MGT 6203 Group Project Progress Report

Optimal Portfolio Using S&P500 Stocks

https://github.gatech.edu/MGT-6203-Fall-2023-Canvas/Team-80

## Team 80:

Anoop Kumar

Daniel Webster

Harsh Ambasta

Hieu Nguyen

Sarah Boyette

Table of Contents

[Background: 3](#_Toc149924070)

[Objective/Approach: 4](#_Toc149924071)

[Data Cleaning: 4](#_Toc149924072)

[Future Work: 5](#_Toc149924073)

[Work Cited: 6](#_Toc149924074)

# Background:

As of 2023, 61% of Americans invest in the stock market in the United States and globally approximately 20%-25% of people invest in the market. With close to a quarter of the world’s population investing, the use of data analytics for optimizing and mitigating risk to the owner’s equity has increased substantially. Although nobody can predict the future of any stock in the stock market, leveraging big data and modeling possible future performance of any given stock can help investors lower their risk and increase their returns.

The three most popular stock indexes are the Dow Jones Industrial Average, S&P 500 Index, and the Nasdaq Composite Index. Using these market indexes, investors and data analysts alike can gauge market movements and use the information to make investment decisions regarding their portfolios. For the continuation of this project our team will be utilizing the S&P 500 Index.

**Share of adults investing money in the stock market in the United States from 1999 to 2023**

A graph with numbers and lines

Description automatically generated

Figure : Percentage of adults who invest in the stock market in the united states

# Objective/Approach:

Given the possible risks that investors face when investing their money in the stock market, our objective is to build the optimal portfolio—a portfolio which maximizes return on investment while minimizing risk. This will be accomplished by using linear optimization programming to identify portfolio weights that will produce an optimal Sharpe ratio, which we will use as an estimate of the stocks’ past, present, and future performance. Historical S&P500 data was taken from the following Kaggle database <https://www.kaggle.com/datasets/camnugent/sandp500>; also important for our analysis is broad S&P500 data taken from <https://www.investing.com/indices/us-spx-500-historical-data> and the federal funds effective rate <https://fred.stlouisfed.org/series/FEDFUNDS>, which are both downloaded for a range of time corresponding to the Kaggle dataset’s date range. The data was then cleaned and explored to ensure that there were no anomalies or outliers present. Once the data was looked over, the average return, compounded return, and standard deviation of each ticker was calculated using Python. Moving forward we will use these calculated returns to determine the weight of each stock and the number of stocks that minimizes risk while maximizing the return of the portfolio. We will set up multiple portfolios with varying numbers of stocks and weights to evaluate the impact of each. Finally, we will plot the expected return and expected standard deviation to determine which portfolio or portfolios perform the best.

A screenshot of a computer

Description automatically generated

Figure : average and compounded returns of S&P500 stocks

# Data Cleaning:

To pull all the necessary data together for computing the various returns, three data frames “stocks”, “ffer”, and “sp500” are loaded from downloaded csv files. The first step in the data cleaning is to match the dates of all the data frames and convert them to the same data type(numeric). The date column is then refactored to be the datetime index, daily returns are calculated, and new monthly returns for each data frame are calculated. On inspection, it is found that some months do not have returns for the whole month, so February 2013 and February 2018 are dropped from the dataset. Similarly, 29 of the 505 tickers in “stocks” don’t span the full date ranges needed, so they are dropped from the dataset as well. Indexes were added to the new monthly returns data frames for ease of data manipulation. Once the data appeared to be cleaned and all missing and outlying data was removed, we proceeded with the return calculations (monthly compounded and average).

# Future Work:

The completion of the data cleaning and return calculations allows us to begin the next steps of determining weights and number of stocks that are optimal for a high performing portfolio. We will also create portfolios of different asset make-ups to compare the expected performance of asset variation, as well as stock weight and number of stocks. As most investors know there is no magical number of stocks or portfolio makeup that is automatically guaranteed to make money. Returns are highly dependent on various factors including the current market conditions. Due to this, an approach of randomizing the weights and number of stocks, analyzing their expected returns, and then homing in on the top performing portfolios to determine what aspects allowed them to perform optimally will be our next steps forward.

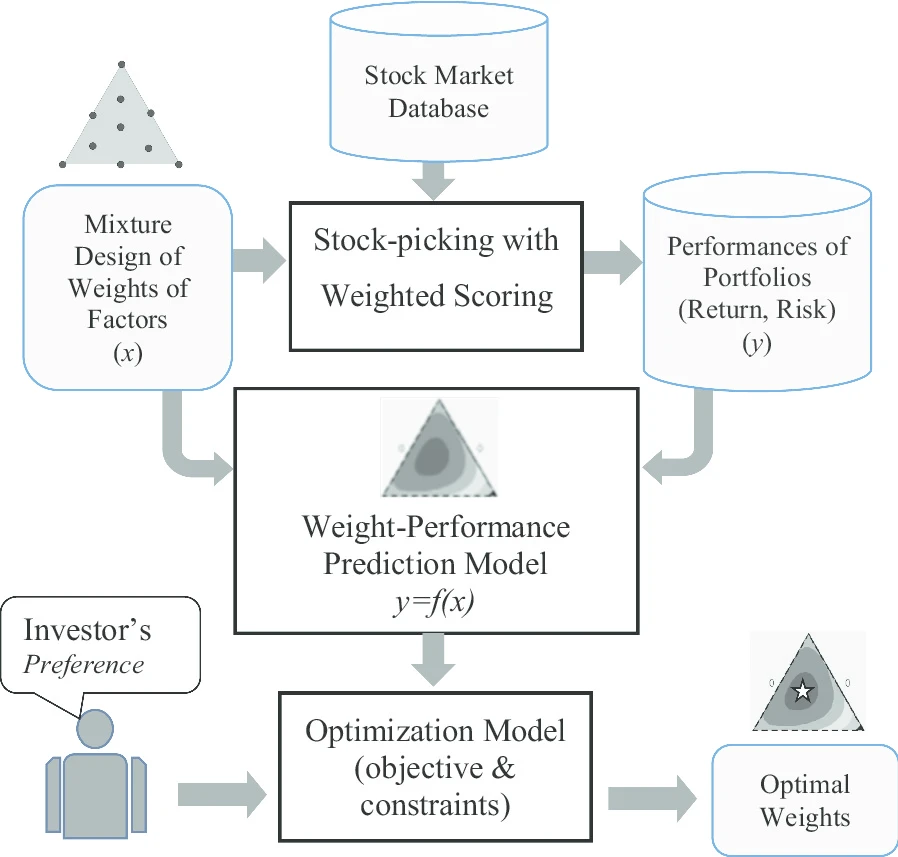


Figure : Weighted Stock Evaluation on Portfolio Performance Flow Chart

# References:

Statista Research Department. (2023, September 18). Share of Americans investing in stocks 2023. Statista. <https://www.statista.com/statistics/270034/percentage-of-us-adults-to-have-money-invested-in-the-stock-market/>

Yeh, IC., Liu, YC. Discovering optimal weights in weighted-scoring stock-picking models: a mixture design approach. *Financ Innov* **6**, 41 (2020). https://doi.org/10.1186/s40854-020-00209-x