Stochastic Analysis for Optimal Asset Portfolio Allocation

Submitted to Professors Trobaugh and Schaettler
Department of Electrical and Systems Engineering
Washington University in St. Louis

ESE 499 Capstone Design Project Report
Submission Date: December 10, 2017
Fall 2017 Semester

Authors:
Jonathan Gross & Susan Luo
Jonathan.Gross@wustl.edu & SusanLuo@wustl.edu
Bachelors of Science in Systems Engineering Candidates

Advisor: Dr. Jr Shin Li, jsli@wustl.edu
Abstract

With the mass accessibility of discount online investment brokerages, society has seen an influx of amateur investors, who commonly invest large portions of their savings with limited financial and mathematical knowledge. This can lead investors to poorly construct portfolios, contrasting the desired low-risk high-return portfolios. This project seeks to aid these amateur investors in more effectively constructing asset portfolios using stochastic analysis of individual assets, and non-linear optimization of sets of assets. We investigate methodologies for optimizing asset portfolios in the Markowitz Mean-Variance sense, in which portfolios should maximize returns and minimize risk as measured by the Sharpe ratio. Stochastic techniques for modeling individual assets, including the Capital Assets Pricing Model (CAPM), Fama-French 3-Factor model and Carhart 4-Factor model are analyzed and compared. Numerical nonlinear optimization techniques are applied to determine the optimal portfolio from assets. Portfolios are considered both without rebalancing, in which portfolios remain the same throughout a simulation, and with rebalancing, in which portfolios are dynamic and changing throughout a simulation. Historical stock prices are used for a limited number of assets, to back test and benchmark simulation predictions against actual results.
Table of Contents

Abstract .................................................................................................................. 1
Table of Contents .................................................................................................... 2
Introduction ............................................................................................................. 3
   A. Background ................................................................................................. 3
   B. Problem Statement ...................................................................................... 4
Project Objectives .................................................................................................. 5
Methods .................................................................................................................. 6
   A. Portfolio Optimization .............................................................................. 6
   B. Efficient Frontier ....................................................................................... 7
   C. Rebalancing .............................................................................................. 8
   D. Stochastic Models ..................................................................................... 9
      1. Capital Assets Pricing Model (CAPM) ............................................... 9
      2. Fama-French 3 Factor ........................................................................ 9
      3. Carhart 4-factor ................................................................................ 10
   E. Numerical Simulation Methods ............................................................... 10
   F. Data Acquisition ....................................................................................... 11
Results ................................................................................................................... 11
Discussion .............................................................................................................. 27
   A. General Stochastic Model Trends ......................................................... 27
   B. Impact of Pre-Simulation Data Length .................................................. 27
   C. Impact of Simulation Length ................................................................. 28
   D. Impact of Varying Stocks & Dates on Stochastic Models ..................... 29
   E. General Optimization Trends ................................................................. 29
   F. Impact of Rebalancing .......................................................................... 30
   G. Comparison of Models ...................................................................... 31
Conclusions ........................................................................................................... 32
Deliverables ........................................................................................................... 33
References ............................................................................................................. 35
Introduction

A. Background

The stock market is unquestionably a driving force in the global economy, with the New York Stock Exchange alone facilitating $1.3 trillion of stock transactions in August 2017 [1]. Due to the stock market’s importance, academics and organizations have continually been researching its behavior hoping to derive new insights that will potentially lead to better investment choices.

Much of the research into present-day investment management techniques roots itself in the work of Nobel-Prize winning economist Harry Markowitz, who in 1952 derived what is commonly referred to as Modern Portfolio Theory [2]. Modern Portfolio Theory develops a framework to help portfolio managers maximize profits while minimizing risk [2]. Typically, this yields the formulation of a nonlinear optimization problem, where the solution set of optimal portfolios is dubbed the efficient frontier [2]. A generic form of this curve is shown in Figure 1. Everything along the efficient frontier is considered optimal, in sense that for a specified return value, it is impossible to find a lower risk portfolio than the one on the curve [2].

![Image](en.wikipedia.org/wiki/File:Capital_Market_Line.png)  

Figure 1. The Efficient Frontier.¹

Markowitz’s original formulation of this problem uses a portfolio’s standard deviation as a measure of risk [2]. Markowitz’s also develops another core principle of asset management – diversification. Diversification postulates that spreading investment capital across many assets will mitigate risk, allowing a portfolio to have less risk than its constituent assets [2].

Modern Portfolio Theory and its subsequent branches require that individual assets have known expected returns and covariances [2,3]. This prerequisite has been the subject of much debate. Early investigators of this matter, dating back to Louis Bachelier in 1900, attempted to model individual assets according to Brownian Motion [3]. However, in the 1960’s Mandelbrot and Fama published a critique of Bachelier’s model, claiming returns did not follow a normal distribution as Brownian Motion suggests [3]. This led to the development of new asset pricing models, notably including the Capital Assets Pricing Model (CAPM) and Arbitrage Pricing Theory [3]. These subsequently evolved into an array of additional models, such as the Fama-French Three Factor model and the Carhart-Four Factor model [3]. To this day, there is still much debate regarding the most effective asset pricing models, and how to apply these models to create an optimal portfolio.

B. Problem Statement

With low- and no-commission self-directed online stock brokerages, such as Robin Hood and E-trade, becoming increasingly prevalent, stock trading is transitioning from an activity previously limited to Wall Street traders into one accessible by everyday citizens. E-Trade alone manages over 3.5 million accounts, equating to just under $50 billion in assets, as of December 2016 [4]. While this recent empowerment of individuals to manage their own finances is permitting lower fee structures and lower minimum account balances, both appealing factors to a would-be investor, it is not without additional risk. Without proper financial education and data to guide investment decisions, poor investments are routinely made, leading people to lose their savings.
This project demonstrates the capabilities of a platform, targeted to everyday investors, that derives insights into investment decisions. Specifically, two essential problems are addressed. First, what future price movement trends should one expect from an asset? Second, how can one take advantage of these trend predictions to construct high-performing asset portfolios? The first problem is addressed through the investigation and implementation of stochastic asset pricing models, while the latter problem integrates these stochastic models with non-linear optimization techniques.

Project Objectives

The overarching objective of this project is to develop a platform that will allow everyday investors to make smarter investment decisions. Since it is impossible to predict the future movement of a stock price with certainty, we are not definitively claiming that a suggested portfolio will produce the highest returns, but instead seeking to estimate how a proposed portfolio may trend in the future, hopefully allowing investors to construct portfolios with above average returns and performance. To achieve this end, three prevalent asset pricing models are investigated and applied in dynamic stochastic simulations: The Capital Assets Pricing Model (CAPM), the Fama-French Three Factor model and the Carhart-Four Factor model. Since there is no universal agreement on which single model is best, tradeoffs of each are explored. Non-linear optimization techniques are then applied to stochastic model results to determine optimal portfolios as per the principles of Modern Portfolio Theory. Results of stochastic models on both individual assets and optimized portfolios are tested against historical data to analyze accuracy and effectiveness.
Methods

A. Portfolio Optimization:

Modern Portfolio Theory suggests that investors should seek to either maximize returns given a targeted risk level, or minimize risk given a targeted return value. This problem can be additionally quantified using the Sharpe ratio, as defined in Equation 1:

\[ \text{Sharpe} = \frac{r_p - r_f}{\sigma_p} \]  

(1)

where \( r_p \) is the expected portfolio return, \( r_f \) is the risk-free rate of return, and \( \sigma_p \) is the portfolio standard deviation [5]. This ratio qualitatively seeks to measure how much additional return is captured from taking on additional risk [5]. In the Modern Portfolio Theory optimization problem formulation, the objective function will be to maximize the Sharpe ratio. The expected portfolio return, \( r_p \), in Equation 1 is calculated as the weighted sum of the individual assets, as shown in Equation 2:

\[ r_p = \sum_{i=1}^{n} (w_i \times r_i) \]  

(2)

where \( r_i \) is the expected return of an individual asset, as determined by a stochastic pricing model, and \( w_i \) is the weight placed on a given asset in the portfolio. The weights placed on individual assets, \( w_i \), are the variables that an investor seeks to optimize and likewise the decision variables in the optimization problem [6]. The sum of all weights is one [7]. When ignoring short sales, each weight is constrained to be between zero and one [6].

The portfolio standard deviation in Equation 1, \( \sigma_p \), is defined by Equation 3:
where $\sigma_p$ is the portfolio standard deviation, $R$ is the covariance matrix of returns and $w$ is the column vector composed of all asset weights [7]. The covariance matrix of returns can be calculated using MATLAB’s `cov()` function from the daily return values of each asset.

To solve the complete nonlinear optimization problem, MATLAB’s built in function, `fmincon()`, will be used.

B. Efficient Frontier:

As an alternative to considering one portfolio with an optimal Sharpe ratio, the efficient frontier can be considered. The efficient frontier is an upper bounding set of portfolios, acting as an envelope to contain all possible portfolios. Every portfolio along the efficient frontier is said to contain an optimal risk-return pair, meaning that for the specified return value, it is impossible to construct a lower risk portfolio from the assets considered.

To construct the efficient frontier, many smaller optimization problems are solved. In the implementation used for this project, the possible range of returns is broken down into a discrete set of 100 equally spaced values. As an example, if returns ranging from -1% to 9% are possible, the discrete return value set {-1%, -0.9%, -0.8% … 8.8%, 8.9%, 9.0%} would be considered. Then for each of the 100 return values, an optimization problem is solved to find the minimum risk portfolio. Instead of utilizing the Sharpe ratio described in Equation 1 as the objective function for this optimization problem, the portfolio standard deviation, $\sigma_p$, is used, as described in Equation 3. An additional equality constraint is added in this optimization problem formulation, to restrict the portfolio return to the targeted value:

$$r_p = r_{target}$$  \hspace{1cm} (4)

where $r_p$ is defined as described in Equation 2 and $r_{target}$ is the targeted return value, which is taken from the return value set.
After all 100 portfolios are determined, the plotted points will represent an approximation of the efficient frontier. MATLAB’s `PortOpt()` function will be used to assist in construction of this efficient frontier. Note that more or less than 100 points could alternatively be used, depending on the accuracy desired in efficient frontier construction.

C. Rebalancing:

Portfolios are also considered with or without rebalancing. Rebalancing dictates whether asset weights will remain constant throughout a simulation. A portfolio without rebalancing indicates that weights placed on assets will remain constant throughout the entire simulation. In this case, each asset return, represented as $r_i$ in Equation 2, is determined from the net return of the asset across the entire simulation period. The covariance matrix of returns, which is necessary to determine the portfolio standard deviation, is constructed using all the daily return values from the simulation.

When rebalancing is considered, the weights placed on assets are now subject to change throughout the simulation. Portfolio rebalancing is done at predefined intervals. At a maximum, a portfolio could be rebalanced daily, indicating that the weights placed on a stock could change every day. At a minimum, a portfolio could be rebalanced once throughout a simulation, indicating that the weight placed on an asset could change halfway through the simulation. Typically, a portfolio that is rebalanced daily would be expected to have the highest returns, as this would equate to selling your stock before you expect the price to go down. Mathematically, the periods between portfolio rebalancing can be treated as separate optimization problems. For example, if daily rebalancing is done, the expected return of an asset for one day will be considered and utilized as the $r_i$ value in Equation 2. Likewise, the covariance matrix of returns will be generated only from the expected returns for that day. This process is then repeated for each day. The portfolios from all days can then be combined to construct an efficient frontier for the entire simulation period.
D. Stochastic Models:

Stochastic models are required to estimate expected returns and the covariance matrix of returns. The models considered are all probabilistic, not deterministic, pricing models. Widely used models, including the Capital Assets Pricing Model (CAPM), Fama-French 3 Factor model and Carhart 4-Factor model, are described below and applied in this project.

1. Capital Assets Pricing Model (CAPM):

The expectation of an asset using the CAPM can be calculated as:

\[ r_i = r_f + \beta_m \times (r_m - r_f) \]  

(5)

where \( r_i \) is the expected return of an individual asset, \( r_f \) is the risk-free rate of return and \( \beta_m \) is the sensitivity coefficient between an asset’s return and the stock market \([6]\). Qualitatively, the CAPM claims that the future movement of an asset can be predicted by how well the asset has correlated to stock market movements in the past \([8]\).

2. Fama-French 3 Factor:

The expectation of an asset using the Fama-French model can be calculated as:

\[ r_i = r_f + \beta_m \times (r_m - r_f) + \beta_{size} \times SMB + \beta_v \times HML \]  

(6)

where \( SMB \), or Small Minus Big, is the difference between the average returns of small cap companies and large cap companies, \( HML \), or High Minus Low, is the difference between the average returns of value companies and growth companies, and \( \beta_m, \beta_{size}, \) and \( \beta_v \) are sensitivity coefficients between the asset’s return and stock market movement, average asset size and average valuation respectively \([8]\). The Fama-French 3 Factor model continues to build upon the CAPM by utilizing the same root equation while also incorporating additional factors. One new claim is that small companies carry more risk than large companies, which will be compensated with higher returns \([8]\). The model also incorporates company valuations, claiming that
companies whose stock prices are low relative to their book value, namely value companies, will lead to higher expected returns [8].

3. Carhart 4-factor:

The expectation of an asset using the Carhart model can be calculated as:

\[ r_i = r_f + \beta_m (r_m - r_f) + \beta_{size} \times SMB + \beta_v \times HML + \beta_{mom} \times UMD \]  

(7)

where UMD, or Up Minus Down, is a market momentum factor, describing difference between the average return of high-return and low-return companies [9]. \( \beta_{MOM} \) is a sensitivity coefficient relating a stock’s return to momentum [9]. The Carhart model builds off the Fama-French model, adding a momentum factor which seeks to incorporate the idea that companies with strong short-term performance will continue to perform well in the immediate future [8].

E. Numerical Simulation Methods:

MATLAB is the primary tool utilized for this project. Dynamic stochastic simulations can only be done in discrete time with MATLAB. Using daily time intervals continuous time simulations across longer periods of time, such as one year, are closely approximated.

The CAPM, Fama-French, and Carhart models are all reliant on calculating sensitivity coefficients (\( \beta \)) that relate a stock's return with market factors. These values can be calculated using a multiple linear regression, which is done through MATLAB’s `fitlm()` function. Values are recalculated at every increment, indicating that simulated data is incorporated into calculations. Since it is assumed that future prices can be predicted from historical data, a period of historical, “pre-simulation” data is used in sensitivity coefficient calculations as well. The length of historical data to be used is a design choice that is investigated. Potential problems with using too much historical data include incorporating irrelevant, too-old data, and decreasing computational performance. In contrast using too little data could result in a model overly dependent on recent short-term trends, leading to under- or over-estimations.
Since these models utilize random variables, individual simulation trials are insufficient for accurate predictions; randomness can lead to large deviations between individual trials. As a result, Monte Carlo simulation techniques are applied, in which a large number of trials are run, then retrospectively aggregated to determine useful parameters and make statistical predictions. Expected returns are identified by taking the mean return values of trials. A 90% confidence interval is also used, which is constructed by identifying the 5th and 95th percentile simulation results. Results from Monte Carlo simulations are benchmarked against real-world historical results.

F. Data Acquisition:

Market trend data, including the \( SMB, HML, UMD, r_f \) and \( r_m \) factors, was downloaded from Dr. Kenneth French’s website. Dr. French is one of the co-founders of the Fama-French model, and is a current professor at Dartmouth University [8]. He maintains an extensive, regularly updated, and publicly available data library containing these factors in a variety of time increments.\(^2\) Data for individual stocks was downloaded from Yahoo! Finance, a site which is ubiquitous with live, accurate and online stock data.\(^3\) Stock price data from Jan 1, 1996 through June 1, 2017 was collected for eleven companies, American Express, Boeing, Caterpillar, Citi Group, Coca-Cola, Disney, ExxonMobil, IBM, 3M, Proctor & Gamble (P&G) and Wendy’s.

Results

Implementations of the CAPM, Fama-French and Carhart stochastic models are shown below. For consistency and comparability, many of the results utilize the Coca-Cola (COKE) stock, although results from other stocks are discussed as well. Parameters of simulations, such as

\(^2\) Dr. French’s Data Library: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

\(^3\) Yahoo Finance: https://finance.yahoo.com/
simulation length, simulated date range, and amounts of pre-simulation, historical data used in the model were varied to investigate the impacts of each decision choice in models.

The results of three-month simulations, from March 1 - June 1, 2017, are show below. The plot on the top incorporates one year of historical data (March 1, 2016 – February 28, 2017) into the simulation, while the plot on the bottom incorporates five years of historical data (March 1, 2012 – February 28, 2017). The expected price for a given day is shown with a solid blue, red or pink with colors corresponding to the CAPM, Fama-French and Carhart models respectively. Both plots also show the actual movement of Coke’s stock price over the same period, shown with a black line, for comparison. The dashed lines indicate the 90% confidence interval for each model, corresponding to the same color lines.
Figure 2. Simulation results from Mar. 1 – Jun. 31, 2017 simulations, comparing the CAPM, Fama-French and Carhart models. Top: Predicted Coca-Cola stock price movement using one year of historical data; Bottom: Predicted Coca-Cola stock price movement using five years of historical data.

As can be seen in Figure 2 (top), the CAPM was the closest when one year of historical data was used, with the final stock price in this model differing from the historical result by approximately $38, or 22%. When five years of historical data was used, all three models resulted in a near identical price as shown by Figure 2 (bottom); here the final difference between the models and the real-world result is approximately $40 or 24%. In both simulations, the final real-world price of coke fell outside of all 90% confidence intervals, indicating that all the models significantly over-estimated the final real-world price for this period. However, it can be noted that the real-world price remained within all confidence intervals until the 50-60 trading day range, where the real-world Coke price rapidly dropped.

The model parameters specified above were modified to use a January 1 – March 1, 2001 simulation period, as opposed to a 2017 simulation period. Historically stocks in this date range
tended to have a price breakout, following the volatility associated with the Y2K. Testing different date ranges allows for a comparison of how historical data and real-world price abnormalities impact model accuracy. Results of the 2001 simulation are shown in Figure 3. The top plot in Figure 3 incorporates one year of historical data (Jan 1, 2000 – Dec 31, 2000), while the figure on the bottom incorporates five years of historical data (Jan 1, 1996 – Dec 31, 2000). The dashed lines represent the 90% confidence intervals for models of corresponding colors. The black line represents the real-world movement of Coke over the same period.
As can be seen in Figure 3, in the simulation using one year of historical data, the CAPM and Carhart models approximately tied for the most accurate model, with the final predicted price underestimating the final real-world price by approximately $7 or 37%. In the simulation using five years of historical data, the CAPM is the most accurate, underestimating the real-world result by approximately $8.50 or 45%. In all cases, the final real-world price fell outside of the 90% confidence intervals, indicating the models significantly underestimated the real-world result. It can be noted that in the top figure, which incorporated one year of historical data, that portions of the real-world price movement did begin to fall inside of the confidence interval as the simulation continued. Furthermore, the real-world price of coke did have a rapid increase within the first 20 trading days, increasing by nearly 50% over this period.
In both Figures 2 and 3 above, when comparing the simulations using one year and five years of historical data, it can be seen the confidence intervals are significantly wider when using only one year of historical data, reflecting the fact that models are expected to more accurately make predictions as more known data is used. Furthermore, in both cases the confidence interval width, equating to a likely price range, increases with time.

The length of the simulations was also extended to examine the impact of simulation length on accuracy. The results of a three-month, a one year, and a two-year simulation, all beginning on Jan 1, 2001 for the Coca-Cola stock, using five years of historical data are shown below in Figure 4. Confidence intervals for each model are shown using dashed lines. The real-world price movement of Coke over the same period is shown in black.
Simulation results comparing the CAPM, Fama-French and Carhart models, with all simulations beginning on Jan 1, 2001. **Top:** Coke stock price prediction in a three-month simulation. **Middle:** Coke stock price prediction from a one-year simulation. **Bottom:** Coke stock price prediction from a two-year simulation.
Figure 4 shows that as simulation periods grow longer, the real-world stock price becomes more firmly within the 90% confidence intervals. At the end of the three-month simulation, the final real-world stock price lies outside of all three confidence intervals. At the end of the one-year simulation, the final real-world stock price lies within the Carhart and Fama-French model confidence intervals, and slightly outside of the CAPM confidence interval. At the end of a two-year simulation, the final real-world price lies within all three confidence intervals. In the three-month confidence interval the final real-world price differs by approximately 45% from the most accurate model’s predicted price. In the one year and two-year simulations, this price difference changes to 50% and 29% respectively.

Simulations were also done using other stocks. Three-month simulations for Boeing and Exxon are shown below, in Figure 5, over the same January 1-March 31, 2001 period that was previously used for Coke. Both simulations use one year of historical data.
Figure 5 shows predictions for Boeing’s and Exxon’s stock prices are both more accurate than the predictions were for Coke’s stock price. In the Boeing simulation, the final price differed from the Carhart simulation’s predicted price by approximately $2.5, or 6%. In the Exxon simulation, the final price differed from the CAPM by approximately $1, or 4%. In comparison, the price difference in the Coca-Cola simulation, yielded an $7 or 37% price difference. Furthermore, in the Boeing and Exxon simulations, the real-world price at almost every day was contained within all confidence intervals.

The aggregate results from stochastic simulations on multiple assets can be compared using a risk-return spectrum, on which each asset’s standard deviation (risk) and return across the entire simulation is plotted. A one-year simulation, from Jan. 1 - Dec. 31, 2014, using the Carhart model and two years of historical data, is shown below in Figure 6. Note that the simulation period changed from the previous 2001 and 2017 simulation periods, as the 2014 year was found
to have more accurate predictions, in turn allowing for more meaningful optimization results. In this simulation eleven assets, Coke, Boeing, Exxon, Citi Group, Disney, IBM, MMM, American Express, Proctor & Gamble (P&G), Wendy’s and Caterpillar are considered. Note that ten out of the eleven belong to the Dow 30, with Wendy’s being the exception. Each asset is represented as a blue circle in the Figure 6.

![Risk-Return Spectrum, Carhart Model](image)

**Figure 6.** Risk-return spectrum for Jan. 1, 2014 – Dec. 31, 2014 simulation. Results generated using the Carhart model, 2 years of historical data and considering eleven assets.

In Figure 6 each of the eleven assets standard deviation range is projected to be between approximately 10% and 22%, and the expected returns are projected to be between approximately -25% and 25%. When comparing individual assets, a company that is directly vertically above another company on the spectrum is expected to have a higher return for the same amount of risk, and would thus be considered by an investor to be a better investment decision. Likewise, an asset that is directly to the left of another asset is expected to have the same return, but less risk. Generally, companies in the upper left corner are typically ideal, as these have a high return and low risk.
After determining the return and risk of each company through simulations, the process of optimization can now be considered. Considering the simplest case, with no rebalancing, the optimization process described by Equations 1-3 is applied. A sample result plot is shown below in Figure 7, with the simulation being run over the 2014 calendar year, using two years of historical data, using the Carhart model and considering the same eleven assets that were used in Figure 6. The efficient frontier is denoted in a blue line, and represents the upper bound of possible portfolios which can be constructed from the assets considered. The risk and return values of each of the eleven considered assets are marked with red circles. A portfolio with the best possible Sharpe ratio, as described in Equation 1, is marked by a black asterisk.

![Efficient Frontier, Carhart Model](image)

**Figure 7.** Efficient frontier for Jan. 1, 2014 – Dec. 31, 2014 simulation. Results generated using the Carhart model, two years of historical data and considering eleven assets.

As seen in Figure 7, all eleven assets are contained under or along the Efficient Frontier, except for one asset serving as an upper endpoint for the efficient frontier. Furthermore, the efficient frontier can reach a lower standard deviation range than any of the individual assets. This indicates that is generally possible to construct higher-return or lower-risk portfolios than can be achieved from putting all your money into one asset, except for the one highest return asset.
Furthermore, the portfolio with the best Sharpe ratio lies along the efficient frontier, as denoted by the asterisk lying along the blue curve.

The results of the simulated efficient frontier can be benchmarked against historical data by applying the portfolio weights to the real-world data from the same period. Note that when portfolios are being compared, three factors are being compared: the portfolio return, the portfolio risk (standard deviation), and the asset weight vector used to construct the portfolio. As a result, it is difficult to fully compare all factors in a single figure. In the following results and discussion, the primary focus is comparing portfolio returns between the simulated and real-world portfolios constructed from the same weight vectors. Thus, while the standard deviation of simulated portfolios is accurately plotted, the standard deviation of the real-world portfolios is not accurately plotted, and portfolios are instead superimposed to horizontally align with simulated portfolios constructed from the same weights. Three sample portfolios are shown below, all using a one-year simulation, a Jan. 1 – Dec. 31, 2014 simulation period, two years of historical data and the same eleven assets that were used in Figures 6 and 7. Each figure represents the result from a different stochastic model, allowing the CAPM, Fama-French and Carhart models to be directly compared.
Figure 8 also shows that in all three models, with some exception, simulated portfolios that were expected to generate higher returns, did in fact generate higher returns in real life. It can also be seen that the CAPM’s return accuracy tends to be worse than the Fama-French and Carhart models’ return accuracy. In the Carhart model plot (Figure 8 bottom) the blue simulated efficient frontier line is often below the pink real-world result line, indicating that the predicted portfolio return was often too low, and actual portfolio often outperformed predictions. In contrast, the Fama-French model tended to have more simulated portfolios that over-estimated performance, with real portfolios often underperforming predictions.

The predicted optimal Sharpe ratio portfolios for each model are denoted by black asterisks in Figure 8. The predicted Sharpe ratios of the optimal Sharpe portfolio in the CAPM, Fama-French and Carhart models respectively are 2.37, 1.64, 1.82. The Sharpe ratios of the corresponding real-world portfolios are 1.94, 1.72, 1.73. This indicates that for these specific three portfolios, the CAPM was the least accurate, and incidentally, also produced the portfolio with the highest Sharpe ratio. The Carhart model was the second most accurate, and the Fama-French model was the most accurate. The Sharpe ratio considers both portfolio returns and portfolio standard deviations, as opposed to purely comparing portfolio returns, as was done in Figure 8, highlighting the fact that a model with the most accurate return projections may not in fact have the most accurate Sharpe ratio projections. It should also be noted that in all three cases, none of the three portfolios predicted to have the highest Sharpe ratio did in fact have the highest Sharpe ratios; better performing portfolios, as measured by the Sharpe ratio were obtainable.

Models are also considered with and without rebalancing. The plots in Figure 9 show the results of simulations with rebalancing once, rebalancing three times, and rebalancing daily. All simulations were run over the 2014 year, with two years of historical data and the Carhart model. Like Figure 8, the standard deviations of real-world portfolios are not represented accurately, with real-world portfolios instead superimposed to align vertically with the corresponding simulated portfolios.
Figure 9. Comparison of efficient frontiers rebalanced at varying increments. Results generated using Carhart model, Jan. 1 – Dec 31, 2014 simulation period, and one year of historical data. *Top:* Efficient frontier with portfolios rebalanced once. *Middle:* Efficient frontier with portfolios rebalanced three times. *Bottom:* Efficient frontier with portfolios rebalanced three times.

As Figure 9 shows, as a portfolio is rebalanced more often, better expected returns are expected. However, it can also be seen that as a portfolio is rebalanced more often, the inaccuracy, denoted by the gap blue and pink lines, grows substantially. The portfolio standard deviation also increases as a portfolio is rebalanced more often. The portfolio rebalanced once has a reasonably accurate result, with the real-world return result generally following the expected trend. The portfolio rebalanced daily in contrast has a real-world result that is far from accurate, indicating that the simulation is far less accurate in this case. When considering the Sharpe ratios, the predicted Sharpe ratios are 1.78, 1.47, and 1.44 for portfolios rebalanced once, three times and daily respectively. This indicates that although portfolios with more frequent rebalancing are expected to achieve higher returns, the risk is growing faster, leading to lower Sharpe ratios, even in simulations.
Discussion

A. General Stochastic Model Trends

For Coke’s three-month simulations in 2001 and 2017, shown in Figures 2 and 3, the final real-world price of Coke’s stock always fell outside of a 90% confidence interval. In the 2017 simulation, the simulated price over-estimated the real-world result, while in the 2001 simulation, the simulated price under-estimated the real-world result. In contrast, the Boeing and Exxon 3-month simulations were relatively accurate, with nearly all the price movement falling within confidence interval, and the final prices having a much smaller percent difference. This indicates that although models can predict prices with some degree of accuracy, none are guaranteed to be accurate.

Additionally, in both of Coke’s three-month simulation periods, the real-world stock price was subject to a very short period of volatility, either experiencing rapid growth or rapid decline over a short period of time. All the models are reliant on longer term trends and correlations; they do not incorporate any factors that attempt to predict day-to-day movements. Even the Carhart model, which incorporates a shorter-term momentum factor, is incapable of predicting day-to-day price movements. As a result, the models would not be expected to be accurate over the two simulated periods, since they include periods of short-term volatility. In contrast, the Boeing and Exxon simulations were likely more accurate since they did not have periods of short-term volatility, with future prices generally following past trends.

B. Impact of Pre-Simulation Data Length

In simulations used to create Figures 2 and 3, it was noted that range of the confidence intervals narrowed as more historical, pre-simulation data was incorporated into models. Qualitatively, all the models assume that future price movement can be predicted from historical movement; this narrower confidence interval range likely corresponds to the increase in prediction certainty that
we gain from incorporating additional known historical data. Conversely, this equates to predictions made using only small amount of data being very uncertain and risky.

C. Impact of Simulation Length

Simulations of different lengths exhibit different behaviors in predicting stock prices, as shown by Figure 4. For Coke’s three-month simulations, the real-world end-of-simulation price fell outside of all 90% confidence intervals. In contrast, in one- and two-year simulations, the real-world end-of-simulation price fell within all confidence intervals. This can likely be attributed to two possible factors.

First, the range of the confidence intervals is growing over time. This means that after one or two years, the price range of the confidence interval is significantly larger than it is over the first three-months. Therefore, if the confidence interval bounds grow or depreciate faster than the actual stock price, the confidence interval will be more likely to include the actual price over longer periods of time. By this regard, inclusivity of the real-world price inside the confidence interval reflects the fact that models have more uncertainty over longer periods of time.

However, one other factor to consider is that much of the short term “noise” and volatility may be balancing, or canceling, out over time. In Figures 2 and 3, the real-world stock price of coke experienced either a 10% increase or decrease in a less than ten trading day period. Stock analysts generally consider this to be abnormal, and highly volatile for a company as large as Coca-Cola. As a result, these volatile periods will likely have less of an impact, over a longer period. Alternatively, a stock could later experience a volatile movement in the opposite direction that would effectively result in a “net zero” effect. Thus, in this sense, the inclusivity of the prices in a confidence interval is reflection of the fact that models are designed to estimate trends, not day-to-day movements.
D. Impact of Varying Stocks & Dates on Stochastic Models

Figure 5 shows that the models can be more accurate for some stocks than others. Similarly, a comparison of Figures 2 & 3, which use the same stock, but non-overlapping sets of date ranges, further reinforce the concept of variant model accuracy. Overall, this further reflects the idea that models cannot be guaranteed to be correct. Looking in hindsight, it was generally observed that stocks that continue to perform in a manner comparable to past performance, have more accurate model results. In contrast stock prices that have rapid swings, such as Coca-Cola did, will be less accurate. However, in a real-world environment, where these models are being applied to real-time data, as opposed to historical data, it would be impossible to guarantee with foresight whether a model will be correct. Thus, when using a stochastic pricing model, it is important for an investor to be vigilant of other quantitative and qualitative factors that were not considered.

E. General Optimization Trends

In Figures 7 and 8 we observe that it is possible to achieve portfolios with a much lower risk than any single asset individually has. This indicates that investing all capital into a single asset is a higher-risk strategy, and it is instead less risky to spread capital around to multiple assets. This observation closely aligns with the principle of diversification.

When we compare the optimization results against real-world data, we note that although the expected returns do not fully align with the real-world return, they do clearly follow the general trends. For example, in Figures 7 and 8, portfolios that are projected to have higher returns via simulation do generally end up having higher returns for the specific periods examined. When we relate this consideration to the stochastic model trends previously discussed, this correlates to the fact that models are unable to perfectly predict prices, but can generally predict general trends if abnormal trading days are not included. One other factor to note however, is that all optimization curves shown are from 2014 simulations. Over this period, the economy was relatively stable and consistent with past years. There were no stock crashes or other factors that
created rapid economic price swings. Periods that did have major unexpected economic events, such as a market crash, were met with far worse accuracy.

F. Impact of Rebalancing

Portfolios that did not rebalance generally had the best accuracy in terms of relating expected returns with real-world returns. Portfolios that rebalanced daily generally achieved the highest expected returns, but had the worst accuracy. Portfolios that rebalanced more sparsely fell in between these two extremes. Generally, they followed a trend of increasing expected returns at the expense of accuracy as portfolio rebalances increased.

When we consider rebalancing daily, it can be observed that an asset manager would logically sell of all his assets that he expects to decrease in value in the next day. Thus, with daily rebalancing, a portfolio could hypothetically avoid almost all, if not all, decreases in value for an asset and leading to a portfolio that is strictly increasing in value. This would resultantly lead the portfolio across the entire simulation period to have an extremely high return.

In terms of accuracy, we noted that rebalancing more frequently decreased accuracy. Each asset’s risk and return value is derived from a stochastic model. It was furthermore observed that due to the random noise associated with day-to-day trading, models are not accurate at predicting day-to-day price movement. Thus, when daily or high frequency rebalancing is performed, the data being optimized, namely the assets returns and covariance, is highly inaccurate and in-turn produces in an inaccurate efficient frontier. The lack of accuracy is also reflected by the fact that portfolio standard deviations increase as rebalancing is done more frequently. This resulted in portfolios that were rebalanced frequently achieving lower Sharpe ratios than those rebalanced less frequently, indicating that portfolios rebalanced too frequently are less preferable in terms of Sharpe ratios.

One other factor that should be noted, although not directly considered within the scope of this project, is that transactions typically have fees and potentially holding periods. This would mean
that daily rebalancing in a real-world situation would likely incur an additional and substantial cost to an individual, offsetting some of the potentially captured additional return.

G. Comparison of Models

Although it is not possible to claim one stochastic model is absolutely the best, there are a few discernable drawbacks and advantages to each. Generally, when determining trends, we utilized data from the efficient frontier, as these plots effectively aggregated key data, showing the accuracy of 100 portfolios that each considered 11 assets, as opposed to considering the simulation results of just one asset at a time, as is done in Figures 2-5.

The CAPM consistently had the worst model accuracy. Although it still captured trends, indicating that some insight into stock price movement can be obtained via the CAPM, it is less preferable to the Fama-French and Carhart models. Both the Fama-French and Carhart models achieved a similar level of accuracy, with either being more or less accurate depending on parameters. When considering the mathematical models, as described in Equations 5-7, it is observed that the CAPM is the basis of the Fama-French and Carhart models, with the latter two incorporating additional factors to improve accuracy. Thus, it appears that the models were generally successful in achieving this goal, and the additional factors do in fact derive additional insight. Furthermore, when comparing just the Fama-French and Carhart models, we observe both are very similar, with the primary difference being the incorporation of a market momentum factor. Mathematically, this would equate to the Carhart model being more accurate in cases where a high-performing stock continues to be a high-performing stock, but less accurate in price swings and trend changes. Overall though, it appears that both the Carhart and Fama-French models are of a similar level of accuracy and useful for aiding in investment decisions.
Conclusions

Using asset pricing models in combination with optimization techniques can lead investors to derive significant and meaningful insights into investment decisions. Although none of the models matched real-world results with perfect accuracy, price trends were predicted with a significant level of accuracy; simulated portfolios followed a very similar trend to their real-world counterparts. This will allow an investor to make far better investment decisions that would have been possible by picking stocks at random, or picking stocks based on name recognition. Furthermore, given the relative ease of the developed platform, with the mathematics being run in the back-end an investor with a very limited mathematical background can utilize a portfolio optimizer such as the one developed here.

It is also crucial for an investor to recognize that the developed platform is most effective under “normal trading conditions” in which there are no major unexpected economic events such as a stock crash, and future price trends are generally comparable to historical ones. Of course, events and trend changes such as these cannot be known in foresight. Resultantly, it is still important for an investor to remain vigilant to qualitative factors not considered by models. There is likewise still risk associated with using a portfolio optimizer such as the one considered here. However, these risk factors are likely to affect an investor, whether he randomly chooses a stock, or utilizes the portfolios mathematically constructed for him.

When comparing all three stochastic pricing models, the Fama-French and Carhart models both tended to have a similar level of accuracy, with either achieving better or worse accurate depending on circumstances. Either model would resultingly be suitable for an investor to derive relevant insights. The CAPM, although still providing some insights, typically was less accurate than other models, and consequentially, would be less preferable to the other two models.

One other key factor that should be noted is that rebalancing portfolios too frequently is not advised for investors. Although it may be theoretically possible for rebalanced portfolios to
achieve better results, this is very difficult to do in practice. Especially when considering that any real-world application would have a trading fee that would detract from investor profits, the actual returns from a portfolio rebalanced too frequently will likely be equal to or worse than a portfolio rebalanced less frequently.

In short, investors, regardless of financial market familiarity, can benefit from using stochastic pricing modelling and optimization techniques. The Fama-French and Carhart stochastic pricing models are typically capable of generating the most accurate insights. When optimizing, rebalancing should not be done excessively, as this leads predictions to become ineffective. Regardless of parameter choices, investors need to be vigilant to qualitative market conditions, as there is still risk not considered by models present. After considering these factors, investors will likely be able to make smarter investment decisions than would have otherwise been achievable.

**Deliverables**

Our final platform is very similar to the one initially proposed, in the sense that a “two-stage” platform, first running Monte Carlo Simulations on assets using a pricing model, then utilizing optimization techniques to construct an asset portfolio, was developed. We additionally did implement functionality to back test simulated results against historical data.

Stochastic pricing models considered include the Capital Assets Pricing Model, Fama-French 3-Factor model and Carhart 4-Factor model. We initially planned to implement all three models and a Brownian Motion model. We later decided to omit the Brownian Motion model, since there is a consensus in the finance community that this model is ineffective relative to others. By omitting this model, we additionally were able to undertake a more in-depth analysis of the other models.

The implementation of optimization techniques did differ from our originally proposed design in order to implement features more applicable to real-world data, although the optimization
problem itself remained consistent with our original proposal. Notably we initially planned to manually develop implementations of different optimization methods such as Steepest Descent and Newton’s Method. However, it seemed unnecessary and a waste of time to redevelop algorithms that already have equivalent result versions in MATLAB. Resultantly, we utilized MATLAB’s built in optimization solver. The optimization problem implemented is designed to optimize portfolios in the Markowitz Mean-Variance sense, as originally planned. Both the portfolio with the optimal Sharpe ratio and the entire efficient frontier are identified. We additionally designed and implemented a rebalancing feature, which was not originally planned. This is of significant use to a real-world investor, since they are likely to rebalance their portfolios and appreciate a feature to facilitate this.

Functionality to compare simulated results with real-world results is also implemented, as planned. Models applied to individual assets can be directly compared to real-world results through mean simulated price values and confidence intervals. Additionally, simulated portfolio results can be compared to real-world portfolios results graphically using the efficient frontier.
References


