University of Notre Dame
College of Engineering
Risk Assessment

Statistical Analysis of Daily Stock Market Trends

Molly Doyle & Chase Saca

June 30 2019

TABLE OF CONTENTS

Abstract

Using three years of daily changes in stock price for Apple, Facebook, and General Electric stock, we will predict a range of potential future stock prices based on gaussian, exponential, and cauchy distributions. After fitting the raw data of daily price changes to each distribution, we will input the starting stock price for each, then generate daily changes over the desired time interval to generate the stock price for each day in the time interval. We will then use this data to generate the mean, 95^{th} and 5^{th} percentile stock prices for each distribution and weigh them equally to produce three possible values for each stocks' future price. The resulting prices will allow us to conclude which stocks were most advisable to invest in and which posed the most risk. We concluded that Apple, while risky, is the best to invest in due to the highest potential for profit and its historical increasing trend. While our results provide rough estimates about potential growth or decline of stocks, limitations in expertise and the natural randomness of stocks, especially in fast growing tech companies, limited the accuracy of our results.

Introduction

When deciding a project idea, we wanted to attempt a project which applied the concepts we learned throughout the course in a useful and meaningful way. We were drawn to analyze stocks because of how prominent the stock market is in professional industry and because we both hope to invest in the stock market once we enter the workforce. By analyzing the stock market using the methods we learned in Risk Assessment, we will have a deeper understanding of the stock market, its trends, and their implications. Although we are both computer science majors, having a more intimate understanding of the trends of the stock market will allow us to have better insight in investing and understanding the meaning of stock trends, which will aid us in having a better understanding of the stocks in terms of their risk and historical trends.

Our project attempts to use the daily change in value of stock prices (based on the daily opening value) over the past three years to predict the average, potential highest, and potential lowest possible stock price based on Gaussian, Exponential, and Cauchy distributions for Apple, Facebook, and General Electric (GE) stocks. Through these predictions, we attempt to give investors insight into which stocks to invest in. By using statistical analysis on historical data, we generate corresponding distributions to each stock's historical changes, then synthesize and weight these distributions to derive equations which produce predictions of potential stock prices according to user desired length of time and the stock starting price. We chose these three stocks because of the diversity of their trends, as Apple has as positive, primarily increasing trend, Facebook has a very tumultuous, unpredictable trend, and GE has a very negative, continuously decreasing trend, as shown in the figures below.

Figure 1. Three Year Historical Trend for Apple, Inc. Stock

Figure 2. Three Year Historical Data for Facebook, Inc. Stock

Figure 3. Three Year Historical Data for General Electric Co. Stock

Recognizing that we picked an extremely difficult, arguably impossible goal of predicting changes in stock prices, we are limited by the sheer randomness of the stock market as well as our lack of expertise in financial trends. The very nature of the stock market is its unpredictability, which is affected by various factors, like possible recessions or stock market booms, which we chose to ignore in order to focus exclusively on the statistical analysis and understanding of the various stocks. In addition, after deriving our distributions and equations, we lacked the expertise to choose certain weights for each distribution to more accurately predict the stock price changes; this expertise comes from familiarity with stocks and finance.

Objective: Based on three years of historical data for daily price changes of Apple, Facebook, and General Electric stocks, we will derive predictions for the average, 95th percentile, and 5th percentile stock prices using Gaussian, Exponential, and Cauchy distributions. We will then use this data to guide potential investment decisions in order to maximize profit.

The Engineering Model

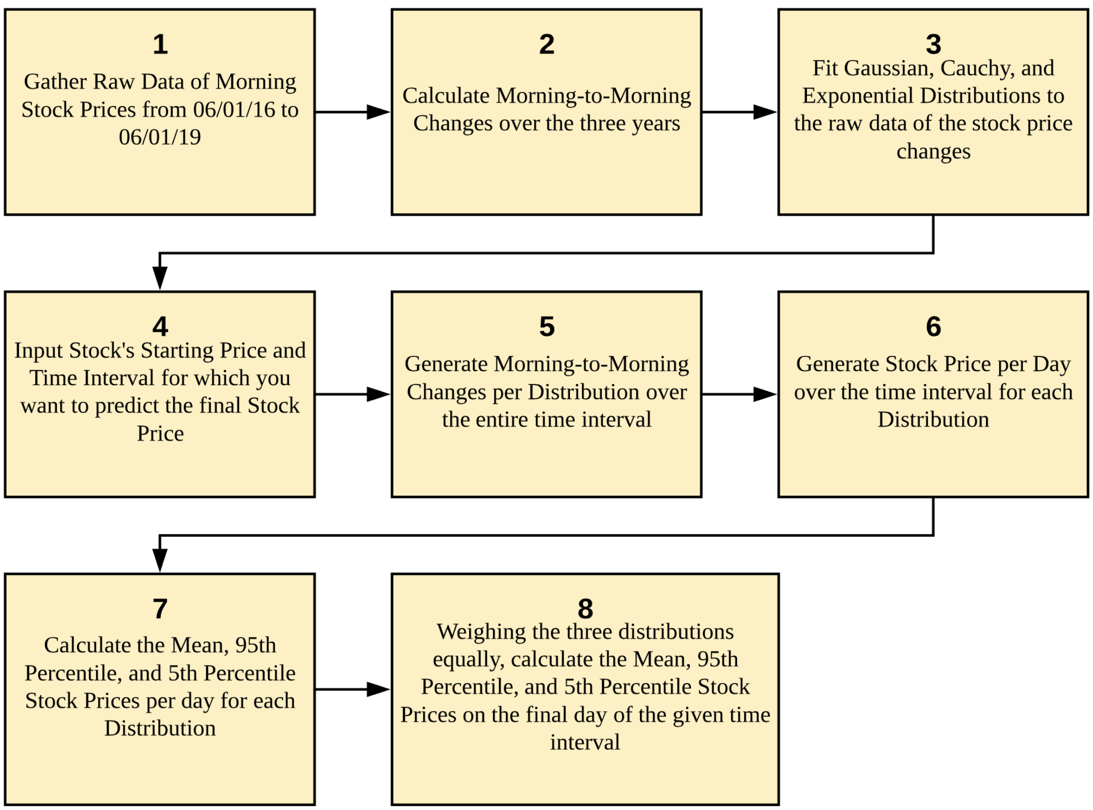

Figure 4. Flow Chart of System on a Per Stock Basis

1. We gathered the raw data for the model, which is the opening stock prices for each day over a span of three years from June 1st, 2016 to June 1st, 2019. We gathered three sets of this raw data, one for each stock: Apple, Facebook, and GE.
2. For each of the three stocks, we calculated the changes between each morning's opening price. Since we are predicting the change in the stock price over a specified time interval, this calculated data is the data necessary to create our distributions. After inputting this

data into MATLAB, we made a histogram for each stock of their corresponding morning-to-morning price changes.

3. Due to the shape of the histograms, we decided that the three best distribution fits for the data are gaussian, exponential, and cauchy distributions, and we generated these using the CDF method. In the following equations, rand is a random number from 0 to 1, while randn is a normally distributed random number, and both the standard deviation and the mean are calculated from the original raw data set per each stock.
   a. When fitting the gaussian distribution, we used Eqn 1 in order to generate a vector of x-values that are normally distributed. From these x-values, we plotted a histogram, which is the gaussian distribution for the data.

$$x = randn * standard\ deviation + mean$$

Eqn. 1

   b. In order to derive Eqns 3 and 4, we set Eqn 2, which is the Exponential CDF equation, equal to rand, and then we solved for the variable x. When fitting the exponential distribution, we used Eqns 3 and 4 in order to generate a vector of x-values that are exponentially distributed on both the positive and negative sides of the x-axis. From these x-values, we plotted a histogram, which is the exponential distribution for the data.

$$P(x) = 1 - e^{-\lambda x}$$ **Eqn. 2**

$$x = \frac{e^{-rand+1}}{-\lambda}$$

Eqn. 3

$$x = -\frac{e^{-rand+1}}{-\lambda}$$ **Eqn. 4**

   c. In order to derive Eqn 6, we set Eqn 5, which is the Cauchy CDF equation, equal to rand, and then we solved for the variable x. When fitting the cauchy distribution, we used Eqn 6 in order to generate a vector of x-values that are cauchy distributed. From these x-values, we plotted a histogram, which is the cauchy distribution for the data.

$$P(x) = \frac{1}{\pi} arctan(\frac{x-x_0}{\gamma}) + \frac{1}{2}$$ **Eqn. 5**

$$x = \gamma * tan(\pi * rand - \frac{1}{2}\pi) + mean$$ **Eqn. 6**

4. We created the variables StockStartingPrice and timeDays, which are inputs from the user, that will be used in order to predict the stock price after the specified time interval. StockStartingPrice is the initial stock price, while timeDays is the number of days after the initial stock price that we want to predict the price.
5. Similarly to Step 3, we predicted the morning-to-morning changes of the stock price over a specified number of days through the CDF Method. First, we generated a vector of random numbers, rand, or randn for the gaussian distribution, that was the size of the specified number of days, timeDays. Next, we plugged these random numbers into one of the derived equations for x, depending on the distribution, where x is the predicted

changes between mornings over the specified time interval. We did this for all three distributions. For the gaussian and exponential distributions, we ran 100 simulations and, for the cauchy distribution, we ran 10,000 simulations.

6. Using the predicted changes from Step 5, we then calculated the actual predicted stock price for each day from each distribution. Eqn 7 shows that the current day's stock price is equal to the previous day's stock price plus the amount of change that happens between the two days. For the gaussian and exponential distributions, we ran 100 simulations and, for the cauchy distribution, we ran 10,000 simulations.

$$StockPrice_{\ n} = StockPrice_{\ n-1} + Change$$

Eqn. 7

7. Utilizing the MATLAB functions mean and prctile, we found the means, 95th percentiles, and 5th percentiles per day of the predicted stock prices. We did this per distribution, which left us with three sets of results.
8. Finally, we weighed the distributions equally, as shown in Eqn 8, in order to calculate three final results, which are a predicted mean stock price, 95th percentile stock price, and 5th percentile stock price on the last day of the specified interval.

Prediction=(⅓)(CauchyVal)+(⅓)(GaussianVal)+(⅓)(ExponentialVal) **Eqn. 8**

In the following sections, we will mostly discuss the final three numbers per stock that we calculated from Eqn 8, which are the predicted mean stock price, 95th percentile stock price, and 5th percentile stock price on the last day of the specified interval. Through these values, we will determine which stocks are the most valuable to invest in, depending on the risk that the investor is willing to take.

Model Implementation

When gathering data for this project, we first gathered our raw data, which was the opening stock prices for each day over a span of three years from June 1st, 2016 to June 1st, 2019, which is 756 values. We gathered three sets of this raw data, one for each stock: Apple, Facebook, and GE. For each of the three stocks, we then calculated the changes between each morning's opening price. Since we are predicting the change in the stock price over a specified time interval, which in turn allows us to predict the final stock price, this calculated data is the data necessary to create our distributions. Therefore, these changes from day-to-day are the raw data that we inputted as a vector of 755 values into MATLAB in order to fit our distributions to the stocks and, eventually, to predict the stock prices at the end of the time interval. Since we are fitting the distributions to the original data set, both the standard deviations and the means in Eqns 1 and 6 are calculated from the original data sets for each stock.

The means of the stock price changes are:

Apple: 0.1015

Facebook: 0.0806

GE: -0.0260

The variances of the stock price changes are:

Apple: 6.2341

Facebook: 9.8013

GE: 0.0957

The standard deviations of the stock price changes are:

Apple: 2.4968

Facebook: 3.1307

GE: 0.3093

In Step 3b, when fitting an exponential distribution to our raw data, we used Eqns 3 and 4 in order to generate a vector of x-values that are exponentially distributed. In order to find λ for Eqns 3 and 4, we took the inverse of the mean of our raw data.

Apple λ: 0.6626

Facebook λ: 0.5499

GE λ: 4.0974

In Step 3c, when fitting a cauchy distribution to our raw data, we used Eqn 6 in order to generate a vector of x-values that are cauchy distributed. In order to find a γ for Eqn 6, we used a trial-and-error method while testing the error of the cauchy histogram's line to the actual line of the raw data. Thus, we found that setting γ to the values below created a low enough error in order to create a close cauchy fit to the data.

Apple γ: 1.2

Facebook γ: 1.2

GE γ: 0.1

Apple Cauchy Error: around 52-120

Facebook Cauchy Error: around 45-115

GE Cauchy Error: around 210-340

Ultimately, in order to implement the steps described in The Engineering Model, we manipulated variables, used FOR loops and the MATLAB random number generator, and analyzed numerous figures, including both histograms and plots.

While performing the CDF Method in Step 3, we ran through a FOR loop for each distribution that created a vector of x-values that would be made into a histogram of the designated distribution. More specifically, in Step 3b, we chose between using either Eqn 3 or 4 by random number generation and logic statements. If the generated random number was >=0.5 then the positive Eqn 3 was used; if not, then the negative Eqn 4 was used, as shown in the code below.

```
U=rand(1,N); % generate vector of random numbers

% iterate through xExp and U in order to generate exponentially
% distributed x-values for the exponential distribution
for i=1:1:N
    % if random number >= 0.5 , take the positive x-value
    if rand(1) >= 0.5
        xExp(i) = log(-U(i)+1) / -lambda;
    else % if random number < 0.5 , take the negative x-value
        xExp(i) = - log(-U(i)+1) / -lambda;
    end
    % shift x-values in order to match raw data
    xExp(i) = xExp(i) + meanStocks;
end
```

Figure 5. Example Code for Step 3

As shown in Figure 6, while performing the CDF Method in Step 5, we ran through two FOR loops for each distribution; one iterated through the specified time interval and was nested inside the other, which iterated through the number of simulations that we had set. For the gaussian and exponential predictions, we ran 100 simulations; though, for the cauchy predictions, we ran 10,000 simulations due to the nature of the cauchy's distribution. The cauchy distribution accounts for large jumps and drops, like out of the ordinary occurrences. In the nested FOR loops, we generated a vector of random numbers that was the length of the specified time interval. As the inside loop iterates, it iterates through the vector of random numbers and plugs in the day's corresponding random number into one of the derived equations for x, depending on the distribution, in order to generate the matrix of predicted day-to-day changes for one simulation. With Y days and 100 simulations, the resulting variable would be a 100xY matrix, where the number of simulation is the row number and the number of the day is the column number.

```
%% cauchy prediction
% initialize matrix that will hold stock price predictions each day
% over entire interval and through 100 simulations of that time interval
stockPriceCauch = zeros(10000,timeDays);

% generate cauchy distributed changes from morning-morning
for i=1:1:10000
    for p=1:1:timeDays
        randNums = rand(1,timeDays);
        cauchPredict(i,p) = gam * tan(pi*randNums(p) - 0.5*pi) + meanStocks;
    end
end

% generate stock price per day over entire time interval for 100 simulations
for i=1:1:10000
    for p=1:1:timeDays
        stockPriceCauch(i,p+1)=stockPriceCauch(i,p)+cauchPredict(i,p);
    end
end
```

Figure 6. Example Code for Step 5 and Step 6

As shown in Figure 6, in Step 6, we ran through two FOR loops; one iterated through the specified time interval and was nested inside the other, which iterated through the number of simulations that we had set. For the gaussian and exponential predictions, we ran 100 simulations; though, for the cauchy predictions, we ran 10,000 simulations again. In the nested FOR loops, we generated a matrix that would hold the predicted stock prices on each day. With Y days and 100 simulations, the resulting variable would be a 100xY matrix, where the simulation number corresponds to the row number and the day number corresponds to the column number. As the inside loop iterates, it also iterates through the matrix that holds the morning-to-morning stock price changes and plugs in the day's corresponding price change into Eqn 7 in order to generate the matrix of stock prices for one simulation.

In Step 8, we first found the means, 95th percentiles, and 5th percentiles, of the 100 possible stock prices on the last day of the specified interval. As shown in the figure below, we scaled the means correctly by adding the stock starting price, and we scaled the percentiles by adding both the stock starting price and the previously calculated prediction means. Using the means, 95th percentiles, and 5th percentiles that we had just calculated, we took (⅓) of each value for each distribution and added their values together. Essentially, we weighed each distribution's prediction the same in order to get three final results: the predicted mean stock price, the predicted 95th percentile stock price, and the predicted 5th percentile stock price.

```
%% predict the stock price average, low, and high on the final day of the
% time interval

% mean prediction
mC = meanCauch(timeDays+1) + StockStartingPrice;
mE = meanExp(timeDays+1) + StockStartingPrice;
mG = meanGauss(timeDays+1) + StockStartingPrice;

predictionMean = (1/3)*(mC) + (1/3)*(mE) + (1/3)*(mG);

% high prediction
hC = kiwi(timeDays+1);
hE = percExpHigh(timeDays+1) + meanExp(timeDays+1) + StockStartingPrice;
hG = percGaussHigh(timeDays+1) + meanGauss(timeDays+1) + StockStartingPrice;

predictionHigh = (1/3)*(hC) + (1/3)*(hE) + (1/3)*(hG);

% low prediction
lC = mango(timeDays+1);
lE = percExpLow(timeDays+1) + meanExp(timeDays+1) + StockStartingPrice;
lG = percGaussLow(timeDays+1) + meanGauss(timeDays+1) + StockStartingPrice;

predictionLow = (1/3)*(lC) + (1/3)*(lE) + (1/3)*(lG);
```

Figure 7. Example Code for Step 8

Results

When calculating our results, we used a starting stock price of $100 to predict the stock price after 30 days for each stock. When the programs run per each stock, each program generates 100 simulations for the predicted change in price per day for the Gaussian and Exponential distributions and 10,000 simulations for the Cauchy. Because a random number generator is used throughout the entirety of the program in each derived equation from the distributions CDFs, the variables change every time, which produces different distributions and price predictions each time as well.

Using the inputted data of daily changes of Apple, Facebook, or GE stock prices, we fit Gaussian, Exponential, and Cauchy distributions to this historical data though the CDF method. Using the normally, exponentially, or cauchy distributed x-values that we generated, we made histograms of and plotted the line of best fit per distribution per stock. The figures below show the three distributions for the three stocks.

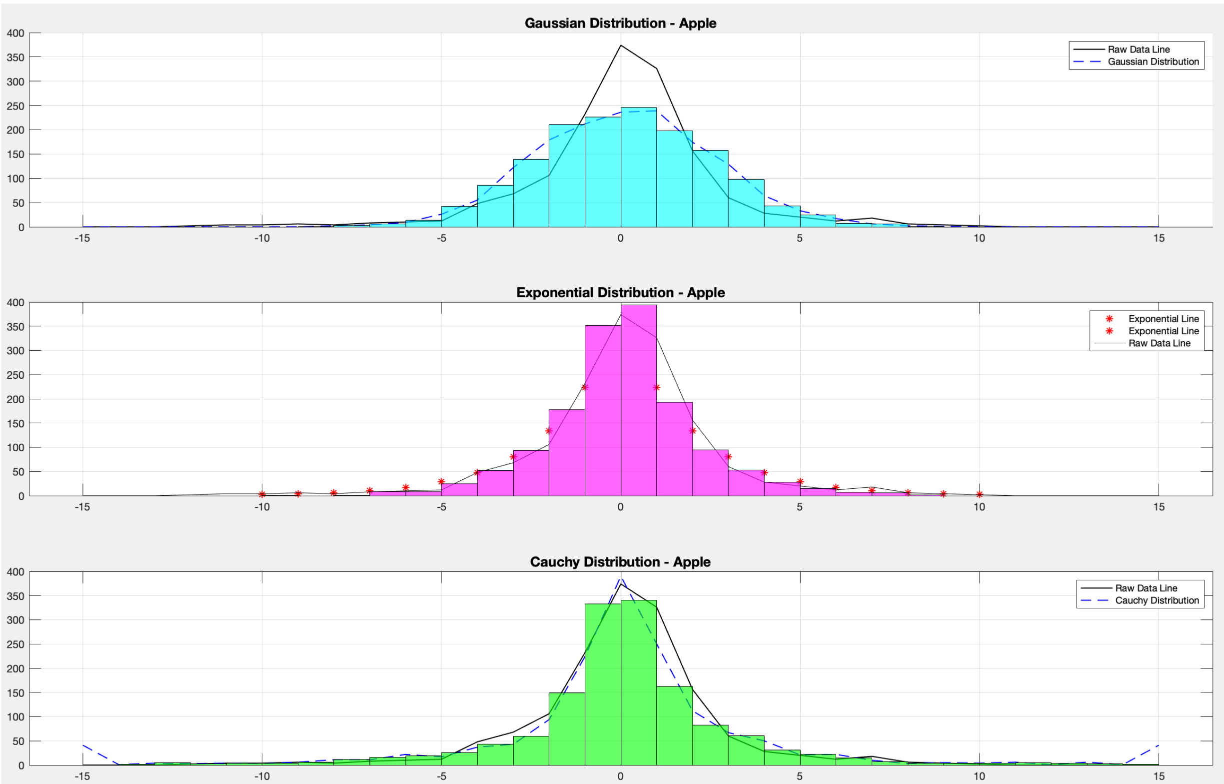

Figure 8. Gaussian, Exponential, and Cauchy Histograms and Trendlines for Apple Data

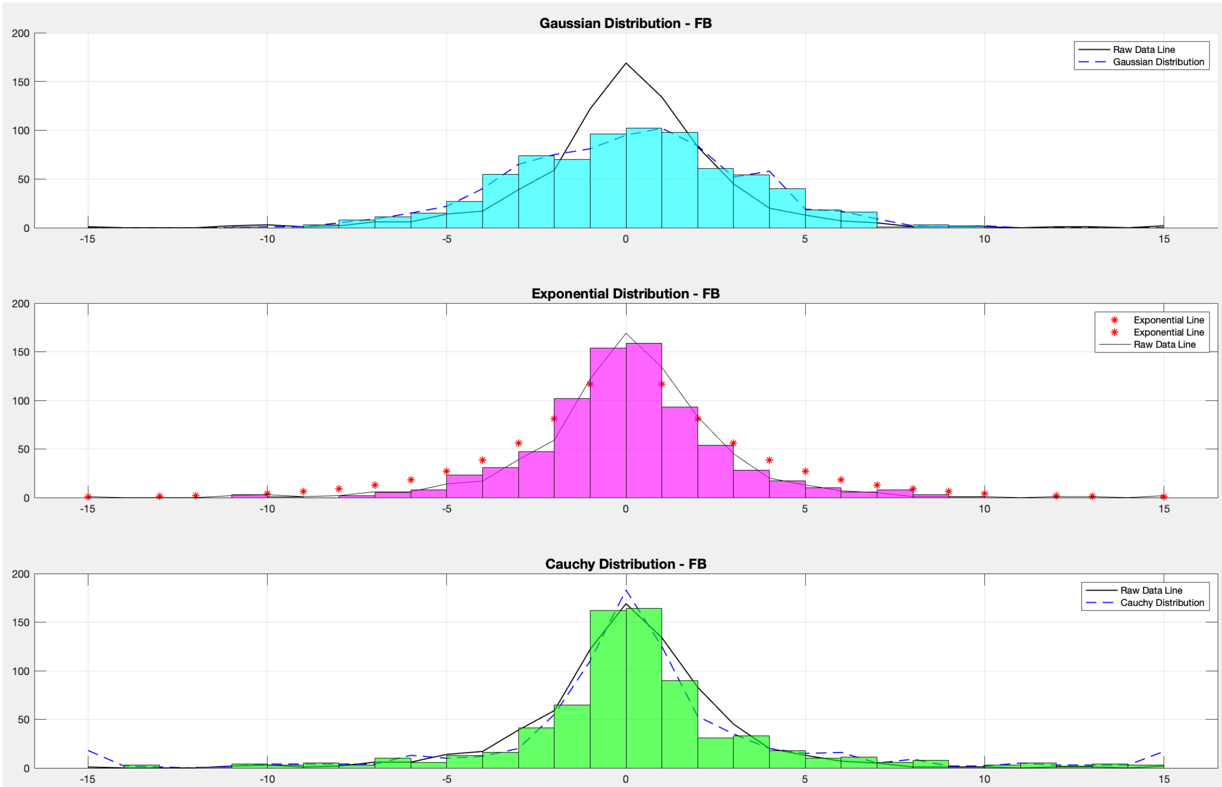

Figure 9. Gaussian, Exponential, and Cauchy Histograms and Trendlines for Facebook Data

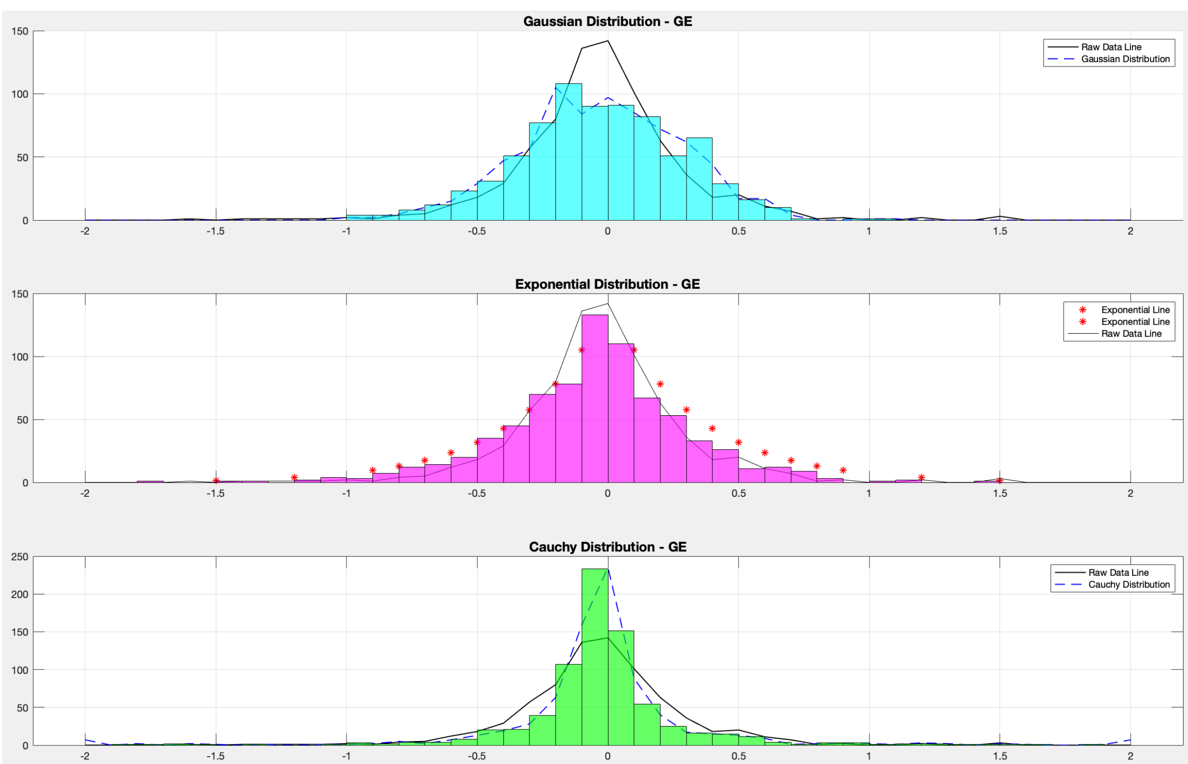

Figure 10. Gaussian, Exponential, and Cauchy Histograms and Trendlines for GE Data

After the distributions were made for each stock, these distributions and their CDF equations were used in order to predict the daily change in each stock's price over a specified interval. Using the stock's starting price at the beginning of the interval and the calculated daily changes, we predicted the final mean, 95th highest percentile, and 5th lowest percentile stocks' prices at the end of the interval.

In single program run for the Apple stock, a mean stock price of $101.25, a 5th percentile stock price of $55.95, and a 95th percentile stock price of $194.51 was predicted. Figure 11, which is shown below, shows the plots for the predicted mean, 95th percentile, and 5th percentile stock prices per day for each distribution.

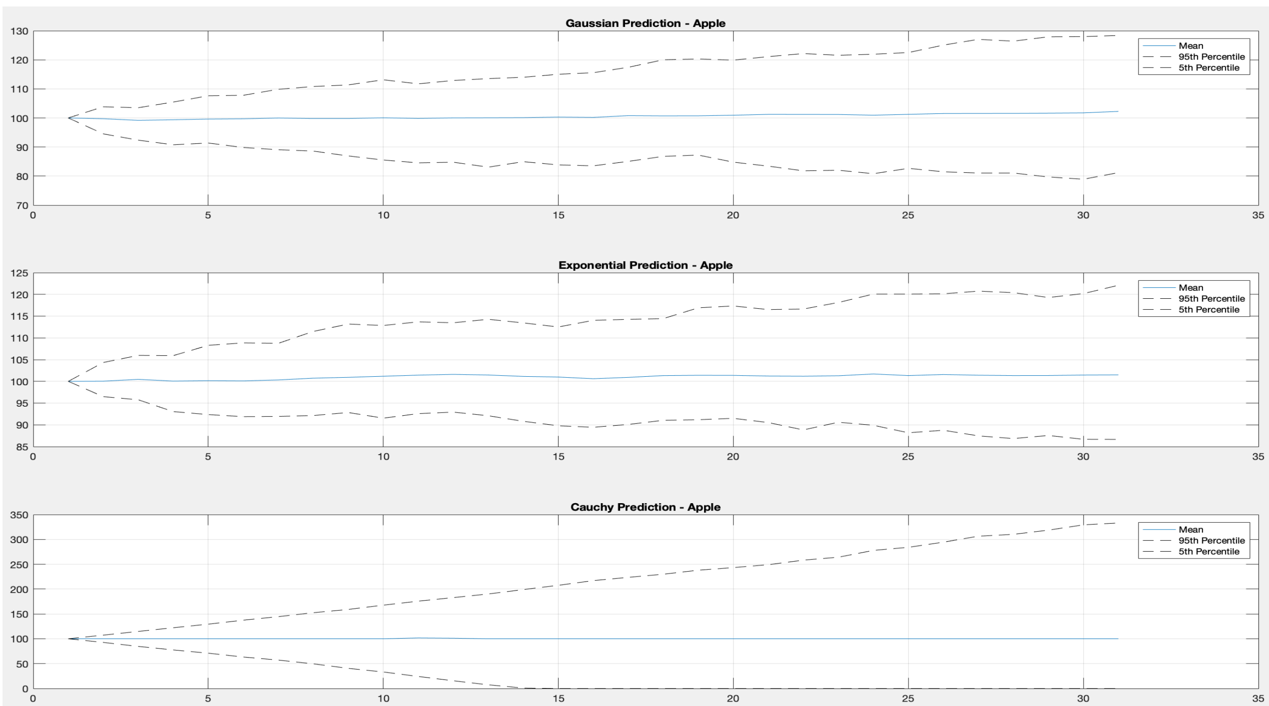

Figure 11. Mean, 95th Percentile, and 5th Percentile Stock Price Predictions Based On Gaussian, Exponential, and Cauchy Data for Apple

In single program run for the Facebook stock, a mean stock price of $99.72, a 5th percentile stock price of $50.42, and a 95th percentile stock price of $195.97 was predicted. Figure 12, which is shown below, shows the plots for the predicted mean, 95th percentile, and 5th percentile stock prices per day for each distribution.

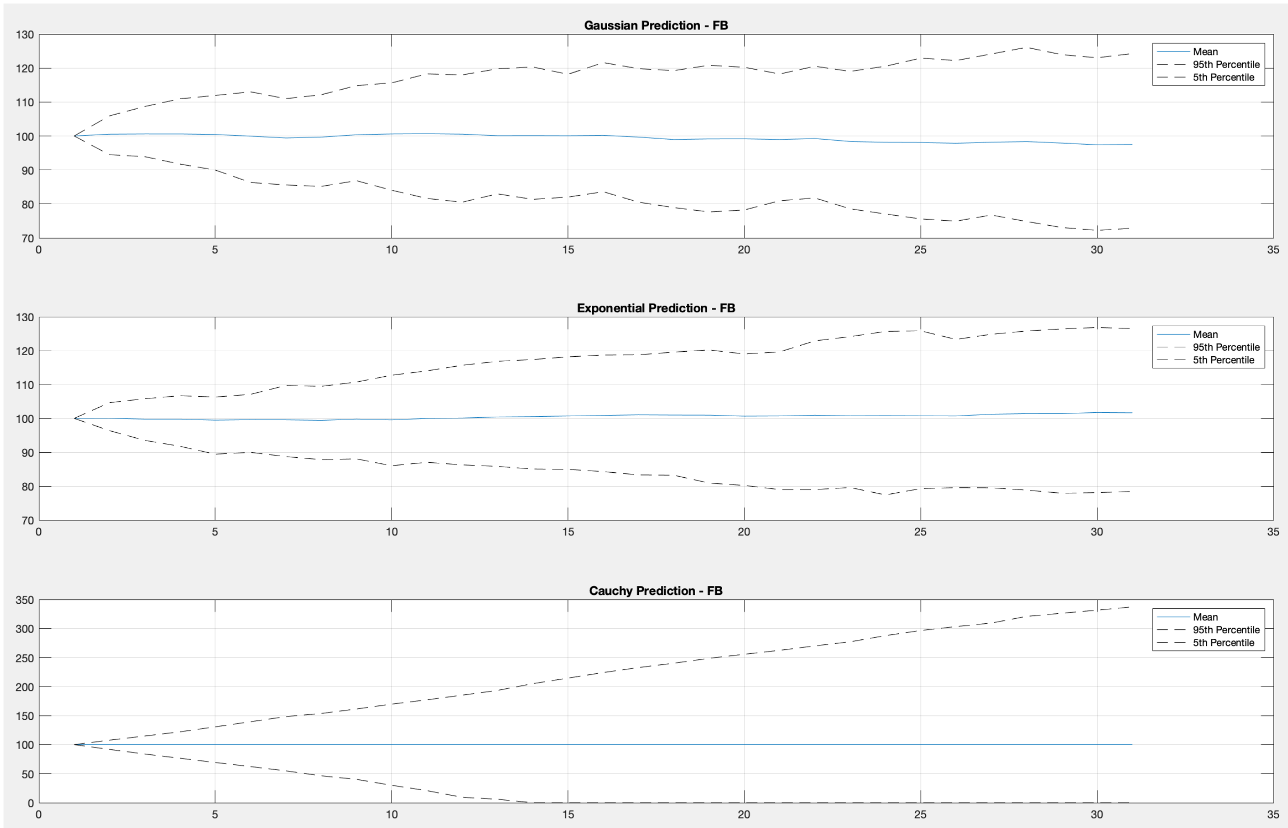

Figure 12. Mean, 95th Percentile, and 5th Percentile Stock Price Predictions Based On Gaussian, Exponential, and Cauchy Data for Facebook

In single program run for the GE stock, a mean stock price of \$99.36, a 5th percentile stock price of \$90.27, and a 95th percentile stock price of \$106.76 was predicted. Figure 13, which is shown below, shows the plots for the predicted mean, 95th percentile, and 5th percentile stock prices per day for each distribution.

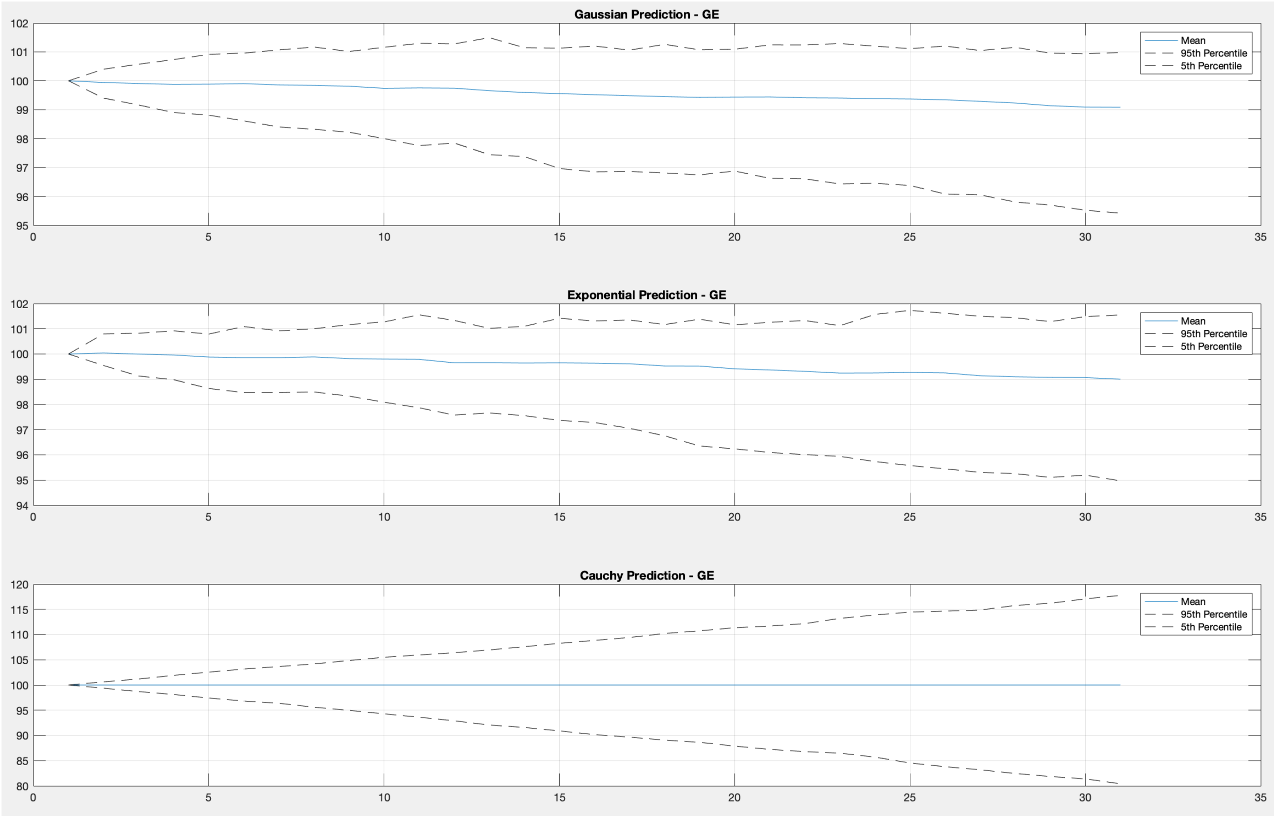

Figure 13. Mean, 95th Percentile, and 5th Percentile Stock Price Predictions Based On Gaussian, Exponential, and Cauchy Data for Apple

<u>Discussion</u>

Using the raw data for daily changes in stock price for each stock, we fit our data into Gaussian, Exponential, and Cauchy distributions. We found that the distribution with the best fit was that of exponential or gaussian, but the distribution with the true best fit could only be determined using trial and error or by an expert. We eliminated cauchy as a possible best fit because, while cauchy distributions account for irregular occurrences, such as huge jumps and drops in the stock market, our cauchy model did not have the data of irregular occurrences inputted. We did not account for recessions, or for buyer attitude (when buyers usually buy or sell their stocks) and we also did not account for the market or company developments, such as the introduction of new products or sudden developments in the market. These limitations affect the accuracy of all of our distributions, but has the greatest impact on that of the cauchy distribution and ultimately makes our cauchy distribution unreliable in the real market. This is evident in our Apple and Facebook predictions, as the cauchy predicts that Apple and Facebook could go bankrupt within the next two weeks, which is extremely unlikely.

Analyzing our results, we found that Apple and Facebook stocks produced predicted stock prices with the largest range between the 5th and 95th percentile stock prices, and that GE stocks had a generally low range of variability. Using this trend, we would use our simulation to advise stock investors to invest in Apple and Facebook stock if they are willing to risk a possible greater loss, but also have the possibility for a significant profit. However, if an investor was uncomfortable investing in a stock with such unpredictable risk, we would advise they invest in GE, which still poses risk for potential profit or loss, but to much less significant extremes. This can be demonstrated through the generated results, in which Apple and Facebook have potential profits of $94.51 and $95.97, respectively, and potential losses of $44.05 and $49.58 based on an investment of $100. In comparison, GE only has potential profit of $6.76 and a potential loss of $9.73.

Further investigating the stocks, we conclude that the best stock to invest in of the three we analyzed would be Apple. When looking at the raw data, Apple stock is the only stock with a continuous, consistent increasing trend over the past three years. In addition, the Apple stock is the only stock which produced a profitable mean predicted value of $101.25, where Facebook and GE resulted in mean stock prices less than the initial investment amount of $100. Although Apple has a higher risk and poses greater potential loss, it has a historical, stable positive trend over the past three years, making it the most advisable stock to invest in.

The results from our simulation were interesting in that they represented the variability and unpredictability of investing in fast growing tech companies compared with another established company. When analyzing our results, we noticed the variability of outcomes that investing in Apple and Facebook could produce. We attested this to the fact that tech companies' stocks have fast dramatic high peaks and corresponding dramatic low peaks, which parallels the quickly occurring changes in the market which results from the introduction of new technologies and entries by new competitors into the market.

Before generating our results, we expected Apple to be the best stock to invest in. Even before applying statistical models to our data, we noticed the consistently positive trend in Apple's data. We were not surprised to see that Apple had the only predicted increasing mean, which made it obvious to us that it would be the best stock to invest in.

However, we were surprised by how well the raw data of changes in daily stock price fit the gaussian, exponential, and cauchy distributions, especially because we chose stocks with such different trends and stocks which grew or fell at different rates. Despite the differences in trends and stability of each stock, each stock had a similarly shaped histograms of raw data, which could easily be applied to the gaussian, exponential, and cauchy distributions. We were also surprised by the similarity in results that Apple and Facebook produced. Despite having different trends, Apple and Facebook had similar mean, 5th percentile, and 95th percentile predictions.

The programs we wrote produce different graphs and predictions each time they are run because of the use of a random number in each derived equation. This produced different distributions and predictions for each distribution, but affected the graphs and predictions by the cauchy distribution most significantly. Cauchy had the most diversity in results and predictions, while gaussian and exponential distributions had the most consistent results and trends.

Our results are useful, but can be made more accurate with further development and expertise. In creating our programs, we focused on analyzing the data from a strictly statistical perspective. We did not account for major influences that could have effected the results such as current trends in the market and developments in each company. In addition, we lacked the expertise to make a more precise cauchy distribution that accounted for sudden changes in our data which could have made that trend more accurate. We also lacked the expertise to create a truly accurate prediction equation with accurate weights for each distribution. We were limited by our knowledge of the market and our limited data set. However, our results provide a rough approximation of potential trends in the market. While the results we produce may not be exact they produce a roughly accurate estimate of a potential stock price based on the historical pool of data we generated.

Conclusion, Summary, and Suggested Future Work

Overall, we found it very interesting to analyze the stock market, as it is extremely prominent in the professional world, regardless of what field one may work in. We both hope to invest in stocks in the future as another source of income and, now after this project, we have a deeper understanding of the stock market, its trends, and their implications in order to somewhat guide our investments. Ultimately, we are glad that we were able to fuse our knowledge of Risk Assessment in the field engineering with the stock market, which generally falls within the business field.

Since we picked an extremely difficult goal of essentially predicting a range of possible stock prices after a certain time interval into the future, we had to make many simplifying assumptions and, thus, we have many limitations to our model. One of the biggest limitations to our model is that we assume that the stock market is completely dependent on random numbers and pure statistics, rather than also on financial and buyer trends. For example, we do not take into account possible upcoming recessions or economic booms. We also do not take into account investor attitude, like certain times of the year when a greater number of people either buy or sell stocks, and investor’s reactions to companies actions, like when a company releases a new product or perhaps gets in trouble with the press. Therefore, in order to improve our model, steps that we would take in the future include analyzing financial and buyer trends, deciding the likelihood of certain trends happening within certain time intervals, and then applying those trends to our stocks in order to predict more accurate stock prices.

A second important limitation to our model is that we assumed that all three predictions from the three distributions, gaussian, exponential, and cauchy should be weighted the same. In reality, one distribution is more accurate than another, and the weights should be assigned accordingly. Since we currently lack expertise in financial and stock market trends, we can not yet make the decision regarding the weights of the distributions. Therefore, in order to improve our model, steps that we would take in the future include analyzing more data and possibly more distributions, deciding which distribution matches each stock with the best fit, and then weighing the distributions accordingly in order to predict more accurate stock prices.

References and Appendices

Graphs of Stock Charts

Figure 1. Three Year Historical Trend for Apple, Inc. Stock

https://www.nasdaq.com/symbol/aapl/interactive-chart?timeframe=1y

Figure 2. Three Year Historical Data for Facebook, Inc. Stock

https://www.nasdaq.com/symbol/fb/interactive-chart

Figure 3. Three Year Historical Data for General Electric Co. Stock

https://www.nasdaq.com/symbol/ge/interactive-chart?timeframe=1y

Raw Data of Morning Stock Prices

Apple:

https://finance.yahoo.com/quote/AAPL/history?period1=1464735600&period2=1559343600&interval=1d&filter=history&frequency=1d

Facebook:

https://finance.yahoo.com/quote/FB/history?period1=1464735600&period2=1559343600&interval=1d&filter=history&frequency=1d

GE:

https://finance.yahoo.com/quote/GE/history?period1=1464735600&period2=1559343600&interval=1d&filter=history&frequency=1d