COMILLAS
UNIVERSIDAD PONTIFICIA

Facultad de Ciencias Económicas y Empresariales (ICADE)

QUANTITATIVE FILTERING TECHNIQUES APPLIED TO DIVERSIFIED PORTFOLIO SELECTION AND PRACTICAL IMPLEMENTATION OF THE SAHM RULE

Author: Gabriel Medem López-Brea
Director: Ramón Bermejo Climent

MADRID | April 2024

Abstract

By designing an algorithmic trading strategy, this paper aims to solve a real problem of the asset management industry: being able to efficiently select a correctly diversified portfolio of stocks from a vast universe of companies consistently beating the benchmark index.

To do this, a backtesting approach has been taken by evaluating thousands of different combinations of parameters to conclude what would be the best possible investing strategy for maximizing either one of two target variables: total return or the Sharpe ratio. These two possibilities ensure that a broad range of retail investors' risk preferences are covered: risk-seeking investors will opt to use the total return strategy while risk-averse clients will likely prefer the more conservative Sharpe ratio option.

Moreover, an innovative macroeconomic parameter has been included in the investment strategy, the Sahm Rule. This factor was created by American economist Claudia Sahm as a method to predict periods of economic recession and automatically trigger a set of stimulus checks with the purpose of reducing the impact of the crisis on consumer demand. This factor has been adapted to be used in the investment strategy as an alert to sell all existing positions and stop investing until the recession has passed, in order to avoid taking unnecessary risks and preserve as much capital as possible.

Key findings show that the designed strategy is consistently capable of beating the benchmark index, the S&P 500. However, since the strategy is also capable of generating positive returns during recessionary periods, the implementation of the Sahm Rule negatively impacts its performance.

INDEX

1. **Introduction and objective**
   1.1. Objective
   1.2. Justification and motivation
   1.3. Expected results
2. **Data**
3. **Investment strategy design**
   3.1. Index selection: S&P 500
   3.2. Macroeconomic environment and indicators: Sahm Rule
   3.3. Factor selection
      3.3.1. Value, profitability, and momentum
      3.3.2. Piotroski F-score
      3.3.3. Altman Z-score
      3.3.4. Combination of factors and scores
   3.4. Number of companies in a correctly diversified portfolio
   3.5. Portfolio rebalancing methodology
   3.6. Stock selection process
      3.6.1. Parameter optimization and optimal strategy selection
4. **Results**
   4.1. Benchmarking against the reference index
   4.2. Optimal parameters
   4.3. Evaluation of the Sahm Rule
5. **Limitations and areas for improvement**
6. **Conclusion**
7. **Declaration of use of Generative Artificial Intelligence**
8. **Bibliography**
9. **Appendices**

1. INTRODUCTION
1.1. Objective

While different well-known authors and investment specialists over time have suggested a range of factor investing techniques for American equities, this paper aims to build on the research carried out by Ramón Bermejo in 2021 which studies the use of value, profitability, and momentum metrics for selecting cheap, profit-generating, and momentum-carrying stocks from a vast universe of European equities. This study aims to apply this technique to the S&P 500 and improve it with the addition of guru selection methods like the Piotroski F-score and Altman Z-score, and also through the implementation of the Sahm Rule as a macroeconomic trigger to avoid large losses due to the non-diversifiable risk.

The objective of this dissertation is to develop an algorithmic trading strategy for efficient portfolio selection based on the use of these quantitative metrics which shall be able to select a diversified portfolio of stocks from a large universe of companies such as the S&P 500 in a fast and efficient manner which would not be possible through traditional stock selection methodologies which do not involve algorithmic trading.

The nature of this work is explicative and not predictive, meaning that through the programming of a strategy backtesting algorithm and using past financial data from the Factset database it will be possible to analyse thousands of different investment strategies and conclude which would have been the best one to maximize the target variable for a given time period, which may vary depending on each client's risk profiles. For instance, risk-averse clients may opt to maximize Sharpe ratio or minimize maximum drawdown throughout the investment period, while risk-seeking clients may choose to maximize total return. The main advantage of the implementation of an algorithmic approach to the investment strategy analysis is the possibility to conduct advanced parameter optimization. Optimizable parameters include:

a) Variable selection: deciding what single combination of financial variables will be a better predictor of stock price performance and, most importantly, be able to outperform the reference index regarding the target variable (e.g., total return or Sharpe ratio). Examples of these variables are financial ratios such as the price-earnings ratio and EV/EBITDA, momentum metrics like the 1-month total return, or aggregated scoring methodologies like the Altman Z-score. Moreover, it will be possible to analyse the level of importance that should be given to each different variable.
b) Rebalancing period: trading strategies may set a specific rebalancing period which could be daily, monthly, semi-annually, or any other time period that may maximize the target variable.

c) Portfolio size: number of companies present in the portfolio at a time. It must be ensured that the portfolio always has the minimum number of stocks to be considered a diversified portfolio.
d) Investment distribution: capital allocation between the selected stocks (equally vs. value weighted portfolios).

The overall study objective will be to design a strategy that beats the benchmark index while maximizing the target variable and with a minimal number of inputs in terms of the optimizable variables to avoid bias.

1.2. Justification and motivation

The paper addresses an existing problem in the asset and wealth management industry: being able to instantly select a diversified investment portfolio from a large stock universe depending on each client's risk profile and investment preferences. With algorithmic trading, we can develop a quantitative, non-biased selection methodology in an efficient and optimized manner, which would not be possible via traditional stock selection techniques.

Although there is an extensive amount of literature on stock selection methodologies, there is still vast room for improvement in the implementation of investment strategies in American equities as investment funds struggle to consistently outperform the market; for instance, 93% of large-cap stock funds in the USA have underperformed relative to the S&P 500 in the past 20 years (Foster, 2024).

1.3. Expected results

The outcome of this study will be an optimized investment strategy capable of generating the highest total return or Sharpe ratio over a given period, with all parameters having been optimized without the influence of human bias. The strategy must be able to steadily outperform the benchmark index, for instance, if the S&P 500 is used for stock selection, the strategy shall outperform that same index for comparison's sake.

Moreover, it will be possible to analyse whether the implementation of the Piotroski F-score, Altman Z-score, and Sahm Rule improve the already tested value, profitability, and momentum strategy.

2. DATA

The development of an algorithmic trading strategy based on the previously mentioned financial and macroeconomic metrics has involved the use of three distinct data sources:

a) List of members of the S&P 500 throughout the investment timeline: the companies who belong in the S&P 500 index continuously change over time on an as-needed basis (Standard & Poor's, n.d.), so it is necessary to obtain a database containing a list of the index members over time. This database was accessed through the Bloomberg terminal provided by Universidad Pontificia de Comillas (ICADE) using the Excel API (Bloomberg, 2024).
b) Financial data: the data used by the stock selection algorithm belongs to the companies' quarterly reports and is sourced from FactSet (2024a).
c) Unemployment data: to calculate the Sahm Rule indicator, US unemployment rate data for the investment period is needed. This data has also been sourced from FactSet (2024b).

3. INVESTMENT STRATEGY DESIGN

To be able to design an adequate investment strategy without the influence of human bias, an algorithmic approach has been implemented which gives the algorithm as much freedom as possible when it comes to variable selection and parameter optimization. Nevertheless, the algorithm has been provided with a given set of options for the different parameters that it must optimize from which it shall iteratively test all different parameter combinations and ultimately choose the one that maximizes whichever is the target variable (total return or Sharpe ratio), always in line with the existing literature on factor investing in equity markets.

These inputs which the algorithm has been provided with cover six topics: stock universe to select equities from (S&P 500), the macroeconomic environment, financial variables which may be used, number of companies to invest in to be considered a diversified portfolio, investment distribution throughout the portfolio and options for rebalancing period schedule, and ultimately the mathematical approach to stock selection once all the other factors have been decided.

3.1. Index selection: S&P 500.

The first decision to be taken regarding portfolio selection is the company universe from which the algorithm may select the outperformers. The use of a structured database, meaning data that is organized in a homogeneous format in lines and columns (Suyash and Anuranjan, 2017), greatly simplifies the programming of an investment strategy, as such, it is primordial that the chosen index is subject to strict legislation which ensures that member companies regularly file structured financial reports which are used to build the database.

The United States Securities and Exchange Commission, the regulatory body of the USA stock market, has a strict reporting regulation which requires public companies to file quarterly reports in a structured format provided by the 10-Q template for quarterly reports and 10-K for yearly reports. The SEC mandates that any company with market capitalization over $700m is categorized as a “Large Accelerated Filer” and therefore must file quarterly reports in 10-Q format 40 days after period end, and for the last quarter of the year they are exempt from 10-Q but instead must file a yearly report in 10-K format 60 days after year end (United States Securities and Exchange Commission, 2002).

Exhibit 1: 10-Q and 10-K filing dates (United States Securities and Exchange Commission, 2002)

Company category	**Revised deadlines for filing Periodic Reports**	
	10-Q deadline	**10-K deadline**
Large Accelerated Filer (over $700m)	40 days	60 days
Accelerated Filer (between $75m and $700m)	40 days	70 days
Non-accelerated Filer (less than $75m)	45 days	90 days

This strict regulatory framework enforced on US equities makes American large-cap indexes like the S&P 500 perfectly suitable for structured data analysis. Since Standard & Poor’s sets the minimum market capitalization for companies to be eligible for the S&P 500 at $12.7bn, all the index members categorize as “Large Accelerated Filers” for the SEC and must therefore comply with the 10-Q and 10-K reporting requirements.

There are several US stock indexes that would comply with the necessary criteria to build a structured database to use for this analysis (e.g., Dow Jones Industrial Average, Nasdaq, Russell 3000, etc.); however, the S&P 500 has been chosen because of it being the most followed one by investors worldwide, and therefore being an appropriate benchmark for any new investment strategy. As stated by Nasdaq (2022), the SPDR S&P 500 ETF, which tracks the S&P 500, is not only the largest Exchange-Traded-Fund mirroring an American stock index, but the overall largest ETF in the USA with over $520bn in assets under management as of April 2024 (State Street Global Advisors, 2024).

By deciding to select a diversified portfolio of stocks from the S&P 500 and not any different universe of companies we are making a first investment decision that inevitably impacts the performance of the strategy. Companies included in large-cap indexes such as the S&P 500 have higher trading volumes than smaller companies and are therefore more prone to reflect market sentiment. The S&P 500 shows an especially strong correlation between trading volume and performance in terms of momentum, implying that being a member of the index positively impacts stock performance (Brown, Crocker and Foerster, 2009). To correctly account for this effect, the investment strategy performance must always be compared against the S&P 500 as a benchmark index and not any other reference.

Moreover, it is necessary to account for the changes in the S&P 500 members list as Standard & Poor's may include and exclude companies on an as-needed basis. The solution has been to obtain a monthly list of members for the whole investment period from the Bloomberg database (Bloomberg, 2024), and only allowing the algorithm to invest in companies which at each rebalancing moment were present in the index. By doing this, we prevent the algorithm from investing in companies before they were included in the index, which would not comply with the criteria of the strategy, and it would also be unrealistic to assume that we have gotten to know about a stock before it being large enough to be included in an index such as the S&P 500. The latter is known as survivorship bias, as Falkenberry (2006) explains, survivorship bias is particularly noticeable when backtesting investing strategies which select companies from indexes whose membership is depicted by market capitalization, as is the case of the S&P 500.

3.2. Macroeconomic environment and indicators: Sahm Rule

Given the explicative nature of this study, a major decision to be taken is the period within which the analysis is carried out as the resulting best-performing investment strategy is inevitably overfitted to that range of time. When implementing backtesting techniques for investment strategy evaluation there is a significant overfitting risk to the macroeconomic environment in which the algorithm is tested, resulting in enhanced results but a strategy which is highly susceptible to failure in different economic scenarios (Bailey et al., 2016). As suggested by Ying (2019), there are different solutions that can be implemented to avoid this phenomenon. One of the solutions is data expansion, which refers to using a vast array of data which includes environments of all different kinds to generate an all-weather strategy capable of outperforming the benchmark in all kinds of macroeconomic conditions.

Catering to this remedy, the chosen period ranges from January 2007 to January 2023. Using such an extensive period ensures that the resulting strategy is not overfitted since it must perform in all the different situations encountered by the US in said period:

- <u>2007-2009: The Great Recession</u>. Thomas (2011) explains how in those years the US suffered the consequences of an inflated housing market and the blow-up of the subprime mortgage market due to banks and financial institutions taking excessive amounts of risk. As a result, the GDP fell by 25% and strong fiscal and monetary policies had to be implemented. Firstly, the Federal Reserve was forced to lower interest rates to almost 0% in an attempt to lower borrowing costs and stimulate the economy. Since this measure alone was not sufficient, the Fed had to implement a Quantitative Easing strategy of pumping money into the economy by purchasing government bonds and mortgage-backed securities (MBS) and therefore effectively lowering the long-term interest rate. A mix of emergency fiscal policies were also implemented by the government, including tax incentives, stimulus packages, and specific programs such as the Troubled Asset Relief Program (TARP) which aimed to stabilize the financial sector.

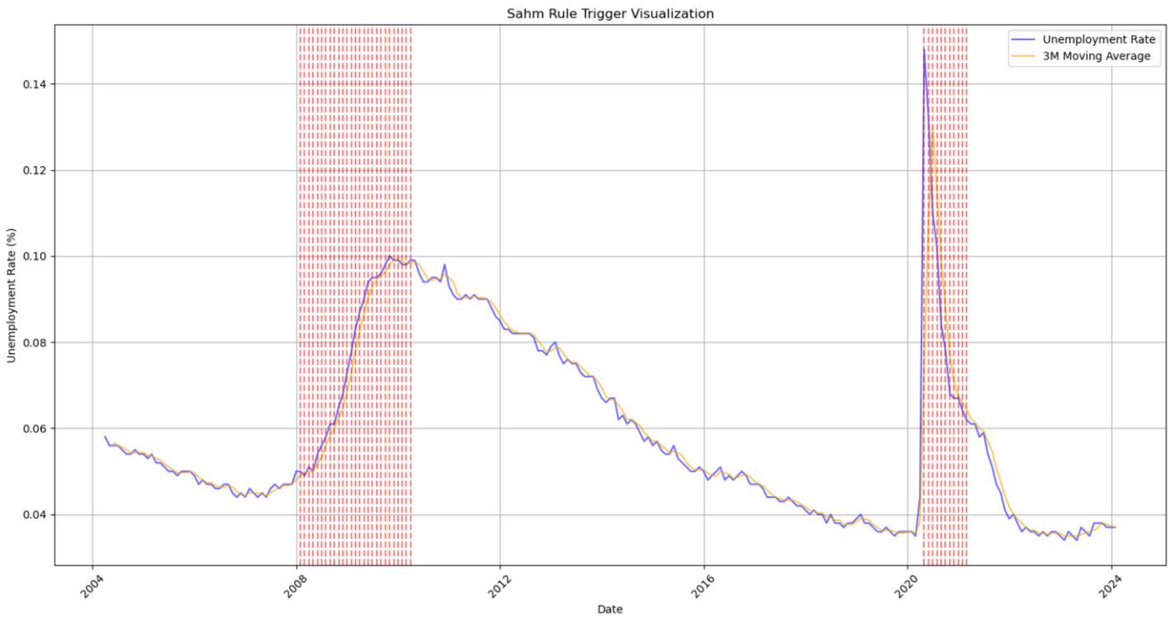
- 2010-2012: Recovery period. Throughout the following years to the Global Financial Crisis the Fed maintained its Quantitative Easing and low interest rate strategy (Thomas, 2011).
- 2013-2019: Economic expansion. As the economy bounced backed and started growing again the Federal Reserve cautiously initiated a slow rate hike starting in 2015, and leaving behind the near-zero rate era of the financial crisis. By 2019 interest rates were already at a business-as-usual level of c.2.0%, and no particular fiscal policies were needed in this time period (Federal Reserve, 2018).
- 2020-2021: Covid-19 pandemic. The pandemic caused the global economy to pause due to inevitable lockdowns which brought business activity to a stop and consequently caused a supply shutdown. The Fed had to implement similar policies to those of the Great Recession, bringing interest rates back down to almost zero and respawning the Quantitative Easing program to try to revitalize the economy. At the same time, strong fiscal measures were implemented, such as providing citizens with a $1,200 per person stimulus package aimed at increasing demand for consumption (Chadha et al., 2021).
- 2022-2023: Supply bottlenecks and inflation concerns. After Covid, the US population had substantial savings as a result of the reduced activity during 2020 and the stimulus check provided by the government, this factor coupled with a global supply constraint caused inflation to skyrocket due to increased demand and restricted supply. In an attempt to halter inflation, the Fed steadily rose interest rates in 2022 and 2023 up to over 5.0%. At the same time, a non-conventional monetary policy of Quantitative Tightening by which the government sells its securities, such as MBS, was implemented to further reduce the money supply and freeze inflation (Federal Reserve, 2023).

Overall, overfitting risk has been diminished through the use of such an economically diverse period of US history. However, using a period with economic downturns implies that the investment strategy is exposed to them and must perform in said situation, as any real-life investor would have. This risk of the whole economy running aground is a non-diversifiable risk, since it is not possible to prevent it through the generation of a well-diversified stock portfolio (Anyika, 2015). To provide the algorithm with a tool which enables it to foresee economic downturns and avoid non-diversifiable risk the Sahm Rule has been implemented.

The Sahm Rule was developed by Claudia Sahm, an American economist and member of the Federal Reserve Board, as a means to predict recessions and trigger an automatic response in the form of stimulus checks with the aim to reduce the impact on families' purchasing power during the economic downturn. The trigger sets off when the moving average of the national unemployment rate increases by at least 0.5 percentage points relative to its low in the previous 12 months (Sahm, 2019).

To mitigate recessionary risk, the Sahm Rule has been used in the investment strategy as a signal to divest all existing positions and stop investing until the trigger-pulling economic conditions set by the unemployment rate disappear. As a result, there are two periods, from January 2008 to March 2010 and April 2020 to February 2021 when the Rule sets off and therefore the algorithm sells all positions and does not invest to avoid potential losses.

Exhibit 2: Implementation of the Sahm Rule from 2004 to 2023.



3.3. Factor selection

As explained before, this study capitalizes on the positive results obtained by Ramón Bermejo (2021) on factor investing techniques using combinations of value, profitability, and momentum metrics to evaluate and select a diversified portfolio of companies from a large index. However, the generated investment strategy aims to improve this strategy with the addition of aggregated scoring methodologies created by gurus of the financial industry such as Stanford accounting professor Joseph Piotroski and New York University Professor of Finance Edward I. Altman.

3.3.1. Value, profitability, and momentum

In 1992, Fama and French introduced the use of the book-to-market ratio as a value metric to incorporate in their three-factor investment model. Fama and French concluded that "value stocks" with a high accounting value in comparison to their market capitalization were more likely to outperform those with opposite characteristics.

The value metrics used as factors in the investment strategy are:

- <u>Market to book value</u>: ratio of market capitalization to accounting value of equity. A higher market to book value implies a company is higher valued compared to its book value.
- <u>Price earnings</u>: ratio of price per share to earnings per share (equal to market capitalization to net income). A higher PE ratio implies overvaluation compared to profits.
- <u>EV / EBITDA:</u> this metric evaluates the ratio of enterprise value to EBITDA. A higher value implies overvaluation compared to EBITDA.
- <u>EV / EBIT</u>: enterprise value to EBIT. A higher value implies overvaluation compared to EBIT.

Since value metrics are implemented in the investment strategy to identify undervalued stocks with potential to increase their share price, the algorithm rewards having lower values for these four ratios.

Novy-Marx (2013) introduced profitability metrics to factor investing as a means to not only capture undervalued investments but also stocks that are outperformers in terms of profit generation. Novy-Marx used the gross-profit-to-assets ratio as an accurate way of measuring profitability in relation to invested capital, avoiding the accounting and reporting intricacies involved in calculating more standard variables such as net income. He was able to demonstrate that the combination of profitability metrics coupled with Fama and French's value factors produced better results than each of them alone. Moreover, Joel Greenblatt (2010) used a combination of EBIT/EV and Return of Capital (defined by him as EBIT divided by the sum of net fixed assets and working capital) to select a combination of undervalued and profitable stocks in a similar approach to that of Novy-Marx.

Contrary to value factors, in profitability factors the algorithm must reward the highest values since they indicate more profit-generating potential. The profitability metrics used as factors in the investment strategy are the following:

- <u>Gross profit to assets</u>: ratio of gross profit (taken as sales minus the cost of goods sold) to total assets. A higher value indicates more profit-generating capacity per monetary unit invested in assets.
- <u>Return on Capital (Greenblatt):</u> EBIT compared to capital invested as defined by Greenblatt (2010). A higher value implies higher profitability.
- <u>Return on Capital (Detailed):</u> extension of Greenblatt's approach to capital invested by including intangible assets. A higher value implies higher profitability.

In 1993, Jegadeesh and Titman first introduced the concept of momentum investing and concluded that past stock performance is a statistically significant variable impacting future stock performance. As such, using total return metrics of different time period lengths (e.g., last 1-month total return, 3-month total return, one-year total return, etc.) allows the algorithm to capitalize on this relationship discovered by Jegadeesh and Titman and improve the value and profitability only strategies of Novy-Marx and Greenblatt.

The momentum metrics used in the investment strategy are the total return values for previous month, quarter, half-year, and full year. Since the relationship between past and future total return is direct, the algorithm rewards stocks with higher momentum values.

3.3.2. Piotroski F-score

The use of the Piotroski F-score has been proven to enrich factor investing strategies in European equities (Tikkanen and Äijö, 2018), and the objective of including it in this study is to confirm whether this is the case for American equities quoted in S&P 500 or not.

As explained by Joseph D. Piotroski (2000), the F-score is an accounting-based stock scoring methodology aimed at providing a well-rounded approach to company analysis and stock selection. In his paper, Piotroski performed a backtest simulating an investment from 1976 to 1996 by selecting the highest quantile book-to-market stocks and from that set of companies separating the winners and losers by applying a combination of 9 factors encompassing profitability, financial leverage, liquidity, and operating efficiency. On each of these requisites a company could score either 1 point or none depending on whether they met the criteria or not, adding up to a maximum 9 points of Piotroski F-score for each company.

The nine requisites proposed in the Piotroski F-score are the following:

1. Positive Net Income.
2. Positive Operating Cash Flow.
3. Return on Assets (ROA) improvement compared to previous year.
4. Cash Flow from Operations Exceeding Net Income.
5. Decrease in Long-term Debt Relative to Assets compared to previous year.
6. Increase in Current Ratio compared to previous year.
7. No New Shares Issued in the year.
8. Improvement in Gross Margin compared to previous year.
9. Increase in Asset Turnover Ratio compared to previous year.

3.3.3. Altman Z-score

Similarly to Piotroski F-score, Altman's Z-score has also been combined with other factor investing techniques and has proven to be a useful predictor of bankruptcy risk (Roy, 2016).

The Altman Z-score as defined by Edward I. Altman (2000) is an aggregated financial score aiming to predict the probability of bankruptcy of a company within two years. The score uses a range of financial variables from the balance sheet and income statement and adds them up in the following way to calculate an aggregated Z-score = 1.2*X1 + 1.4*X2 + 3.3*X3 + 0.6*X4 + 1.0*X5, where each variable is equal to a different financial ratio:

- X1: working capital to total assets.
- X2: retained earnings to total assets.
- X3: EBIT to total assets.
- X4: market value of equity to total liabilities.
- X5: sales to total assets.

3.3.4. Combination of factors and scores

Overall, the algorithm is given 13 different factors aggregated into five groups:

- Value: market to book value, price earnings, EV/EBIT, and EV/EBITDA.
- Profitability: gross-profit-to-assets, Return on Capital (Greenblatt), Return on Capital (Detailed).
- Momentum: 1-month total return, quarterly total return, semi-annual total return, annual total return.
- Piotroski F-score.
- Altman Z-score.

Given the task of automatic optimization of the investment strategy and avoiding human bias, the algorithm is allowed to test investment strategies with any combination of 5 factors including only 1 of each group. For instance, one combination would be: EV/EBIT, gross-profit-to-assets, 1-month total return, Piotroski F-score, and Altman Z-score. Overall, the total number of factor combinations which the algorithm may test is 48.

3.4. Number of companies in a correctly diversified portfolio

Another very impactful variable that must be optimized is the portfolio size. At every portfolio rebalancing date, the algorithm must sell all existing positions, reevaluate all stocks in the S&P 500, and select a given number of them to invest in until the next rebalancing date.

Existing literature on the optimal number of stocks to make a diversified portfolio capable of mitigating unsystematic or idiosyncratic risk, which is the risk related to company and industry specific characteristics (Mokkelbost, 1971), is very varied and does not draw to a consensual conclusion or a specific number, but rather provides a range of what should be generally acceptable. For instance, a recent literature review carried out by Zaimovic, Omanovic, and Arnaut-Berilo (2021) suggests that while previous authors in the 1950's to 2000's recommended an optimal size of c.10 stocks, more recent studies suggest a larger number, closer to the 40-stock mark. Overall, most studies suggest that a diversified portfolio is achieved with 20-30 stocks, and although including a larger number of them further decreases volatility, the marginal benefit of them decreases significantly (Bermejo, 2021).

In order to maintain a correctly diversified portfolio as suggested by the literature but at the same time allow for automatic optimization of the investment strategy, the algorithm has been allowed to use either 15, 25, or 35 companies as the number of stocks to invest in at each rebalancing date.

Nevertheless, it will not be 15, 25, or 35 stocks from the total of 500 members of the S&P 500 since financial industry companies (e.g., banks and insurance companies) have been excluded from the investment strategy due to their particular accounting characteristics which make them incomparable with the rest. Banks and insurance companies operate under different regulatory frameworks which make it mandatory for them to report financial statements in a unique manner. This unique accounting techniques are designed to minimize specific risk exposure that these companies face and are tailored to the complex financial products that they offer (Tinta, 2015).

3.5. Portfolio rebalancing methodology

The rebalancing schedule of an investment strategy can greatly impact the total returns of the investment over time. Although it may seem like constant rebalancing as the market changes is the best option to capture as much growth as possible and avoid persistent losses, the truth is that investors must adapt to the timings of information availability as companies file their quarterly reports. Nevertheless, the price per share of stocks does update constantly and can be considered in real time to make investment decisions.

The strategy developed in this study is based on financial factors which are published in companies' quarterly reports as mandated by the US Securities and Exchange Commission; therefore, for the sake of avoiding human-bias and pursuing simplicity, the rebalancing periods that have been tested are quite standardised while remaining optimizable at the same time. The algorithm is allowed to test four different rebalancing schedules in its pursue of selecting the best investment strategy possible within the given parameters: monthly, quarterly, semi-annually, and annually.

Past research suggests that longer rebalancing periods (e.g., annually) provide a more balanced solution in terms of cost optimization and total return, due to the high costs incurred when carrying out the rebalancing of a portfolio (Tokat and Wicas, 2007). However, with an algorithmic trading strategy such as the one developed here these costs are minimized by eliminating the need for human processing of all the data, so shorter periods may also be a good option and could benefit from a faster reaction time to sharp changes in the stock market.

When using a backtesting approach like the one taken in this study to test an investment strategy one must take special care to not incur in forward looking. This mistake refers to using future data which would not have been available at the time of rebalancing. Since financial reports of S&P 500 companies are published with a 40-day delay to the end-of-quarter date and 60-days to end-of-year date as mandated by the Securities and Exchange Commission, at each rebalancing date the algorithm must use financial data from the financial report which would have been available at that date to simulate the real-life availability of financial reports. A common error would be to take the financial data from the rebalancing date which is now available after enough time has passed from that date for companies to publish the report containing that period.

The only data that would be available at the rebalancing date is the price per share, and therefore the algorithm must calculate all ratios involving price per share with the rebalancing date price but with older financial information found in the latest available financial report.

Exhibit 3: Information availability simulation used by the algorithm to avoid forward looking.

REBALANCING DATE	AVAILABLE REPORT
31/12	Q3
31/01	Q3
28/02	Year-end
31/03	Year-end
30/04	Year-end
31/05	Q1
30/06	Q1
31/07	Q1
31/08	Q2
30/09	Q2
31/10	Q2
30/11	Q3

Moreover, a major decision regarding the investment strategy and in particular the rebalancing process is whether to implement a long-only strategy as opposed to a long-short strategy where short-selling is also allowed. In this matter, advice provided by Novy-Marx and Velikov (2015) has been followed, as they explain that long-short strategies do not outperform traditional long-only strategies when accounting for transaction costs. Also, a majority of investors are restricted to long-only strategies due to legislation (Bermejo, 2021), and therefore it is more realistic to approach a long-only strategy.

The last decision to be taken regarding portfolio rebalancing is the investment distribution across the portfolio, which may be equally weighted or value weighted. Existing literature observes that equally weighted portfolios tend to outperform value weighted ones, and although value weighted ones seem to produce portfolios with lower volatility (Fama and French, 1992) they seem to offer limited returns (Chen, Chen and Li, 2012). Following this research the investment strategy here designed is equally weighted.

3.6. Stock selection process

Once the index, time range, factors, portfolio size, and rebalancing methodology has been set, the only remaining area of the strategy design is the mathematical stock selection algorithm; this is, the mathematical process by which at each rebalancing date the algorithm assigns a score to each of the available companies in the index and ranks them to select the top ones for investing.

As explained in section 3.3., the algorithm tests strategies with any combination of 5 factors, one of each of these five types: value, profitability, momentum, Piotroski F-score, and Altman Z-score. However, as suggested by Ramón Bermejo (2021), there are two ways to combine these factors and select stocks:

- Iterative method: applying the factors sequentially. For example, out of the S&P 500, first keeping the top 50% in terms of value, then profitability, then momentum, from the remaining set excluding those with Altman Z-score lower than 1.5, and from the few remaining selecting those with Piotroski F-score equal to 9.
- Mixed: combining the factors without a specific order, through standardization of the metrics and averaging them to obtain a unique score per company.

Ramón Bermejo concludes in his thesis (2021) that iterative strategies yield higher returns than mixed strategies. However, he does not test an advanced version of the mixed approach which is to assign different weights to the factors. For example, assigning value 50% of the weight, profitability 10%, momentum 10%, Piotroski F-score 25%, and Altman Z-score 5% for a total of 100%. In the standard mixed strategy, all the variables have the same importance (e.g., 20% weight each).

The algorithmic approach taken in this study allows for automatic optimization of these weights, allowing the algorithm to detect what weight combination can better predict stock performance.

Due to computational limitations, the proposed algorithm has only been able to test weights of 0%, 10%, 20%, 30%, 40%, and 50% for each factor. Overall, the total number of weight combinations is 651 for each of the 48 factor combinations explained in section 3.3., adding up to a total number of factor and weight combinations of 31,248.

In conclusion, the stock scoring and selection method at each rebalancing date is the following:

1. The Sahm Rule trigger is checked, if a) the trigger is activated due to macroeconomic conditions, the algorithm will divest all existing positions and not invest until the next rebalancing date when the alert is over.

2. The tested factors (value, profitability, momentum, Piotroski F-score, and Altman Z-score) are standardized across the whole S&P 500 for comparability purposes. This results in a Z-score (not Altman's) for each factor and company, calculated as the factor (e.g., EV/EBITDA) minus the mean of that factor across the whole index and divided by its standard deviation.
3. A weighted average of the five Z-scores of each stock is calculated using the combination of weights being tested (out of the 651 possibilities) to result in one single final score for each company.
4. Depending on the number of companies in the portfolio being tested, the highest ranked companies up to that number are selected and equally invested in until the next rebalancing period, when a full divestment will take place and the process restarts.

3.6.1. Parameter optimization and optimal strategy selection

Overall, the final objective of taking an algorithmic approach is allowing for the optimization of as many variables as possible. The way the code has been programmed, the algorithm does not really optimize the variables to reach a final result but rather tests all the possible combinations of variables and returns a list of the results of all the different tested strategies in terms of total return and Sharpe ratio, where:

a) Total return is a comprehensive measure of financial performance which accounts both for capital gains and any dividends received from stocks. As such, it serves as an accurate return metric better than simply stock price changes because it includes dividends.
b) The Sharpe ratio was defined by William F. Sharpe (1994) as a measure of return in relation to volatility of an investment strategy, so, the higher the Sharpe ratio the higher the returns per unit of volatility. This is a measure of performance preferred by more risk-averse investors who care about volatility.

There are four variables for which the algorithm tests different possibilities:

- <u>Factor combination</u>: 48 different combinations of value, profitability, momentum, Piotroski F-score, and Altman Z-score.
- <u>Factor weight combinations</u>: 651 weight combinations used to make a weighted average of the standardized factors for each stock and calculate a final score.
- <u>Portfolio size</u>: 3 possibilities for the size of the portfolio (15, 25, or 35 stocks).
- <u>Rebalancing period</u>: 4 possible rebalancing schedules (monthly, quarterly, semi-annually, and annually).

Overall, the total number of strategies that have been tested by the algorithm is the multiplication of the combinations of these variables: 48 x 651 x 3 x 4 = 374,976 different investment strategies.

Exhibit 4: Examples of possible strategies that have been tested.

Factors	Weights	Portfolio size	Rebalancing period
('PE', 'ROC GREENBLATT', 'SEMI TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.5, 0.2, 0.3, 0.0, 0.0)	15	1M
('EVEBITDA', 'GPA', 'ANNUAL TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.2, 0.3, 0.5, 0.0, 0.0)	25	3M
('EVEBIT', 'ROC DETAILED, 'MONTHLY TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.2, 0.2, 0.5, 0.1, 0.0)	35	6M
('BOOK TO MARKET', 'ROC GREENBLATT', 'QUARTERLY TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.2, 0.3, 0.4, 0.1, 0.0)	15	12M

The final output of the algorithm is a list of the 374,976 tested strategies from January 2007 to January 2023 and the total return and Sharpe ratio generated. Therefore, there will be a winning combination of factors, weights, size, and rebalancing for maximizing the total return and another one for the Sharpe ratio.

4. Results

4.1. Benchmarking against the reference index

The output obtained after running the backtesting algorithm is a dataset containing the 374,976 tested strategies with the total return and Sharpe ratio which they would have achieved if put into practice between January 2007 and January 2023. Given the high optimization achievable by testing different options for four parameters, the top strategies significantly outperform the benchmark (S&P 500) both in total return and Sharpe ratio. A list of the top strategies in terms of a) total return and b) Sharpe ratio is shown in Appendices 1 and 2.

The strategy delivering the maximum total return managed to get 891% return and 1.17 Sharpe ratio, and it had the following parameters:

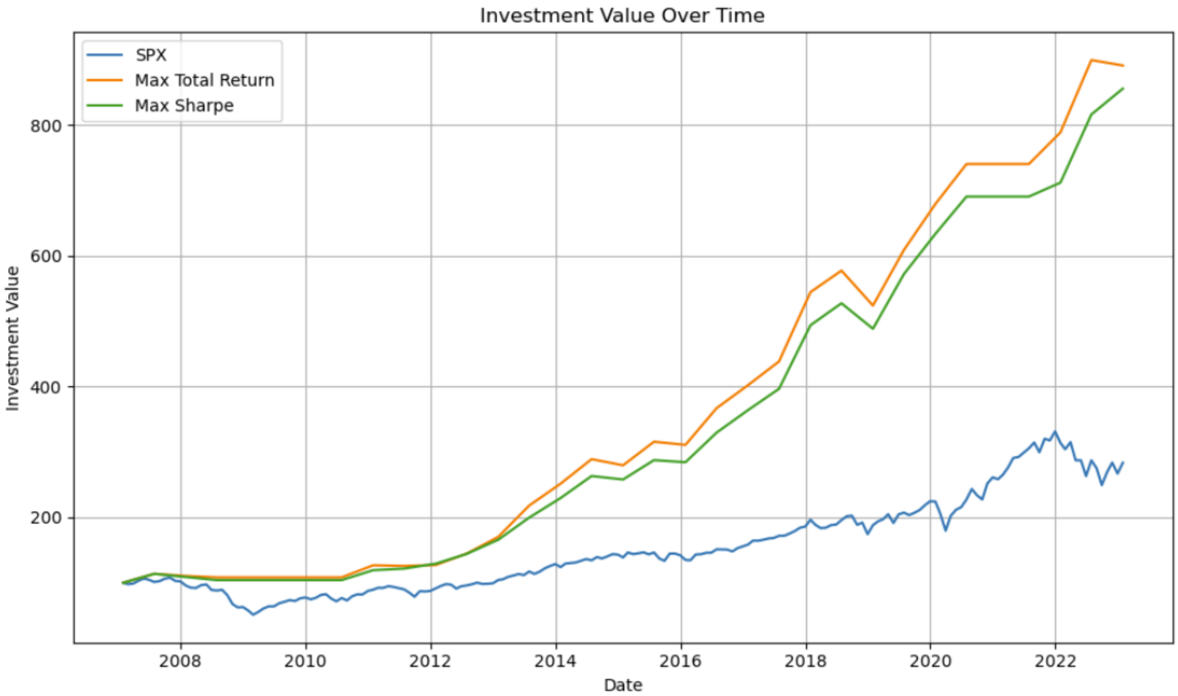
- Variables and weights: price earnings ratio (20%), Return on Capital Greenblatt (30%), and semi-annual total return (20%).
- Portfolio size: 15 companies.
- Rebalancing period: semi-annual.

Moreover, the strategy which achieved the highest Sharpe ratio obtained 856% total return and 1.28 Sharpe ratio. Interestingly, the only change compared to the maximum total return strategy is the weights given to each factor, changing 10% of the price earnings weight to Piotroski’s F-score. The strategy had the following parameters:

- Variables and weights: price-earnings (10%), Return on Capital Greenblatt (30%), Semi-annual total return (50%), and Piotroski F-score (10%).
- Portfolio size: 15 companies.
- Rebalancing period: semi-annual.

In the same time period, S&P 500 investors would have gotten a total return of 292% and a Sharpe ratio of 0.5, meaning that these strategies greatly outperform their benchmark.

Exhibit 5: Top total return and Sharpe ratio maximizing strategies vs. S&P 500.



This strong outperformance can be explained by the overfitting achieved by these strategies to this specific period, and although this is not the case for the majority of the tested strategies, overall results are positive: 25% of the strategies were able to achieve over 380% total return, and c.75% managed to beat the benchmark both in Sharpe ratio and total return.

4.2. Optimal parameters

Although interesting conclusions can be drawn from the top-performing strategies, having such a large list of strategies and their performance from 2007 to 2023 provides the possibility to analyse the individual effectiveness of each of the parameters.

Exhibit 6: Weight distribution in the strategies ordered by total return.

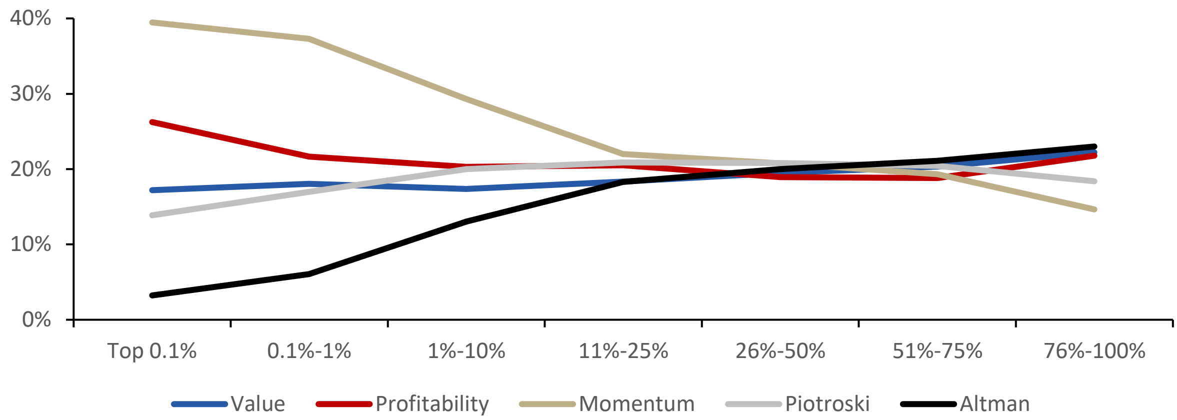
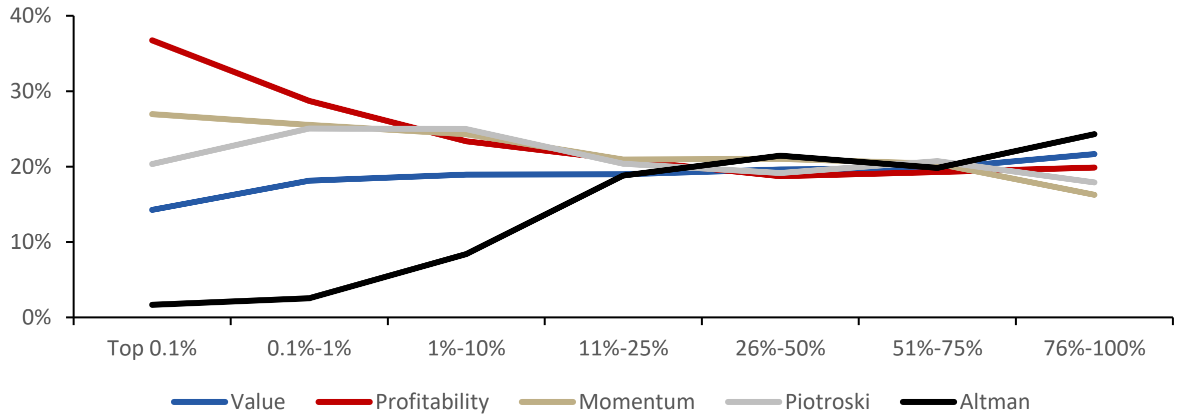
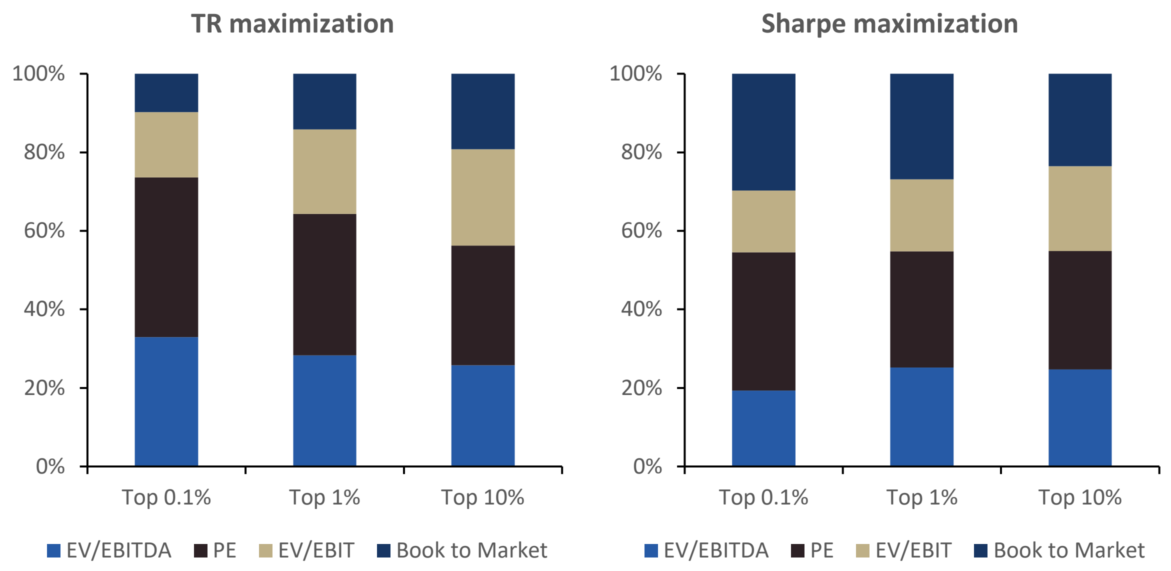


Exhibit 7: Weight distribution in the strategies ordered by Sharpe ratio.



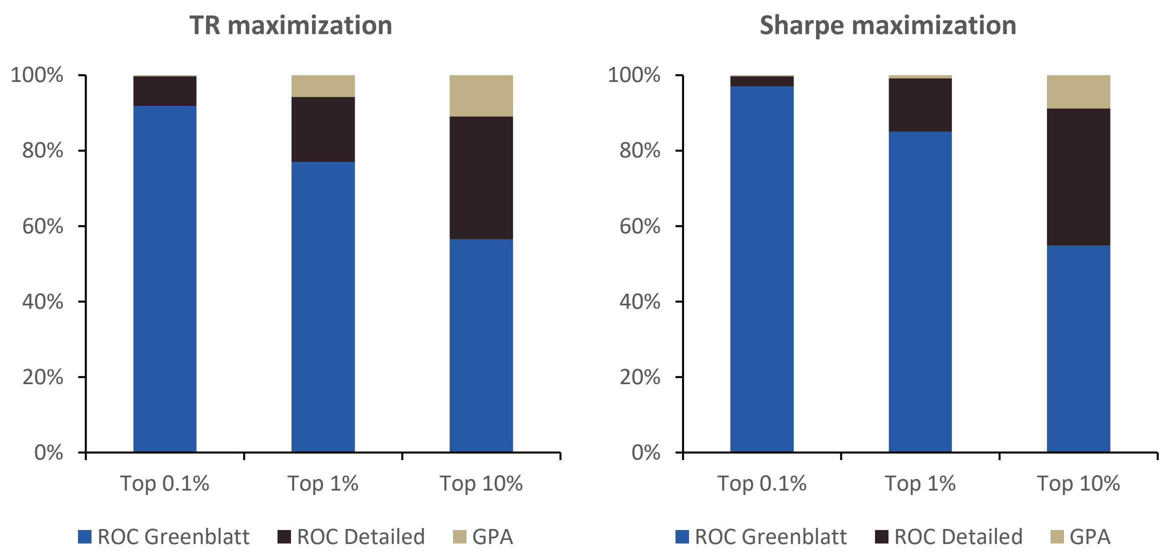
In Exhibit 6, it can be appreciated that momentum is given the highest importance in the 10% top-performing strategies in terms of total return, followed by profitability, value metrics, Piotroski F-score, and very little weight given to Altman Z-score. This implies that momentum metrics are by far the best predictors of future stock price performance, and, on the other hand, the guru metrics developed by Joseph Piotroski and Edward I. Altman are not as good predictors when used in combination with value, profitability, and momentum. However, when seeking to reduce volatility of the returns and maximizing the Sharpe ratio, the profitability metrics turn out to be much more effective. This implies that profitable companies are more prone to have a stable stock performance than companies whose valuation might be more based on speculation and future expectations than current profit-making capacity. Momentum metrics are also a good indicator for volatility strategies, and Piotroski F-score seems to perform better for these kinds of investors too. However, the Altman Z-score does not perform well in either scenario of high-return- or low-volatility-seeking strategies.

Exhibit 8: Value metrics used in the top strategies.



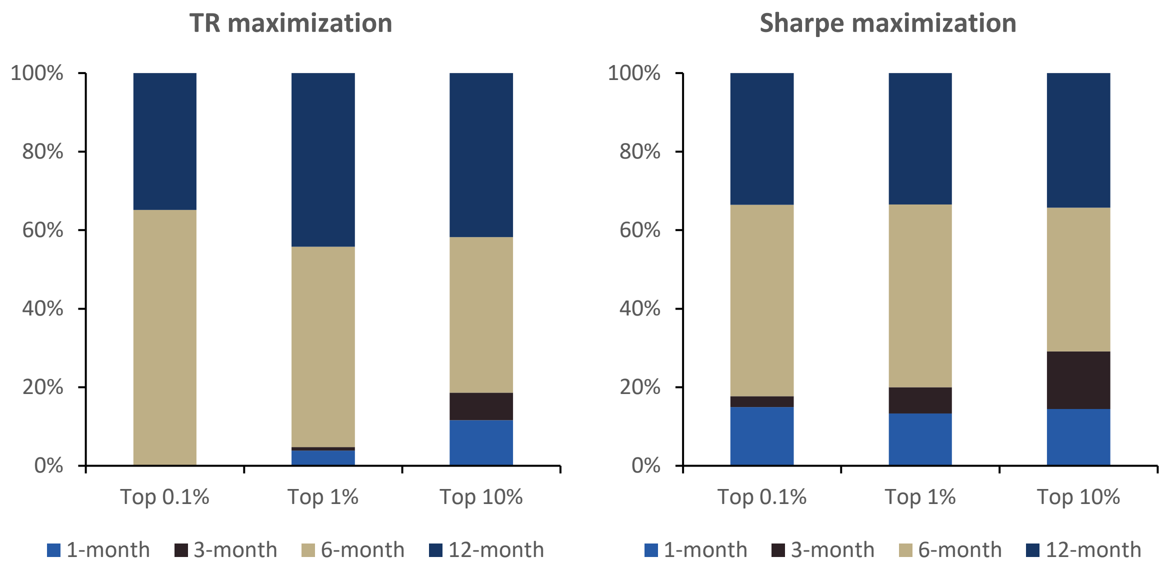
Individually, as shown in Exhibit 8, the value metrics which better predict price-increasing trends are EV/EBITDA and price earnings. Interestingly, the book-to-market ratio is used in less than 10% of the top 0.1% total return strategies, but it is, together with the price earnings ratio, the best identifier of less volatile yet profitable strategies. This can be explained by the ratio's inherent ability to detect heavily mispriced stocks which are subject to sudden stock price changes (Aras and Kemal Yilmaz, 2008). The price earnings ratio clearly stands out as the best value metric due to the strong correlation between share price and financial results.

Exhibit 9: Profitability metrics used in the top strategies.



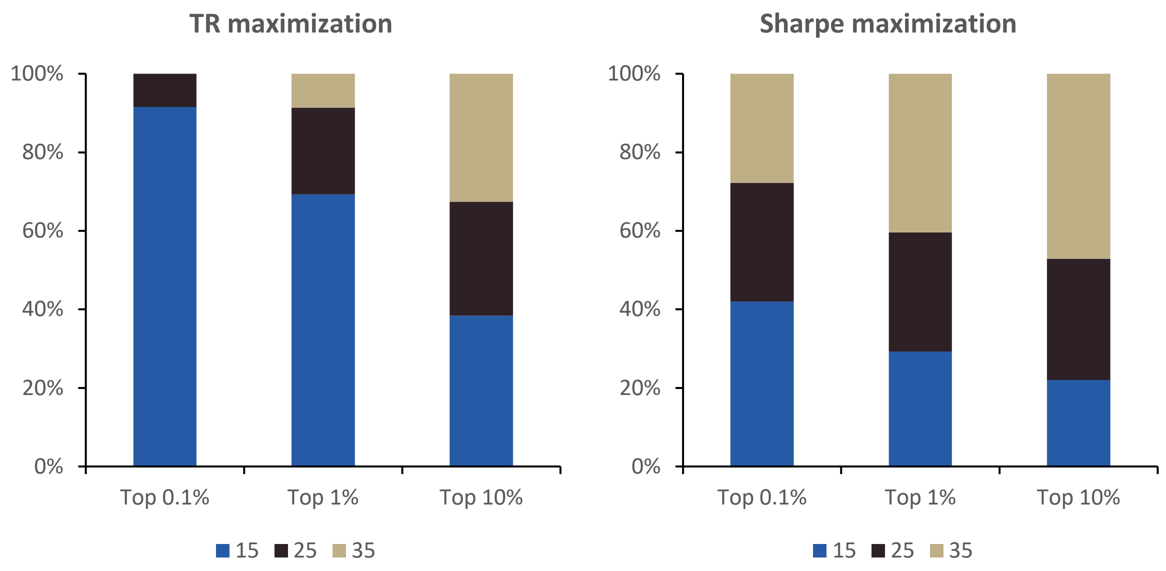
As Greenblatt (2010) explains, the Return on Capital is the best profitability metric for predicting share price performance and stable returns. The ratio depicts a company's ability to turn capital investments into financial returns, and therefore a better Return on Capital implies a better ability to manage resources, optimized production processes, and intrinsic competitive advantages. Moreover, Exhibit 9 clearly shows that Greenblatt's approach to Return on Capital calculation, which excludes intangible assets, is superior to the Detailed version of Return on Capital ratio which includes them, likely due to the questionable valuations awarded to the intangible assets (e.g., patents and intellectual property). Nevertheless, even including intangibles, Return on Capital is very much preferred by the best performing strategies over the Gross Profit on Assets ratio.

Exhibit 10: Momentum metrics used in the top strategies.



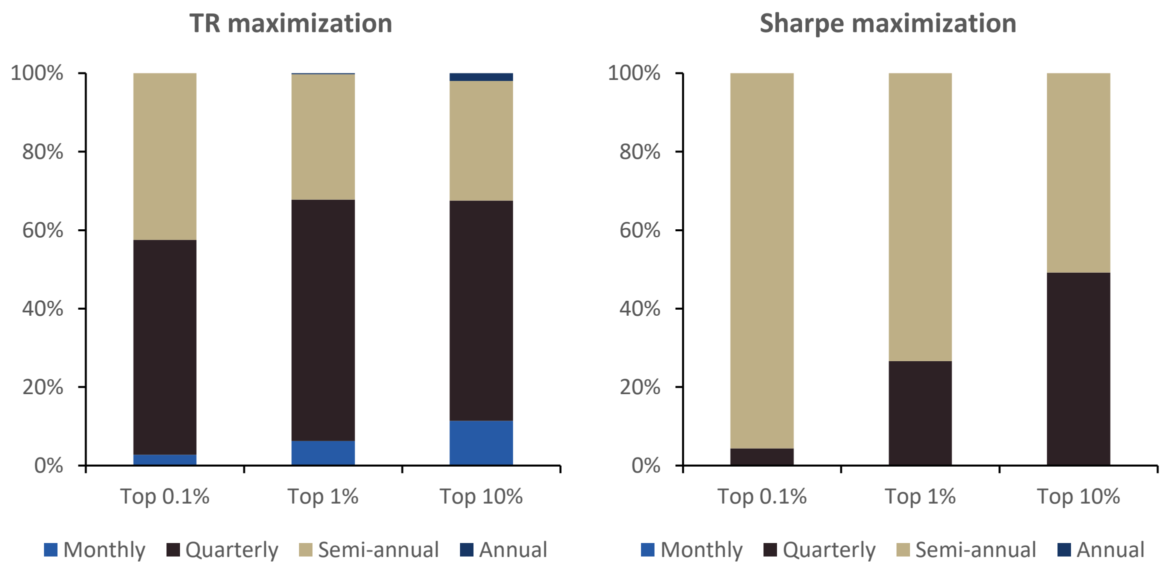
Regarding the appropriate length of momentum metrics, exhibit 10 depicts how longer metrics (6-month and 1-year total return vs. 1-month and 3-month) are very good predictors of stock price performance and stability, while shorter versions of the total return ratio are not even used in the top 0.1% total return strategies. This phenomenon can be explained by longer metrics being able to capture sustained price trends and being less exposed to short-term noise in share price changes (Sapp, 2010).

<u>Exhibit 11: Portfolio size used in the top strategies.</u>



Portfolio size is the variable which shows the biggest difference between total return seeking strategies and Sharpe ratio strategies. The number of companies included in the portfolio is effectively the level of diversification and, inevitably, a more diversified portfolio is more stable but has less profit generation potential. As such, most top total return strategies only use 15-stock portfolios while Sharpe ratio strategies are also inclined to use 25 and 35-stock portfolios.

<u>Exhibit 12: Rebalancing period used in the top strategies.</u>



In line with existing literature, quarterly and semi-annual rebalancing periods show much better results than monthly and annual periods. Firstly, while the algorithm is capable of identifying outperforming stocks within the S&P 500 index, a monthly rebalancing schedule does not allow for the selected stocks to capitalize on their potential and increase share price; and secondly, an annual schedule is too long for selected companies to remain as competitive as they were in their selection date, and the strategy is more exposed to changing macroeconomic conditions and can't sell its positions until a whole year has passed even though the market might have turned bearish.

4.3. Evaluation of the Sahm Rule

The tested strategies are heavily influenced by the Sahm Rule trigger which aims to provide a loss-prevention mechanism by accurately predicting economic crises. A comparative test has been carried out to calculate what would have been each of the tested strategies' results if the Sahm Rule had not been part of the algorithm. Contrary to the expectations, total return and Sharpe ratio would have been an average 41% and 6% higher respectively.

These negative results can be attributed to the investment strategies' ability to generate positive results also during periods of technical economic recession.

<u>Exhibit 13: Comparison of the same strategy with and without using the Sahm Rule</u>

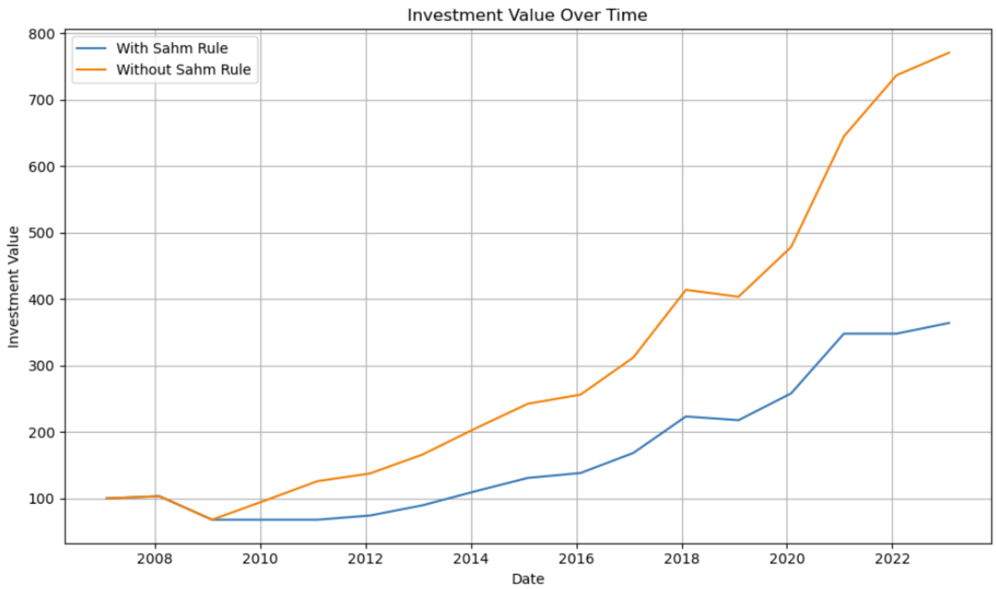


Exhibit 13 shows an example of the performance of the same strategy with and without using the Sahm Rule. It can be appreciated how during the two periods when the trigger was activated the strategy that kept investing still managed to generate positive returns while the other one maintained its investment value until the macroeconomic conditions improved. Over time, the opportunity cost of not investing in those two periods is much higher than the individual returns that would have been produced in that time due to the exponential nature of returns.

5. Limitations and areas for improvement

The Python code generated for the backtesting algorithm has several limitations regarding further parameter optimization and optimal replicability of a real-life investing scenario.

Firstly, further adjustment to SEC mandatory reporting dates could be achieved to invest in the announcement date as soon as the data is available. In the proposed strategy, the rebalancing date is delayed to the end of the publication month for homogeneity purposes throughout the year, as the end-of year report is published at the end of February. Since quarterly reports are published the 10th day of the month after quarter-end, for quarterly reports the rebalancing date could better be set at the 10th and initial stock price adjustments after results announcements would be captured.

Secondly, further optimization could be carried out in the rebalancing methodology. Some authors such as Jung and Kim (2017) recommend using dynamic rebalancing triggers which set off the portfolio rebalancing process whenever a significant change in the stock market or any of the portfolio companies is identified. This approach enables the investor to adapt to changing market or macroeconomic conditions as soon as they are detected rather than whenever the upcoming rebalancing date is set. Ultimately, a dynamic rebalancing schedule should be able to yield higher returns by capturing stock price increases and avoiding decreases earlier.

Moreover, the rebalancing approach can also be improved by allowing for different weight distribution across the portfolio. Although significant literature suggests that value weighted portfolios (by market capitalization) offer lower returns than equally weighted equivalents (Chen, Chen and Li, 2012), there is also evidence that in specific circumstances they could deliver better performance. For instance, Hillion (2019) explains that equally weighted portfolios tend to perform better because they allocate a larger percentage of capital to smaller companies which have more volatile stock prices and therefore in certain market conditions outperform larger companies, but that is not always the case.

Lastly, the programmed strategies assume that whenever the Sahm Rule trigger is activated no investment is made and the overall investment value remains constant until the alarm turns off and allows for a new portfolio rebalancing. However, a better approach could be taken by investing in fixed income securities, such as government bonds, whenever stock investing is vetted. This would grant the investor a moderate but stable return in periods of economic crisis and ultimately would improve both the total return and Sharpe ratio.

6. Conclusion

After thorough analysis of the results, it can be concluded that the initial objective of creating an investment algorithm capable of selecting a diversified portfolio of stocks from a vast universe of companies and consistently beating the benchmark in total return and Sharpe ratio has been achieved.

Results show that most of the tested strategies for the period ranging from January 2007 to January 2023 are well capable of beating the performance of the S&P 500, which is the most followed stock index in the world and therefore is a good comparable for any new investing strategy. Moreover, good conclusions can be drawn from the predictive ability of each of the parameters that have been included in the strategies:

- Momentum is the best predictor of future stock price performance; however, profitability is the characteristic that shows the most correlation with share price stability.
- The price earnings ratio has proven to be the best and most polyvalent value metric, and the book to market ratio is quite effective for stability-seeing strategies but is not as correlated with increasing share prices.
- Greenblatt's approach to Return on Capital calculation is by far the best profitability factor.
- Altman's Z-score does not significantly improve strategy performance when used in combination with value, profitability, and momentum metrics. The Piotroski F-score is also not a good predictor when seeking to maximize total return, but it adds value when maximizing the Sharpe ratio.
- Stocks showing sustained positive performance over 6 to 12 months are likely to maintain that price hike, whereas shorter momentum indicators such as 1-month total return are too affected by short-term noise.
- Rebalancing the portfolio once or twice per year is a more efficient approach than doing it monthly or quarterly. This approach allows for sufficient time for the selected stocks to grow within expectations.

Probably the most relevant conclusion is that using a macroeconomic indicator such as the Sahm Rule to predict eras of technical recession and trigger a stop signal to the investing algorithm is a safe approach to ensure that no losses are materialised during such periods; but since the designed strategies are mostly able to perform during crises, the Sahm Rule has a negative effect on overall total return.

7. Declaration of Use of Generative Artificial Intelligence

Por la presente, yo, Gabriel Medem López-Brea, estudiante de doble grado en Administración de Empresas y Business Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "Quantitative Filtering Techniques Applied to Diversified Portfolio Selection and Practical Implementation of the Sahm Rule", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Brainstorming de ideas de investigación:** Utilizado para idear y esbozar posibles áreas de investigación.
2. **Crítico:** Para encontrar contra-argumentos a una tesis específica que pretendo defender.
3. **Referencias:** Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
4. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
5. **Interpretador de código:** Para realizar análisis de datos preliminares.
6. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.
7. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
8. **Traductor:** Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 13 de abril de 2024

Firma: Gabriel Medem López-Brea

8. BIBLIOGRAPHY

Altman, E.I. (2000). Predicting Financial Distress of Companies: Revisiting the Z Score and ZETA Models.

Anyika, E. (2015). Non-Diversifiable Risk in Investment Portfolios - an Aid to Investment Decision Making. *Journal of Risk Analysis and Crisis Response*, [online] 5(1), p.31. doi:https://doi.org/10.2991/jrarc.2015.5.1.3.

Aras, G. and Kemal Yilmaz, M.K.Y. (2008). Price-Earnings Ratio, Dividend Yield, and Market-to-Book Ratio to Predict Return on Stock Market: Evidence From Emerging Markets. *Journal of Global Business and Technology*, [online] 4(1), pp.18–30. Available at: https://mkyilmaz.com/wp-content/uploads/2014/11/price-earnings-ratio.pdf [Accessed 13 Apr. 2024].

Bailey, D.H., Borwein, J.M., Salehipour, A., López de Prado, M. and Zhu, Q. (2016). Backtest overfitting in financial markets. [online] Available at: https://www.davidhbailey.com/dhbpapers/overfit-tools-at.pdf [Accessed 6 Apr. 2024].

Bermejo, R. (2021). *Factor Investing, ESG Metrics and Market Timing Rules*. Doctoral Thesis.

Bloomberg (2024) *S&P 500 Index Members*. Available at: Bloomberg Terminal (Accessed: January 2024).

Brown, J.H., Crocker, D.K. and Foerster, S.R. (2009). Trading Volume and Stock Investments. *Financial Analysts Journal*, [online] 65(2), pp.67–84. doi:https://doi.org/10.2469/faj.v65.n2.4.

Chadha, J., Corrado, L., Meaning, J. and Schuler, T. (2021). Monetary and Fiscal Complementarity in the Covid-19 Pandemic. *European Central Bank*, [online] 2588. doi:https://doi.org/10.2139/ssrn.3923209.

Chen, H. (Jason), Chen, S.J. and Li, F. (2012). Empirical Investigation of an Equity Pairs Trading Strategy. *SSRN Electronic Journal*. [online] doi:https://doi.org/10.2139/ssrn.1361293.

FactSet (2024a) *Monthly Financial Data of S&P 500 members*. Available at: www.FactSet.com [Accessed January 2024].

FactSet (2024b) *US Unemployment Rate*. Available at: www.FactSet.com [Accessed January 2024].

Falkenberry, T.N. (2006). Survivorship Bias in the Development of Equity Trading Systems. *The Journal of Trading*, 1(2), pp.62–68. doi:https://doi.org/10.3905/jot.2006.628196.

Fama, E.F. and French, K.R. (1992). The Cross Section of Expected Stock Returns. *The Journal of Finance*, [online] 47(2), pp.427–465. Available at: https://www.ivey.uwo.ca/media/3775518/the_cross-section_of_expected_stock_returns.pdf [Accessed 7 Apr. 2024].

Federal Reserve (2018). *Monetary Policy Report*. [online] Available at: https://www.federalreserve.gov/monetarypolicy/files/20180713_mprfullreport.pdf [Accessed 6 Apr. 2024].

Federal Reserve (2023). *Monetary Policy Report*. [online] Available at: https://www.federalreserve.gov/monetarypolicy/2023-06-mpr-summary.htm [Accessed 6 Apr. 2024].

Foster, L. (2024). Most Active Large-Cap Funds Couldn’t Beat the Market in 2023. *Barron’s*. [online] 6 Mar. Available at: https://www.barrons.com/articles/stock-market-active-mutual-large-cap-funds-54ce55fc.

Greenblatt, J. (2010). *The little book that still beats the market*. Hoboken, N.J.: Wiley, Chichester.

Hillion, P. (2019). The Specialness of Equally-Weighted Portfolios: The Rebalancing Fallacy. INSEAD Working Paper. [online] doi:https://doi.org/10.2139/ssrn.3466059.

Jegadeesh, N. and Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, [online] 48(1), pp.65–91. doi:https://doi.org/10.2307/2328882.

Jung, J. and Kim, S. (2017). Developing a dynamic portfolio selection model with a self-adjusted rebalancing method. Journal of the Operational Research Society, [online] 68(7), pp.766–779. doi:https://doi.org/10.1057/jors.2016.21.

Mokkelbost, P.B. (1971). Unsystematic Risk over Time. *The Journal of Financial and Quantitative Analysis*, [online] 6(2), p.785. doi:https://doi.org/10.2307/2329714.

Nasdaq (2022). What Are the Three Most Quoted U.S. Stock Indexes? [online] 25 Jan. Available at: https://www.nasdaq.com/articles/what-are-the-three-most-quoted-u.s.-stock-indexes [Accessed 6 Apr. 2024].

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, [online] 108(1), pp.1–28. Available at:

https://www.sciencedirect.com/science/article/abs/pii/S0304405X13000044?via%3Dihub [Accessed 7 Apr. 2024].

Novy-Marx, R. and Velikov, M. (2015). A Taxonomy of Anomalies and Their Trading Costs. *Review of Financial Studies*, [online] 29(1), pp.104–147. doi:https://doi.org/10.1093/rfs/hhv063.

Piotroski, J.D. (2000). Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers. *Journal of Accounting Research*, 38, pp.1–41.

Roy, A. (2016). A Model of Security Selection and Portfolio Building through Z-scores. *Global Business Review*, [online] 17(2), pp.389–399. doi:https://doi.org/10.1177/0972150915619815.

Sahm, C. (2019). *Direct Stimulus Payments to Individuals*. [online] Available at: https://www.brookings.edu/articles/direct-stimulus-payments-to-individuals/#:~:text=The%20Proposal&text=Sahm's%20proposal%20would%20offer%20lump,in%20the%20previous%2012%20months. [Accessed 6 Apr. 2024].

Sapp, T.R.A. (2010). The 52-week high, momentum, and predicting mutual fund returns. *Review of Quantitative Finance and Accounting*, [online] 37(2), pp.149–179. doi:https://doi.org/10.1007/s11156-010-0199-7.

Sharpe, W.F. (1994). The Sharpe Ratio. *The Journal of Portfolio Management*, [online] 21(1), pp.49–58. Available at: https://www.pm-research.com/content/iijpormgmt/21/1/49 [Accessed 7 Apr. 2024].

Standard & Poor's (2023). *S&P Dow Jones Indices Announces Update to S&P Composite 1500 Market Cap Guidelines and Results of S&P Composite 1500 Index Consultation on Market Capitalization and Liquidity Eligibility Criteria*. [online] Available at: https://press.spglobal.com/2023-01-04-S-P-Dow-Jones-Indices-Announces-Update-to-S-P-Composite-1500-Market-Cap-Guidelines-and-Results-of-S-P-Composite-1500-Index-Consultation-on-Market-Capitalization-and-Liquidity-Eligibility-Criteria [Accessed 6 Apr. 2024].

Standard & Poor's (n.d.). *The S&P 500 and The Dow | S&P Dow Jones Indices*. [online] www.spglobal.com. Available at: https://www.spglobal.com/spdji/en/research-insights/index-literacy/the-sp-500-and-the-dow/ [Accessed 31 Mar. 2024].

State Street Global Advisors (2024). *SPDR S&P 500 ETF Trust.* Available at: https://www.ssga.com/us/en/intermediary/etfs/funds/spdr-sp-500-etf-trust-spy [Accessed April 2024].

Suyash, M. and Anuranjan, M. (2017). Structured and Unstructured Big Data Analytics. *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*, [online] pp.740–746. doi:https://doi.org/10.1109/CTCEEC.2017.8454999.

Thomas, L.B. (2011). The Great Crisis and Great Recession of 2007-2009. *The Financial Crisis and Federal Reserve Policy*, [online] pp.93–112. doi:https://doi.org/10.1057/9780230118072_6.

Tikkanen, J. and Äijö, J. (2018). Does the F-score improve the performance of different value investment strategies in Europe? *Journal of Asset Management*, 19(7), pp.495–506. doi:https://doi.org/10.1057/s41260-018-0098-3.

Tinta, A. (2015). Financial Risk Evaluation in the Context of International Accounting Normalization. [online] Available at: https://ideas.repec.org/a/rau/journl/v9y2015i1p103-112.html [Accessed 7 Apr. 2024].

Tokat, Y. and Wicas, N.W. (2007). Portfolio Rebalancing in Theory and Practice. *The Journal of Investing*, [online] 16(2), pp.52–59. doi:https://doi.org/10.3905/joi.2007.686411.

United States Securities and Exchange Commission (2002). *Acceleration of Periodic Report Filing Dates and Disclosure Concerning Web Site Access to Reports*. [online] Available at: https://www.sec.gov/rules/2002/09/acceleration-periodic-report-filing-dates-and-disclosure-concerning-web-site-access [Accessed 6 Apr. 2024].

Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, [online] 1168(2), p.022022. doi:https://doi.org/10.1088/1742-6596/1168/2/022022.

Zaimovic, A., Omanovic, A. and Arnaut-Berilo, A. (2021). How Many Stocks Are Sufficient for Equity Portfolio Diversification? A Review of the Literature. *Journal of Risk and Financial Management*, [online] 14(11), p.551. doi:https://doi.org/10.3390/jrfm14110551.

9. APPENDIX

9.1. Appendix 1: Top 10 strategies in terms of total return

Rank	Combination	Weights Combination	Portfolio size	Rebalancing period	TR	Sharpe
1.	('PE', 'ROC GREENBLATT', 'SEMI TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.2, 0.3, 0.5, 0.0, 0.0)	15	6M	891%	1.17
2.	('PE', 'ROC GREENBLATT', 'SEMI TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.2, 0.3, 0.4, 0.1, 0.0)	15	6M	888%	1.25
3.	('PE', 'ROC GREENBLATT', 'SEMI TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.2, 0.2, 0.5, 0.1, 0.0)	15	6M	878%	1.25
4.	('PE', 'ROC GREENBLATT', 'SEMI TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.2, 0.4, 0.4, 0.0, 0.0)	15	6M	869%	1.13
5.	('PE', 'ROC GREENBLATT', 'SEMI TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.1, 0.3, 0.5, 0.1, 0.0)	15	6M	856%	1.28
6.	('PE', 'ROC GREENBLATT', 'SEMI TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.3, 0.2, 0.4, 0.1, 0.0)	15	6M	853%	1.24
7.	('PE', 'ROC GREENBLATT', 'SEMI TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.1, 0.4, 0.5, 0.0, 0.0)	15	3M	850%	1.02
8.	('EVEBITDA', 'ROC GREENBLATT', 'SEMI TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.5, 0.2, 0.3, 0.0, 0.0)	15	6M	843%	1.08
9.	('PE', 'ROC GREENBLATT', 'SEMI TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.3, 0.2, 0.4, 0.0, 0.1)	15	6M	840%	1.19
10.	('PE', 'ROC GREENBLATT', 'SEMI TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.2, 0.2, 0.5, 0.0, 0.1)	15	6M	821%	1.13

9.2. Appendix 2: Top 10 strategies in terms of Sharpe ratio

Rank	Combination	Weights Combination	Portfolio size	Rebalancing period	TR	Sharpe
1.	('PE', 'ROC GREENBLATT', 'SEMI TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.1, 0.3, 0.5, 0.1, 0.0)	15	6M	856%	1.28
2.	('PE', 'ROC GREENBLATT', 'SEMI TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.2, 0.2, 0.5, 0.1, 0.0)	15	6M	878%	1.25
3.	('MARKET TO BOOK', 'ROC GREENBLATT', 'SEMI TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.3, 0.5, 0.2, 0.0, 0.0)	35	6M	596%	1.25
4.	('PE', 'ROC GREENBLATT', 'SEMI TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.2, 0.3, 0.4, 0.1, 0.0)	15	6M	888%	1.25
5.	('PE', 'ROC GREENBLATT', 'SEMI TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.3, 0.5, 0.1, 0.1, 0.0)	15	6M	604%	1.25
6.	('EVEBITDA', 'ROC GREENBLATT', 'ANNUAL TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.1, 0.2, 0.2, 0.5, 0.0)	15	6M	691%	1.25
7.	('PE', 'ROC GREENBLATT', 'SEMI TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.3, 0.2, 0.4, 0.1, 0.0)	15	6M	853%	1.24
8.	('PE', 'ROC GREENBLATT', 'ANNUAL TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.3, 0.1, 0.2, 0.4, 0.0)	15	6M	722%	1.24
9.	('MARKET TO BOOK', 'ROC GREENBLATT', 'SEMI TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.3, 0.5, 0.2, 0.0, 0.0)	25	6M	567%	1.24
10.	('EVEBIT', 'ROC GREENBLATT', 'MONTHLY TR', 'PIOTROSKI', 'ALTMAN Z SCORE')	(0.0, 0.5, 0.2, 0.3, 0.0)	35	6M	483%	1.24

9.3. Appendix 3: S&P 500 constituents data structure for code parts 1 and 2 (Bloomberg, 2024).

31/01/2007			28/02/2007		
ID			ID	ID_ISIN	NAME
0111145D UN Equity	US6540861076	Nicor Inc	0111145D UN Equity	US6540861076	Nicor Inc
0202445Q UN Equity	US6010731098	EMD Millipore Corp	0202445Q UN Equity	US6010731098	EMD Millipore Corp
0203524D UN Equity	US6376401039	National Semiconductor Corp	0203524D UN Equity	US6376401039	National Semiconductor Corp
0226226D UN Equity	US1523121044	Centex LLC	0226226D UN Equity	US1523121044	Centex LLC
0574018D UN Equity	US0395831094	Archstone-Smith Trust	0574018D UN Equity	US0395831094	Archstone-Smith Trust
0772031D UN Equity	US7491211097	Qwest Communications Internati	0772031D UN Equity	US7491211097	Qwest Communications Internati
0848680D UN Equity	US8520611000	Sprint Communications LLC	0848680D UN Equity	US8520611000	Sprint Communications LLC
0910150D US Equity	US4328481092	Park Hotels & Resorts Inc/Old	0910150D US Equity	US4328481092	Park Hotels & Resorts Inc/Old
0961514D UN Equity	US7432631056	Progress Energy Inc	0961514D UN Equity	US7432631056	Progress Energy Inc
0964591D UW Equity	US0906131000	Biomet Inc	0964591D UW Equity	US0906131000	Biomet Inc
1086832D UN Equity	US86764P1093	ETC Sunoco Holdings LLC	1086832D UN Equity	US86764P1093	ETC Sunoco Holdings LLC
1255173D UN Equity	US9299031024	Wachovia Corp	1255173D UN Equity	US9299031024	Wachovia Corp
1280712D UW Equity	US6700081010	Novellus Systems Inc	1280712D UW Equity	US6700081010	Novellus Systems Inc
1284849D UN Equity	US0184901025	Allergan Inc/United States	1284849D UN Equity	US0184901025	Allergan Inc/United States
1293405D UN Equity	US9149061023	Univision Communications Inc	1293405D UN Equity	US9149061023	Univision Communications Inc

9.4. Appendix 4: Financial data structure for code part 2 (Factset, 2024a). Example of monthly total return data.

ID_ISIN	31/01/2005	28/02/2005	31/03/2005	30/04/2005	31/05/2005	30/06/2005	31/07/2005	31/08/2005	30/09/2005
US6540861076		1.00	0.74	-0.32	6.84	5.39	-0.85	1.45	2.64
US8520611000		-0.63	-3.41	-2.15	6.42	6.46	7.21	-3.61	-8.20
US53217V1098		1.82	-1.09	5.88	8.28	4.98	2.98	-1.21	-11.21
US0184901025		-0.88	-7.60	1.32	9.98	10.26	4.85	3.11	-0.47
US6117401017		7.02	36.97	-5.46	31.51	13.66	9.07	7.47	-5.18
US4783661071		-0.10	-5.24	-1.60	3.26	-0.14	1.97	4.42	3.87
US2605431038		10.97	-8.99	-7.86	-1.39	-0.94	7.68	-9.91	-2.76
BMG982941046		0.29	-2.85	-2.86	7.08	-0.48	-3.49	-3.23	-1.40
US74005P1049		3.89	7.18	-2.15	0.09	-0.20	5.99	-2.21	-0.39
US5024131071		1.14	-1.50	-0.07	-0.08	8.19	2.15	4.83	-3.43

9.5. Appendix 5: Dataset containing each company's sector for use in code part 2 (Factset, 2024a).

ID_ISIN	SECTOR	NAME
US6540861076	Utilities	Nicor Inc
US8520611000	Communications	Sprint Nextel Corporation
US53217V1098	Commercial Services	Life Technologies Corporation
US0184901025	Health Technology	Allergan, Inc.
US6117401017	Consumer Non-Durables	Monster Beverage Corporation
US4783661071	Producer Manufacturing	Johnson Controls, Inc.
US2605431038	Process Industries	DuPont de Nemours, Inc.
BMG982941046	Finance	XL Group Ltd
US74005P1049	Process Industries	Linde plc
US5024131071	Electronic Technology	L3 Technologies Inc
US8865471085	Consumer Durables	Tiffany & Co.
US4698141078	Technology Services	Jacobs Solutions Inc.
US00846U1016	Health Technology	Agilent Technologies, Inc.
US6010731098	Commercial Services	Millipore Corp

9.6. Appendix 6: US unemployment data (Feb-2004 to Nov-2025 extract) (Factset, 2024b).

27/02/2004	5.6%
31/03/2004	5.8%
30/04/2004	5.6%
31/05/2004	5.6%
30/06/2004	5.6%
31/07/2004	5.5%
31/08/2004	5.4%
30/09/2004	5.4%
31/10/2004	5.5%
30/11/2004	5.4%
31/12/2004	5.4%
31/01/2005	5.3%
28/02/2005	5.4%
31/03/2005	5.2%
30/04/2005	5.2%
31/05/2005	5.1%
30/06/2005	5.0%
31/07/2005	5.0%
31/08/2005	4.9%
30/09/2005	5.0%
31/10/2005	5.0%
30/11/2005	5.0%

9.7. Appendix 7: Python code.

1. Aggregate list of all S&P 500 constituents at any time of the investment period.

```
import pandas as pd

# File path for the Excel file containing Bloomberg dataset with S&P 500 constituents
file_path = 'C:/Users.../dataset.xlsx'

# Load the data, skipping the first three rows
data = pd.read_excel(file_path, skiprows=3)

# Filter the dataframe to keep only columns with the title 'ID_ISIN'
id_isin_columns = data.filter(like='ID_ISIN')

# Combine all the 'ID_ISIN' columns into a single column
all_id_isin = pd.Series(id_isin_columns.values.ravel())

# Remove any NaN values and duplicate values
unique_id_isin = all_id_isin.dropna().unique()

# Convert the unique ID_ISIN values to a DataFrame
unique_id_isin_df = pd.DataFrame(unique_id_isin, columns=['ID_ISIN'])

# Specify the path for the new Excel file
excel_file_path = 'C:/Users.../FINAL_LIST_SPX_MEMBERS.xlsx'

# Save the DataFrame to an Excel file
unique_id_isin_df.to_excel(excel_file_path, index=False)
```

2. Database management

```python
import pandas as pd

# Load the Excel file containing the financial data from Factset
file_path = 'C:/Users.../dataset.xlsx'

# Read all sheets into a dictionary of dataframes for better indexing effectiveness
xls = pd.ExcelFile(file_path)
sheets = xls.sheet_names

# Process each sheet and store in a list and include variable names
all_data = []
for sheet in sheets:
    df = pd.read_excel(file_path, sheet_name=sheet)
    melted_df = df.melt(id_vars=['ID_ISIN'], var_name='Date', value_name='Value')
    melted_df['Variable'] = sheet
    all_data.append(melted_df)

# Concatenate all dataframes along rows to create a single dataframe
concatenated_df = pd.concat(all_data, axis=0)

# Pivot the concatenated DataFrame to have financial variables as columns
final_df = concatenated_df.pivot(index=['ID_ISIN', 'Date'], columns='Variable', values='Value')

# Reset the column index to remove the multi-level index caused by pivoting
final_df.columns = final_df.columns.get_level_values(0)

# Import the dataset generated in code Part 1

file_path = 'C:/Users.../FINAL_LIST_SPX_MEMBERS.xlsx'

constituents_df = pd.read_excel(file_path)
constituents_list = constituents_df['ID_ISIN']

# Keep only the companies in the constituents dataframe just in case there are extra ones in the dataset

final_df = final_df[final_df.index.get_level_values('ID_ISIN').isin(constituents_list)]

# Exclude companies with no data

financial_variable_columns = final_df.columns.difference(['ID_ISIN', 'Date'])
filtered_df = final_df.groupby('ID_ISIN')[financial_variable_columns].sum().any(axis=1).reset_index()
deleted_companies = filtered_df.loc[filtered_df[0] == False, 'ID_ISIN'].tolist()
final_df = final_df[~final_df.index.get_level_values('ID_ISIN').isin(deleted_companies)]


#Exclude companies in the financial sector to avoid accounting heterogeneity using the dataset with the company sectors

second_file_path = 'C:/Users.../dataset.xlsx'
second_df = pd.read_excel(second_file_path)
finance_companies = second_df[second_df['SECTOR'] == 'Finance']['ID_ISIN'].tolist()
final_df = final_df[~final_df.index.get_level_values('ID_ISIN').isin(finance_companies)]

# CREATION OF NEW FINANCIAL VARIABLES FOR LATER USE IN THE INVESTMENT ALGORITHM

# Shift the 'PRICE PER SHARE' data by two months since we use data from two months before to adjust for SEC reporting date re
final_df['PRICE PER SHARE Shifted'] = final_df.groupby('ID_ISIN')['PRICE PER SHARE'].shift(-2)

# Calculate EPS
final_df['EPS'] = final_df['NET INCOME'] / final_df['SHARES OUTSTANDING']

# Recalculate 'BOOK TO MARKET' and 'PE' using shifted 'PRICE PER SHARE'
final_df['MARKET TO BOOK'] = final_df['PRICE PER SHARE Shifted'] / final_df['BOOK VALUE PER SHARE']
final_df['PE'] = final_df['PRICE PER SHARE Shifted'] / final_df['EPS']

# Define the other financial variables
final_df['EVEBIT'] = final_df['EV'] / final_df['EBIT']
final_df['EVEBITDA'] = final_df['EV'] / final_df['EBITDA']
final_df['GPA'] = final_df['GROSS INCOME'] / final_df['TOTAL ASSETS']
final_df['ROC GREENBLATT'] = final_df['EBIT'] / (final_df['WORKING CAPITAL'] + final_df['PPE'])
final_df['ROC DETAILED'] = final_df['EBIT'] / (final_df['WORKING CAPITAL'] + final_df['PPE'] + final_df['INTANGIBLES'])
```

```python
# CALCULATION OF PIOTROSKI F SCORE

final_df['PIOTROSKI'] = 0

# 1: Positive Net Income
final_df.loc[final_df['NET INCOME'] > 0, 'PIOTROSKI'] += 1

# 2: Positive ROA
final_df['ROA'] = final_df['NET INCOME'] / final_df['TOTAL ASSETS']
final_df.loc[final_df['ROA'] > 0, 'PIOTROSKI'] += 1

# 3: Positive Operating Cash Flow
final_df.loc[final_df['OPERATING CF'] > 0, 'PIOTROSKI'] += 1

# 4: Cash Flow from Operations > Net Income
final_df['CFO_TO_NET_INCOME'] = final_df['OPERATING CF'] / final_df['NET INCOME']
final_df.loc[final_df['CFO_TO_NET_INCOME'] > 1, 'PIOTROSKI'] += 1

# 5: Lower Long-Term Debt
final_df['LT DEBT_YEARLY_CHANGE'] = final_df.groupby('ID_ISIN')['LT DEBT'].diff(12)
final_df.loc[final_df['LT DEBT_YEARLY_CHANGE'] < 0, 'PIOTROSKI'] += 1
final_df.drop(columns=['LT DEBT_YEARLY_CHANGE'], inplace=True)

# 6: Higher Current Ratio
final_df['CURRENT_RATIO_YEARLY_CHANGE'] = final_df.groupby('ID_ISIN')['CURRENT RATIO'].diff(12)
final_df.loc[final_df['CURRENT_RATIO_YEARLY_CHANGE'] > 0, 'PIOTROSKI'] += 1
final_df.drop(columns=['CURRENT_RATIO_YEARLY_CHANGE'], inplace=True)

# 7: No New Shares Issued
final_df['SHARES_OUTSTANDING_YEARLY_CHANGE'] = final_df.groupby('ID_ISIN')['SHARES OUTSTANDING'].diff(12)
final_df.loc[final_df['SHARES_OUTSTANDING_YEARLY_CHANGE'] <= 0, 'PIOTROSKI'] += 1
final_df.drop(columns=['SHARES_OUTSTANDING_YEARLY_CHANGE'], inplace=True)

# 8: Higher Gross Margin
final_df['GROSS_MARGIN_YEARLY_CHANGE'] = final_df.groupby('ID_ISIN')['GROSS MARGIN'].diff(12)
final_df.loc[final_df['GROSS_MARGIN_YEARLY_CHANGE'] > 0, 'PIOTROSKI'] += 1
final_df.drop(columns=['GROSS_MARGIN_YEARLY_CHANGE'], inplace=True)

# 9: Higher Asset Turnover Ratio
final_df['ASSET_TURNOVER_YEARLY_CHANGE'] = final_df.groupby('ID_ISIN')['ASSET TURNOVER'].diff(12)
final_df.loc[final_df['ASSET_TURNOVER_YEARLY_CHANGE'] > 0, 'PIOTROSKI'] += 1
final_df.drop(columns=['ASSET_TURNOVER_YEARLY_CHANGE'], inplace=True)


# KEEPING ONLY THE NEW VARIABLES

variables_to_keep = [
    'PIOTROSKI', 'ALTMAN Z SCORE', 'MARKET TO BOOK', 'PE', 'EVEBIT', 'EVEBITDA',
    'GPA', 'ROC GREENBLATT', 'ROC DETAILED', 'MONTHLY TR', 'QUARTERLY TR',
    'SEMI TR', 'ANNUAL TR'
]

final_df = final_df[variables_to_keep]

# DOWNLOADING THE GENERATED DATASET

# Specify the path for the new dataset to be later used in the investment algorithm
excel_file_path = 'C:/Users/.../MANAGED_DATAFRAME.xlsx'

# Save the DataFrame to an Excel file
final_df.to_excel(excel_file_path, index=True)
```

3. Dataset with monthly index constituents

```python
# IMPORT SAME DATASET AS IN PART 1

# File path for the Excel file containing Bloomberg dataset with S&P 500 constituents
file_path = 'C:/Users.../dataset.xlsx'

# Load the data, skipping the first three rows
data = pd.read_excel(file_path, skiprows=3)

# IMPORT DATASET GENERATED IN PART 2

managed_df_path = 'C:/.../MANAGED_DATAFRAME.xlsx'
managed_df = pd.read_excel(managed_df_path)

# Extract unique identifiers of the companies in the managed dataframe
unique_ids = managed_df['ID_ISIN'].dropna().unique()

# CREATE MONTHLY LIST OF S&P 500 MEMBERS WITH THE INFORMATION IN THE TWO DATASETS
# We need the MANAGED_DATAFRAME from part 2 to eliminate from the members list the financial companies and other errors

# Clean data and keep only the used companies
for column in data.columns:
    filtered_column = data[column][data[column].isin(unique_ids)]
    compacted_column = filtered_column.dropna().reset_index(drop=True)
    data[column] = compacted_column

# Create new dataframe with the monthly list of members
id_isin_columns = data.filter(like='ID_ISIN')
date_range = pd.date_range(start='2007-01-31', periods=len(id_isin_columns.columns), freq='M')
id_isin_columns.columns = date_range.strftime('%Y-%m-%d')

# Save the DataFrame to an Excel file
excel_file_path = 'C:/.../FINAL_LIST_CONSTITUENTSBYMONTH.xlsx'
id_isin_columns.to_excel(excel_file_path, index=False)
```

4. Investment algorithm

```python
# LOAD THE FINANCIAL DATA IN MANAGED_DATAFRAME CREATED IN PART 2

import pandas as pd
final_df = pd.read_excel('C:/.../MANAGED_DATAFRAME.xlsx', index_col=[0, 1], engine='openpyxl')

# LOAD DATA OF MONTHLY S&P 500 MEMBERS FROM THE DATASET CREATED IN PART 3

file_path = 'C:/.../FINAL_LIST_CONSTITUENTSBYMONTH.xlsx'
constituents_df = pd.read_excel(file_path, header=None)

# Transpose the DataFrame so that dates become the index
constituents_df = constituents_df.transpose()

# Convert the first column to datetime and set it as index
constituents_df.index = pd.to_datetime(constituents_df.iloc[:, 0], format='%Y-%m-%d')
constituents_df = constituents_df.drop(columns=0)

# GENERATE THE DIFFERENT COMBINATIONS OF VARIABLES

from itertools import product
value_metrics = ['MARKET TO BOOK', 'PE', 'EVEBIT', 'EVEBITDA']
profitability_metrics = ['GPA', 'ROC GREENBLATT', 'ROC DETAILED']
momentum_metrics = ['MONTHLY TR', 'QUARTERLY TR', 'SEMI TR', 'ANNUAL TR']
piotroski = ['PIOTROSKI']
altman = ['ALTMAN Z SCORE']

simple_combinations = list(product(value_metrics, profitability_metrics, momentum_metrics, piotroski, altman))

# GENERATE THE DIFFERENT VARIABLE WEIGHT COMBINATIONS

import numpy as np

step = 0.1
weights = np.arange(0, 1 + step, step)
combinations = []

for w1 in weights:
    for w2 in weights:
        for w3 in weights:
            for w4 in weights:
                for w5 in weights:
                    if np.isclose(w1 + w2 + w3 + w4 + w5, 1.0):
                        combination = tuple(round(w, 2) for w in [w1, w2, w3, w4, w5])
                        combinations.append(combination)

filtered_combinations = [combo for combo in combinations if all(0.0 <= weight <= 0.5 for weight in combo)]
```

```python
# IMPORT DATA FOR SAHM RULE

file_path = 'C:/.../US unemployment.xlsx'
sahm_data = pd.read_excel(file_path)

# Correct the dataframe by adding column headers
sahm_data.columns = ['Date', 'Unemployment Rate']
sahm_data['Date'] = pd.to_datetime(sahm_data['Date'])

# Calculate the three-month moving average of the unemployment rate
sahm_data['3M Moving Average'] = sahm_data['Unemployment Rate'].rolling(window=3).mean()

# Calculate the lowest unemployment rate over the previous 12 months
sahm_data['12M Min'] = sahm_data['Unemployment Rate'].rolling(window=12).min()

# Calculate the difference between the 3M Moving Average and the 12M Min
sahm_data['Difference'] = sahm_data['3M Moving Average'] - sahm_data['12M Min']

# Identify Sahm Rule triggers
sahm_data['Sahm Trigger'] = (sahm_data['Difference'] >= 0.005).astype(int)

sahm_data.set_index('Date', inplace=True)

## INVESTMENT CODE

rebalance_periods = ['M', '3M','6M', '12M']

return_calc_periods = {
    'M':'MONTHLY TR',
    '3M': 'QUARTERLY TR',
    '6M': 'SEMI TR',
    '12M': 'ANNUAL TR'
}

import time
start_time = time.time()
results = []


for combo in simple_combinations:
    value_metric, profitability_metric, momentum_metric, piotroski, altman = combo

    for weights_combo in filtered_combinations:
        value_weight, profitability_weight, momentum_weight, piotroski_weight, altman_weight = weights_combo

        weights = {
            'value_weight': value_weight,
            'profitability_weight': profitability_weight,
            'momentum_weight': momentum_weight,
            'piotroski_weight': piotroski_weight,
            'altman_weight': altman_weight
        }

        for num_companies in range(15,36,10):

            for rebalance_period in rebalance_periods:

                # Track the investment value each month
                investment_values = []
                # Prepare to track investment over time
                investment_timeline = []

                # Define the date range for the backtest
                start_date = pd.Timestamp('2007-01-31')
                end_date = pd.Timestamp('2023-01-31')
                date_range = pd.date_range(start=start_date, end=end_date, freq=rebalance_period)

                investment_value=100
                investment_timeline = []
                returns = []

                for current_date in date_range:
```

```python
                for current_date in date_range:

                    two_months_prior = current_date - pd.DateOffset(months=2)
                    one_month_prior = current_date - pd.DateOffset(months=1)
                    one_month_prior = pd.Timestamp(one_month_prior.year, one_month_prior.month, 1) + pd.offsets.MonthEnd(0)

                    # Adjust the date to be the end of the month for the two-months-prior date
                    selection_date = pd.Timestamp(two_months_prior.year, two_months_prior.month, 1) + pd.offsets.MonthEnd(0)

                    if current_date != start_date:
                        previous_sahm_trigger = sahm_trigger
                        sahm_trigger = sahm_data['Sahm Trigger'].get(one_month_prior, default=0)
                    else:
                        previous_sahm_trigger = sahm_data['Sahm Trigger'].get(one_month_prior, default=0)
                        sahm_trigger = previous_sahm_trigger

                    # Get the list of companies that are in the index at the current date
                    companies_in_index = constituents_df.loc[current_date].dropna().values

                    # Select data for the selection date (two months before the current date)
                    monthly_data = final_df.loc[(slice(None), selection_date), :].reset_index(level='Date', drop=True)

                    # Select data for the CURRENT date (for the total returns)
                    current_monthly_data = final_df.loc[(slice(None), current_date), :].reset_index(level='Date', drop=True)

                    # Filter monthly_data to only include companies in the index
                    monthly_data = monthly_data[monthly_data.index.get_level_values('ID_ISIN').isin(companies_in_index)]
                    current_monthly_data = current_monthly_data[current_monthly_data.index.get_level_values('ID_ISIN').isin(c

                    if current_date!=start_date and previous_sahm_trigger == 0:
                        investment_returns = (final_df.loc[(top_companies, current_date), return_calc_periods[rebalance_peric
                        investment_value = investment_value * (1 + investment_returns)
                    else:
                        investment_returns = 0

                     # Append the investment value for this month
                    investment_timeline.append((current_date, investment_value))
                    returns.append(investment_returns)

                    # Calculate scores
                    value_std = monthly_data[value_metric].sub(monthly_data[value_metric].mean()).div(monthly_data[value_metr
                    profitability_std = monthly_data[profitability_metric].sub(monthly_data[profitability_metric].mean()).div
                    tr_std = current_monthly_data[momentum_metric].sub(current_monthly_data[momentum_metric].mean()).div(curr
                    piotroski_std = monthly_data[piotroski].sub(monthly_data[piotroski].mean()).div(monthly_data[piotroski].s
                    altman_std = monthly_data[altman].sub(monthly_data[altman].mean()).div(monthly_data[altman].std())

                     # Calculate the aggregate score with weights
                    aggregate_score = (weights['value_weight'] * value_std +
                                       weights['profitability_weight'] * profitability_std +
                                       weights['momentum_weight'] * tr_std +
                                       weights['piotroski_weight'] * piotroski_std +
                                       weights['altman_weight'] * altman_std)

                    # Select the top companies based on the aggregate score
                    top_companies = aggregate_score.nlargest(num_companies).index.get_level_values('ID_ISIN')

                # Calculate sharpe ratio and store results
                sharpe = np.average(returns[1:])/np.std(returns[1:])
                results.append((combo, weights_combo, num_companies, rebalance_period, investment_value, sharpe))

# Convert results to a DataFrame
results_df = pd.DataFrame(results, columns=['Combination',
                                            'Weights Combination',
                                            'Number of companies',
                                            'Rebalancing period',
                                            'Final Investment Value',
                                            'Sharpe'])

# Download the output
excel_file_path = 'C:/.../FINAL_OUTPUT.xlsx'
sorted_results_df = results_df.sort_values(by='Final Investment Value', ascending=False)
sorted_results_df.to_excel(excel_file_path, index=False)
```