Forecasting Stock Prices and Constructing an Optimal Portfolio using Prophet: A Comparative Analysis Against Benchmark Indices

KHAN, Asif

Supervised by: Dr. David Rossiter

COMP4971C - Independent Work

Department of Computer Science and Engineering

The Hong Kong University of Science and Technology

Summer 2023

Abstract

This project aims to explore the construction of an optimal portfolio using stock price forecasting and portfolio rebalancing methodologies with the help of Prophet, a predictive open-source instrument developed by Facebook. The study was primarily focused on the Dow Jones Industrial Average (DJIA), and utilises certain algorithms to apply the methodologies to compete with and even exceed the growth rate of the benchmark index. This strategy is put through multiple backtesting phases with each being an attempt to improve and enhance the previous portfolio rebalancing algorithm used in the phase prior. Through extensive backtesting, it is consistently demonstrated that the optimal portfolio, rebalanced regularly over a fixed time interval throughout a backtesting period of 180 days from June 2022 to December 2022, was able to outperform the DOW 30, achieving nearly triple the growth at 20.3% return compared to that of 7.9% in the DOW 30.

This study ultimately showcases the effectiveness of pairing a well-developed and comprehensive forecasting tool with the principle of portfolio rebalancing in developing an optimised portfolio with consistent performance, even throughout economic crises. Hence, these findings can prove to have valuable insights for investors seeking to maximise their returns through minimal market research in a dynamic and fast-paced stock market environment.

Table of Contents

Table of Figures

1. Introduction

1.1 Stock market

The stock market is where people from various social and economic classes gather to achieve one common goal; to earn money through being a shareholder of a company's assets, primarily by owning their stocks. People involved in the stock market have various investment profiles, with some owning high-risk stocks to maximise their return, some with blue-chip stocks to find a stable means of earning passive income over time.

However, one very common advice that goes around within the community based around the stock market, especially among those who do not invest and trade for a living, is to have a diversified portfolio. That way, the risk of losing all their money solely due to the poor performance of a bad decision in one stock is minimised.

1.2 Benchmark indices

The construction of a profitable and sound investment portfolio greatly depends on a thorough understanding of the market, as well as the impact individual companies have on the stock market. For the general investor who wouldn't usually take their time out to study and investigate each and every asset, the best way to construct a diversified portfolio is to simply explore and invest in index funds.

In simple terms, an index fund is a portfolio that tracks and attempts to follow the components of a financial index. This is particularly useful for casual investors, as most of the market research is done by experts in the field, saving them the time and effort in finding stocks that are worth investing into. As a result, there has been a surge in popularity lately in terms of financial advice surrounding the investment of index funds, where the top index funds usually track well-known benchmark indices, such as S&P 500, Dow Jones Industrial Average, NASDAQ Composite and the MSCI World Index.

1.3 Portfolio rebalancing

With every investment, there is a certain risk and reward that an investor is exposed to. As a result, it is important that their stock portfolio stays in line with their investment ambitions. To illustrate the importance and effectiveness of rebalancing one's portfolio, an example is used below:

- Suppose someone has a total portfolio value of $3000, with $2000 (66.7%) invested in Stock A, $600 (20%) in Stock B, and $400 (13.3%) in Stock C.
- After a certain period of time, the market changes and now his portfolio has a total value of $3400, and is distributed as follows: $2020 (59.4%) invested in stock A, $900 (26.5%) in Stock B, and $380 (11.2%) in Stock C.
- We now look at the scenario where Stocks A, B, C subsequently experience a 25% increase, 15% decrease and a 5% increase respectively.

Scenario 1: Unbalanced portfolio

Simply adding the performance change to the stocks, we can calculate the value of assets owned in Stock A, B and C to come down to $2525, $765 and $399 respectively, which comes down to a total portfolio value of $3689.

Scenario 2: Rebalanced portfolio

The investor decides to rebalance his portfolio before stocks have gone through any changes, by trying to maintain a similar distribution of assets owned initially. This would imply that his portfolio worth $3400 would have $2267 in Stock A, $680 in Stock B, and $453 in Stock C after rebalancing.

Applying the calculation used earlier to find the new value of his assets and total portfolio, we can see that the values have come down to $2834 in Stock A, $578 in Stock B, and $476 in stock C, which sums up to a total portfolio value of $3888.

Hence, by referring to the above example, it is evident that there are certain benefits to rebalancing a portfolio as this ensures that your portfolio stays true to the risk profile you choose in your investments, mitigating potential losses and providing you with more opportunity for financial gains in certain scenarios.

1.4 Price forecasting

Along with the principle of portfolio rebalancing, most investors also would like to optimise their investment earnings by predicting the movement of stock prices. This is usually done through studying the historical trends of the stock markets, keeping up to date with current affairs, and even looking into certain financial indicators or correlations between multiple stock listings.

Forecasting techniques have been around for a long time, with many mathematical methods developed to forecast different patterns in modelling problems. As this is a very hot topic in the field of research, there is a wide variety of tools and programming libraries that attempt to apply mathematical models in various scenarios to obtain valuable information in the context of prediction.

1.5 Objective

In this study, we would like to investigate how we could, with the right forecasting tool and risk profile, construct an optimal portfolio. This investigation would also employ the principle of portfolio rebalancing by utilising an algorithm and by applying the algorithm through a specific library to do so.

This trading strategy will be analysed and evaluated against benchmark indices in terms of its effectiveness and reliability, with metrics such as the growth rate as well as the maximum drawdown.

2. Methodology

2.1 Data collection

In most studies and projects involving financial analysis, the key step in getting started is to gather all the necessary data in carrying out said analysis. In our situation, as we are employing forecasting tools to construct an optimal portfolio, it is important that we gather enough historical data to feed into our forecasting model.

With tens of thousands of stocks in the market spanning over many benchmark indices, it is crucial to choose the right benchmark index that comprises a suitable number of stocks. Initially, the S&P 500 and its constituent stocks was considered as the source of data collection, but with the frequent listing and delisting of stocks, and the large amount of information that would have been analysed over 500+ stocks would have required a lot of thorough analysis.

Through some research done regarding portfolio diversification, it was found that it is best to have around 25-30 stocks. Any more than that, it would be more likely to mimic a mutual fund and would not be able to perform significantly better than a market, and managing such an over-diversified portfolio would be far too much work. Any fewer than that, our portfolio would be highly dependent on top performers, and any poor performing sectors would greatly cause a loss in the performance of the investment portfolio.

Keeping this in mind, the best benchmark index that would fit the objective of the project would be the DOW 30 index. The DOW 30 is a widely recognized stock market index, and is often dubbed as a "barometer" of the overall stock market, and is usually a good reflection of the US stock market. The companies chosen in the DOW 30 are selected by the editors of The Wall Street Journal, and are considered large and well-established companies that are based in the United States.

Extracting the information for each stock on DOW 30 can be done through the yfinance library, a Python library that allows for a very efficient and seamless method of scraping financial data from the Yahoo Finance's database. With the information of the stocks listed in the DOW 30, their tickers can be used to extract data through the library as shown below:

```
[3]: hist = yf.download('AAPL', start="2018-01-01", end="2022-09-25",  interval = "1d")
     hist
```

```
[**********100%**********]  1 of 1 completed
```

[3]:

Date	Open	High	Low	Close	Adj Close	Volume
2018-01-02	42.540001	43.075001	42.314999	43.064999	40.831593	102223600
2018-01-03	43.132500	43.637501	42.990002	43.057499	40.824471	118071600
2018-01-04	43.134998	43.367500	43.020000	43.257500	41.014103	89738400
2018-01-05	43.360001	43.842499	43.262501	43.750000	41.481056	94640000
2018-01-08	43.587502	43.902500	43.482498	43.587502	41.327000	82271200
...	...	...	...	...	...	...
2022-09-19	149.309998	154.559998	149.100006	154.479996	153.776352	81474200
2022-09-20	153.399994	158.080002	153.080002	156.899994	156.185318	107689800
2022-09-21	157.339996	158.740005	153.600006	153.720001	153.019806	101696800
2022-09-22	152.380005	154.470001	150.910004	152.740005	152.044266	86652500
2022-09-23	151.190002	151.470001	148.559998	150.429993	149.744797	96029900

1191 rows × 6 columns

2.2 Data preprocessing

With the yfinance library, all the data concerning the stock companies listed in the DOW 30 were effectively gathered and consolidated. To facilitate smoother modelling performance and for a more structured method of constructing and rebalancing a portfolio, the data extracted from yfinance were filtered and extracted in the format that would best suit the forecasting tool in this study. The specifications required by the forecasting tool of choice is to have two columns labelled ‘ds’ for the time period, and ‘y’ for the value to be forecasted, which is considered the closing price of a stock share at that time.

	ds	y
0	2018-01-02	235.639999
1	2018-01-03	235.630005
2	2018-01-04	238.710007
3	2018-01-05	240.570007
4	2018-01-08	239.789993
...	...	...
1186	2022-09-19	116.639999
1187	2022-09-20	116.519997
1188	2022-09-21	114.809998
1189	2022-09-22	114.139999
1190	2022-09-23	112.989998

1191 rows × 2 columns

2.3 Forecasting method

The forecasting tool of interest is Prophet, a time-series forecasting library developed by the Data Science Team from Facebook (now Meta) based in R and Python, which is a very effective tool that utilises an additive model based on non-linear trends, while factoring in seasonality. The library stands out among other forecasting libraries for two primary reasons:

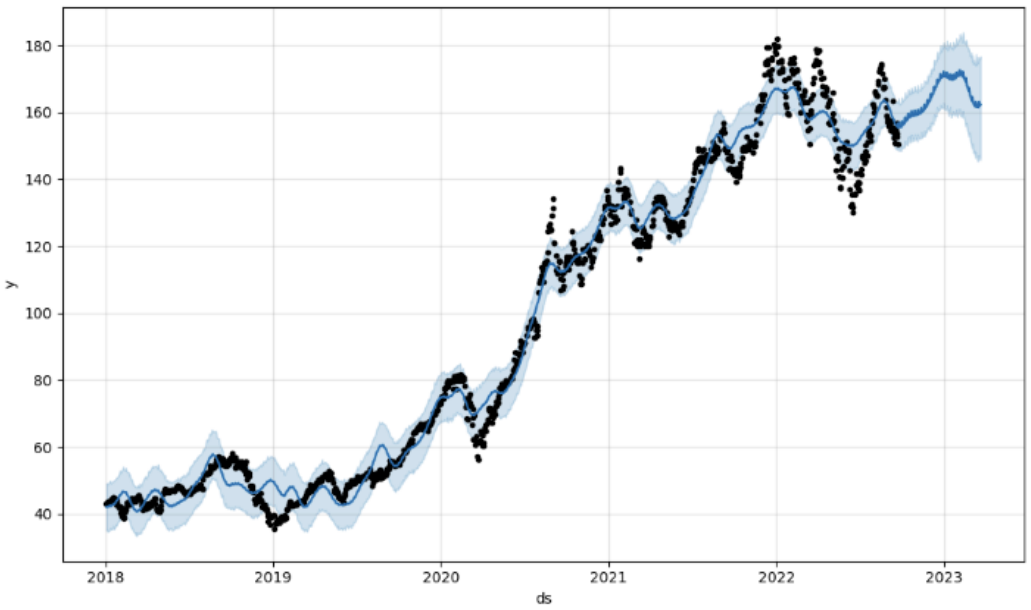
- Holiday inclusion: Prophet factors in holidays and their respective effects in time series forecasting, which makes for a more robust and realistic prediction
- Seasonality and Trend: Instead of a simple forecasting method by using mathematical regression analysis, there are various seasonalities that could be considered based on the users’ input, which could result in a more consistent forecasting prediction, rather than a plain upwards or downwards trend.

The additive model that Prophet is based on is effectively the summation of three functions, shown below:

$$y(t) = g(t) + s(t) + h(t) + e\_t$$

The three functions can be explained as follows:

- ***g(t)* [Growth function]:** This is a function that looks at the overall growth of the data points, and can be broken down into three kinds of growth modes.
    a. Flat growth: *g(t)* takes on a constant value in this case, where in any given period of time, no growth is observed.
    b. Linear Growth: This is similar to the formula of $y = mx + b$, which is a basic equation that describes linear growth. Prophet takes this a step further, with the slope differing at change points (points where an external circumstance influences the direction of the data points) that are cleverly determined by the tool
    c. Logistic Growth: This is a growth pattern that usually comes with an upper or lower limit, which can often be seen in population or bacteria modelling problems, and the final value cannot exceed the provided upper limit.
- ***s(t)* [Seasonality function]:** This is a function that captures the fluctuations that occur periodically in intervals, ranging from daily, weekly, monthly or even yearly. This function is a Fourier Series, which means that it can be used to approximate most curves that occur when plotting the data points. If one would like to tweak this function, it requires a higher level of mathematical understanding. Prophet has taken care of this by automatically being able to detect the right Fourier equation for the approximation.
- ***h(t)* [Holiday function]:** This function is what uniquely sets Prophet apart from many forecasting libraries, where public holidays are factored in forecasting the future data points. Certain holidays such as Christmas, Halloween, Thanksgiving or New Years could influence the growth of different data, which is factored in through *h(t)*. It could be particularly useful in our study, as the holiday function could be greatly relevant to the movement of stock prices based on historical data that could be learnt through Prophet.



An example figure of the trend of closing prices of Apple's stock (AAPL) is shown above, with the solid data points indicating the true prices extracted from the data preprocessing stage, and the blue solid data points indicating the price fitted and predicted by Prophet's model. The data points are also in the middle of a bluish region, with the upper and lower bound of the region indicating the confidence intervals of the forecasting model, which would eventually prove to be useful when factoring in investors' risk profiles.

2.4 Portfolio construction

As the goal of this project is to carry out a comparative analysis against benchmark indices, or in our case, the DOW 30, the portfolio is designed with the same companies that are listed in the DOW 30. In addition, this portfolio could also provide further insights as both DOW 30 and our portfolio contain the exact same companies listed, only with a customised investment strategy in our portfolio.

For the initial setting of the portfolio, we assume the distribution of our investment portfolio to follow the same distribution as that in DOW 30, which is through a price-weighted technique. In essence, any stock component within the DOW 30 is weighted based on its stock price, which means that higher-priced stocks (e.g. Goldman Sachs) would have a higher weighting in the index and ultimately have a stronger influence.

To achieve this distribution, we calculate the sum of share prices in all the 30 companies at a given time point, and the proportion of a single stock to share price sum, and invest the calculated proportion of our capital in the company at that time point.

2.5 Portfolio rebalancing algorithm

In order to perform better than a benchmark index, we would need to diverge from their usual price-weighting distribution strategy through rebalancing the investment portfolio. To do this, an algorithm has been devised, which is paired up with the forecasting done by the models fitted through Prophet. The working mechanism of the rebalancing algorithm is listed below:

1. Iterate through every stock ticker listed in the DOW30 and do the following:
    a. Extract the relevant historical data (Date, Closing Price) within a fixed timeframe
    b. Instantiate a forecasting model through Prophet, with seasonality factored in
    c. Fit the model to the historical data
    d. Forecast the model several periods ahead of the fixed timeframe, and save the forecasted stock prices (stored in the 'yhat' column)

2. For all the forecasted prices, calculate the rebalanced portfolio distribution, which is given by: Forecasted price of stock / Sum of forecasted price of all stocks in DOW30

3. Buy the required number of shares for all the stocks that meet the new rebalanced distribution, observe the actual price trend of all the stocks and track the portfolio balance

2.6 Backtesting configurations

In order to ensure the algorithm is tested in a controlled environment, data analysis would be carried out in the following conditions and methods:

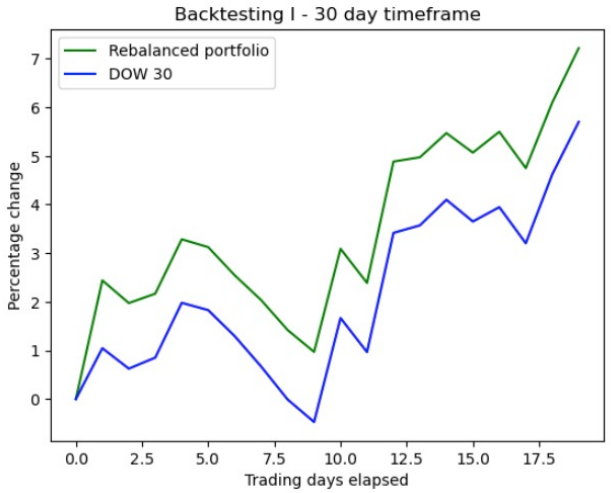
- The historical data extracted from all stocks listed in the DOW30 would land within the timeframe of January 2018 to June 2022
- The portfolio would then be forecasted ahead with the following timeframes (30 to 150 days, in increments of 30) and be rebalanced accordingly based on the forecasts

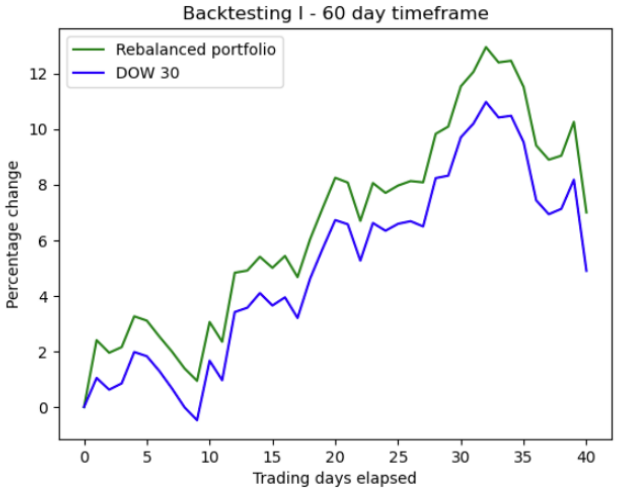
- For the sake of an error-free comparison, the trend of DOW30 would be extracted starting from June 2022 for the duration of the aforementioned timeframes, and be compared against the growth of our rebalanced portfolio.

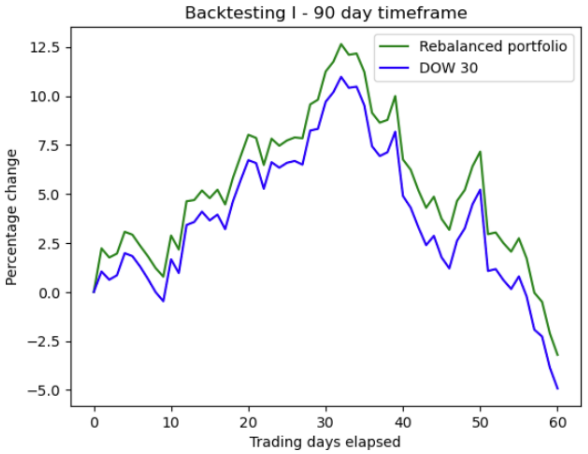
3. Data analysis

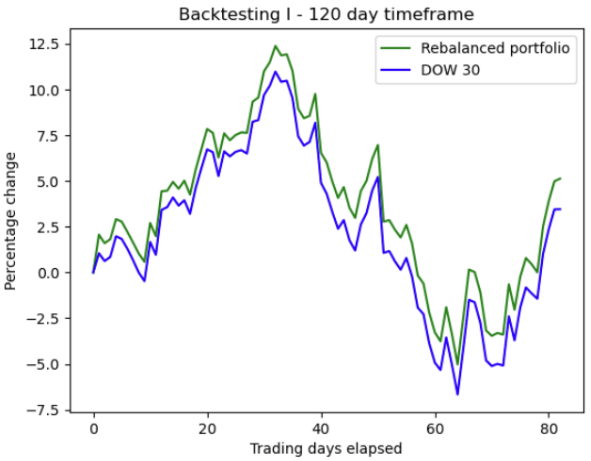
3.1 Backtesting I: Single iteration portfolio-rebalancing

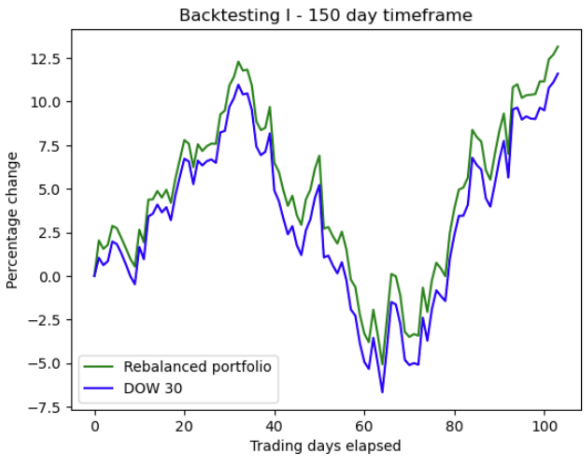
In the first backtesting scenario, we use a simple approach to test the effectiveness of our algorithm. In this approach, we attempt to do a single forecast through the Prophet forecasting model and then rebalance the portfolio once based on the forecasts put out by Prophet.

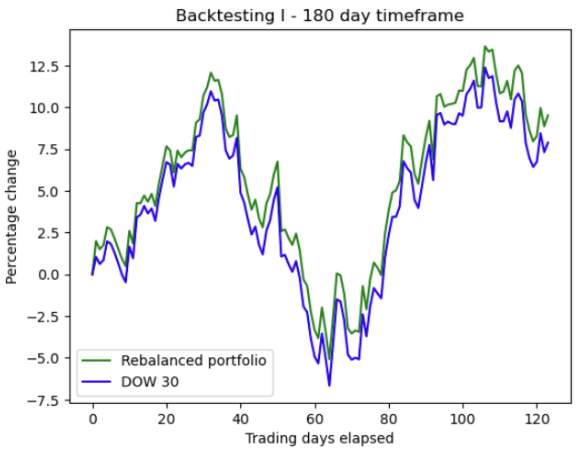






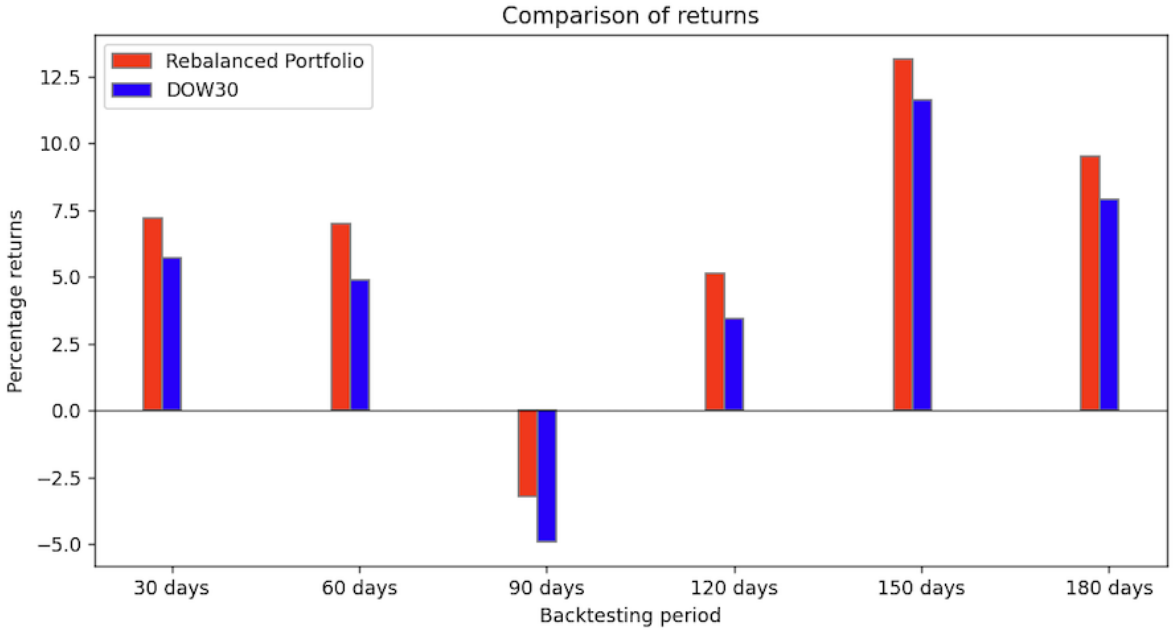






3.2 Evaluation of Backtesting I

Based on the line graphs in Section 3.1, it can be seen that the growth of our rebalanced portfolio very much mimics that of the growth of DOW 30. This can be explained due to the rebalancing being done only once at 0 trading days elapsed, so any increase factor with a constituent stock in DOW 30 should also closely affect that of our rebalanced portfolio. This essentially means that the strategy used in Backtesting I is very similar to that of a buy-and-hold strategy, with the only consideration being the rebalancing of assets based on the Prophet forecasts.

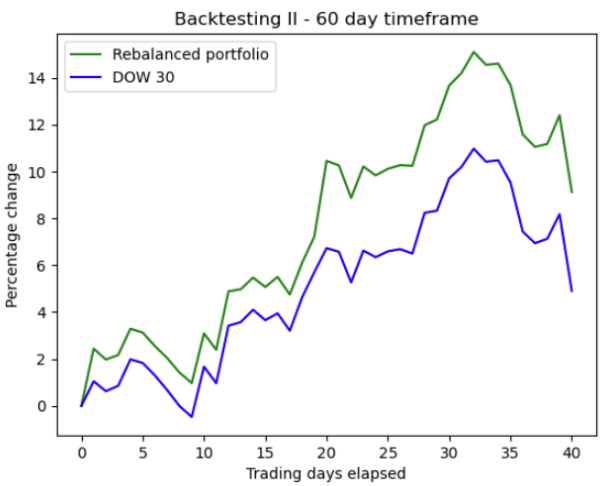


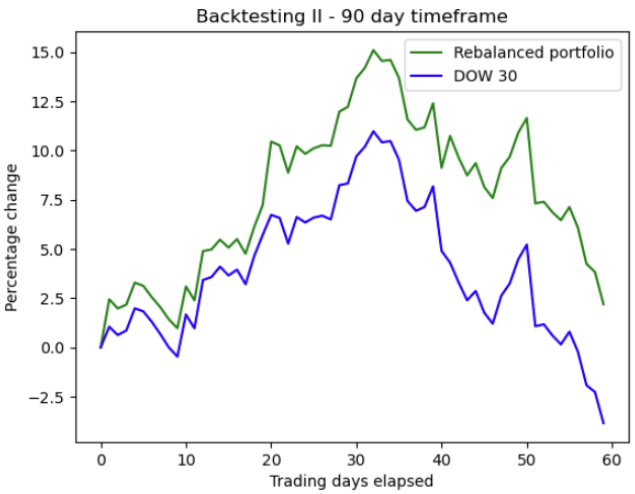
Based on the bar chart shown above comparing the overall returns between DOW 30 and our rebalanced portfolio, it is evident that for all the gains, our portfolio has performed considerably better than DOW 30, meaning that the rebalancing has proved to be effective. However, the negative values in the 90-days backtesting period does raise a concern, and further investigation reveals that this was the aftermath effect of both the instability stemming from the end of the COVID-19 pandemic, along with the Russian-Ukraine conflicts. This ultimately triggered a global financial instability, and a large majority of stocks have experienced a crash along with the DOW 30 index.

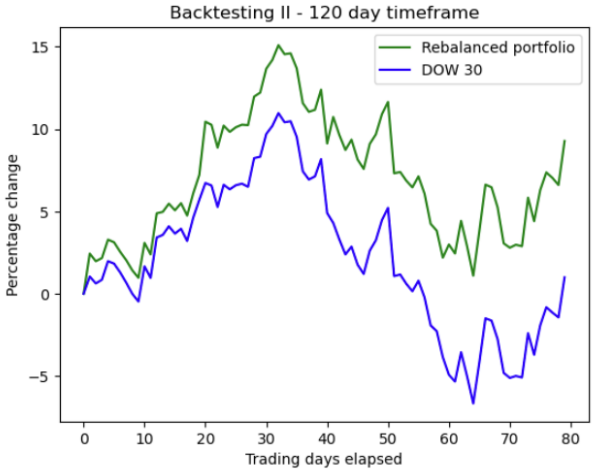
Overall, it can be concluded that factoring in the forecasts from Prophet could potentially prove to give investors an edge over benchmark indices, even if it is through a one-time rebalancing.

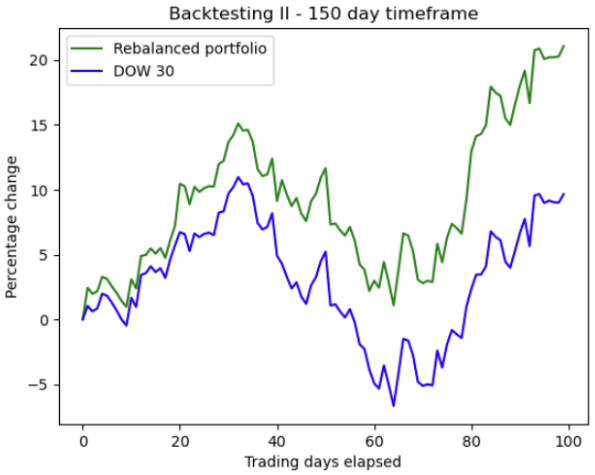
3.3 Backtesting II: Multiple iteration portfolio rebalancing

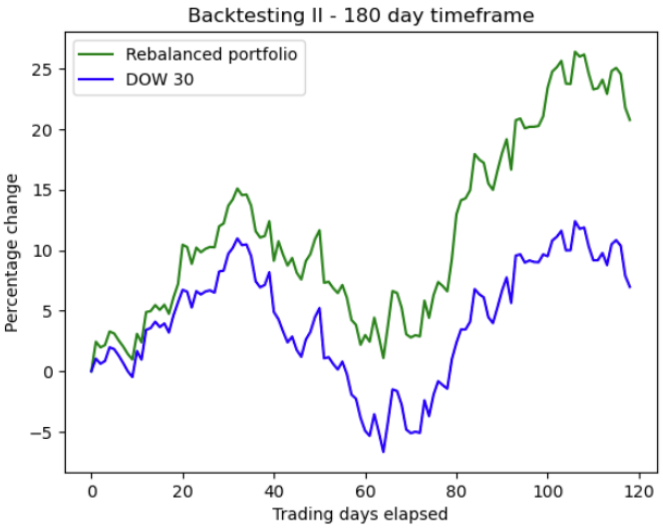
In the second backtesting scenario, we take our algorithm a step further and attempt to make it much more robust. In the real world, most investors would look into rebalancing their portfolio regularly as mentioned in the earlier section of this paper. As a result, we attempt to replicate such a scenario by carrying out forecasting every 30 days for the longer backtesting timeframes, and then rebalancing it accordingly.





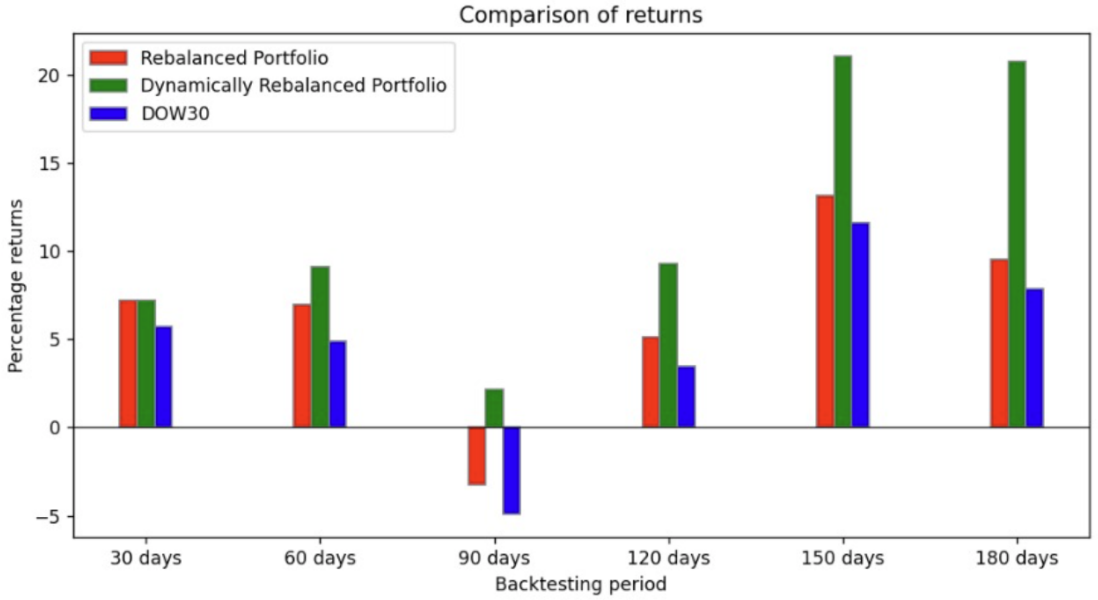






3.4 Evaluation of backtesting II

Based on the line graphs in Section 3.3, it can be seen that the behaviour of our rebalanced portfolio generally follows that of the DOW 30 index, but as the timeframes get longer, it can be seen that the patterns start to move differently. For instance, in the figure showing the 180 day timeframe, it can be seen that days 90 to 100 did not observe much movement within DOW 30, but it has seen a significant increase in our rebalanced portfolio. This is primarily due to the fact that our gains are then re-invested in the stocks that are forecasted by Prophet to do well in the market, which significantly increases our gains compared to if we were to put our investments in a general DOW 30 index fund. This further proves the effectiveness of the portfolio rebalancing theory mentioned in Section 1.3 of this paper.

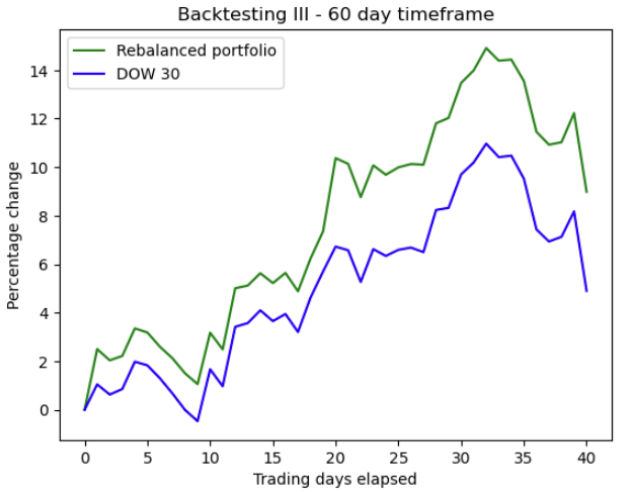


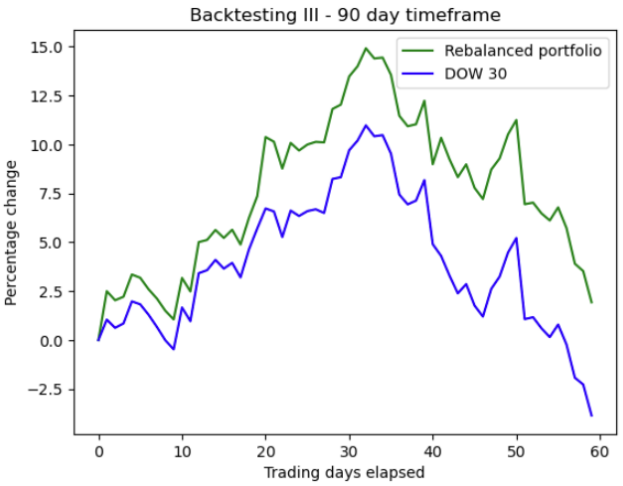
Observing the bar chart above that not only compares our new approach in Backtesting II, but also the earlier results from Backtesting I. It can be seen that the portfolio used in Backtesting II has consistently provided us with returns, even during the economic downturn that occurred during the 90-day backtesting period. The portfolio performs exceptionally well when there is significant growth in the economy, as can be seen in the 180 day backtesting period, with the portfolio experiencing more than double growth of that of the DOW 30. This effect is due to the fact that the portfolio has more capital to purchase shares, and a longer backtesting period allows more information to be fed into Prophet for a more reliable forecasting.

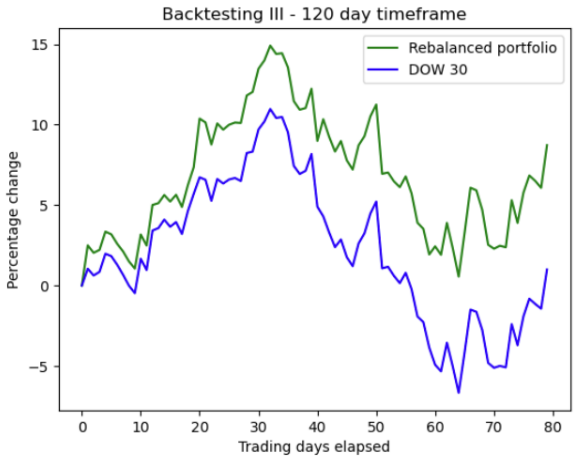
Overall, this trading strategy is very favourable in generating marginally higher returns, and with the long trading timeframe, it can certainly be considered feasible if an investor is willing to make their predictions through models like Prophet and go against the general population's strategy of following benchmark indices in investments.

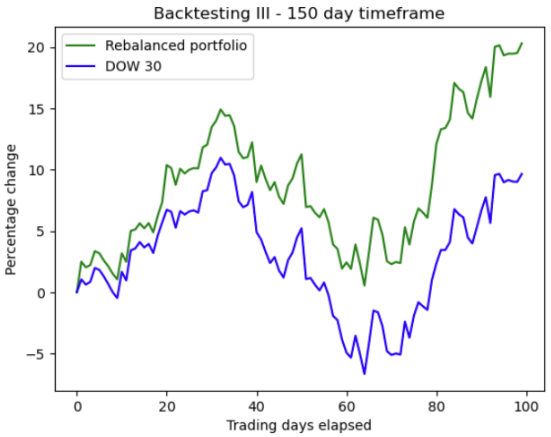
3.5 Backtesting III: Multiple iteration portfolio rebalancing (optimised)

The third and final backtesting scenario attempts to leverage the forecast features of Prophet. As investors would have various risk profiles, we try to introduce a high risk investment portfolio that tries to maximise returns, but also exposes the investor to greater losses. This is done through accessing the upper and lower bound of the Prophet forecasts, and factoring these results into our rebalancing algorithm.



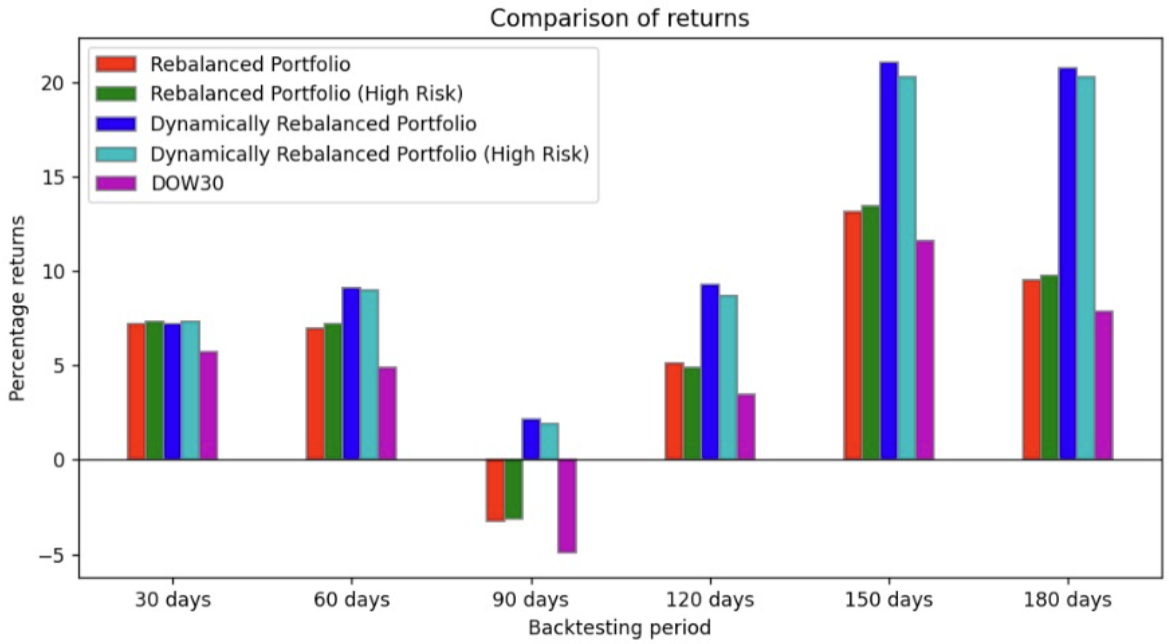






3.6 Evaluation of backtesting III

Based on the line graphs in Section 3.4, we notice that the optimised portfolio follows a similar pattern to that of the DOW 30 and it behaves very similarly to the earlier backtesting scenarios. This behaviour is expected, as the constituent stocks are identical to that of DOW 30, and it could be concluded that the optimization algorithm does not construct a portfolio that has a very different distribution from that of the previous investment portfolios due to the tight confidence bounds from the Prophet models.



Looking at an overall comparison of returns from all the portfolios through the backtesting stages, the unoptimised portfolio that is regularly rebalanced is the best-performing portfolio, with the single-iteration rebalanced portfolio performing the worst in general. When considering the optimisations done to construct a higher-risk portfolio, a single-rebalanced portfolio may benefit slightly more, though this would essentially be sticking close to the forecasts done by Prophet and heavily relying on the predictions done through the models.

4. Conclusion

4.1 Conclusion

The purpose of this study was to see if an optimal portfolio can be constructed using the right tools and methods, and based on our findings through the backtesting phases done earlier, it has been achieved consistently. The portfolios that were often rebalanced optimally through the forecasts done by Prophet performed considerably well, with a return of 21.05% over a 180-day backtesting time frame from June 2022 to December 2022 with 4 years of historical training data being fed into a forecasting model for each constituent stock.

The best performing portfolio performed twice as well as a DOW 30 index fund, never observing a loss, even throughout an economic crisis in the second half of 2022. In essence, the portfolios constructed would generally experience similar positive or negative

effects that would be experienced by DOW 30, meaning that the portfolio is reliant on the performance of the top companies based in the US economy.

To conclude, this trading strategy could certainly be looked into for investors who are hoping to skip the tedious step of going through extensive market research and simply use mathematics and programming to enhance their trading methods. However, it does take time and a lot of training and testing data to fine tune a forecasting model, but with the user-friendliness and clever interfaces in Prophet, the process has become a whole lot easier.

4.2 Limitations and potential improvements

It goes without saying that timing a market is not entirely possible, and forecasting models cannot predict sudden events that trigger economic recessions, such as COVID-19 and international conflicts. Thus, this strategy is not exactly “invincible”, but rather a more nuanced buy-and-hold strategy with some forecasting factored within.

Although this model fared rather well against DOW 30, this is a study that should be expanded to other benchmark indices, such as S&P 500 or the NASDAQ Composite to ensure that this approach is not specific towards a single benchmark index. In addition, this model could perform much better with more factors being considered when forecasting the movement of a stock price, such as economic indicators, current affairs and even positive/negative relationships between stocks.

Lastly, it is also important to note that in the real world, portfolio rebalancing is not done after a fixed amount of time elapsed, as this would not be reasonable from the point of view of a trader. To decide when a portfolio needs to be rebalanced, one must factor in the current distribution of assets, as well as the forecasted growth and whether the investor would be confident enough to distribute more of their portfolio to a stock that is expected to grow at a rapid rate. All of this could be considered in order to improve the rebalancing algorithm, building more confidence in it from an investors’ perspective.

All in all, investors are expected to approach stock trading with a very cautious, yet open mindset, and they would greatly benefit, both financially and educationally, from their experiences by taking a rather novel approach towards making investments.

References

[1] D. Gura, “Stocks sink, sending the S&P 500 to a bear market,” NPR, https://www.npr.org/2022/06/13/1104552530/stocks-sink-s-p-500-bear-market

[2] Facebook, “Facebook/prophet: Tool for producing high quality forecasts for time series data that has multiple seasonality with linear or non-linear growth.,” GitHub - Facebook/prophet, https://github.com/facebook/prophet

[3] M. Statman, “How many stocks make a diversified portfolio?`,” How many stocks make a diversified portfolio?, https://www.jstor.org/stable/2330969

[4] CNBC, “Dow 30,” CNBC, https://www.cnbc.com/dow-30/

[5] S. C. Heyford, “Rebalance your portfolio to stay on track,” Investopedia, https://www.investopedia.com/investing/rebalance-your-portfolio-stay-on-track/#:~:text=Portfolio%20rebalancing%20is%20the%20process,in%20check%20and%20minimize%20risk

[6] M. Krieger, “Time series analysis with Facebook Prophet,” Medium, https://towardsdatascience.com/time-series-analysis-with-facebook-prophet-how-it-works-and-how-to-use-it-f15ecf2c0e3a

[7] S. Taylor and B. Letham, “Forecasting at scale,” Prophet - Forecasting at scale, https://peerj.com/preprints/3190.pdf