MTD 350

Two stage Portfolio optimisation using Ball Mapper Algorithm

Project Members

Anshuman Panda
2019MT10463

Ritvik Gupta
2019MT10512

Supervisor

Prof. Aparna Mehra
Department of Mathematics
Indian Institute of Technology, Delhi

Contents

1 Introduction

In finance, portfolio construction or asset allocation is one of the most frequent problems faced by practitioners. This was first introduced by Markowitz in his Portfolio Theory in which he used Mean-Variance as optimisation problem, specifically reducing portfolio variance (or risk) for a given expected return and vice versa (maximising expected return for a given variance). Beating the market consistently is very difficult. In fact, the efficient market hypothesis is a theory that claims share prices of stocks always trade at fair value because they incorporate all relevant information. This hypothesis therefore implies that it is impossible for one to sell stocks that are too inflated or buy stocks that are undervalued. An investor, therefore, can only obtain higher returns by partaking in riskier investments.

Additionally, there are many biases involved when a person chooses stocks manually. Illogical decision making has been observed in studies on human behavior when financial choices are being made. Negativity bias (overweighting bad news more than good news) and tendency to weight tail events, i.e., events that are highly unlikely to happen (like winning the lottery). Due to these biases and the many more that influence the decision-making of investors, a more quantitative and formulaic way to choose investments is necessary to reduce biases in judgement, as well as reduce risk and maximize returns despite uncertainty in the market.

Modern portfolio theory is the most widely used practice by individuals to develop portfolios. It is based on a principle of attempting to maximize expected return for a given amount of risk or equivalently minimizing risk for a given amount of return. A highly utilized method to reduce risk is diversification. The idea of diversification is to split investment between varying companies so that if a few securities one owned were to take a downturn, the others would not, reducing the loss. Risk can be diversified by picking assets that are different from each other with respect a particular aspect about the assets themselves. This aspect could be related to industry, country, type of asset, or more. An automated method to classify or cluster assets would be very useful and essential to investment decision-making and the practice of diversification. The stocks would be separated into groups via a clustering method that maximizes similarity within groups and minimizes similarity between groups. Doing this with securities would allow one to figure out what combination of assets could make up a well diversified portfolio.

Topological Data Analysis is an approach to the analysis of datasets using rigorous techniques of topology, like mapper graphs as well as persistent homology, to analyze complicated, large, noisy and high dimensional data. The Ball Mapper algorithm [1] is a new addition to the TDA tools and techniques. It is a mapper-inspired method to obtain descriptors of the local and the global structure of given data sets, which can be used in exploratory data analysis. It has been used previously to study the relation between financial ratios and stock returns [2], and to detect corporate failure [5].

2 Modern Portfolio Theory

MPT is a practical method developed first by Harry Markowitz for allocating funds to assets in order to maximize their overall returns within an acceptable level of risk.
This can be formulated as an optimization problem as follows:

$$\begin{aligned} \min_{w} \quad & \frac{1}{2} w^t \Sigma w \\ \text{s.t.} \quad & w^t \mu = p \\ & w^t 1 = 1 \end{aligned}$$

2.1 Expected Return

It is the amount of profit or loss from an asset gets on the basis of historical data of return rates. It can be measured by average value of the returns of the asset.

2.2 Expected Risk

It is defined as the chance that an asset's actual gains are different from the expected rewards of the asset. Standard deviation of the gains is a common metric associated with financial risk. It is measure of the risk associated with the asset prices in comparison to their historical averages for some period of time.
There are two main types of risk:
Systematic Risk: inherent to the market and affects the entire market
Idiosyncratic Risk: specific to a particular asset or a group of assets

2.3 Sharpe Ratio

It is used to help investors understand the return of an investment compared to its risk.

$$SharpeRatio = \frac{R_p - R_s}{\sigma_p}$$

R_p = return of portfolio

R_s = risk-free rate

σ_p = standard deviation of portfolio's excess return

2.4 Efficient Frontier

It is the set of optimal portfolios that maximise returns for a given amount of risk or those portfolios with minimum risk for a given amount of expected returns. Portfolios that lie below that efficient frontier are sub-optimal as they do not maximise the returns or minimise the risk.

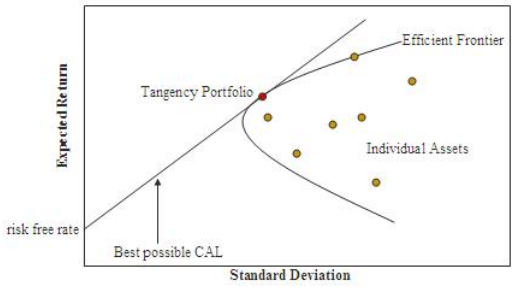


Figure 1: Efficient Frontier

2.5

3 Topological Data Analysis

In order to get an understanding of the ideas of Topological Data Analysis, we first cover a few concepts related to it.

3.1 Simplex

Simplex is the generalisation of the of the notion of a triangle or tetrahedron to arbitrary dimensions. A k-dimensional simplex is a k-dimensional polytope which is the convex hull of $k+1$ vertices in $\mathbb{R}^k$.
If $v_1, v_2, \cdots v_k$ are $k+1$ vertices which are affinely independent, i.e. $v_2 - v_1, v_3 - v_1, \cdots v_{k+1} - v_1$ are linearly independent. Then, the simplex created by them is-

$$S = \left\{ \sum_{i=1}^{k+1} c_i v_i \middle| \sum_{i=1}^{k+1} c_i = 1,\ c_i \geq 0, \forall i = 1, 2, 3 \cdots k+1 \right\}$$

Simplices in 0, 1, 2, and 3 dimensions are a point, a line segment, a triangle and a tetrahedron respectively.

3.2 Simplicial Complex

A simplicial complex $\mathcal{K}$ is a set of simplices that satisfies the following conditions-

- Every face of a simplex from $\mathcal{K}$ is also in $\mathcal{K}$
- The non-empty intersection of any two simplices $\sigma_1, \sigma_2 \in \mathcal{K}$ is a face of both σ_1 and σ_2.

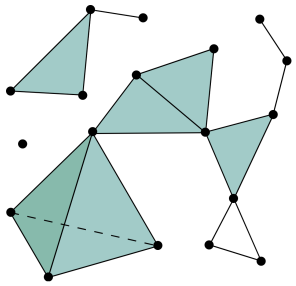


Figure 2: A Simplicial Complex

3.3 Abstract Simplicial Complex

A collection Δ of non-empty finite subsets of a set S is called a set-family.
A set-family Δ is called an abstract simplicial complex if, for every set X in Δ, and every non-empty subset $Y \subseteq X$, the set Y also belongs to Δ. Or the set-family Δ is closed under taking subsets.
Abstract Simplicial Complex are the combinatorial counterparts of Simplicial complexes and for every abstract simplicial complex, there exists a corresponding simplicial complex and vice-versa.

3.4 Simplicial Homology

Simplicial homology formalizes the idea of the number of holes (voids or their higher dimensional counterparts) in a simplicial complex.
This generalizes the number of connected components in the case of dimension 0.

3.5 Persistant Homology

Persistent homology is a method for computing topological features of a space at different spatial resolutions. In simple terms, persistent homology describes the changes in homology that occur to an object which evolves

with respect to a parameter. It identifies topological features of the underlying data at different resolutions and clubs all the information obtained. It is stable from perturbations in input as desired by any algorithm.

3.6 The Ball Mapper Algorithm

The Conventional Mapper algorithm uses a number of parameters, like the function $f : X \to R$, the cover of $\mathbb{R}$, and the clustering algorithm to find an overlapping cover C of a point cloud X. The Ball Mapper Algorithm aims to create similar outputs by using a single parameter instead of multiple parameters as in the case of the Conventional Mapper algorithm. The main idea is to first construct ϵ-nets and then constructing a nerve using this. This allows us to project point clouds in high-dimensions in the form of a nerve complex for better interpretability.
Besides this, points that are close in BM graph are also close in the corresponding CM graph, but the converse is not true. Thus, BM graph brings along more accurate information about the point cloud, but at times this can come at the cost of its interpretability.

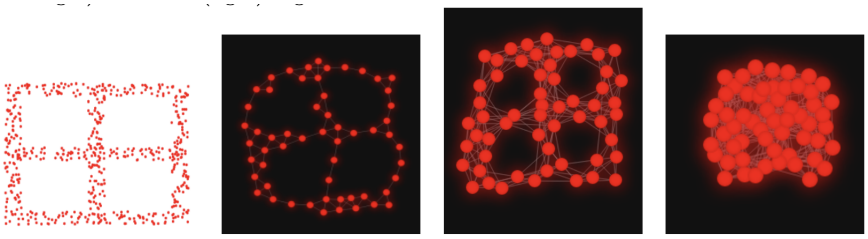


Figure 3: The Ball Mapper Algorithm

3.7 Ball Mapper and Portfolio Optimisation

The aim of the project is to cluster companies in order to build a diversified portfolio. Companies with similar financial ratios (debt to assets, asset turnover, return on assets) tend to perform similar to each other, even in times of crisis.
The Ball Mapper algorithm can be used to identify and cluster companies with similar values of the aforementioned ratios (these companies in an n-dimensional field represented by financial ratios will be within an epsilon neighborhood). Investing in companies from different clusters will lead to diversified portfolio, hence minimising idiosyncratic risk of the portfolio.

3.8 Optimizations in Ball Mapper

The CRAN package of Ball Mapper has some suboptimal code blocks that were optimised: In a part we needed to eliminate duplicate edges from the ball mapper graph, that was $O(n^2)$ in the CRAN package and was optimised to $O(nlogn)$ The whole code was deployed in C++, which dramatically decreased the runtime by 100 times

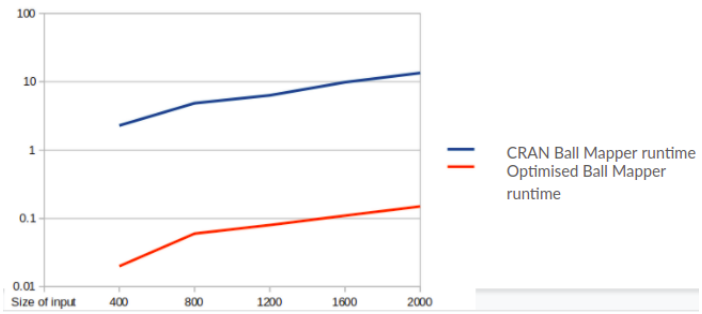


Figure 4: The Ball Mapper Algorithm Optimisation

4 Related Work

Automated methods of classifying assets for the purpose of diversification are recent innovations. In 2005, Zhiwei Ren [3] in Portfolio Construction Using Clustering Methods uses cluster analysis to group highly correlated stocks and then uses those clusters to run mean-variance portfolio optimization. This paper and most other typically use clustering algorithms like K-means or some hierarchical clustering algorithm. The similarity measure they use are based on correlation between assets. The rationale is that if one cluster's stock price decreases, it is likely that the other cluster's stock price will not decrease, therefore creating a hedge and reducing loss. However, in times of financial stress, correlations can reverse and the clusters formed do not remain structurally sound. Furthermore, the increased risk would be coming at a time of financial stress, which is the worst possible time. Similar work has been done in [4].

Other approaches use measure of similarity that are related to the previous success or potential for growth of the companies. They use financial ratios like $\frac{Revenue}{Asset}$ or $\frac{Net\ Income}{Asset}$ as the similarity measure for clustering that are more inherent to the companies rather than correlation of stock prices. A clustering method is run on the data to partition it into groups. Then, a stock is picked from each cluster. Since the clusters are (ideally) very different from each other, in case one of the stocks goes down, the other stocks do not loses value as much, and make up for the losses. In the interest of having high returns and low risks, stocks are picked to maximize the return to risk ratio. The Sharpe ratio, or the average return earned in excess of the risk-free rate per unit of volatility, is a measure of calculating risk-adjusted return. The stock from each cluster with the highest Sharpe ratio is therefore picked to be in the portfolio. The portfolio is then diversified and comprised of historically high performing stocks. Actual partition of capital allocation to each of these diversified assets is done using a traditional model like MVO (Mean Variance Optimisation).

All the literature on using Clustering for portfolio optimisation so far has used classical clustering algorithms like K-means, Average Linkage (AL), Complete Linkage (CL), Ward's Method (WM) etc. In this work we explore a different type of clustering algorithm derived from Topological Data Analysis called Ball Mapper algorithm.

5 Formulation

This work aims at a two-stage portfolio optimisation: first stage will shortlist potential assets to invest in from a list of assets indexed in a typical stock exchange like NASDAQ, NIFTY, etc. The next stage will refine the choice of assets to invest in, and produce an algorithm that optimises the capital allocation to each of these assets based on various financial ratios.

The first stage will shortlist potential assets to invest in from a list of assets indexed in a typical stock exchange like NASDAQ, NIFTY, etc. Following this, we cluster the assets using the Ball Mapper algorithm based on input parameters like financial ratios and the fundamentals of companies. In the project, we have used the SP 500 stocks for the list of assets ad used Return on Assets and Debt to Assets ratio as parameters for distance metric while clustering the data.

The second stage refines the choice of assets to invest in, and optimises the capital allocation to each of these assets based on MVO. From each cluster, we pick some assets based on their Sharpe ratio and create a shortlist of assets from the 500 stocks available to us. We then apply MVO to the shortlisted stocks to find weights for asset allocation for portfolio with the best Sharpe ratio.

6 Data Source

- Company financial statements: Obtained from AlphaVantage API
- Prices of SP 500 Stocks: Obtained from AlphaVantage API
- Prices of SP 500 Index: Obtained from FRED

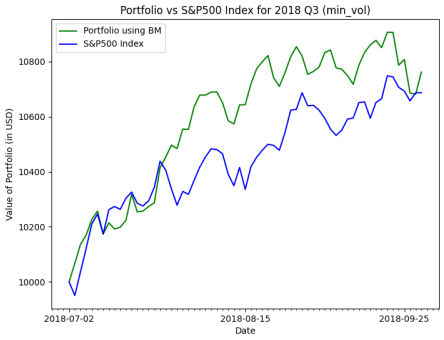
7 Evaluation

We first shortlist the stocks using financial data of SP 500 companies from a quarter. We use the company balance sheet data for getting the values Returns and Debt for the quarter. In order to normalise these values on the basis of company size, we divide them by the total assets of the company.

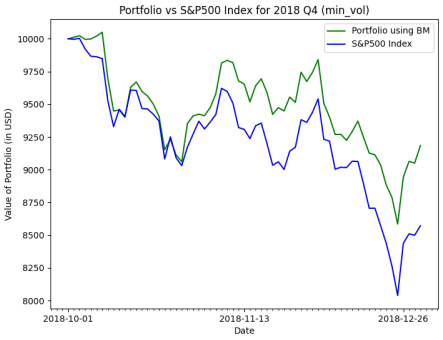
After using these values for running the two stages of our proposed method, we then test the performance of the portfolio in comparison to the SP Index for the following quarter by considering a portfolio of $10,000. We create the portfolio in two ways: by maximising sharpe ratio, or by minimum volatility.

8 Results

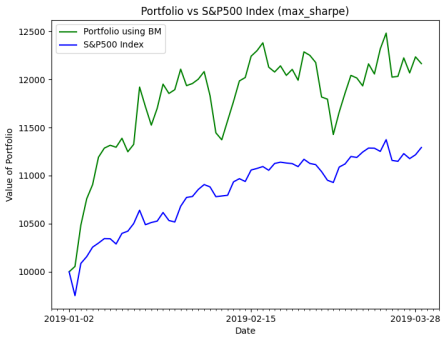
In the following graphs we have plotted our portfolio values against SP500 index values. In most of the quarters our portfolio seems to beat the index. Blue line indicates SP500 index value and green line indicates our portfolio value. Both start from $ 10000.



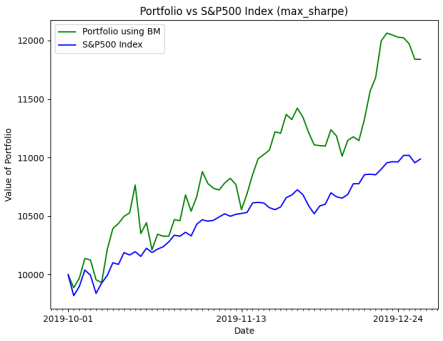
(a) 2018 Q3



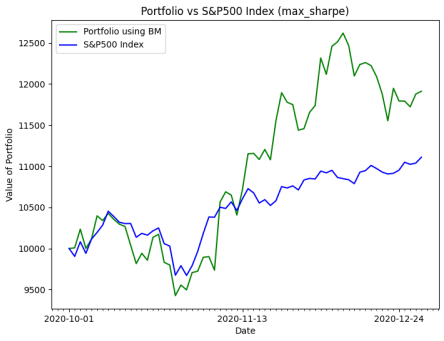
(b) 2018 Q4



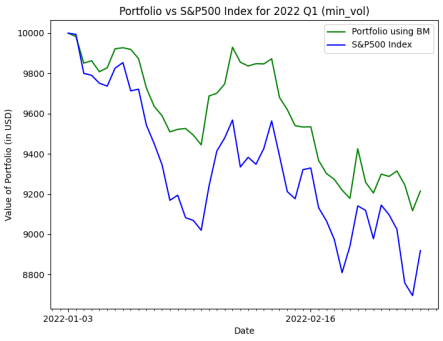
(c) 2019 Q1



(d) 2019 Q4

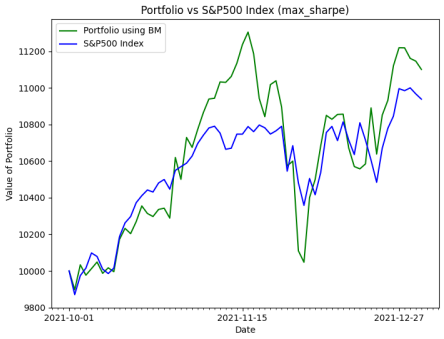


(e) 2020 Q4

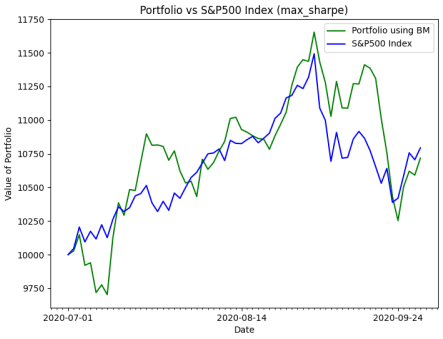


(f) 2022 Q1

Figure 5: Results where our strategy beats the index

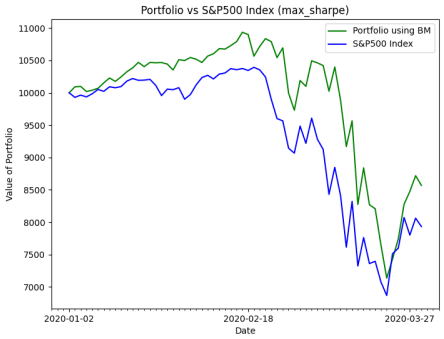


(a) 2021 Q4

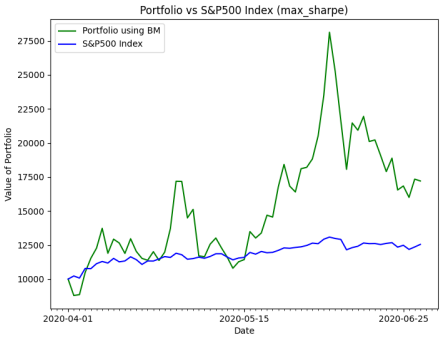


(b) 2020 Q3

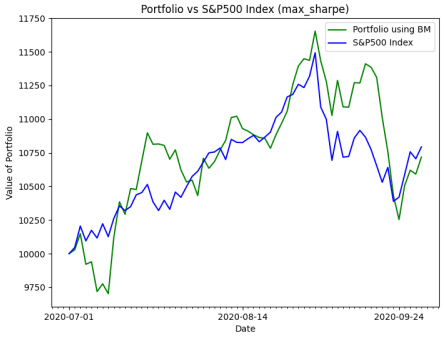
Figure 6: Results where our strategy follows the index



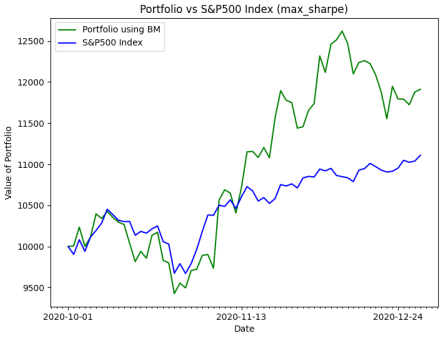
(a) 2020 Q1



(b) 2020 Q2

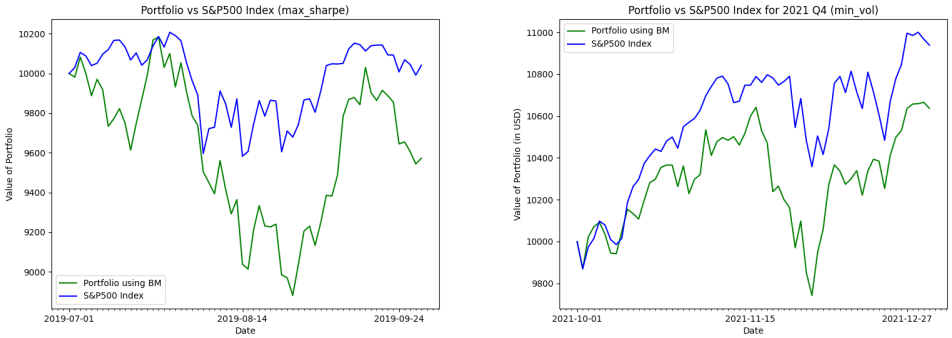


(c) 2020 Q3



(d) 2020 Q4

Figure 7: Results during the Covid-19 Pandemic



(a) 2019 Q3 (b) 2021 Q4

Figure 8: Results where our strategy couldn't keep up with the index

9 Future Work

- Currently, we have used Return to Asset and Debt to Assets ratio for refining using Ball Mapper Algorithm. We can explore other similar company fundamental ratios like Price to Equity, Return on Equity, Quick Ratio, etc.
- We can also use PCA to project high dimensional data incorporating multiple fundamental values to a lower dimension to allow proper clustering.
- Compare the performance of Ball Mapper Algorithm w.r.t. general clustering algorithms like k-mean clustering.
- Try out different models for portfolio selection.
- We have done the testing for S&P 500 companies. We could test the model for various other markets like NIFTY 500, BSE 500, etc. Currently, this was not possible due to lack of clean data.

References

[1] Paweł Dłotko. *Ball mapper: a shape summary for topological data analysis*. 2019. arXiv: 1901.07410 [math.AT].

[2] Pawel Dlotko, Wanling Qiu, and Simon Rudkin. *Financial ratios and stock returns reappraised through a topological data analysis lens*. 2019. arXiv: 1911.10297 [q-fin.ST].

[3] Diego León et al. "Clustering algorithms for Risk-Adjusted Portfolio Construction". In: *Procedia Computer Science* 108 (2017). International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland, pp. 1334–1343. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2017.05.185. URL: https://www.sciencedirect.com/science/article/pii/S187705091730772X.

[4] Karina Marvin and Swati Bhatt. "Creating Diversified Portfolios Using Cluster Analysis". en. In: (), p. 26.

[5] Wanling Qiu, Simon Rudkin, and Pawel Dlotko. *Refining Understanding of Corporate Failure through a Topological Data Analysis Mapping of Altman's Z-Score Model*. 2020. arXiv: 2004.10318 [q-fin.GN].