Portfolio Optimization: Mitigating Risk through use of Shapley Values in Genetic Algorithms

William S. Ventura

Johns Hopkins University

Author Note

Correspondence concerning this article should be addressed to William S. Ventura, Whiting School of Engineering, Johns Hopkins University, 3400 N Charles St, Baltimore, MD 21218. E-mail: wventur1@jh.edu

Portfolio Optimization: Mitigating Risk through use of Shapley Values in Genetic Algorithms

Introduction

In modern financial markets, portfolio allocation is one of the major problems faced by investors and fund managers. The goal of choosing several assets among thousands of available assets in order to simultaneously maximize returns and minimize risk remains a daunting task. It is critical for investors to curate the most optimal portfolio - however higher returns are associated with higher risk. As such, investors with varying degrees of risk aversion will demand different levels of returns for the different degrees of risk taken - leading to the study of portfolio management and optimization. Gunjan and Bhattacharyya (2022)'s review of the varying types of portfolio optimization techniques suggests evolutionary optimization methods to be most useful when it comes to optimized allocations of portfolios – risk management being the next critical component of curating an optimal portfolio.

Portfolio risk decomposition can provide crucial information for investors engaged in risk measurement and analysis of portfolio risk management decisions. Even so, decomposition into individual risk contributions is not a straightforward process by only using the portfolio variance expression for σ_p^2. While modern portfolio theory proposes the usage of the market model to decompose returns into systematic and unsystematic components - with systematic risks being capable of reducing the global risk of a portfolio, it cannot indicate the contribution of each asset to the global risk of the portfolio (Elton & Gruber, 1997; Mussard & Terraza, 2008). This paper takes the Shapley value to decompose the covariance matrix of a portfolio and optimal portfolios to estimate the systematic risk of individual assets (Shapley, 1953). The Shapley value, is an idea that stems from cooperative game theory which aims to allocate payoff of a coalition to individual players, with the underlying assumption that they all collaborate. The application of the Shapley value in investments and portfolio theory remains novel with Mussard and Terraza (2008) and Shalit (2021) establishing the theoretical foundations for the Shapley value to decompose the risk of portfolios and optimal portfolios. This area of interest is further developed by Hagan et al. (2021) who proposed Shapley value scheme for allocating risk to non-orthogonal greeks in a portfolio of derivatives as well as leveraging Shapley values as a natural method of interpreting components of enterprise risk measures such as Value-at-risk (VAR) and Expected shortfall (ES - also known as Conditional Value-at-Risk). Moreover, Morelli (2023) focused on clustering companies of the Euro Stoxx 50 according to the Environmental Score (E) – building a portfolio that minimizes the Conditional Value-at-risk (CVaR) for each cluster with the constraint placed on E for the optimization problem. Computing the Shapley value for each portfolio using the minimized CVaR and the optimized E as characteristic functions to yield the contribution of the portfolios to the public welfare.

The Shapley value is the only method that possesses four properties that is believed to be crucial for satisfactory portfolio attribution i.e fairness, correct baseline, full attribution, and monotonicity. In comparison to other attribution methods e.g One-at-a-time, Leave-one-out, and sequential attribution, the only challenge with Shapley is computational. The number of simulations required to carry out Shapley attribution is exponential in the number of features to attribute to. To overcome its computational challenge, Moehle et al. (2021) and Hagan et al. (2021) both proposed the use of Monte Carlo simulations, which was first considered in the context of voting games by Mann and Shapley (1960). This paper aims to decompose a portfolio's overall risk to individual components that make up said portfolio and optimize it by replacing riskier components through means of a Genetic Algorithm.

Motivation

Portfolio management and optimization remains to be a challenging and complex problem with increasing interest among researchers in technical and financial domains. The objective being to allocate assets to maximize returns while simultaneously minimizing risk. In spite of multiple advancements in new theories and computational power, portfolio optimization still remains a challenging problem to solve. Per the Bureau of Labor Statistics, the employment of financial and investment analysts and portfolio managers is expected to increase at a faster than average rate by 9% and 6% respectively from 2021 to 2031 (Bureau of Labor Statistics, 2023) – hence, introducing the need for better methodologies to perform portfolio optimizations. Some of the key factors which make it an area of interest include: (i) leveraging portfolios as a cooperative game model, (ii) application of novel evolutionary optimization methods, (iii) leveraging Shapley values for risk allocation. The main contribution of this paper is that it builds upon the previous works of Shalit, Morelli, and Hagan et al. to introduce a novel approach that incorporates evolutionary algorithms.

Background

Before examining the Shapley value, it is important to cover classical systematic risk. Systematic risk in portfolio theory is the measure of an asset's risk relative to market performance. One notable equilibrium theory, the Capital Asset Pricing Model (CAPM) was made possible by utilizing a special case of Von Neumann and Morgenstern's expected utility (Elton & Gruber, 1997).

Beta analysis

Markowitz (1952) established the foundation for the theory of portfolio choice by considering the case where investors are only concerned with the mean and variance of payoffs of a chosen portfolio - its main result being that diversification of holdings is optimal and the benefit that can be obtained depends on the covariances of asset returns. Using Markowitz's theory of portfolio choice as the basis for CAPM, Sharpe (1964) and Lintner (1956) defined it as such:

$$Er_i = r_f + \beta_i(Er_m - r_f)$$

where Er_i is the expected return on asset i, r_f is the return on the risk free asset, Er_m is the expected return on the market portfolio. Beta is then defined as:

$$\beta_i = cov(r_i, r_m)/\sigma_m^2$$

Where r_i are the stock returns, r_m the market returns, and σ_m^2 the variance of market returns. The underlying assumption is that the entire stock market is used as the investor's portfolio of assets and as such faces limitations.

The first is that the risk-free premium r_F is generated from short-term government securities which can change in as little as a matter of days. CAPM also makes the assumption that investors can borrow and lend at a risk-free rate, however in reality individual investors are unable to borrow/lend at the same rate of the U.S. government. Lastly, its major drawback is that it is difficult to determine a β value that is truly reflective of the asset in interest, resulting in a proxy β being used in most cases. The alternative approach proposed would be to express systematic risk as the relative risk an individual asset contributes to a portfolio for a number of assets as expressed by the Shapley value.

Shapley value

Using concepts and language of game theory, a *cooperative game* is a pair $(\mathcal{N}, v)$ consisting of the set of *players* $\mathcal{N}$ and the *characteristic function* $v : 2^{\mathcal{N}} \longrightarrow \mathbb{R}$, with $v(\emptyset) = 0$. A *coalition* is defined as any subset $S \subset \mathcal{N}$. The *characteristic function* v has the following meaning - if S is a *coalition* of *players*, then $v(S)$ is the worth of *coalition* S and describes the total expected sum of payoffs the members of S can obtain by cooperation. The Shapley value is a means of distributing the total gains to the players, with the underlying assumption that they all collaborate. The amount that player i is given in a cooperative game $(\mathcal{N}, v)$ as defined by the Shapley, 1953 value as:

$$\varphi_i(v) = \sum_{S \subseteq \mathcal{N} \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} (v(S \cup \{i\}) - v(S))$$

$$\varphi_i(v) = \sum_{S \subseteq \mathcal{N} \setminus \{i\}} \binom{n}{1, |S|, n - |S| - 1}^{-1} (v(S \cup \{i\}) - v(S))$$

Where n is the total number of players and the sum extends over all subsets S of N not containing player i.

Financial Risk Measures

The financial risk measures of a portfolio is derived from the following equations:

$$\sigma_w^2 = \sum_{i=1}^{n} \omega_i^2 \sigma_i^2$$

$$\sigma_b^2 = 2 \sum_{i=2}^{n} \sum_{j=1}^{i-1} \omega_i \omega_j cov(r_i, r_j)$$

$$\sigma_p^2 = \sum_{i=1}^{n} \omega_i^2 \sigma_i^2 + 2 \sum_{i=2}^{n} \sum_{j=1}^{i-1} \omega_i \omega_j cov(r_i, r_j)$$

$$\sigma_p^2 = \sigma_w^2 + \sigma_b^2$$

Here the σ_w^2 is the specific risk (within-security risk) and σ_b^2 is the systematic risk (between-security risk). ω_i and ω_j are the weight attributed to the ith and jth security, respectively, σ_i^2 & σ_j^2 are the variances of the securities and $cov(r_i, r_j)$ is the covariances between the ith and jth shares. From these equations alone it is not possible to measure the contribution of risk each security has in relation to the overall portfolio variance.

Shapley Application

The Shapley value can be used to decompose the risk of the portfolio to each security, this decomposition was proposed by Mussard and Terraza, 2008 - it is described as the following. Let I be the overall risk indicator and r_k $(k \in K = \{1, 2, ...n\})$ the security returns of the portfolio and this would represent a set of contributory factors that would together then account for the I value and would be written as:

$$I \equiv F(K) \equiv F(r_1, r_2..., r_k, r_n)$$

Here F is a function that properly accounts all the individual returns and returns the overall portfolio risk. The decomposition consists of assigning contributions C_k to each one of the factors of r_k with I being the sum of such contributing factors. Therefore the Shapley value for each factor would be expected marginal impact when all other possible combinations are considered. Here $F(S)$ is defined as the value that I can take when factor r_k is removed, therefore $S = K \setminus \{r_k\}$ and since Shapley's value considers all possible combinations then set of returns $S(S \subseteq K)$ is the domain of securities remaining after security elimination, s the number of factors remaining after removal of different factors (i.e. returns). Here the risk contribution of each k-th factor to the global risk indicator I is defined as below:

$$C_k(K, F) = \sum_{s==0}^{n-1} \sum_{S \subseteq K \setminus \{r_k\}} \frac{(n-1-s)!s!}{n!} \Delta_k F(S)$$

where:

$$\Delta_k F(S) = F(S \cup \{r_k\}) - F(S)$$

$$F(\emptyset) = 0$$

and

$$I = \sum_{k=1}^{n} C_k(K, F)$$

From here it is possible to decompose the contribution of each security to the between-security risk. Let $F : R^t X R^t$ be the weighted covariances between i-th and j-th securities on t observations (i.e. days, minutes), then it be described as:

$$F(r_i, r_j) \equiv w_j w_i cov(r_i, r_j)$$

From here it is possible to get the contribution of security to the between security risk, where $F(r_i) \equiv w_i^2 cov(r_i, r_j) = \sigma_i^2$ and $F(\emptyset) = 0$. Since we are only looking at two securities it would decompose as following:

$$C_i = 1/2(F(r_i, r_j) - F(r_j)) + 1/2(F(r_i) - F(\emptyset)$$

$$C_j = 1/2(F(r_i, r_j) - F(r_i)) + 1/2(F(r_j) - F(\emptyset)$$

and obtain

$$C_i + C_j = F(r_i, r_j) = w_j w_i cov(r_i, r_j)$$

therefore decomposing systematic risk σ_b^2 as:

$$\sigma_b^2 = 2\sum_{i=2}^{n}\sum_{j=1}^{i-1}(C_i + C_j)$$

and the individual contribution of a security as:

$$\sigma_{bi}^2 = 2\sum_{i=2}^{n} C_i$$

with risk σ_w^2 is naturally decomposable as:

$$\sigma_w^2 = \sum_{i=1}^{n} w_i^2 \sigma_i^2$$

and consequently individual contribution being:

$$\sigma_{wi}^2 = w_i^2 \sigma_i^2$$

From here we are able to obtain the overall contribution of i-th security to the portfolio risk σ_p^2:

$$\sigma_{pi}^2 = \sigma_{wi}^2 + \sigma_{bi}^2$$

This allows us to get the relative contribution of the i-th portfolio to the overall portfolio risk σ_p^2.

Model

Through decomposing a portfolio's overall risk to individual i-th securities it is possible to implement this as a means of selection in a Genetic Algorithm (GA). The genetic algorithm will start by initiating a population of portfolios of N size, with each portfolio being of size s, where s is randomly picked amongst the top 25 securities with the largest market cap. Monte Carlo (MC) simulations will be used to assess the best distribution of weights to attribute to the initial portfolio. MC simulations will be based on historical returns data for the previous 3 years (2020-2023) to run 10,000 simulations of returns for a period of 1000 days.

Each portfolio will then be assessed based on the 1000-day projected distribution of returns - each portfolio starting out with an initial value of $10000 dollars. The Value at Risk(VaR) and Conditional Value at Risk (CVaR) VaR quantifies the extent of possible losses for the portfolio over n-days for a confidence level of $\alpha = 5$. CVaR is calculated as the average of values the fall beyond VaR and defined as such:

$$CVaR = \frac{1}{1-\alpha} \int_{-1}^{VaR} xp(x)dx$$

where $p(x)$ is the probability density getting a return with value x, and α the cut-off point on the distribution (i.e. confidence interval).

The Genetic Algorithm will then run for 30 generations, at each generation assessing the CVaR, Median Expected Returns, VaR,and the Shapley values for the portfolio. It will then replace n number of securities with the highest Shapley Value for the portfolio's overall risk with new securities from the top 25 largest market cap securities. At the end of the GA optimization the goal would be to minimize both VaR and CVaR while increasing the rate of returns, and obtain a Shapley-Pareto optimal solution. See Appendix for testing results of the model.

Results

After running the GA here are some of the results for the portfolios and the overall CVaR and VaR for each generation, as well of the compositions of the portfolios and their Shapley values. As shown in Figure 1, through this means it was possible to reduce the overall CVaR and VaR as it progressed through each generation.

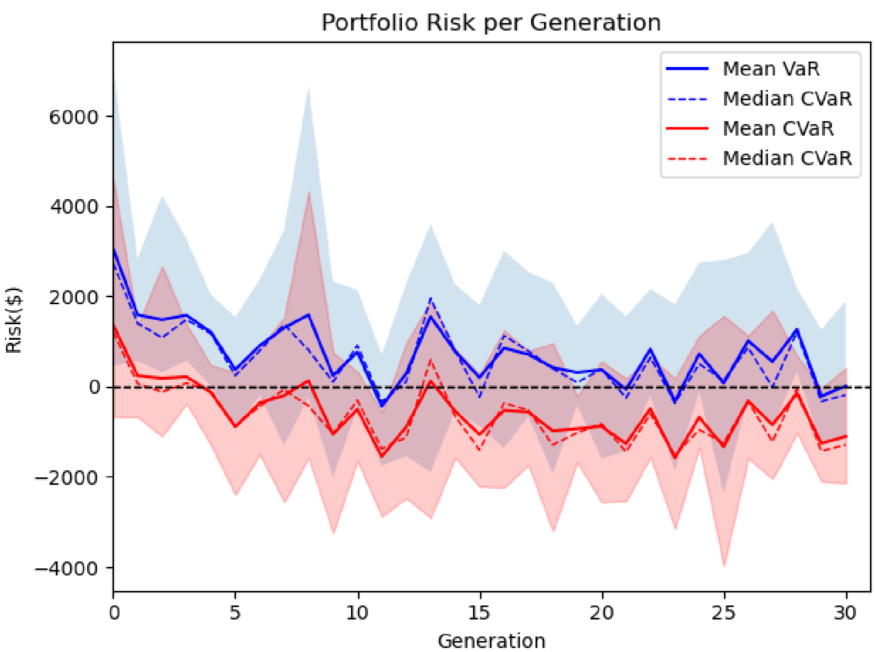

Figure 1

CVaR and VaR per Generation for all Portfolios

For portfolio 4, the initial conditions showed that it had an annualized return rate of 28.24% and that was increased in the last generation to 29.48%, with median projected return going from \$19767.74 to \$20294.81. The Shapley chart on Figure 4 shows us that it was composed of LLY, PG, META, MA, V, HD, and JNJ with LLY, and PG contributing the most to the portfolio's risk at the last generation.

Portfolio 0's, initial generation showed that it had an annualized return rate of 20.09% and that was increased in the last generation to 25.8%, with median projected return going from \$16513.31 to \$18756.18. Also for portfolio 0, the Shapley chart on Figure

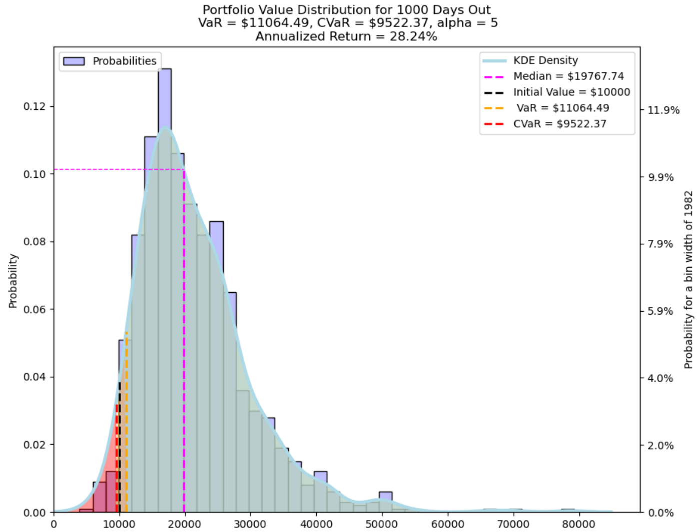

Figure 2

Portfolio 4 Initial

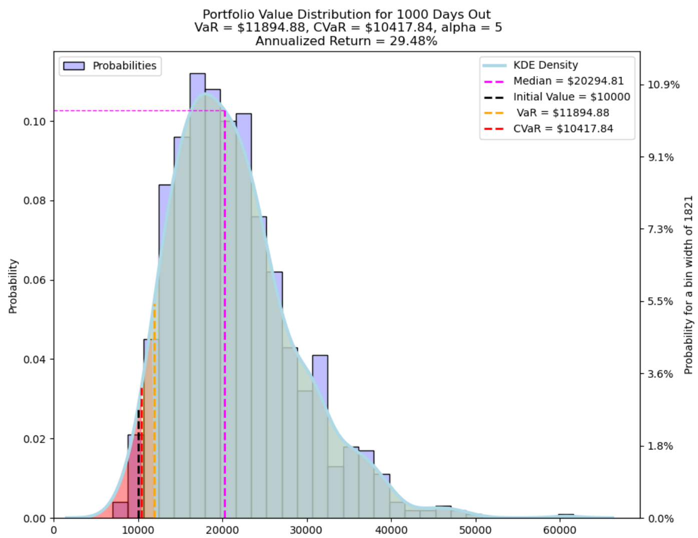

Figure 3

Portfolio 4 Final

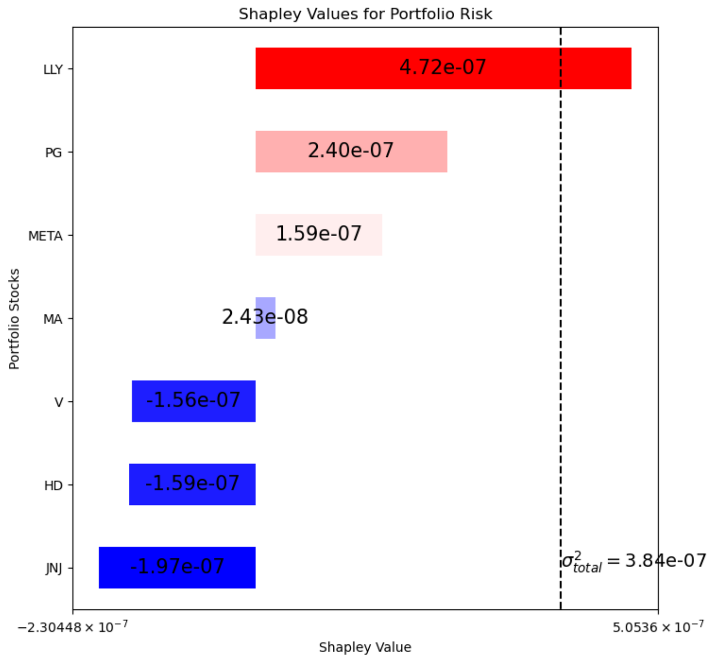

Figure 4

Portfolio 4 Composition and Shapley Values

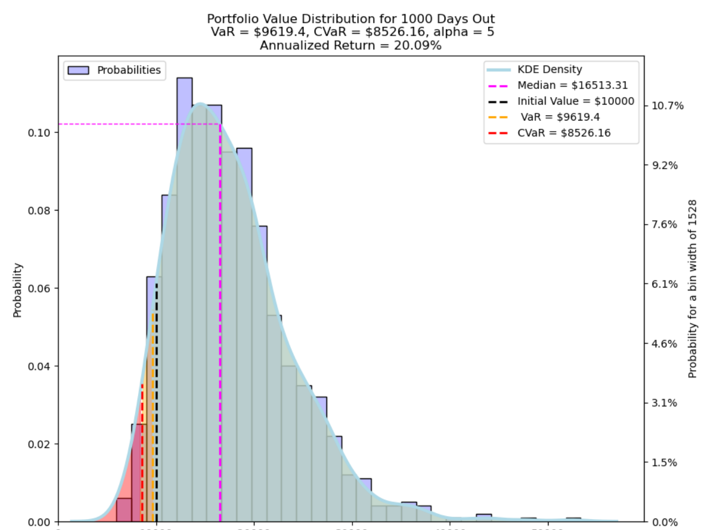

Figure 5

Portfolio 0 Initial

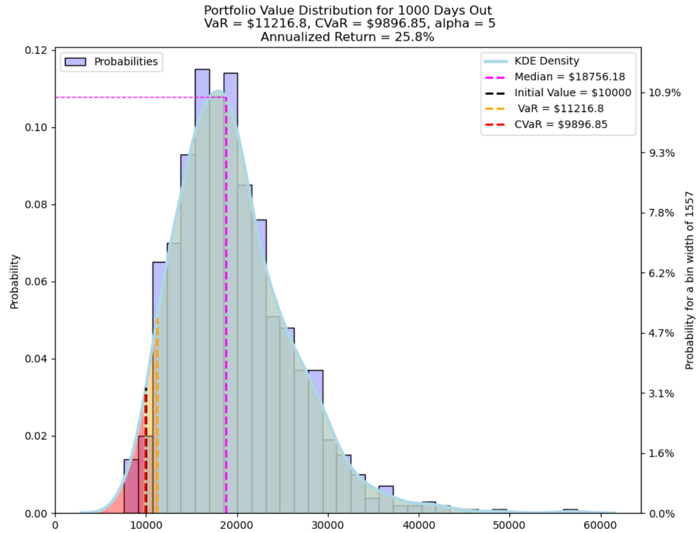

Figure 6

Portfolio 0 Final

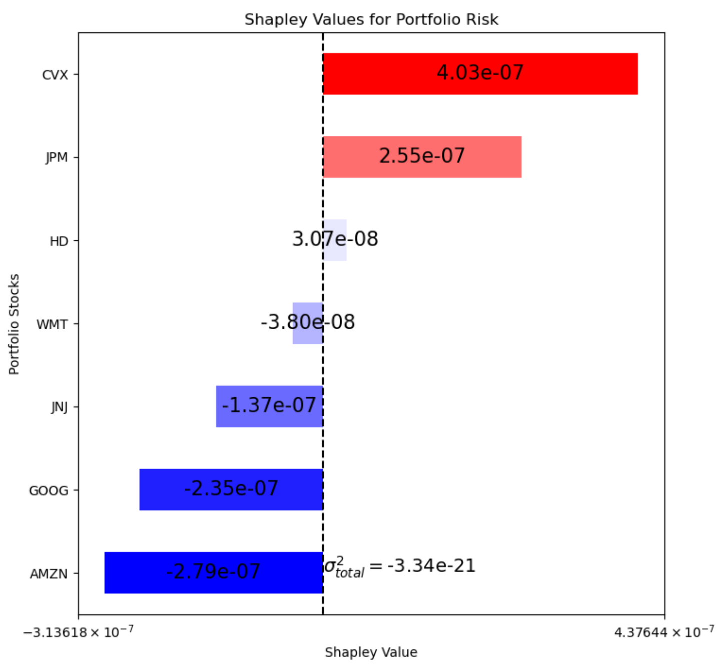

Figure 7

Portfolio 0 Composition and Shapley Values

7 shows us that it was composed of CVX, JPM, HD, WMT, JNJ, GOOG, AMZM with CVX and JPM contributing the most to the portfolio's risk at the last generation.

Conclusion

Portfolio risk decomposition can provide crucial information for risk management and allocation decisions – however it is not a straightforward process by simply using the expression for portfolio variance, σ_p^2. This paper takes a look at using the Shapley value to decompose a portfolio's between and within–variance to individual components that make up that portfolio. In doing so, it is possible to allocate risk of the portfolio to individual components. Through means of genetic algorithm methods and Monte Carlo simulations it is possible to obtain a Shapley-Pareto optimal solution that decreases VaR, and CVaR while increasing annualized rate of returns.

References

Bureau of Labor Statistics, U. (2023). Financial analysts : Occupational outlook handbook. https://www.bls.gov/ooh/business-and-financial/financial-analysts.htm

Elton, E. J., & Gruber, M. J. (1997). Modern portfolio theory, 1950 to date. *Journal of Banking & Finance*, *21*(11), 1743–1759. https://doi.org/https://doi.org/10.1016/S0378-4266(97)00048-4

Gunjan, A., & Bhattacharyya, S. (2022). A brief review of portfolio optimization techniques. *Artificial Intelligence Review*, 1–40. https://doi.org/10.1007/s10462-022-10273-7

Hagan, P. S., Lesniewski, A., Skoufis, G. E., & Woodward, D. E. (2021). *Portfolio risk allocation through Shapley value* (Papers No. 2103.05453). arXiv.org. https://ideas.repec.org/p/arx/papers/2103.05453.html

Lintner, J. (1956). Distribution of incomes of corporations among dividends, retained earnings, and taxes. *The American Economic Review*, *46*(2), 97–113. Retrieved February 10, 2023, from http://www.jstor.org/stable/1910664

Mann, I., & Shapley, L. S. (1960). *Values of large games, iv: Evaluating the electoral college by montecarlo techniques*. RAND Corporation.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91. Retrieved February 10, 2023, from http://www.jstor.org/stable/2975974

Moehle, N., Boyd, S., & Ang, A. (2021). Portfolio performance attribution via shapley value.

Morelli, G. (2023). Responsible investing and portfolio selection: A shapley - cvar approach. *Annals of Operations Research*. https://doi.org/10.1007/s10479-022-05144-x

Mussard, S., & Terraza, V. (2008). The shapley decomposition for portfolio risk. *Applied Economics Letters*, *15*(9), 713–715. https://doi.org/10.1080/13504850600748968

Shalit, H. (2020). Using the shapley value of stocks as systematic risk. *The Journal of Risk Finance*, *21*, 459–468.

Shalit, H. (2021). The shapley value decomposition of optimal portfolios. *Annals of Finance*, *17*, 1–25. https://doi.org/10.1007/s10436-020-00380-2

Shapley, L. S. (1953). 17. a value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the theory of games (am-28), volume ii* (pp. 307–318). Princeton University Press. https://doi.org/doi:10.1515/9781400881970-018

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*(3), 425–442. Retrieved February 10, 2023, from http://www.jstor.org/stable/2977928

Appendix

Test Calculations of Shapley Values

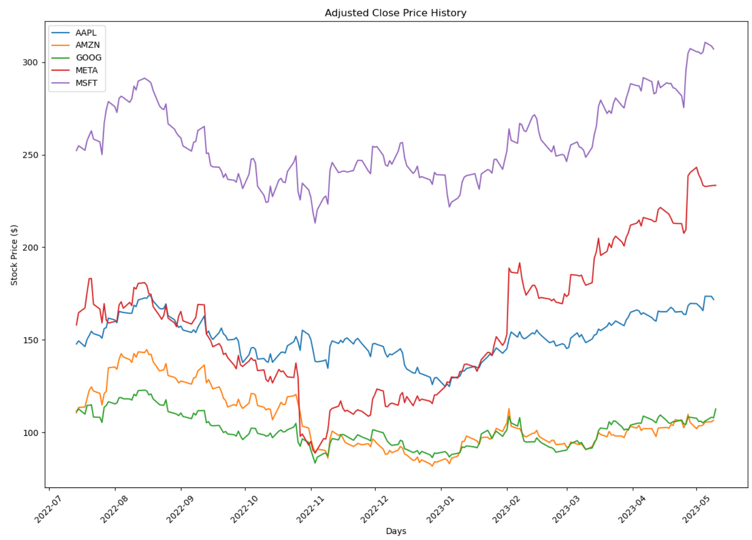

Figure A1

Prices

Testing getting the best weights (highest Sharpe Ratio) from MC Simulations

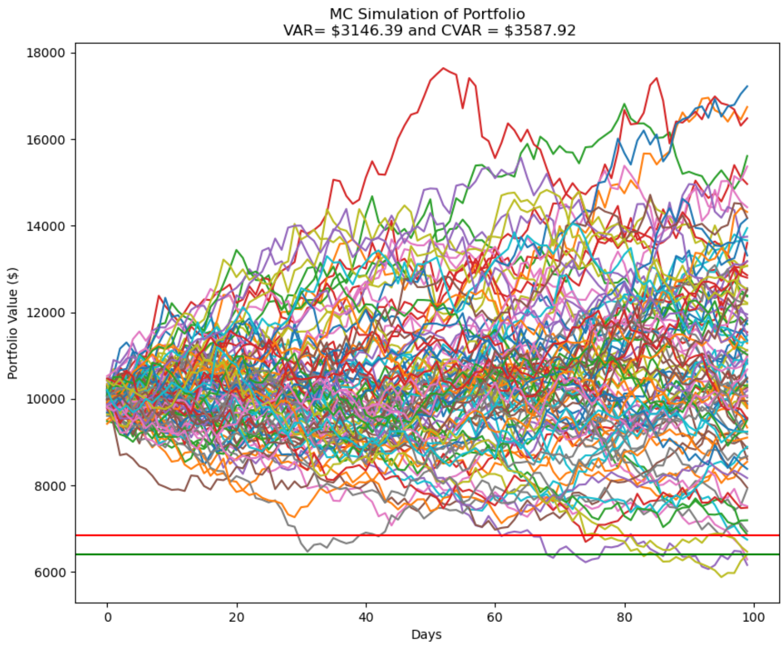

Figure A2

Monte Carlo Simulations

```
        AAPL      MSFT      AMZN      GOOG      META Expected Return  \
30  0.247515  0.015379  0.005523  0.440934  0.290647        0.437967

   Portfolio Variance Portfolio Std Sharpe Ratio
30           0.155738      0.394637     1.059119
```

Figure A3

Best Weights via Sharpe Ratio

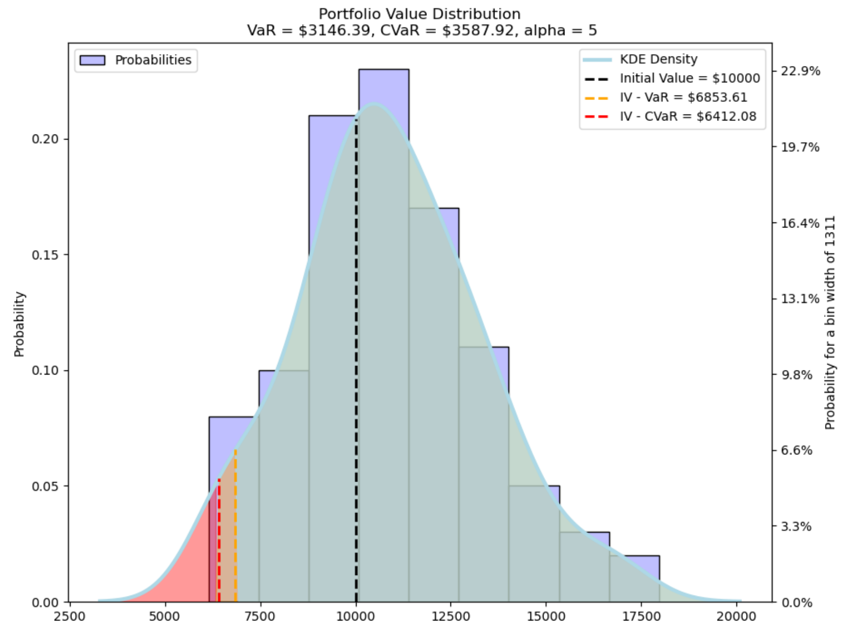

Figure A4

Getting Probability Distribution Graph with Var & CVaR

```
Risk within
 {
    "AAPL": 5.476940184093452e-05,
    "AMZN": 3.55932402688336e-05,
    "GOOG": 1.0284268495183721e-05,
    "META": 4.557463955379237e-05,
    "MSFT": 4.326558863914317e-06
}
```

Figure A5

Calculating Within Risk

```
Weighted Between Covariance
 {
    "wCov_": 0,
    "wCov_AAPL": 1.2451714129188924e-19,
    "wCov_AMZN": 1.1641576650401931e-19,
    "wCov_GOOG": -1.7963954999356906e-19,
    "wCov_META": 9.610889272788566e-20,
    "wCov_MSFT": 7.909758018963831e-20,
    "wCov_AAPL/AMZN": 2.8379543184250174e-05,
    "wCov_AAPL/GOOG": 1.7326460291538388e-05,
    "wCov_AAPL/META": 2.785927537800464e-05,
    "wCov_AAPL/MSFT": 1.1452747318593081e-05,
    "wCov_AMZN/GOOG": 1.388878729511898e-05,
    "wCov_AMZN/META": 2.3323592430504e-05,
    "wCov_AMZN/MSFT": 9.078287024955224e-06,
    "wCov_GOOG/META": 1.4346342264465848e-05,
    "wCov_GOOG/MSFT": 5.430811720550321e-06,
    "wCov_META/MSFT": 8.2816110035763e-06,
    "wCov_AAPL/AMZN/GOOG": 3.474954169401786e-08,
    "wCov_AAPL/AMZN/META": 7.990806113160119e-08,
    "wCov_AAPL/AMZN/MSFT": 2.535616515011491e-08,
    "wCov_AAPL/GOOG/META": 4.1076882473420737e-08,
    "wCov_AAPL/GOOG/MSFT": 1.5695933587524974e-08,
    "wCov_AAPL/META/MSFT": 2.689413561883843e-08,
    "wCov_AMZN/GOOG/META": 4.548851066393586e-08,
    "wCov_AMZN/GOOG/MSFT": 9.900913240052167e-09,
    "wCov_AMZN/META/MSFT": 2.677871897270257e-08,
    "wCov_GOOG/META/MSFT": 1.3970879761248641e-08,
    "wCov_AAPL/AMZN/GOOG/META": 2.1032313303801738e-09,
    "wCov_AAPL/AMZN/GOOG/MSFT": 7.266839396497606e-10,
    "wCov_AAPL/AMZN/META/MSFT": 1.3925346872630386e-09,
    "wCov_AAPL/GOOG/META/MSFT": 7.522993505594092e-10,
    "wCov_AMZN/GOOG/META/MSFT": 6.413283808895614e-10,
    "wCov_AAPL/AMZN/GOOG/META/MSFT": 7.166443500057388e-12
}
```


Figure A6

Between Risk Decomposition

```
Shapley Values
 {
    "AAPL": 1.7753578952481798e-06,
    "AMZN": 9.128872319756759e-07,
    "GOOG": -1.0655317945741618e-06,
    "META": 8.422725041193763e-07,
    "MSFT": -2.4649786703255703e-06
}
Sum of Shapley Values
 7.166443500046792e-12
Between Risk Total
 7.166443500057388e-12
```

Figure A7

Testing if Shapley Values Add up