Results for Period 3: Corresponding to baseline date of Mid. Dec. 1996 (Row 500 in Stock Data Files)

	Linear Regression	Average	Last Known Value	Random Walk (State 0)
NASDAQ	117.01	119.27	225.97	189.03
AAPL	2.22	1.58	1.17	1.98
BBBY	1.28	0.46	1.19	0.83
CA	14.45	9.12	4.42	17.67
CMSCA	0.49	0.53	1.08	1.39
DELL	0.32	1.93	2.74	0.75
ERTS	1.95	1.59	1.05	1.15
FLEX	1.87	0.42	0.58	1.01
HOLX	4.58	1.30	3.19	3.50
IACI	0.21	0.11	0.12	0.22
INFY	3.09	11.48	11.29	8.24
INTC	3.08	4.84	7.94	6.59
MSFT	1.13	3.26	4.26	1.37
PCAR	1.18	1.32	2.07	1.89
PDCO	0.40	0.91	0.58	1.11
QCOM	1.05	0.75	0.65	0.96
ADBE	0.77	0.51	1.15	1.49
CTAS	3.78	0.79	1.73	4.13
FWLT	14.19	4.06	5.00	6.51
LLTC	1.19	2.18	3.51	1.64
MCHP	1.83	1.95	3.17	3.38
VRTX	4.34	6.35	8.79	5.20
XLNX	0.81	2.21	3.24	0.73

Comparison Results for Period 4: Corresponding to baseline date of Early. Dec. 2006 (Row 3000 in Stock Data Files)

	Linear Regression	Average	Last Known Value	Random Walk (State 0)
NASDAQ	213.21	169.82	258.67	287.87
AAPL	16.12	16.84	34.11	45.25
BBBY	1.57	0.57	0.43	8.33
CA	1.08	2.89	6.41	1.31
CMSCA	3.91	3.28	6.39	8.78
DELL	3.29	2.26	1.64	6.79
ERTS	14.90	2.48	6.25	12.46
FLEX	1.20	0.56	1.32	2.26
HOLX	2.01	5.16	5.59	2.18
IACI	2.24	8.15	10.93	7.90
INTC	4.20	1.24	2.22	3.39
MSFT	5.26	2.89	5.93	7.72
PCAR	1.95	9.27	11.59	3.13
PDCO	2.22	3.41	3.88	7.47
QCOM	2.80	4.61	4.82	2.30
ADBE	9.34	5.18	12.98	9.78
CTAS	8.86	2.02	4.27	9.92
FWLT	1.76	7.53	10.15	9.13
LLTC	2.01	2.21	2.95	1.84
MCHP	1.68	3.33	3.60	0.97
VRTX	17.76	5.19	4.40	19.23
XLNX	5.91	2.54	4.24	4.56

Statistical Analysis Results

Period 3: 95% Confidence Interval

T-Test for RANDOM WALK State 0 Vs. LINEAR REGRESSION

	Variable 1	*Variable 2*
Mean	7.878	11.338
Variance	580.481	1515.080
Observations	23.000	23.000
Pearson Correlation	0.995	
Hypothesized Mean Difference	0.000	
df	22.000	
t Stat	-1.097	
P(T<=t) one-tail	0.142	

t Critical one-tail	1.717
P(T<=t) two-tail	0.284
t Critical two-tail	2.074

T-Test for RANDOM WALK State 0 Vs. AVERAGE

	Variable 1	*Variable 2*
Mean	7.69	11.34
Variance	599.93	1515.08
Observations	23.00	23.00
Pearson Correlation	1.00	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	-1.20	
P(T<=t) one-tail	0.12	
t Critical one-tail	1.72	
P(T<=t) two-tail	0.24	
t Critical two-tail	2.07	

T-Test for RANDOM WALK State 0 Vs. LAST KNOWN VALUE

	Variable 1	*Variable 2*
Mean	12.821	11.338
Variance	2167.220	1515.080
Observations	23.000	23.000
Pearson Correlation	0.997	
Hypothesized Mean Difference	0.000	
df	22.000	
t Stat	0.849	
P(T<=t) one-tail	0.203	
t Critical one-tail	1.717	
P(T<=t) two-tail	0.405	
t Critical two-tail	2.074	

Period 3: 85% Confidence Interval

T-Test for RANDOM WALK State 0 Vs. LINEAR REGRESSION

	Variable 1	*Variable 2*
Mean	7.88	11.34
Variance	580.48	1515.08
Observations	23.00	23.00
Pearson Correlation	1.00	
Hypothesized Mean Difference	0.00	
df	22.00	
t Stat	-1.10	
P(T<=t) one-tail	0.14	
t Critical one-tail	1.06	
P(T<=t) two-tail	0.28	
t Critical two-tail	1.49	

Period 3: T-Test for RANDOM WALK State 0 Vs. AVERAGE

	Variable 1	*Variable 2*
Mean	7.693	11.338
Variance	599.926	1515.080
Observations	23.000	23.000
Pearson Correlation	0.998	
Hypothesized Mean Difference	0.000	
df	22.000	
t Stat	-1.198	
P(T<=t) one-tail	0.122	
t Critical one-tail	1.061	
P(T<=t) two-tail	0.244	
t Critical two-tail	1.492	

Period 3: T-Test for RANDOM WALK State 0 Vs. LAST KNOWN VALUE

	Variable 1	*Variable 2*
Mean	12.821	11.338
Variance	2167.220	1515.080
Observations	23.000	23.000
Pearson Correlation	0.997	
Hypothesized Mean Difference	0.000	
df	22.000	
t Stat	0.849	
P(T<=t) one-tail	0.203	
t Critical one-tail	1.061	
P(T<=t) two-tail	0.405	

t Critical two-tail	1.492

Period 4: 95% Confidence Interval

T-Test for RANDOM WALK State 0 Vs. LINEAR REGRESSION

	Variable 1	*Variable 2*
Mean	14.694	21.026
Variance	1991.254	3639.493
Observations	22.000	22.000
Pearson Correlation	0.995	
Hypothesized Mean Difference	0.000	
df	21.000	
t Stat	-1.793	
P(T<=t) one-tail	0.044	
t Critical one-tail	1.721	
P(T<=t) two-tail	0.087	
t Critical two-tail	2.080	

T-Test for RANDOM WALK State 0 Vs. AVERAGE

	Variable 1	*Variable 2*
Mean	11.882	21.026
Variance	1257.375	3639.493
Observations	22.000	22.000
Pearson Correlation	0.994	
Hypothesized Mean Difference	0.000	
df	21.000	
t Stat	-1.691	
P(T<=t) one-tail	0.053	
t Critical one-tail	1.721	
P(T<=t) two-tail	0.106	
t Critical two-tail	2.080	

T-Test for RANDOM WALK State 0 vs. LAST KNOWN VALUE

	Variable 1	*Variable 2*
Mean	18.308	21.026
Variance	2930.331	3639.493
Observations	22.000	22.000
Pearson Correlation	0.996	

Hypothesized Mean Difference	0.000
df	21.000
t Stat	-1.583
P(T<=t) one-tail	0.064
t Critical one-tail	1.721
P(T<=t) two-tail	0.128
t Critical two-tail	2.080

Period 4: 85% Confidence Interval
T-Test for RANDOM WALK State 0 Vs. LINEAR REGRESSION

	Variable 1	*Variable 2*
Mean	14.694	21.026
Variance	1991.254	3639.493
Observations	22.000	22.000
Pearson Correlation	0.995	
Hypothesized Mean Difference	0.000	
df	21.000	
t Stat	-1.793	
P(T<=t) one-tail	0.044	
t Critical one-tail	1.063	
P(T<=t) two-tail	0.087	
t Critical two-tail	1.494	

T-Test for RANDOM WALK State 0 Vs. AVERAGE

	Variable 1	*Variable 2*
Mean	11.882	21.026
Variance	1257.375	3639.493
Observations	22.000	22.000
Pearson Correlation	0.994	
Hypothesized Mean Difference	0.000	
df	21.000	
t Stat	-1.691	
P(T<=t) one-tail	0.053	

t Critical one-tail	1.063
P(T<=t) two-tail	0.106
t Critical two-tail	1.494

T-Test for RANDOM WALK State 0 Vs. LAST KNOWN VALUE

	Variable 1	*Variable 2*
Mean	18.308	21.026
Variance	2930.331	3639.493
Observations	22.000	22.000
Pearson Correlation	0.996	
Hypothesized Mean Difference	0.000	
df	21.000	
t Stat	-1.583	
P(T<=t) one-tail	0.064	
t Critical one-tail	1.063	
P(T<=t) two-tail	0.128	
t Critical two-tail	1.494	