Portfolio Optimization: Creating the Optimal Index Fund

Jin Choi, Setu Madhavi Namburu, Wei Min Li, Soma Szabo

MSDS 460

15 March 2020

Table of Contents

Abstract

Index funds are a common investment vehicle in which the stocks or bonds that are purchased closely track the overall performance of a financial market benchmark. During the index fund creation process, a financial manager must select how to optimally allocate assets across the portfolio while meeting threshold criteria required by the investor and maximizing returns. As more allocation criteria are added, complexity increases and optimization problems become more complex. This paper outlines a novel application of machine learning to predict future returns and a portfolio optimization model that, using linear programming, optimally allocates the funds across industry sector, investment payback duration, and credit rating in order to achieve the highest returns possible. A numerical example is provided for the model.

Keywords

Portfolio optimization, index fund, linear optimization, multi-criteria optimization, machine learning in finance

Introduction

Investing is an important part of life for many working adults who are building their portfolios for retirement. Some are very confident about building their portfolio with their own security selection. However, not every individual has knowledge about investing. An investment fund is the best one-stop-shop solution for them. Therefore many retirement plans are offering predefined index funds for people to contribute. These funds have many different objectives such as all equity, all fixed income, or retirement target funds. Many of them are using a benchmark as a standard performance measurement to inform their investors of the performance over time compared to the fund they contributed to. In addition, investors are attracted to them due to their low management fees since they do not require a finance professional to reallocate and reinvest funds between assets after the fund's inception.

A security index is a pool of securities actively trading in an exchange. Companies such as Standard & Poor's create a security benchmark to represent the market performance in real time, so that investors can get a sense of how the market is doing. There are many types of benchmarks available in the market. Benchmarks can be stocks, bonds, commodities, or even a mix of each. Each benchmark serves its own purpose to represent the specific investment objective.

There is a cost in investing in an individual security as opposed to buying an index fund. The performance of any individual security can be impacted by event risk or financial risk such as bad quarterly earnings or natural disaster. The other cost of investing in an individual security is that you need to pay significantly more if you want to get more exposure across different market segments. Diversification is always the key to reduce risk from buying an individual security. The popular phrase in investment "Don't put all your eggs in one basket" reminds us of the importance of diversification when investing. Therefore, index funds are becoming more and more popular in recent years as people want to get more market exposure and more diversification in their portfolio.

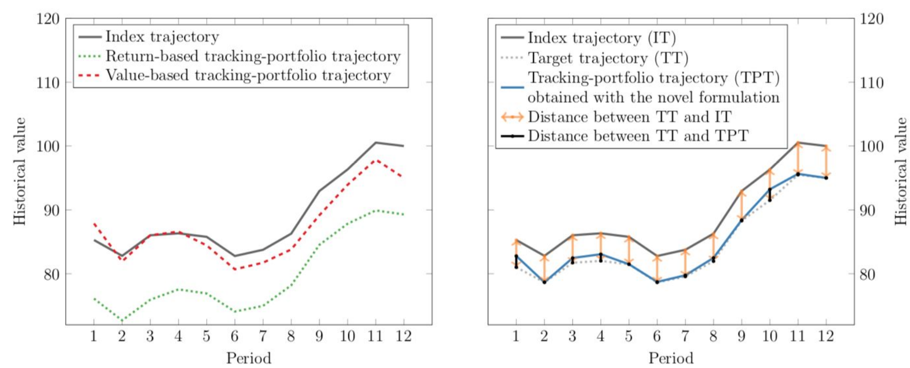
A portfolio manager is the person who is actively monitoring the investment portfolios for the fund. Their responsibility is to make the best decisions for an investment portfolio that achieves the best return possible while keeping the risk as low as possible. Working as a portfolio manager is not an easy task and in order to attract more investors to contribute to their fund, portfolio managers need to find a way to achieve alpha as high as they can. Alpha represents the excess return from the general market return. Just like the linear function: $y = m*x + b$, where m represents the return of the market, and b represents the excess return generated by the fund manager.

There are many ways for portfolio managers to gain alpha. One way is through better asset allocation. That means how much money do you invest that would result in high returns. Optimization is one way of doing this by letting the portfolio manager set up parameters and the target objective. There are quite a few variables for the portfolio manager to consider and without rigorous analysis, a suboptimal allocation strategy could potentially be selected. Our study investigates how to achieve maximum returns while tweaking the allocation of funds between a baseline index and a target index. We consider four criteria for the allocation of funds in our model: market sector, duration of the security, and the credit rating of the underlying company, and a maximum limit on how much is invested into one company. The objective is to develop a model to determine the optimal security combinations with the allocation constraints that could potentially generate greater returns than the general market.

Literature Review

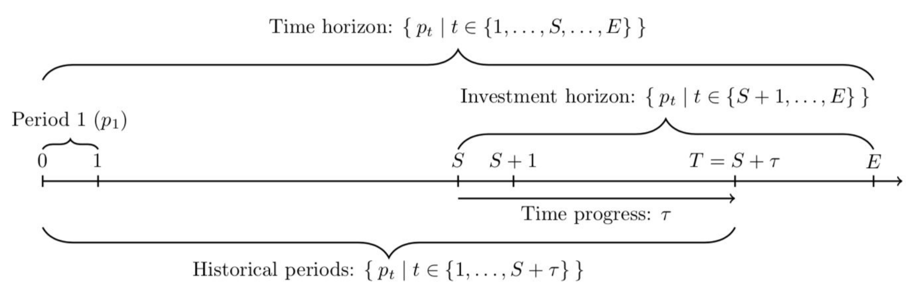
The essence of forming a portfolio is to allocate funds to various investment alternatives, so the risk of investing as a whole can be minimized, and the profit rate of return can be maximized (Sarkar et al., 2013). The risk of investment is measured by the amount of variance or standard deviation from expected return. Traditionally, portfolio optimization's focus has been on finding the proper balance for allocations to different asset classes based on the mean-variance tradeoff. Traditional techniques such as efficient frontier is an example of a top-down approach to achieving an optimal portfolio. Linear programming is widely used in asset/fund allocation and portfolio optimization by incorporating influential factors and parameters as constraints – when multiple factors are at play, the constraint design can quickly become complex. Sukono et al (2018) presented an investment portfolio optimization linear programming model (assuming short selling is not allowed) based on genetic algorithms. The investors of an index fund demand a high tracking accuracy, i.e., small deviations between the future value developments of the benchmark portfolio and the index fund. Various mathematical formulations determine a tracking portfolio by minimizing an objective function, referred to as tracking error, that measures the difference between the historical performance of the tracking portfolio and the index. Based on the specific tracking-error function used, existing formulations are divided into the two classes value-based formulations and return-based formulations (Gaivoronski et al., 2005). The most intuitive index-tracking approach is to invest in all index constituents according to the index composition, which is known as full replication (Strub et al., 2018). However, in the presence of market frictions, full replication causes a substantial amount of transaction costs that reduce the fund's capital, especially for indices with many constituents (Beasley et al., 2003).

Figure A:



(a) Existing formulations: large distance or different shapes

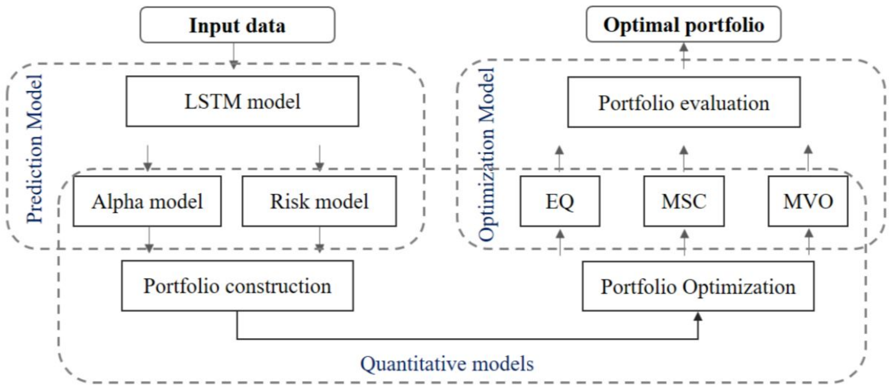
(b) Novel formulation: small distance and similar shapes



Strub et al (2018) proposed a new value-based mixed-integer linear programming (MILP) formulation. The main feature of this formulation is that it determines a target trajectory internally instead of using the historical index trajectory as done by existing value-based formulations. The target trajectory has the same returns as the index, but its level may vary. As shown in Figure A above, the objective function then minimizes both, the distance between the tracking-portfolio trajectory and the target trajectory, and the distance between the target trajectory and the index trajectory. Minimizing the former distance can be viewed as maximizing a similarity measure between the normalized shape of the tracking portfolio and the normalized shape of the index.

While theoretical modeling methods deal with mechanics of trading to accomplish mean-variance tradeoff, microstructure deals with issues of market structure and design, price formation and price discovery, transaction and timing cost, information and disclosure, and market maker and investor behavior (e.g., short selling) – which are uncontrollable variable factors. The Stereoscopic Portfolio Optimization Framework introduces the idea of bottom-up optimization via the use of machine learning ensembles applied to some market microstructure components. The bottom-up techniques are combined with top-down approaches to create Stereoscopic Portfolio Optimization (SPO) Framework (https://blog.quantinsti.com/). This idea presents the unlimited opportunities for quantitative trading where stock prediction plays an important role. Ta et al (January 2020), presented a remarkable idea of using deep learning models (LSTM, a special kind of RNN) to predict the stocks with high accuracy. In order to construct an efficient portfolio, multiple portfolio optimization techniques, including equal-weighted modeling (EQ), simulation modeling Monte Carlo simulation (MCS), and optimization modeling mean variant optimization (MVO), are used to improve the portfolio performance as shown in Figure B. Ta et al, compared the performance of portfolios constructed using these three optimization techniques based on predictions from LSTM, LR and SVR based models.

Figure B:



Modeling and Methodology

We assume a portfolio manager just got a job to build a fixed income portfolio that closely tracks the performance of the ICE Bank of America US Corporate Index (C0A0); a realistic task for a portfolio manager. This index contains all corporate fixed income securities from issuers based in the U.S. The portfolio manager has $680 million to spend and he would like to build a portfolio that could outperform the index. We will run an optimization based on all securities within the C0A0 index, with some distribution modification on market sector, securities duration, issuers' credit rating. In addition, to avoid over allocation to a single company, we will limit the total amount of contribution to $100 million per issuer. For example, if there are two or more securities issued by Apple Inc. within the index, we want to invest no more than $100 million for Apple in total.

Traditionally, financial analysts look at performance metrics that focus on the past results, such as return. Many optimization models use this kind of descriptive information as the target variable to perform optimization. Therefore, the results of the optimization can only represent the past performance and they do not necessarily have predictive value. In order to advance and build on these statistical models, we utilize a machine learning model to make predictions of the future return based on historical data. In our model, we are using XGBoost to come up with a Relative Value score as a return projection. The score is from 1 to 10, the higher the score the better the security is. The input data for the model is based on daily securities trading history for similar type of asset class, i.e. all corporate bonds for our example. The input variables are including security's terms and conditions, price performance and valuation analytics measures as a daily snapshot for each security. The target variable for the model is the ratio between security yield and duration, which measures the per unit return per year (duration is measured in year). The outcome of the model returns a Relative Value score per each security. We can use results for predictive models to project future returns based on current market data and optimally allocate our portfolio.

Optimization Formulation

Let i be fixed income security 1, 2, 3, …, n.

Objective:

$$\textit{Maximize Projected Return} = \sum \textit{Quantity}_i \times \textit{Projected Return}_i$$

Subject to:

$$\textit{Total Investment} = \$680{,}000$$

$$\sum \textit{Quantity}_i \times \textit{Sector Weight}_i \leq \textit{Total Investment}$$

$$\sum \textit{Quantity}_i \times \textit{Duration Weight}_i \leq \textit{Total Investment}$$

$$\sum \textit{Quantity}_i \times \textit{Credit Weight}_i \leq \textit{Total Investment}$$

$$\textit{Issuer}_i \leq \$100{,}000$$

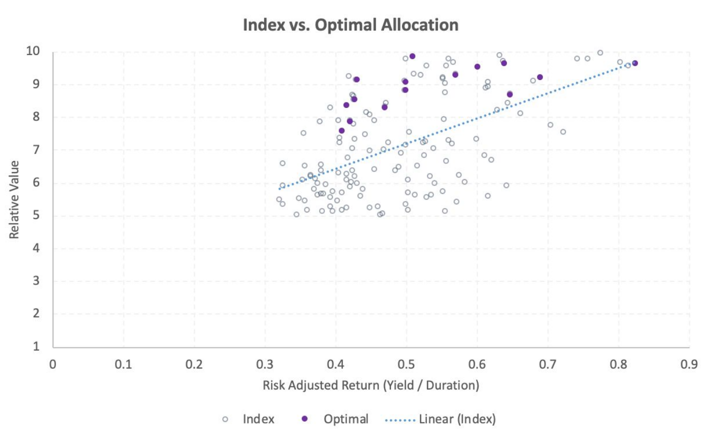
Model Results

Below, we show the summary statistics for the performance between the index (market benchmark) and the optimal (our model) portfolio. The result shows that the portfolio weighted average relative value score increased from 7.028 based on the all securities within the index to 8.818. The optimization is improving the yield from 3.061 to 3.334, while the duration of the portfolio only increased slightly from 6.46 to 6.58. The per unit return improved from 0.474 to 0.507.

	Yield to Worst	Duration	Yield/Duration	Relative Value
Optimal	3.334	6.576	0.507	8.818
Index	3.061	6.456	0.474	7.028

As mentioned in the earlier discussion, better asset allocation can help portfolio managers to gain more alpha, or excess return over the market. Below we demonstrate alpha and beta between the original index and the optimal security selection with a regression. That the optimal portfolio results more than double of the alpha from the index, in the meantime reducing beta by half.

	Beta	Alpha
Optimal	3.54	6.98
Index	7.65	3.38



Discussion and Conclusion

Overall, our methodology for portfolio optimization shows promise for achieving returns on a risk adjusted basis versus a broad index of fixed income securities. Specifically our model is able to optimize our four constraints of sector, duration, credit rating, and diversification in a way that makes our portfolio's expected yield higher than the index's yield in a similar amount of time. The distribution tables in the appendix show in detail how our model allocates weight to each of the categories versus the index.

While the initial results of this analysis provide a promising start to a portfolio optimization product, there are several limitations with the approach. First, we chose to focus our efforts on optimizing the model on four major constraints discussed previously. We decided to do this because the model was designed as a potential product for a portfolio manager to offer to customers looking for an index fund to diversify their investments. Accordingly, these constraints attempt to address the most frequently asked questions that a customer may have about an index fund. However, there are a large number of other constraints that may significantly impact the yields of a portfolio. For example, other constraints include the revenue, profitability, and price to earnings multiples of the companies. The incorporation of these and other constraints into the model may have a significant impact on the projected yields of the portfolios. As more constraints are added to the model, we would also need to take into account the sensitivity of the model to tweaking each of these constraints. Second, our model is based on a limited time frame of historical market data. Therefore, the predictive capabilities of the model are uncertain and dependent on market conditions. All models have this restriction but portfolio managers and financial analysts can work to mitigate this risk by selecting the optimal training data for their models. When the training data for a model is too limited it tends to cause overfitting, meaning the model is only predictive of the training data and not the broader data set. We could build on our experiment by training our portfolio optimization model on a larger set of historical data. This may incorporate several economic cycles and improve the predictive capabilities of our model. Third, since our portfolio is an index fund it has a high beta, or correlation with overall market performance. This means that it is still subject to larger market downturns and volatility. Therefore, it is not a good hedge against broader market movements.

Overall, we believe the incorporation of machine learning to create a relative ranking of securities is an innovative approach to portfolio optimization. While there are several limitations with this modeling approach, it provides a step forward to incorporating machine learning methods into the decision problem of portfolio optimization.

References

Elton, E., Gruber, M., & Blake, C. (1996). The Persistence of Risk-Adjusted Mutual Fund Performance. *The Journal of Business,69*(2), 133-157. Retrieved March 15, 2020, from www.jstor.org/stable/2353461

Sukono, Hidayat, Y., Lesmana, E., Putra, A. S., Napitupulu, H., & Supian, S. (2018). Portfolio optimization by using linear programing models based on genetic algorithm. *IOP Conference Series: Materials Science and Engineering*, *300*, 012001. doi: 10.1088/1757-899x/300/1/012001

Sarker, M.R. (2013). Markowitz Portfolio Model: Evidence from Dhaka Stock Exchange in Bangladesh. *IOSR J. of Business and Management* (IOSR-JBM. e-ISSN: 2278- 487X 8 6 (Mar. - Apr. 2013), pp 68-73. http://www.iosrjournals.org.

Gaivoronski, A. A., Krylov, S., & Van der Wijst, N. (2005). Optimal portfolio selection and dynamic benchmark tracking. *European Journal of Operational Research* 163 (1), 115– 131.

Beasley, J. E., Meade, N., Chang, T.-J. (2003). An evolutionary heuristic for the index tracking problem. *European Journal of Operational Research* 148 (3), 621–643.

Strub, O. & Baumann, P. (2018). Optimal construction and rebalancing of index-tracking portfolios. *European Journal of Operational Research*, Elsevier, vol. 264(1), pages 370-387.

Ta, V-D., Liu, C-M., & Tadesse, D.A. (2020). Portfolio Optimization-Based Stock Prediction Using Long-Short Term Memory Network in Quantitative Trading. *Applied Sciences*. 10. 437. 10.3390/app10020437.

https://blog.quantinsti.com/optimal-portfolio-construction-machine-learning/

Appendix

Table 1: Sector Distribution

	Sector	Index	Target	Optimal	Optimal (%)
S1	Banking	22%	15%	120	18%
S2	Financial Services	2%	3%	24	4%
S3	Insurance	3%	6%	48	7%
S4	Automotive	7%	2%	16	2%
S5	Basic Industry	7%	8%	56	8%
S6	Capital Goods	4%	5%	40	6%
S7	Consumer Goods	3%	6%	48	7%
S8	Energy	13%	12%	96	14%
S9	Healthcare	14%	5%	40	6%
S10	Leisure	0%	1%	-	0%
S11	Media	0%	5%	-	0%
S12	Real Estate	2%	2%	16	2%
S13	Retail	2%	0%	-	0%
S14	Services	1%	5%	40	6%
S15	Technology & Electronics	9%	12%	96	14%
S16	Telecommunications	6%	5%	-	0%
S17	Transportation	1%	3%	-	0%
S18	Utility	5%	5%	40	6%
	Total			**680**	

Table 2: Duration Distribution

	Duration	Index	Target	Optimal	Optimal (%)
D1	0-1	0%	2%	-	0%
D2	1-2	0%	11%	-	0%
D3	2-3	0%	12%	-	0%
D4	3-5	13%	25%	72	11%
D5	5-7	53%	18%	308	45%
D6	7-10	34%	15%	300	44%
D7	10+	0%	17%	-	0%
	Total			**680**	

Table 3: Credit Rating Distribution

	Rating	Index	Target	Optimal	Optimal (%)
C1	AAA	20%	35%	-	41%
	AA1			-	
	AA2			-	
	AA3			-	
	A1			-	
	A2			-	
	A3			280	
C2	BBB1	80%	50%	96	59%
	BBB2			144	
	BBB3			160	
C3	BB1	0%	15%	-	0%
	BB2			-	
	BB3			-	
	B1			-	
	B2			-	
	B3			-	
C4	CCC1	0%	0%	-	0%
	CCC2			-	
	CCC3			-	
	CC			-	
	C			-	
	Total			680	