Springboard–DSC Program // Capstone Project 3 Proposal

Portfolio Optimization

By Steve Walters

May, 2023

Problem Statement (Hypothesis): What's the best allocation of assets within a portfolio? This project aims to answer that question by creating models that optimize portfolios based on annual return or risk.

Investing is an excellent way to increase capital, and it has been shown that a diversified portfolio can provide the best risk-adjusted returns. However, it isn't always easy to determine the best mix of stocks and other assets to maximize returns and/or minimize risk in a diversified portfolio.

Having data-driven methods for optimizing the allocation of stocks or other assets in a portfolio can be applied in use cases, for a broad section of investors, from individuals to portfolio managers, to fund managers.

This project will use the PyPortfolioOpt library to look at diversification in stock portfolios to either maximize annual return or minimize risk based on a covariance matrix. We will look at this through three different optimizers:

1. Efficient Frontier
2. Black-Litterman
3. Hierarchical Risk Parity

The Efficient Frontier optimizer uses the work of Harry Markowitz and its key insight is that by combining assets with different expected returns and volatilities, one can decide on a mathematically optimal allocation which minimizes the risk for a target return.

Black-Litterman asset allocation combines a prior estimate of returns (e.g the market-implied returns) with your own views to form a posterior estimate. This results in much better estimates of expected returns than just using the mean historical return.

Hierarchical Risk Parity uses clustering algorithms to choose uncorrelated assets. It is a portfolio diversification technique that uses machine learning (single linkage clustering) to allocate weights to asset classes based on their risk contributions. The algorithm was introduced in 2016

by Marcos López de Prado in his paper, Building Diversified Portfolios that Perform Well Out-Of-Sample.

Once we have optimized a sample portfolio of five randomly selected stocks we will then look to validate the results and determine the most effective portfolio optimization method through the use of regression models. We will model using linear regression, OLS, and Ridge on all three of the portfolio optimizations to further refine the optimizations and find the best combination of optimizer and regression model.

Data sources: This project will use a csv of all S&P 500 stocks obtained from Github along with the yfinance Python library to retrieve S&P 500 historical stock data from Yahoo Finance. Because the model will encompass the entirety of the S&P 500 it is possible (though not guaranteed) that data for all 500 stocks in the index will be used in the models. By using the same portfolios with each optimizer we will also be able to determine if one optimizer outperforms the others.

Deliverables:

- A GitHub repo containing the work completed for each step of the project, including:
  - All python code
  - A slide deck
  - A project report