Using mean absolute deviation as a measurement of risk in financial portfolio optimization

MIDN 1/C Gabby Lavin

MIDN 1/C Peyton Smith

MIDN 1/C Tyler Zancanella

Abstract

This study provides a methodology to minimize the risk and maximize the return of a financial portfolio, which is a collection of an individual's assets. Using linear programming techniques implemented in Python, this study models the choice of assets and the amount to invest in each asset that minimizes portfolio risk while meeting a predetermined benchmark return, based on historical data. In particular, this study uses the mean absolute deviation of the portfolio as the measure of risk. The data is composed of the prices of different stocks from various economic sectors from 2012 to 2023. We used the model to generate efficient frontiers along which risk is minimized given a desired return.

1 Introduction

Due to national and global events, the stock market can drastically rise and fall within a single day, making predictions for the best stocks to invest in difficult. A portfolio, which is composed of assets such as stocks, must be resilient enough to withstand shocks and adjust to the changing market landscape. Through operations research techniques, it is possible to generate a method to create a portfolio for an investor that responds to market volatility by minimizing risk given a benchmark return.

The ultimate goal of our study is to develop and analyze an optimization model that determines how much to invest in different stocks to minimize the risk associated with the portfolio while ensuring the investor receives a certain amount of return on their investment. This problem has been studied widely, and our project focuses on one means of financial portfolio optimization that uses a linear program to minimize a measurement of risk known as mean absolute deviation. In our project, we study the opening and closing prices of the stock market, which is split up into 11 different sectors as specified by the S&P 500: IT, consumer discretionary, communication services, healthcare, consumer staples, financials, industrials, energy, materials, real estate, and utilities (CFI Team 2022). In each of these sectors, we selected three prominent stocks, with the respective owners of the affiliated companies having large market shares, to represent that sector in our portfolio. We gathered data on these stocks from Yahoo! Finance using the yfinance Python package. With this data, we formulated a linear program that minimizes the mean absolute deviation, a measure of portfolio volatility, subject to a constraint that sets the benchmark return of our portfolio. We implemented the linear program using Pyomo, an optimization modeling package in Python. Through the linear program, we generate an efficient frontier and find the percentage of the investor's budget he or she should invest in each stock.

2 Literature Review

There currently exists extensive literature pertaining to financial portfolio optimization, specifically detailing measurements of risk and linear and non-linear ways to model various aspects of the problem. Below, we detail some of this research and how it pertains to our own study.

In their study, Henriksson and Leibowitz (1989) examined using shortfall constraints and benchmark portfolios as constraints for portfolio optimization. The researchers crafted examples of return distributions, using values similar to what would be expected for the standard deviation and expected return, such as 10 percent and 8 percent respectively. The researchers determined that shortfall constraints were an effective tool in a mean-variance optimization process, providing a means of setting a limit on acceptable risk and return. Using a benchmark portfolio was also effective in portfolio optimization, as potential portfolios that have a higher correlation with the benchmark are more likely to have the same or greater returns as the benchmark portfolio. Constraints on optimization involved eliminating the portfolios that did not perform as well as the benchmark portfolio a certain percentage of the time. The researchers did not use these concepts in a real-world example with a specific dataset, but the concepts themselves provide background on how to optimize a portfolio using a shortfall constraint and benchmark portfolio, similar to the benchmark return constraint in our study.

Another measure used to quantify risk is the Sharpe ratio, which is the ratio of adjusted returns of the portfolio based on the risk-free rate to the standard deviation of the portfolio itself (Gökgöz & Atmaca 2016). Gökgöz's and Atmaca's research aimed to provide background on portfolio optimization using the mean-variance, down-side, and semi-variance models, searching for the optimal portfolio with the highest Sharpe ratio. Three models were used: (1) mean-variance, using variance as the measure of risk, (2) down-side, focusing on direct deviations below the target return amount, and (3) semi-variance, which is similar to down-side in focusing on negative deviations from the target, but evaluates the square of those deviations. The researchers used data on Turkish electricity market prices over a two-year period from 2014 to 2016, assuming each 24-hour day was a risky asset and calculating the rate of return for customers using the three different models. The researchers produced efficient frontiers, with standard deviation on the horizontal axis and rate of return on the vertical axis, that showed the frontier portfolios generated from each method reaching their maximum Sharpe ratio around a similar point. When varying risk aversion, standard deviation, and rate of return, all three models produced similar results in terms of the value of the Sharpe ratio, signifying that each method is effective in creating a portfolio with a high Sharpe ratio. Further analysis needs to be conducted to determine if other risk measurements are superior to the Sharpe ratio in portfolio optimization, and if one specific model is most effective and accurate in portfolio optimization.

Konno and Yamazki (1991), on the other hand, applied an absolute deviation function, denoted as L1, to measure risk in the Markowitz portfolio optimization model, which normally uses a standard deviation function to measure risk. The absolute deviation risk function developed reduces the computational load by eliminating the quadratic terms associated with the standard deviation risk function. Historical data from the Tokyo stock market, 224 stocks over 60 months, was used to test the absolute deviation risk function in the Markowitz model against the standard deviation risk function in the Markowitz model and Sharpe's single-factor risk model. All three models performed comparably to each other; no model was significantly superior in terms of return and computation time. The researchers demonstrated that the absolute deviation function within the Markowitz model performs in line with the other models and takes less time to compute. The researchers were limited in their ability to incorporate substantially more stocks, given the computing capabilities in 1991. The model in this paper serves as the basis of our project. The paper gave us the insight to use absolute deviation to measure risk and showed us how to incorporate it into our model in order to easily measure risk in the portfolio.

Gondzio and Grothey (2005) demonstrated that computing has evolved enough to efficiently solve portfolio optimization problems modeled as stochastic programs with non-linear constraints. The researchers' model incorporates numerous non-linear constraints, including risk and downside risk. The model can be solved using the primal-dual interior point method. The model was tested using randomly generated data for periods, realizations per period, and assets. The researchers used a SunFire 15K machine to test their model and found that the hardest problem required slightly over 7 hours. The researchers proved that large non-linear programming problems could be solved with modern computing infrastructure. Since the release of the SunFire 15K in 2002, new computers have been released that are cheaper with increased computing speeds, allowing for these kinds of problems to be solved even quicker.

The approach taken by Boyd, Fazel, and Lobo (2002) involved selecting a portfolio by maximizing the expected return while considering transaction costs and risk constraints. Transaction costs are what an investor pays when rebalancing their portfolio and are an important factor to consider when looking to optimize a financial portfolio. The authors considered both linear and fixed transaction costs. Modeling linear transaction costs resulted in a convex optimization problem, which is easy to solve. The authors ran into problems when they included fixed transaction costs because they result in a non-convex optimization problem. These non-convex optimization problems can be solved by solving many convex optimization problems, but they found that the number of problems that needed to be solved grows exponentially with the number of assets added and therefore are not efficient for large portfolios. In order to solve for much larger portfolios, the authors approximated a solution by developing a model that produces a suboptimal portfolio and a guaranteed upper bound on the optimal portfolio. Their model does not necessarily give an optimal solution to the portfolio optimization problem with fixed transaction costs, but gives an effective way to solve the problem for large portfolios that have nonlinear aspects.

Young (1998) developed a simple linear programming model to maximize a financial portfolio's average rate of return by using minimum return for risk instead of variance. This tactic allowed Young to keep his model linear because he did not need to deal with the quadratic functions that mean-variance optimization requires. The model Young created had the overall objective to maximize the minimum return on the portfolio, subject to the restriction that the average portfolio return exceeds some minimum level and that the sum of the portfolio allocations does not exceed the total allocation budget. Young found this model to be very successful and useful when in action because it was not complex and easy to use with many different types of data. Young's research relates to our project because it demonstrates that a portfolio optimization model can be linear and still take on complexities.

3 Input Data

Our data is composed of the daily price and volume for stocks in the following 11 sectors of the economy: IT, consumer discretionary, communication services, healthcare, consumer staples, financials, industrials, energy, materials, real estate, and utilities. These sectors and the three assets we chose from each sector can found in Table 1.

Sector	**Company 1**	**Company 2**	**Company 3**
IT	Microsoft (MSFT)	Apple (AAPL)	Meta (META)
Consumer Discretionary	Starbucks (SBUX)	Tesla (TSLA)	McDonald's (MCD)
Communication Services	Google (GOOGL)	Comcast Corporation (CMCSA)	AT&T (T)
Healthcare	Pfizer (PFE)	UnitedHealth Group Inc (UNH)	Vertex Pharmaceuticals (VRTX)
Consumer Staples	Coca Cola (KO)	Proctor and Gamble (PG)	Nestle (NSRGY)
Financials	Visa (V)	Berkshire Hathaway (BRK-A)	JPMorgan Chase & Co (JPM)
Industrials	Northrop Grumman Corp (NOC)	UPS (UPS)	Lockheed Martin (LMT)
Energy	Schlumberger (SLB)	Devon Energy (DVN)	Exxon Mobil (XOM)
Materials	BHP Group ltd (BHP)	Linde PLC (LIN)	Rio Tinto PLC (RTNTF)
Real Estate	SBA Communications Corps (SBAC)	Prologis (PLD)	American Tower Corporation (AMT)
Utilities	NextEra (NEE)	Duke Energy Corporation (DUK)	Southern Company (SO)

Table 1: List of sectors of the economy and assets within each sector

The data comes from Yahoo! Finance, a source for financial news and market data. The data contains information on the selected 33 stocks, with each row corresponding to one day from open to close of the stock market. The columns provide the ticker of the stock and information on the adjusted closing price, closing price, high price for the day, low price for the day, opening price, all of which are in U.S. dollars, and the volume of stock. Our data ranges from January 1, 2012 to December 31, 2022. We also collected data from January 1, 2023 to April 1, 2023 to test our model. We computed the yearly return of each stock by subtracting the first closing price for a given year from the first closing price for the previous year:

$$Yearly\ Return\ =\ \frac{Current\ Year's\ Closing\ Price - Last\ Year's\ Closing\ Price}{Last\ Year's\ Closing\ Price}$$

We also computed the quarterly return from each stock, by subtracting the first closing price for a given quarter from the first closing price for the previous quarter:

$$Quarterly\ Return\ =\ \frac{Current\ Quarter's\ Closing\ Price - Last\ Quarter's\ Closing\ Price}{Last\ Quarter's\ Closing\ Price}$$

The data we have consists of 33 separate dataframes, one for each asset. For the yearly model, each dataframe has 11 rows and 4 columns each corresponding to year, date, closing price, and yearly return. A sample of this data is shown below in Table 2.

Year	Date	Close	Return
2012	2012-01-03 00:00:00-05:00	45.03	
2013	2013-01-02 00:00:00-05:00	43.78	-0.03
2014	2014-01-02 00:00:00-05:00	40.69	-0.07
2015	2015-01-02 00:00:00-05:00	49.37	0.21
2016	2016-01-04 00:00:00-05:00	47.03	-0.05
2017	2017-01-03 00:00:00-05:00	49.04	0.04
2018	2018-01-02 00:00:00-05:00	47.17	-0.04
2019	2019-01-02 00:00:00-05:00	43.72	-0.07
2020	2020-01-02 00:00:00-05:00	62.62	0.43
2021	2021-01-04 00:00:00-05:00	59.36	-0.05
2022	2022-01-03 00:00:00-05:00	68.17	0.15

Table 2: Sample of input data

The input data for the quarterly model has an additional column that accounts for the month that the data is taken; data is taken every three months in the quarterly model.

4 Model

Our optimization model contains two sets. One set, ASSETS, is composed of the individual assets. The other set, TIME, contains the years for which we have financial data, which is from

2012 to the first quarter of 2023. The parameters within our model are BUDGET, which is the investor's budget in dollars, R, which is the benchmark return values for the portfolio, r_{jt}, which is the return of asset j in period t, and r_j, which is the average return of asset j over the time horizon.

The first set of decision variables in our model, x_j for $j \in$ ASSETS, represents the dollar amount invested in each asset j. Another set of decision variables in our model, z_t for $t \in$ TIME, represents the absolute value of the portfolio's deviation from the average return in period t. The final decision variable in the model, y, represents the amount put into a risk-free option. The objective function and constraints are as follows:

Objective Function:

$$min \quad \frac{1}{T}\sum_{t=1}^{T} z_t$$

Subject to:

$$z_t \geq \sum_{j \in ASSETS} (r_{jt} - r_j) * x_j \qquad (1)$$

$$z_t \geq -\sum_{j \in ASSETS} (r_{jt} - r_j) * x_j \qquad (2)$$

$$\sum_{j \in ASSETS} (x_j * r_j + x_j) + y \geq BUDGET * (1 + R) \quad (3)$$

$$\sum_{j \in ASSETS} x_j + y = BUDGET \qquad (4)$$

$$x_j \leq BUDGET \quad \forall\, j \in ASSETS \qquad (5)$$

$$x_j \geq 0 \quad \forall\, j \in ASSETS \qquad (6)$$

Constraints (1) and (2) represent the absolute deviation value of the portfolio at period t. The absolute deviation value allows us to quantify the risk associated with the portfolio at each time period, the goal of our model being to drive these values to the lowest values possible. Absolute values of decision variables cannot be directly included in linear programs, but constraints (1) and (2) create the same effect by forcing z_t to be equal to the positive value of the deviation in time period t. Constraint (3) ensures the total return of the portfolio is greater than or equal to (1+R) times the BUDGET. Constraint (4) ensures that the total amount invested in all assets has to equal the parameter BUDGET. Constraint (5) ensures that the dollar amount invested in each asset cannot exceed the set maximum value for each asset, with that maximum value being set to BUDGET. The purpose of this constraint is for redundancy. Lastly, constraint (6) ensures the amounts invested in each asset are non-negative. In our model, we assumed only certain assets were available, and that historical data are an accurate representation of the future. We also calculated the return for each asset on a yearly basis.

5 Experimental Setup

In our study, we used Pyomo to implement our optimization model with GLPK as the solver. We used the yfinance Python package to import financial data on the stocks we selected for our portfolio. We also used Pandas to wrangle the data imported using yfinance, and to generate Excel files that stored our results, which were used in our analysis. All code was run on a computer with an Intel Core i7-8650U CPU at 1.90 GHz and an installed physical memory of 16.0 GB using Microsoft Windows 11 Enterprise.

We ran two variants of this model: a yearly version and a quarterly version. The yearly version has 10 time periods, one for each year between 2012 and 2022. The quarterly version has 42 time periods, starting in June 2012 and ending in December 2022. These versions of the model correspond to different instances of the set TIME. In running our optimization model, we chose different values for the benchmark return R, specifically 5%, 6%, 7%, 8%, 9%, 10%, and 15%. We also assumed BUDGET equals 100,000.00, and we kept all other input data the same.

6 Results

6.1 Yearly Model

The time it took for all variants of our model to run and be solved to optimality was 30 seconds. Figure 1, below, shows the benchmark portfolio return values plotted against the corresponding minimum mean absolute deviation values for the yearly model. The plotted line represents the efficient frontier, providing the minimum risk at a specific return.

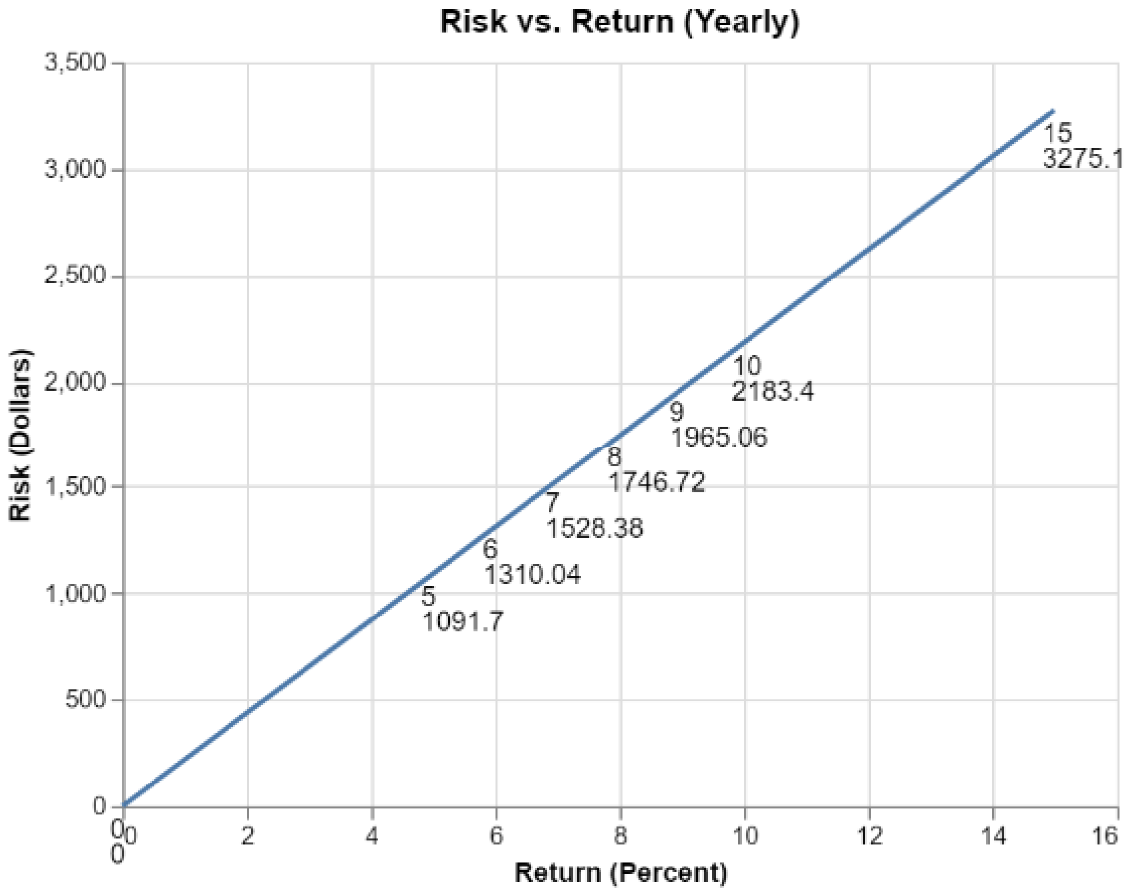

Figure 1: Efficient frontier displaying the tradeoff between risk and return for the yearly model

Figure 1 shows the linear relationship between risk and return in our model. The trend shows that an increase in the benchmark return corresponds to an increase in the minimum possible risk. For each benchmark, the same four assets were chosen for investment: TSLA (Tesla), SBUX (Starbucks), CMCSA (Comcast), UNH (UnitedHealth Group Inc.). The investment in each asset for each benchmark return is shown in Table 3.

	TSLA	SBUX	CMCSA	UNH	Cash
5%	0.19%	6.84%	5.06%	9.91%	77.99%
6%	0.23%	8.21%	6.07%	11.90%	73.59%
7%	0.27%	9.58%	7.08%	13.88%	69.19%
8%	0.31%	10.95%	8.10%	15.86%	64.78%
9%	0.35%	12.32%	9.11%	17.85%	60.38%
10%	0.39%	13.68%	10.12%	19.83%	55.98%
15%	0.58%	20.53%	15.18%	29.74%	33.97%

Table 3: Optimal portfolios based on the yearly model

As the benchmark return increases, investment in SBUX, CMCSA, and UNH increase, with UNH having the highest increase in the investment, while the amount of cash held from investment decreases.

Actual returns for the portfolios described in Table 3 were generated using the real data from the first quarter of 2023. The results are shown in Table 4. The projected total was calculated by dividing the desired yearly return at the given budget by 4, given that the portfolio was applied to only a quarter's worth of data. Upon comparing the actual and projected returns for each portfolio in the first quarter of 2023, every portfolio underperformed against the projected return.

	5%	6%	7%	8%	9%	10%	15%
TSLA	154.51	185.41	216.31	247.22	278.12	309.02	463.53
SBUX	272.80	327.36	381.92	436.48	491.04	545.60	818.40
CMCSA	347.12	416.55	485.97	555.40	624.82	694.25	1041.37
UNH	-467.41	-560.90	-654.38	-747.86	-841.34	-934.83	-1402.24
Actual Total	307.02	368.42	429.83	491.23	552.63	614.04	921.06
Projected Total	1250	1500	1750	2000	2250	2500	3750

Table 4: Returns for optimal portfolios in Q1 of 2023

6.2 Quarterly Model

Figure 2 shows the benchmark portfolio return values plotted against the corresponding minimum mean absolute deviation values for the quarterly model. The plotted line in Figure 2 represents the efficient frontier.

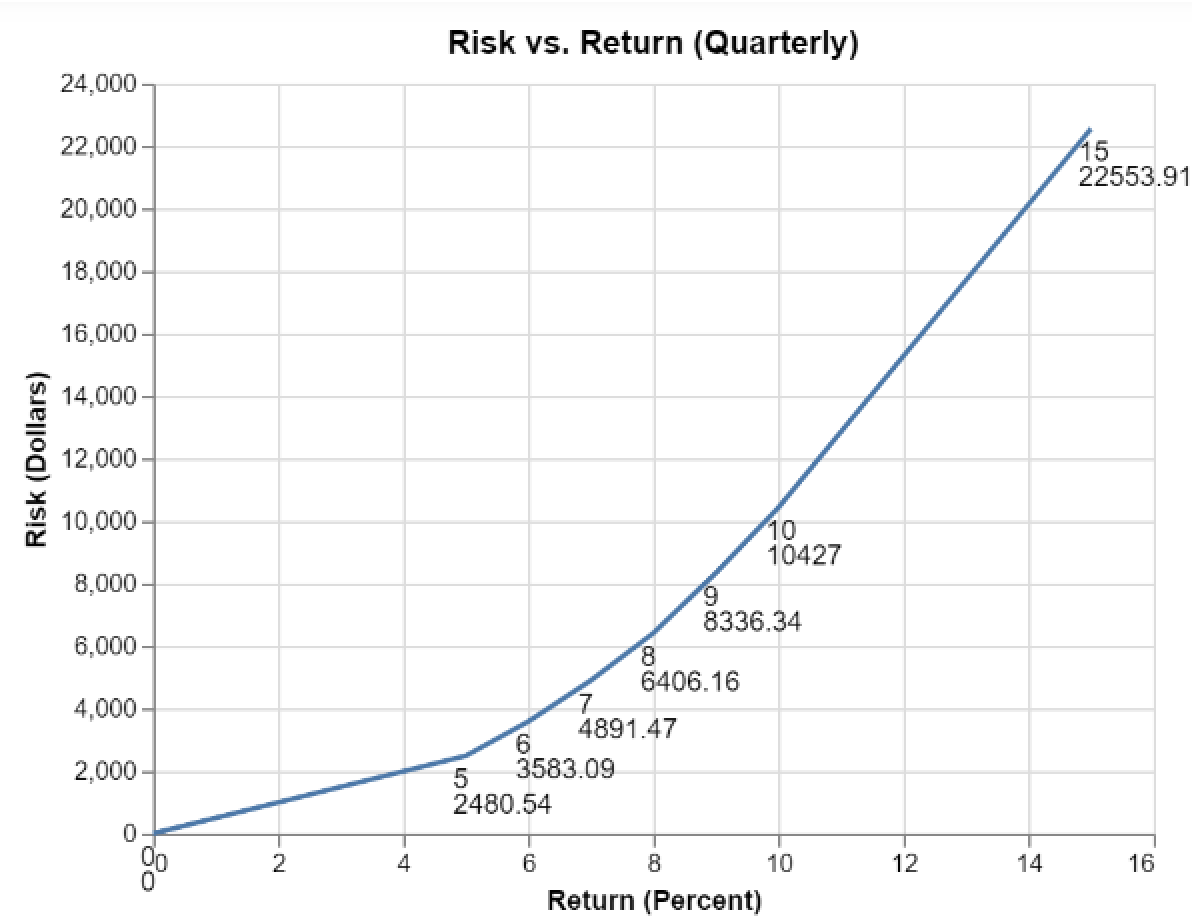

Figure 2: Efficient frontier displaying the tradeoff between risk and return for the quarterly model

Figure 2 shows the same trend between risk and return as in Figure 1, but the efficient frontier is non-linear. The slope on the efficient frontier in Figure 2 is increasing, signifying that as the return increases, the change in risk increases. In the quarterly model, we observe deviations from the mean return more frequently and capture short-term volatility, which may explain the higher magnitudes of risk for the quarterly model at the same benchmark return.

Unlike the yearly model, the optimal portfolios consisted of a wider variety of assets and no cash was withheld from investment. These results are shown in Table 5.

	5%	6%	7%	8%	9%	10%	15%
MSFT	0%	6.90%	15.71%	8.77%	0%	0%	0%
TSLA	1.99%	6.89%	10.72%	17.18%	24.76%	32.18%	70.40%
SBUX	5.11%	6.80%	4.69%	0%	0%	0%	0%
UNH	32.02%	38.17%	37.77%	34.34%	26.56%	10.71%	0%
VRTX	0.61%	0%	0%	0%	0%	0%	0%
V	10.12%	14.30%	19.96%	16.07%	14.51%	9.89%	0%
JPM	2.18%	0%	0%	0%	0%	0%	0%
NOC	4.24%	5.70%	9.35%	23.64%	34.16%	47.22%	29.60%
LMT	5.64%	0%	0%	0%	0%	0%	0%
SLB	0.07%	0%	0%	0%	0%	0%	0%
DVN	0%	0.43%	1.80%	0%	0%	0%	0%
XOM	8.35%	6.73%	0%	0%	0%	0%	0%
RTNTF	0.88%	0%	0%	0%	0%	0%	0%
SBAC	2.71%	0%	0%	0%	0%	0%	0%
NEE	26.07%	14.09%	0%	0%	0%	0%	0%

Table 5: Optimal portfolios based on the quarterly model

The quarterly model identified more stocks for investment than the yearly model. The quarterly model also did not have any risk-free assets chosen for investment. Actual returns for the portfolios described in Table 5 were generated for the first quarter of 2023, as shown in Table 6.

	5%	6%	7%	8%	9%	10%	15%
MSFT	0.00	1371.47	3125.21	1744.17	0.00	0.00	0.00
TSLA	1598.80	5525.08	8592.99	13775.16	19854.69	25799.63	56443.96
SBUX	203.93	271.09	186.86	0.00	0.00	0.00	0.00
UNH	-1509.36	-1799.60	-1780.58	-1618.92	-1252.23	-504.91	0.00
VRTX	64.96	0.00	0.00	0.00	0.00	0.00	0.00
V	1054.39	1489.82	2079.75	1674.22	1512.30	1030.33	0.00
JPM	-79.89	0.00	0.00	0.00	0.00	0.00	0.00
NOC	-558.85	-749.91	-1231.28	-3112.59	-4497.53	-6217.38	-3897.11
LMT	107.43	0.00	0.00	0.00	0.00	0.00	0.00
SLB	1.13	0.00	0.00	0.00	0.00	0.00	0.00
DVN	0.00	-35.45	-149.03	0.00	0.00	0.00	0.00
XOM	753.77	607.44	0.00	0.00	0.00	0.00	0.00
RTNTF	-42.19	0.00	0.00	0.00	0.00	0.00	0.00
SBAC	-257.30	0.00	0.00	0.00	0.00	0.00	0.00
NEE	-2226.44	-1203.69	0.00	0.00	0.00	0.00	0.00
Acutal Return	-889.61	5476.26	10823.92	12462.04	15617.23	20107.68	52546.85
Projected Return	5000	6000	7000	8000	9000	10000	15000

Table 6: Returns for optimal portfolios in Q1 2023

The portfolios for the 5% and 6% benchmark underperformed in their actual return in the first quarter of 2023 compared to their projected return. When compared to their projected return, the other portfolios overperformed in their actual return.

The data used to build this model was collected during a prosperous time for the economy, however the COVID-19 pandemic disturbed the historically positive trend and the current recovery makes it difficult to identify stocks that have positive returns because the stocks included in the model are not matching their historical performance.

With the model formulation, we did not add any constraints to force the diversification of each portfolio, and yet the portfolios contained assets in the consumer discretionary, communication services, and healthcare sectors. We also noticed that the efficient frontier became non-linear when we calculated the returns on a quarterly basis.

7 Conclusion

Our research dives into one approach of finding an optimal portfolio by minimizing the mean absolute deviation subject to a benchmark return constraint. Without any constraint limiting investment in a particular sector, the model diversified the portfolios in the yearly model across the consumer discretionary, communication services, and healthcare sectors. In addition, we determined that there is a direct tradeoff between risk and return. If an investor desires a certain return on their investment, they must accept a certain level of risk inherent in uncertainty, as assets with higher returns tend to be associated with greater risk and volatility. It is important to note that we assumed past performance was an accurate predictor of future performance in our model.

There are many ways to improve and build upon our current model. Our research uses data from 2012 to 2022 due to companies like Tesla starting in the early 2010s. With more time, we could have found companies with large market shares that have existed for more than 20 years, choosing to include only these companies in our list of assets. We only viewed the performance of the assets in our model over the past 10 years, which may have prevented us from obtaining a more accurate representation of their volatility. We could have also included constraints to ensure that the investor invests a certain amount or percentage of their budget in certain assets or in specific sectors. We could also add a constraint that sets a lower bound on the number of assets and sectors that have to make up the investor's portfolio in order to ensure a certain level of diversification and as an additional way of managing risk. In the current model, we have a constraint that sets an upper bound on the amount that can be invested in each asset. This constraint is redundant given the other constraints in the model, but it can be modified to create another version of the model where the upper bound on the amount invested in each asset is a certain percentage of BUDGET.

With our model, we made the assumption that previous performance of an asset will reflect future performance, though this assumption can be difficult to justify given the unpredictability of the financial markets. We also allowed the investor to only choose from 33 different assets for their portfolio, but with more time, we could increase that number to provide more options to the investor and allow for greater portfolio diversification.

Our model is only one way to approach portfolio optimization; each investor should pick the approach that works best for them depending on their goals.

References

Boyd, S., Fazel, M., & Lobo, M. S. (2006). Portfolio optimization with linear and fixed transaction costs. *Annals of Operations Research, 152,* 341-365.

CFI Team. (2022, November 27). *The S&P Sectors.* CFI. https://corporatefinanceinstitute.com/resources/valuation/the-sp-sectors/

Gondzio, J. & Grothey, A. (2007). Solving nonlinear portfolio optimization problems with the primal-dual interior point method. *European Journal of Operational Research, 181(3)*. 1019-1029.

Gökgöz, F. & Atmaca, M.E. (2016). Financial portfolio optimization in electricity markets: evaluation via sharpe ratio. *International Journal of Economics and Management Engineering, 10(11).*

Henriksson, R. D. & Leibowitz, M. L. (1989). Portfolio optimization with shortfall constraints:

A confidence-limit approach to managing downside risk. *Financial Analysts Journal, 45(2),* 34-41.

Konno, H. & Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market. *Management Science*, *37(5)*, 519-531.

Young, M. (1998). A minimax portfolio selection rule with linear programming solution. *Management Science, 44(5),* 673-683.