Optimal Risky Portfolio

Group 4: Fuyu Liu, Yuehan Yu, Zhuoping Fan

**Introduction**

Portfolio optimization plays a crucial role in maximizing returns while managing risk, making it a pivotal aspect of investment strategies. In the following report, we embark on the interpretation of the data we use specifically and how we use Python to do a comprehensive exploration of the returns generated by constituent companies within the S&P 500 index. The goal is to pinpoint and assemble an optimal investment portfolio that strikes an ideal equilibrium between risk and return. By delving into the intricacies of individual company returns, we seek to uncover the Sharpe ratio that will create a well-balanced and high-performing investment portfolio. After obtaining the results, we make a summary and discuss the implications of the study.

**Description of Data**

The first step in the analysis is data preparation. The 'sp500constituent.csv' file includes 'PERMNO' (permanent number) and 'MthRet' (monthly return) variables, while the index data encompasses 'mthtotret' and the treasury bill dataset contains '30-day return'. The dataset comprises S&P 500 constituent companies' monthly returns obtained from the 'sp500constituent.csv' file covering the period from 2010 to 2022. Additionally, S&P 500 index monthly returns and 30-day treasury bill rates are sourced from WRDS, spanning the same period. Following are the purpose and utilization of each dataset:

1. Constituent Monthly Returns:

This dataset encapsulates the monthly returns of individual constituent companies within the S&P 500 index, identified by their unique permanent numbers ('PERMNO') and their corresponding monthly returns ('MthRet'). It serves as the foundation for analyzing the performance of each S&P 500 constituent company over the period from 2010 to 2022. This data facilitates the evaluation of individual stock performances within the index, aiding in the assessment of their potential inclusion in an optimized investment portfolio.

1. S&P 500 Index Monthly Returns:

This dataset records the monthly returns of the broader S&P 500 index. It offers insights into the overall market performance represented by the S&P 500 index. This data provides a broader market perspective.

1. 30-Day Treasury Bill Monthly Returns:

This dataset comprises the monthly returns of 30-day treasury bills, which are often considered a proxy for the risk-free rate in investment analysis. The treasury bill rates serve as a benchmark for evaluating the performance of investment portfolios, enabling the calculation of risk-adjusted returns and Sharpe ratios.

**Methodology**

The methodology employed in this project revolves around utilizing Python functions to process data related to S&P 500 constituent companies. The primary objective is identifying the optimal investment composition by determining the component weights and Sharpe ratio.

Initially, upon importing the data, we transformed it into a pivot table, excluding stocks with incomplete data to ensure ease and accuracy in subsequent model fitting. The average monthly return and the covariance matrix of S&P 500 constituent companies were calculated using simple functions such as .mean() and .cov(). The variable "N" was defined to represent the number of components.

Subsequently, in preparation for the following iterations, three crucial functions were developed, with Numpy's np.dot() being a key element due to matrix multiplication:

1. The function to compute the expected return of the portfolio: F\_PortRtn(r\_i, w\_i). Its original formula is E(Rp) = Σ (r\_i \* w\_i).

2. The function to calculate the portfolio standard deviation: F\_PortStd(cov, w\_i). Its original formula is σp = sqrt(Σ (σ\_i^2 \* w\_i^2)).

3. The function to determine the Sharpe ratio of the portfolio: F\_Sharpe(r\_p, r\_f, s\_p). Its original formula is S.R. = 𝐸[𝑟𝑝 − 𝑟𝑓]/𝜎𝑝.

Finally, a set of (1000, N) uniformly distributed random numbers between 0 and 1 was generated using random.uniform(). Within each set, the numbers were normalized by dividing them by the total sum of the set. This process yielded 1000 sets of different investment proportions. Subsequently, a "for" loop was implemented to calculate the Sharpe ratio for each simulated investment portfolio. The loop leveraged the previously processed data and the three functions developed earlier.

In theory, a higher Sharpe ratio indicates a more favorable outcome, signifying higher investment returns at a certain level of risk. Therefore, we sought the maximum value within the 1000 Sharpe ratios, facilitating the identification of the optimal Sharpe ratio and corresponding investment composition weights.

**Result**

The primary objective of our project is to compute and compare the Sharpe Ratios across these simulated portfolios, thereby identifying the stocks offering optimal monthly returns. In our analysis, we successfully generated 1,000 distinct portfolio sets using Python, each comprising random weights of stocks in a particular portfolio. We attained a list that contains Sharpe Ratios of 1000 portfolios. The portfolio with the largest Sharpe Ratio is chosen as the final result.

Distinguished by its unique combination of stock weights and expected returns, this portfolio reached the optimal balance between risk and reward within our simulated parameters. This optimal portfolio constituents’ allocation, alongside its Sharpe Ratio, underscores its potential for higher risk-adjusted returns compared to other portfolios.

**Conclusion**

This analysis conducted using Python offers valuable insights into portfolio optimization in the context of the S&P 500 index. Our methodology, which involves the generation of 1,000 random portfolios and the calculation of their respective Sharpe Ratios, has led to the identification of an optimal investment strategy within the simulated constraints. This outcome showcases the efficacy of using computational methods and statistical models in the financial decision-making process, particularly in portfolio investment.

While the results are promising, it is crucial to consider the limitations of simulation-based analysis. The reliance on historical data and the assumptions in the process of Sharpe Ratio calculation may not fully encapsulate future market dynamics. Therefore, while the identified optimal portfolio presents a theoretically sound investment strategy, actual market performance may vary due to other unforeseen factors.