ISDS – 570
Business Data Transformation

MV Portfolio Optimization

Group Project Report – Team 3

Team Members:
Anirudh Chandnani
Deepak Nayak
Jagruthi B Jethwani
Madhusudhan Unnava
Krishna Pratheek Seshabhattar

CONTENTS

INTRODUCTION

Portfolio Optimization:

The process of creating an ideal portfolio that reduces risk and maximizes returns across the group of portfolios is known as portfolio optimization.

It is necessary to select a range of assets from several categories, such as stocks, bonds, and commodities, in order to create a broad portfolio. In this paper, we look at how to select stocks more wisely using Mean-Variance Portfolio optimization to enhance portfolio performance. The major objective is to achieve the best feasible balance between risk and profit by using statistical analysis and mathematical models. The study thoroughly examines the possibilities of this method to optimize our investing strategy and possibly increase the overall performance of our portfolio.

In this assignment, we investigate the "stock market" case study that was provided during the course by using state-of-the-art portfolio optimization techniques. The stock trading dataset spans the period from December 31, 2015, to March 26, 2021, and was sourced from Yahoo Finance. The initials of each team member's last name were used to select their tickers. The data goes through a thorough ETL procedure before the Portfolio is optimized and analyzed in its entirety. During the optimization process, our main goal is to increase diversification by varying the weights of the tickers that we have already chosen. This will minimize overall risk and maximize rewards.

This methodical technique helps in making well-informed investment selections and provides insightful information for portfolio management. Through the application of sophisticated strategies, we aim to attain the best possible balance between returns and risk, which enhances the portfolio's overall performance.

SUMMARY

In this report we shall see how we devised a plan to assemble a portfolio of stocks that are traded on US stock exchanges. We offer the outcomes of our stock return along with a list of the most effective ways to get there. We begin by outlining the main process that is employed to identify the optimal portfolio.

A mathematical framework known as mean-variance optimization is used to balance the expected return and risk of an asset collection in order to create the ideal portfolio. The goal of mean-variance optimization is to find the asset allocation that minimizes risk for a given level of expected return or maximizes expected return for a given level of risk.

Following the extraction of raw data from Yahoo's SP500 index and the NASDAQ, NYSE, and AMEX exchanges, the data is imported into the Postgres database. Once the data's accuracy has been established, we make changes to it, including merging different Postgres data tables, renaming rows and columns, adding any missing data, and removing any unnecessary information. The optimization procedure is then carried out when we load the data into R-Studio. As for the optimization procedure, we use the portfolio performance from the first 58 trading days of 2021 against the SP500, and we train the data from 2016 to 2020 for back-testing.

With a portfolio ROI of 97.43%, we outperform the SP500 by a factor of 3.3. Our portfolio has a slightly higher risk due to the larger annualized Sharp and standard deviation, even though the returns might be much higher.

Stock Name	Symbol

EXTRACT, TRANSFORM & LOAD

An Extract, Transform, Load (ETL) pipeline is used in the portfolio optimization process to prepare the stock data for analysis. The ETL procedures for locating and converting the required data are described in this report.

EXTRACT:

Combining the equities from the NASDAQ, NYSE, and AMEX exchanges yields the stock data. From Yahoo Finance, the SP500 index data was obtained for the years 2016 through 2021. The corresponding database table, which has a similar structure, received both datasets upon import. To make sure the data is accurate and comprehensive, a custom calendar is also made using Microsoft Excel for the specified date period. Using the NETWORKDAYS.INTL function, the calendar marks the trade days and imports them into the database. More columns that are important for calculating returns, like the end-of-month flag and the preceding trading day, are added to the table structure. The data in PostgreSQL are loaded into R using the RPostgres package and the required credentials, and the optimization is done in R.

TRANSFORM:

R is utilized in the data transformation process. First, the data completeness is checked to see if every ticker has at least 99% data, after which the remaining tickers are filtered away. The reshape2 package is used to further change the data. The dcast function is used to pivot the data frame. To further check any data loss, the pivot table and calendar are then combined. Additionally, any data with three or more missing values are imputed by filling in the last observation value using the last observation carried forward (LOCF) function. Stocks with incorrect entries and returns greater than 100% are also disqualified.
The last change is to convert the results for the benchmark (index) and the chosen 12 assets to extensible time series data (XTS), since the performance analytics and portfolio analytics features need this format to function.

LOAD:

The XTS converted data is fetched to the Portfolio Analytics package, which includes packages needed for portfolio optimization, as part of the data loading process.

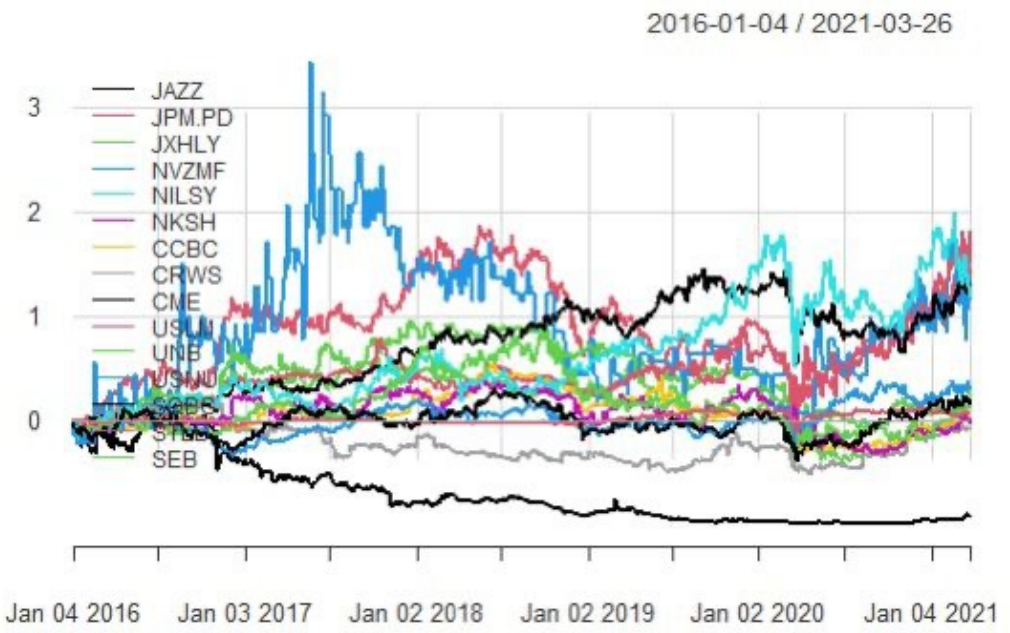
PORTFOLIO OPTIMIZATION

The Markowitz-based Mean Variance portfolio is the subject of the portfolio optimization process. The actions listed below are taken:

- ROI.plugin.quadprog and Portfolio Analytics are the necessary packages.
- The stock returns from 2016 to 2020 make up the training data.
- The mean of the S&P 500's daily returns is used to determine the minimum acceptable return (mar). PortfolioAnalytics has a return target thanks to this value.
- The goal of reducing risk is added along with the asset names to create the portfolio specifications. In order to ensure full investment and satisfy the minimum acceptable return, further limits are added.
- The RIO optimization technique, PortfolioAnalytics configurations, and the portfolio training data are used to optimize the portfolio.
- The weights from the optimization are applied to the test data for 2021 as part of back testing in order to mimic the performance of the portfolio.
- The 2021 SP&500 data is compared with the portfolio performance numbers.

CUMULATIVE RETURN CHART OF INDIVIDUAL STOCKS

First, each stock is represented by a separate line on the cumulative return chart for the five years that the stocks were selected for. With a positive ------- total return, ---------has the highest cumulative return of all the chosen stocks, as can be shown. With a negative cumulative return of -------, ------- has the lowest return.
The best positive cumulative returns, with a positive ------, are displayed by ------. The chosen stocks have a positive mean cumulative return of -------.
The chosen stocks have a cumulative return that is significantly higher than the benchmark, which is ------- for SP500TR stocks.
The cumulative prices are dispersed, and the largest cumulative return is exhibited by --------.



OPTIMIZED PORTFOLIO WEIGHTS

We add a few specifications for optimization, such as the objective and restrictions. Standard Deviation is the goal, and the stocks must have a "Full Investment" and cross the desired MAR (Minimum Average Return) number in order to be considered.
For this analysis,
MAR value is 0.000636.
The weights of our optimized portfolio are now visible:

SYMBOL	VALUE
CCBC	0.007169784
CRWS	0.025997718
CME	0.160384994
JAZZ	0.007145986
JPM.PD	0.406358665
JXHLY	0.065340738
NVZMF	0.085962452
NILSY	0.122657197
NKSH	- 0.094243713
SQBG	- 0.020891258
STLD	0.038623592
SEB	0.030675028
USLM	0.135378699
UNB	- 0.031247818
USNU	0.060687937

With a positive 0.4063, we consider JPM.PD's optimized weights to be the highest. The stocks NKSH and SQBG all have negative indications, indicating that these were shorted.
We aggregate the weight totals to determine if the sum equals unity(1) in order to make sure our optimization is balanced. This results in a balanced distribution of the optimum weights that were acquired.

PORTFOLIO'S CUMULATIVE RETURN CHART

As we can see from the previous section, our portfolio's chosen stock has a mean cumulative return of 155%, while the SP500TR, which serves as a benchmark for comparison, has a mean cumulative return of 115%. This means that our portfolio has a positive return rate that is 40% higher than the SP500TR stocks. The portfolio's cumulative return chart is shown below. Even if both stock categories appear to have similar seasonality, it is clear that the chosen stock portfolio outperforms the benchmark returns.

ANNUALIZED RETURNS of 2021

Our portfolio had a return rate that was three times higher than the SP500TR index, similar to the cumulative returns. We used our portfolio's stock data and the SP500TR from prior years (2016–2020) as our training data, utilizing the train test split. A model that allowed us to calculate the annualized return using our test data was created using the training data. With only three months' worth of data, the annualized return is encouraging, with a 97% rate.

	SP500TR	PORTFOLIO
Annualized return	0.2987	0.3839
Annualized Std Dev	0.1623	0.1556
Annualized Sharpe	1.8399	2.4678

CONCLUSION

While portfolio optimization alone can greatly enhance portfolio performance, it can yield even more remarkable outcomes when coupled with an ETL pipeline. Our portfolio significantly beats the SP500TR index overall, according to our portfolio optimization study of historical stock data.

Our portfolio's weights revealed which stocks were being shorted and which were worth investing in. Furthermore, our train test split model showed that our portfolio would outperform the SP500 index in terms of annualized and cumulative return.
Investors can improve their ability to make decisions, possibly earn larger returns, and successfully manage risk by putting these ideas into practice. In the realm of finance, portfolio optimization is still a useful tool that can help people and businesses looking to maximize their investment portfolios.

APPENDIX

APPENDIX A

ETL PROCESS: Extract, transform & load.

- Extract: We located the data and retrieved it from numerous sources during the extraction step.
- Transform: To guarantee data correctness, consistency, and suitability for analysis, the extracted data is cleaned, verified, and reorganized.
- Load: After transforming the data, we sent it into the targeting system.

APPENDIX B

DAILY RETURNS

The following formula has been used to the computation of daily returns. Daily Return is calculated as follows: (Today's Price - Last Trading Day Price) / Last Trading Day Price

APPENDIX C

SQL Code: