

Predicting ExxonMobil's Stock Price

Conrad Polkosnik

May 8, 2018

1 Abstract

This project will create a prediction for ExxonMobil's stock price. Using historical returns, the distribution of the data will be analyzed, a Random Walk will be implemented, and Monte Carlo simulation will complete the final estimation by re-iterating the Random Walk Method n-number of times. Altogether, the same methods and procedures can be followed and manipulated to predict any stock price or asset of interest with historical returns.

2 Introduction

The stock market is one very intriguing place. Stories upon stories can be found about investors making millions and even billions of dollars. The exact opposite can be found as well, where investors have lost millions and even billions of dollars. But how does the stock market really work, one may ask? By definition, the stock market “Refers to public markets that exist for issuing, buying and selling stocks that trade on a stock exchange or over-the-counter” [CFI]. Essentially, the stock market is not just one market, but multiple. Within this market, stocks, “Represent fractional ownership in a company, and the stock market is a place where investors can buy and sell ownership of such investible assets” [CFI]. Once an investor purchases a share of stock in a company, that investor becomes part-owner in that company.

The stock market is extremely important for two major reasons. The first being that it allows for companies to go “public” on the market. Companies can raise capital from investors which allows for continued growth as well as providing the opportunity to pay short term debts and obligations. Second, investors can purchase shares of stock providing them the opportunity to share in the profits of the companies they invest in. Since the inception of the first American stock exchange in 1790, known as the Philadelphia Stock Exchange, brokers, traders, and investors have been fixated on predicting stock prices to increase their profits. The theoretical strategy known around the financial sector is to simply buy low and sell high or short the stock, meaning to bet against the stock from increase in price, by short-selling high and buying back low. Optimizing

this strategy is much more complicated than it seems. There are thousands upon thousands of different empirical models and analyzations that have been created by investment bankers, mathematicians, statisticians, and traders with all sorts of backgrounds to analyze risk and predict the stock prices in today's stock market. Within the contents of this paper, it will be found that analyzing historical returns using the theory of a Random Walk, and simulating trading days with Monte Carlo Simulation can produce a straightforward prediction of the stock price for ExxonMobil (XOM).

3 Literature Review

Modeling stock price predictions have continuously become more intricate and complex as technology has continued to improve. Computers have become so powerful that simulations and iterations of such simulations to produce predictions can be done in minutes, sometimes even seconds depending on the model created. In a general sense, a stock price model can be produced using variables of any kind to find some sort of intuition or aid in the prediction process. The art in producing the “right” model is mainly dependent upon the creator and their definition of what exactly “right” is. In the machine learning world today, there are five master algorithm’s that surround the base of model creation. These five models are known as tree models, neural networks, genetic algorithms, naïve bayes, and nearest neighbor or support vector machine models. With a non-stationary time-series analysis, as this project will facilitate, machine learning models look to maximize the out-of-sample prediction. Essentially, this means using some sort of discrete-time data to extract meaningful statistics and characteristics from a given data-set to produce a prediction. In retrospect, there have been many theses, articles, and even tutorials explaining the creation and implementation of such models using the basis of these five models mentioned above for the purpose of stock price prediction. Some examples include trading based on classification and regression trees, creation of a deep learning models, training neural networks, and the creation of Artificial Intelligence (AI) machine learning models using tensor flow to predict stock prices. Most of these models seem to produce some great predictions specifically based on what the creator

believes to be a good prediction using specific data-sets.

Many traders in today's markets are using technical analysis to make predictions on when stock prices will rise and fall using technical analysis. Technical analysis is "The study of past market action to try to gauge what the market might do in the future" [**Fidelity**]. Technical analysis is the study of price action using different ranges of time on charts. As one will see in this project, some aspects of interest in typical technical analyses include the use of Bollinger Bands, Volume, and Moving Average Convergence Divergence (MACD) lines. Bollinger Bands were created by the famous technical trader John Bollinger. The central concept behind these bands are to identify what traders call, "squeeze" opportunities. A squeeze occurs when the bands move closer together and makes the moving average narrower. Essentially, a "squeeze" signals a period of time with low trading volatility where traders may find trade opportunities and take advantage. The Moving Average Convergence Divergence (MACD) is a trend-following momentum indication. The MACD shows a relationship between two different moving averages. A famously used pair would be the EMA 26 and EMA 12, with "EMA" meaning exponential moving average, and the "26" and "12" known as the 26-day period and 12-day period. The MACD is then calculated by subtracting the 26-day EMA from the 12-day EMA producing a signal line of which is used as a buy or sell signal for traders. Lastly, integrating and analyzing the volume traded on the day or in live intra-day trading can give a trader the visual of how correlated or uncorrelated the price action is to the volume traded during a time period of interest.

As one will find within this project, the model used to create a stock price prediction implies assumptions that many models in the world may not use today. The main assumption taken on within this project is known as the Efficient Market Hypothesis. This hypothesis “asserts that the price of a stock reflects all information available and everyone has some degree of access to the information” [Xu 2007]. Essentially, this hypothesis implies that it is impossible to beat the average price in the market as the “market efficiency causes existing share prices to always incorporate and reflect all relevant information” [Investopedia]. Investors who follow this hypothesis believe, because of the randomness of the market with constant information being “priced in” to the stock price, there is essentially no way to make trades to put an investor ahead of the market average price over time. With these assumptions in mind, using the historical prices of a stock is an important piece of invaluable information that should be used in prediction models today. According to the EMH, by using these prices a potential investor can quantitatively measure the total average risk over a specified time period while also finding the return on investment as stock prices take into account all information about the company. By using the Random Walk theory along with Monte Carlo Simulation, this project will present the importance of using the daily close prices of a stock as a base for building a model. This project will go so far as to present a simulated prediction of the stock price in the future using just the historical returns alone. But what exactly is the Random Walk Theory and Monte Carlo Simulation, one may ask? The Random Walk Theory “Suggests that stock price changes have the same distribution and are

independent of each other, so the past movement or trend of a stock price or market cannot be used to predict its future movement” [**Investopedia**]. This theory takes on the assumption that the stock price is constantly changing at random and has a path of movement that cannot be predicted. On the other hand, Monte Carlo Simulation is “Used to model the probability of different outcomes in a process that cannot easily be predicted due to the intervention of random variables. It is a technique used to understand the impact of risk and uncertainty in prediction and forecasting models” [**Investopedia**]. This project will use the Random Walk Theory to produce a prediction of XOM’s stock price and then use Monte Carlo Simulation to iterate the Random Walk 100 times to analyze whether using the Random Walk Theory just once is enough to produce a proper prediction.

4 Data

Gathering proper and unbiased data is crucial to any analyzation and model to be run for any prediction process in practice. The major assumption used for this project was the Efficient Market Hypothesis. This hypothesis, in regard specifically to this project, takes on the theory that all existing stock prices always incorporate and reflect all relevant and current information. Using this theory, it would make sense to say that over time, the historical returns have continued to reflect all information about every company has ever been public. For the purpose of this project, this assumption was taken on, and the historical returns were gathered. The historical returns on XOM were gathered from Yahoo Finance since January 2nd, 1970 up until April 20th, 2018. The data itself contains the open, high, low, close and adjusted close prices for XOM for each day, totaling 73,116 observations. More specifically, the adjusted close price after each day since January 2nd, 1970 was used for analysis. Yahoo Finance is an extremely trusted source in the financial sector and can be trusted for the prediction of the stock price in this model. In regard to data used in financial modeling, the amount of data one has available as well as the time period of data by which one uses can be severely influenced by events during that time period. The data used in this project spans a range by which catastrophic events, as well as positive events, have occurred within the stock market itself by which XOM was a part of. Using this data can help to intuitively take into account such events which have influence on the risk assessment and performance of the stock itself.

5 Methods

To begin the project, it was important to load in all of the packages required in R to perform the correct functions by using the “install.packages” function built into R. The packages required to perform the functions for this project consisted of quantmod, xts, tidyverse, and stringr. The first method began with downloading all of the historical data on XOM from Yahoo Finance since January 2nd, 1970 up until April 20, 2018. Using the “getSymbols” function in the quantmod package, the stock data was easily scraped in an automated and orderly fashion. Next, an important aspect to understanding the historical returns for stock price prediction is to visualize the data. Referring to Figure 1, using the “chartSeries” function in R, the overall performance and trend of XOM can be seen. Referring to figure 2, in regard to Technical Analysis, a second visualization can be seen with additive properties using the “chartSeries” function in R. More specifically, one can change the subset property within the function to analyze a date range of interest. Within the source code, one will find it to be “2018” as the day in writing this project is April 20, 2018. With regard to prediction and possible investment from a trading perspective, it would be important to analyze the Bollinger Bands, Volume, and Moving Average Convergence Divergence based on more recent data. Such analysis would help the investor buy shares of XOM at a decent price and time in order to maximize profits.

In order to simulate data through a Random Walk and iterate with Monte Carlo Simulation, it is important to analyze the distribution of the data. Referring to

figure 3, after automating the log daily returns and using the `ggplot2` function in the `tidyverse` package to produce a histogram, we see that the distribution of log daily returns is in fact a normal distribution, which verifies that the data can be used for a Random Walk method and Monte Carlo Simulation. By simply analyzing the mean log daily return and standard deviation in the log daily return, a percentage in daily return can be found if one was to invest in XOM today.

The next process used was the Random Walk model. “A random walk is a statistical phenomenon where a variable follows no discernible trend and moves seemingly at random” [CFI]. Such a theory is applied to trading stocks in the market and is exactly what has been done in the next part of this project. Within the parameters of the model built, one will find the number of trading days, N , to be 252 as there are roughly 252 trading days on a per year basis. In regard to the scraped data pulled for the model, the adjusted historical returns were used. Using a random seed number, or random number generation for reproducibility, a simulation was produced. Referring to figure 4, a plot was created using the `ggplot` function to visualize the simulated data from day 1 to day 252. After analyzing the simulation, it was found that an investor would lose money by investing in XOM today. Although, an important question to ask is could this one simple simulation be trusted? This question leads into the next method, which can be thought of as simulating the simulation itself. This iteration process is better known as Monte Carlo Simulation. Within the parameters of this model, we have the same number of stock price simulations,

252, but a new parameter M , for the number of Monte Carlo Simulations. For the purpose of better visualization, M , was set to 100 to iterate the random walk simulation 100 times. Using a for-loop, the data was stored together within matrices. Referring to figure 5, a plot was then graphed using the ggplot function to visualize the findings. Referring to figure 6, an estimate of the XOM stock price was then found based on setting likelihood probabilities.

6 Findings

For the first finding using the `chartSeries` function, one can visualize how the stock price of XOM has performed over time. One can analyze over time that the stock price of XOM continues to increase in the long-term. With regard to Technical Analysis, one can find useful information of any time frame desired. For the purpose of the project, the subset used for analysis was data in the year, 2018. For example, one take-away from figure 2 can be seen that between January 22nd and February 12th, there was a large sell off in volume, along with a convergence of the two Bollinger Bands. Due to the sell off and “squeeze” play that occurred in regard to the convergence of the Bollinger Bands, the stock price decreased throughout the period of time.

In the next finding, the gathering of the daily log returns was produced. This analysis allowed for the analyzation of the distribution of the data set. After plotting the data set using the daily log returns, we were able to find a normal distribution. By finding a normal distribution, we can analyze the data set using probability characteristics to find the likelihood of a daily average return if one was to invest in XOM. After finding the mean and standard deviation of the return data, the exponential of the mean log return was taken. It was found that the average daily return on investment in XOM would be 0.0658 percent.

In the second half of the experiment, the analysis of the data using the Random Walk Theory along with Monte Carlo Simulation was produced. To begin, the Random Walk Theory was created to simulate 252 trading days (random walks) in order to find the estimated stock price approximately one year’s worth of

trading from today. Because the mean and standard deviation of the log return data was found, this random walk experiment can in fact be produced. After running a for-loop in R can be found here:

```
for(i in 2:N) {  
  Price[i] <- Price[i-1] * exp(rnorm(1, mu, sigma))  
}
```

Following the simulation of prices, a visualization was created to analyze the experiment. It was found that if one was to invest in XOM and look to sell in exactly one year's time, the investor would actually lose money on the investment. This is quite interesting as the average return found previously from the data itself was positive and not negative. In a sense, there is no way that this contradiction in running the random walk experiment just one time can be trusted. In order to test this experiment with better accuracy, Monte Carlo Simulation was produced. The following set of equations were used to create a matrix as well as simulate:

```
MonteCarloMatrix <- matrix(nrow = N, ncol = M)  
  
for (j in 1:M) {  
  MonteCarloMatrix[[1, j]] <- InitialPrices  
  
  for(i in 2:N) {  
    MonteCarloMatrix[[i, j]] <- MonteCarloMatrix[[i - 1, j]]  
    *exp(rnorm(1, mu, sigma))  
  }  
}
```

After running the Monte Carlo Simulation 100 times, a visualization of the experiment was created. The visualization shows the extent of the Random Walk experiment done 100 times, producing 100 different possible outcomes. Using this data and setting up probabilities of interest, it was found that the most likely estimated stock price for XOM was 89.17 using the 95 percent confidence interval. In just one year's time, if an investor was to purchase shares of XOM today, the investor would make just over a 12.5 percent return.

7 Conclusion

The stock market is an action-packed and intriguing place to say the least. Many believe the market to be somewhat of a casino, guessing when to buy and guessing when to sell. On the other hand, many have referred to the stock market as a “get-rich-quick scheme”, building complex statistical models with all sorts of variables to produce an intricate prediction of the stock price. In regard to the prediction produced in this project, by taking the range of data over the last 48+ years, one can measure the average risk and return on investment. The oil industry tends to have its ups and downs based on events in specific time periods. By analyzing the data over a wide range of time, the “priced in” metric also represents events of which have been detrimental and positive to the influence on the stock price. Such a measurement can take on the assumption that outliers of data which do in fact exist in the historical data pulled in this project can represent that same measure going into the future when producing a prediction.

Furthermore, after scraping the data from a trusted source, analyzing the distribution from January 2nd, 1970 to April 20th, 2018, and producing the prediction itself, it was found that the expected stock price after 252 simulated trading days was 89.17, based specifically on the parameters created in this model alone. If one was to take a different range of data to analyze, the prediction would be much different. However, it is with personal belief that using more data would allow for a better and more conservative measure of risk related to investment, taking into account outliers. The ultimate goal of this project, not only for this

class but for myself, was to improve my coding skills. After completing this project, I better understand statistical modeling and simulation which can be helpful when moving into industry as an investment banker. As for this course, I touched on quite a few concepts we learned throughout the course and found them to be extremely positive to my learning as a whole. Also, this project allowed me to find real-world application using statistical theory learned in the classroom. As I plan to move into industry in the near future, I found this aspect of critical thinking to be of extreme importance throughout the creation of this project.

8 References

“Random Walk Theory - Definition, History, Implications of the Theory.” Corporate Finance Institute, corporatefinanceinstitute.com/resources/knowledge/trading-investing/what-is-the-random-walk-theory/.

Staff, Investopedia. “Efficient Market Hypothesis (EMH).” Investopedia, 28 Sept. 2016, www.investopedia.com/terms/e/efficientmarkethypothesis.asp.

Staff, Investopedia. “Monte Carlo Simulation.” Investopedia, 18 Mar. 2018, www.investopedia.com/terms/m/montecarlosimulation.asp.

Staff, Investopedia. “Philadelphia Stock Exchange.” Investopedia, 25 Nov. 2003, www.investopedia.com/terms/p/phlx.asp.

Staff, Investopedia. “Random Walk Theory.” Investopedia, 14 Nov. 2015, www.investopedia.com/terms/r/randomwalktheory.asp.

“Stock Market - What Is the Stock Market and How It Works.” Corporate Finance Institute, corporatefinanceinstitute.com/resources/knowledge/trading-investing/stock-market/.

“What Is Technical Analysis?” Fidelity, 1998, www.fidelity.com/learning-center/trading-investing/technical-analysis/introduction-technical-analysis/what-is-technical-analysis.

Xu, Selene Yue. “Stock Price Forecasting Using Information from Yahoo Finance and Google Trend.” UC Berkeley.

9 Source Code

```
title: "Final Project"
author: "Conrad Polkosnik"
date: "4/20/2018"
output: word_document
```

Loading Packages

```
```{r}

library(quantmod) # get stock prices & useful stock analysis functions
library(xts) # extensible time series
library(tidyverse) # ggplot2, purrr, dplyr, tidyr, readr, tibble
library(stringr) # working with strings
```
```

Downloading Stock Data

```
```{r}

Calling quantmod
library("quantmod")

Gathering historical stock data for Exxon Mobil since 1970
```

```
getSymbols("XOM", src='yahoo', from="1970-01-01")
```

```
'''
```

```
'''{r}
```

```
Understanding data pulled
```

```
library(magrittr) #for pipe operator
```

```
XOM %>% class()
```

```
XOM %>% str()
```

```
XOM %>% head()
```

```
'''
```

```
Visualize Trend
```

```
'''{r}
```

```
XOM %>% Ad() %>% chartSeries()
```

```
'''
```

```
Fundamental Technical Analysis
```

```

'''{r}

XOM %>% chartSeries(TA='addBBands();

 addBBands(draw="p");

 addVo();

 addMACD() ',

 subset='1970-2018',

 theme="white"

)

```

'''

Gathering log daily returns for XOM

```

'''{r}

XOM %>% Ad() %>% dailyReturn(type = 'log') %>% head()

```

'''

Using log returns to analyze distribution

```

'''{r}

XOMLogReturns <- XOM %>% Ad() %>% dailyReturn(type = "log")

```

```

names(XOMLogReturns) <- "XOM.Log>Returns"

```

```

Plotting the log returns using a histogram

```

```

library(ggplot2)

XOMLogReturns %>% ggplot(aes(x = XOM.Log>Returns))

+ geom_histogram(bins = 100) + geom_density() + geom_rug(alpha = 0.5)

'''

Setting probabilities for study of liklihood

'''{r}

Probabilities <- c(.005, .025, .25, .5, .75, .975, .995)

DistributionalLogReturns <- XOMLogReturns %>%

quantile(probs = Probabilities, na.rm = TRUE)

DistributionalLogReturns

'''

'''{r}

MeanLogReturns <- mean(XOMLogReturns, na.rm = TRUE)

StDevLogReturns <- sd(XOMLogReturns, na.rm = TRUE)

MeanLogReturns

StDevLogReturns

'''

```

Finding daily average increase or decrease in return

```
''{r}
```

```
MeanLogReturns %>% exp()
```

```
''
```

Analysis: On average, the investor should expect 0.0657% increase daily in return.

Setting Parameters for Random Walk

```
''{r}
```

```
Parameters
```

```
N <- 252
```

```
mu <- MeanLogReturns
```

```
sigma <- StDevLogReturns
```

```
day <- 1:N
```

```
InitialPrices <- XOM$XOM.Adjusted[[nrow(XOM$XOM.Adjusted)]]
```

```
set.seed(386) # for reproducibility
```

```
Price <- c(InitialPrices, rep(NA, N-1))
```

```
for(i in 2:N) {
```

```
 Price[i] <- Price[i-1] * exp(rnorm(1, mu, sigma))
```

```
}
```

```

library(tidyverse)

PriceSimulation <- cbind(day, Price) %>% as_tibble()

Visualize price simulation

PriceSimulation %>%

 ggplot(aes(day, Price)) +

 geom_line() +

 ggtitle(str_c("XOM: Simulated Prices for ", N," Trading Days"))

'''

Monte Carlo Simulation for Random Walk

'''{r}

Parameters

N <- 252 # Number of Stock Price Simulations

M <- 100 # Number of Monte Carlo Simulations

mu <- MeanLogReturns

sigma <- StDevLogReturns

day <- 1:N

InitialPrices <- XOM$XOM.Adjusted[[nrow(XOM$XOM.Adjusted)]]

Simulate prices using for loop and Monte Carlo Matrix

set.seed(386)

```

```

MonteCarloMatrix <- matrix(nrow = N, ncol = M)

for (j in 1:M) {

 MonteCarloMatrix[[1, j]] <- InitialPrices

 for(i in 2:N) {

 MonteCarloMatrix[[i, j]] <- MonteCarloMatrix[[i - 1, j]] * exp(rnorm(1, mu, sigma))

 }

}

Organizing data frame and format

PriceSimulation <- cbind(day, MonteCarloMatrix) %>% as_tibble()

nm <- str_c("Sim.", seq(1, M))

nm <- c("Day", nm)

names(PriceSimulation) <- nm

PriceSimulation <- PriceSimulation %>%

 gather(key = "Simulation", value = "Stock.Price", -(Day))

Visualization

PriceSimulation %>%

 ggplot(aes(x = Day, y = Stock.Price, Group = Simulation)) +

 geom_line(alpha = 0.1) +

 ggtitle(str_c("XOM: ", M,

 " Monte Carlo Simulations for Prices Over ", N,

 " Trading Days"))

```



```

'''

Stock price estimation and likelihood

'''{r}

EstimatedStockPrices <- PriceSimulation %>% filter(Day == max(Day))

Probabilities <- c(.005, .025, .25, .5, .75, .975, .995)

DistributionalEstimatedStockPrices <-
quantile(EstimatedStockPrices$Stock.Price, probs = Probabilities)

DistributionalEstimatedStockPrices %>% round(2)

'''

```

## 10 Appendix



Figure 1: Stock Price Performance



Figure 2: 2018 Technical Analysis

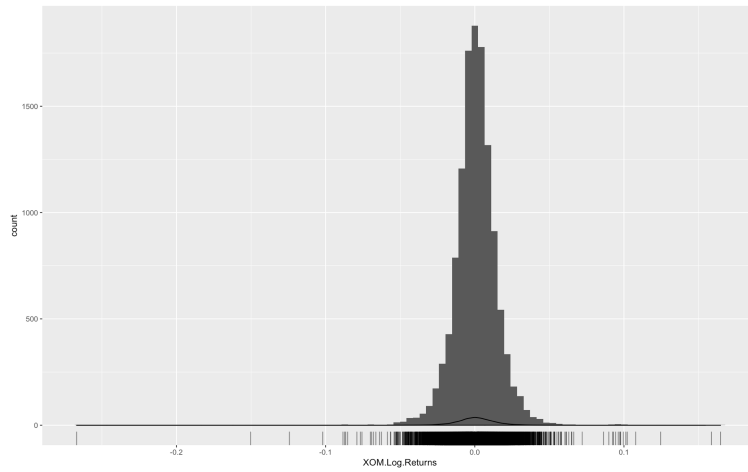


Figure 3: Log Daily Returns

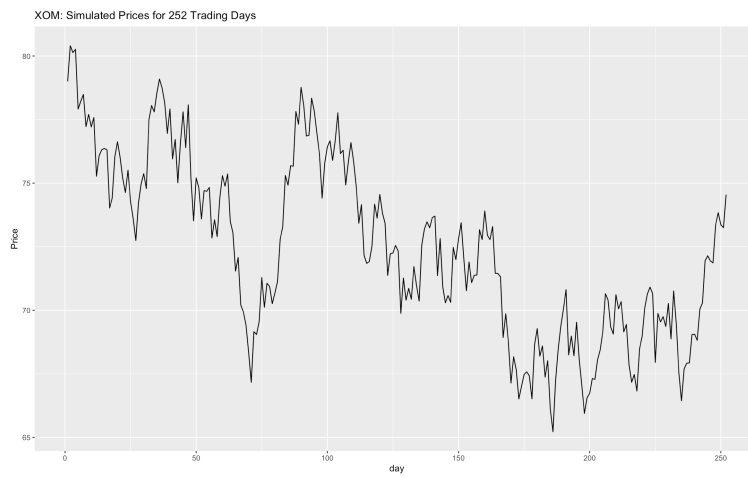


Figure 4: Random Walk

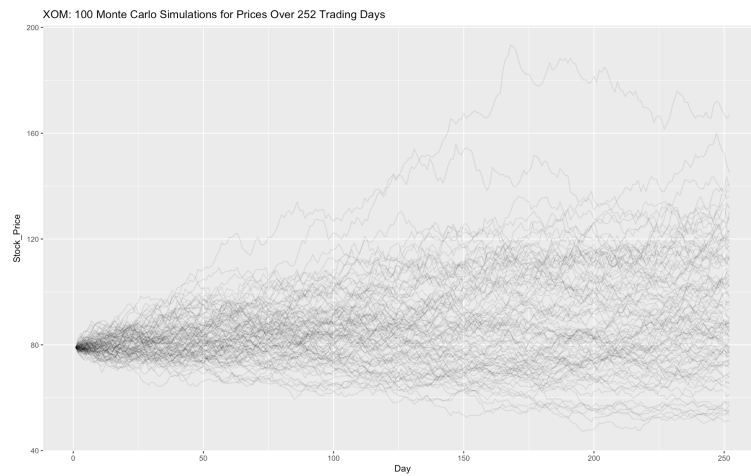


Figure 5: Monte Carlo Simulation

0.5%	2.5%	25%	50%	75%	97.5%	99.5%
52.29	54.72	77.17	89.17	112.95	140.11	156.13

Figure 6: Estimated Prices