UNIVERSITY OF MACEDONIA

MASTER THESIS

A simple stock screener framework for portfolio optimization

Author:
Ioannis BANATAS

Supervisor:
Dr. Dimitris HRISTU - VARSAKELIS

A thesis submitted in fulfillment of the requirements for the degree of Maters Thesis

in the

Artificial Intelligence & Data Analytics
Department of Applied Informatics

September 4, 2022

Declaration of Authorship

I, Ioannis BANATAS, declare that this thesis titled, "A simple stock screener framework for portfolio optimization" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"Nothing has such power to broaden the mind as the ability to investigate systematically and truly all that comes under thy observation in life."

Marcus Aurelius

UNIVERSITY OF MACEDONIA

Abstract

MSc in Artificial Intelligence & Data Analytics
Department of Applied Informatics

Maters Thesis

A simple stock screener framework for portfolio optimization

by Ioannis BANATAS

This thesis proposes a stock portfolio optimization method that is simple, scalable, and efficient compared to other proposed strategies from the literature, while significantly outperforming the market. We discuss the survivor bias effect that affects datasets composed of historical information on stock prices and how that can distort results and hinder the proper evaluation of any portfolio optimization strategy. Our approach uses a screening tool to select stocks out of a large pool. The screener's parameters are optimized on a training dataset. We then construct a portfolio which weights stocks so as to minimize the correlation of the selected stocks. We also incorporate a "trigger" mechanism for identifying downturns in stock prices in a way which informs our trading decisions. Using multiple testing periods of 14, 17 and 20 years, our strategy surpassed the S&P500 index and outperformed many similar studies. Overall, this work shows that a simpler, more fundamental approach can oftentimes perform better than complex models.

Acknowledgements

I would like to thank my supervisor Prof. Dimitrios Hristu-Varsakelis for providing professional guidance to complete this thesis after a creative, interactive, and stimulating process with innovative ideas expanding my knowledge limits. I am filled with gratitude to him, who has supported me in the challenges that I have faced as well as for his invaluable encouragement inspiring me to be the best version of myself.

Also I would like to thank my sister Anneta, my uncle Ilias, my father Panagiotis and my girlfriend Nikoleta for their patience, support, and kindness.

Contents

List of Figures

List of Tables

For/Dedicated to/To my…

Chapter 1

Introduction

The stock market, stock forecasting, and portfolio selection are considered as some of the most important and fascinating topics in the financial world. The stock market enables companies to increase their revenue by issuing shares and using the obtained capital to expand their production [43]. At the same time, investors can profit via capital gains (selling and asset for more than what they originally paid for it) and dividends (distribution of some of company's earnings back to investors). Over the past 200 years, investing in the stock market has outperformed other types of investment like bonds, treasury bills, gold, and outpaced inflation. Despite the ups and downs historically, a long-term investor could grow his capital with an annual return of 7% [7].

Nevertheless, the stock price rarely represents the actual fundamental value of a company. The price of a stock is subject to fluctuations and is determined by supply and demand in the market, which are driven by investors willing to buy or sell shares. As a result, a good approximation of the actual value of a share could be an investment opportunity. For example, an investor can buy an under-priced stock[1] and achieve capital gains when fluctuations in the market will drive the stock price closer to its fundamental value. As argued in [35], the fundamental analysis can identify companies that have the worst and best prospects. Stock screening is a fundamental analysis method which can identify stocks that satisfy specific metrics given a pool of stocks. Such metrics are, for example, price-to-earnings ratio, capitalization, current and historical price, technical indicators, etc. A stock screener, when tuned correctly, is able to identify and invest in under-priced stocks.

Since stock prices are volatile, the risk of capital loss is another characteristic of the portfolio that needs to be considered. Investing only in one security seems reasonable when the stock seems to outperform the market. However, the investor is then subjected to unsystematic risk that cannot be avoided. Instead, aiming for diversification in portfolio construction could mitigate that risk by counterbalancing bad and good performing stocks. Portfolio optimization is the process of finding and selecting the "best" of the available assets (stocks) and then determining how the invested capital will be allocated in these companies. The process of allocating the invested capital in a set of available stocks is called weighting. The objective of this weighting could be the maximization of the expected return for a given level of risk.

Portfolio optimization is an essential subject of study for small retail investors up to large financial institutions. A commonplace of portfolio construction that minimizes unsystematic risk are market indices. Stock market indices consist of many stocks, ranging from 30 to 3,700. Typical weightings of index models are based on the market capitalization of the assets, the current stock prices, or simply equal-weighted, which distributes capital equally in all companies. Due to their size,

[1] A stock which is priced too low compared to its fundamental value

market indices are considered less volatile, safer investments that reflect the current overall market value. Indices include the S&P500, Dow Jones Industrial Average, Nasdaq 100, Russell 2000 Index, NYSE Composite Index, Wilshire 5000 Total Market Index and more [27].

Leveraging state-of-art financial technologies, one can build and manage a stock portfolio by buying and selling stocks in real-time, using a mobile app. Proper knowledge of portfolio optimization and risk management surrounding such ventures is necessary when trying to beat the market. The financial world and the scientific community have been trying to calculate, handle, minimize the risk of loss and maximize return. Many frameworks are published each year in the literature, with the goal of achieving high returns with a low risk. Since return (profit) and variance (risk) are correlated, the proper portfolio selection strategies have to consider both of them.

1.1 Aims and Contribution

Along with introducing a novel strategy and presenting its various components, this thesis includes a brief review of state-of-the-art approaches to portfolio optimization, while comparing their performance to that of our proposed approach. Overall, this thesis creates a novel framework and evaluates its performance on the stock selection problem while emphasizing on modelling and computational efficiency. The goal and contribution of this thesis is to showcase key aspects of the portfolio optimization problem. In more detail, this thesis:

- Introduces a novel low-complexity portfolio optimization framework, which is scalable and efficient compared to other approach found in the literature. Our proposed approach generates high returns in a test sample comprising of historical price data from 03/01/2000 to 31/12/2020, considering a set of 543 stocks.

- Proposes an effective stock screener for selecting the stocks that will be part of the portfolio. Three fundamental measurements were tested, namely the cumulative return, Sharpe ratio and volatility, along with a historical window parameter ranging from 1 to 90 previous trading days. The proposed stock screening method, as well as the window size parameter, were selected based on a four-year training sample (ending before the start of the out-of-time testing sample) which includes an equally weighted portfolio of the top 50 stocks selected by the screener.

- Highlights the importance of reducing as much as possible the survivor bias effect in stock data used for evaluating any strategy. In contrast with many other works in the literature, this thesis examines and calculates the survivor bias effect in the testing sample of the proposed framework.

1.2 Structure of the study

The remainder of this thesis is organized as follows: Chapter 2 discusses the stock portfolio optimization problem and key factors that affect stock price movements, and includes a literature review of optimization strategies. Moreover, Chapter 3 introduces and describes our proposed approach to portfolio optimization. We start by discussing the dataset used and the survivor bias effect. Subsequently, we introduce a stock screener, which will filter available stocks and a risk-minimization technique for optimizing the weights in the portfolio. Chapter 4 discusses the performance of our proposed approach, tested for the years 2000 - 2020, and a comparison with other strategies from the literature. Finally, Chapter 5 summarizes the results, discusses the implications and limitations of the proposed method, and outlines future extensions and applications of this work.

Chapter 2

Background and Literature Review

This chapter discusses stocks and stock price movement, as well as the fundamentals of stock allocation and portfolio optimization. We also briefly discuss optimization frameworks as proposed in the current literature.

2.1 Stock price movement

A stock is a security that represents the ownership of a fraction of a company. The company provides the market with a number of shares outstanding, which are traded on stock exchanges, and investors can buy or sell stocks based on the current market price. This stock price fluctuates with time according to a balance between sellers and buyers of the stock in the open market [43]. Therefore, stocks and stock price movement can be seen and treated as time series.

A time series is a sequence of data ordered by reporting time and can be constructed by collecting any data that is measured over time. Examples of time series include periodically sampled values of heights of ocean tides, global pollution, daily temperatures, etc. In the case of stocks, time series can be created by collecting daily closing price value of a stock over a period of time. Such time series can be plotted in a two-dimensional graph where the y-axis represents stock price of the data and the x-axis represents discrete time. Figure 2.1 is an example of a stock's daily closing price movement represented as time series.

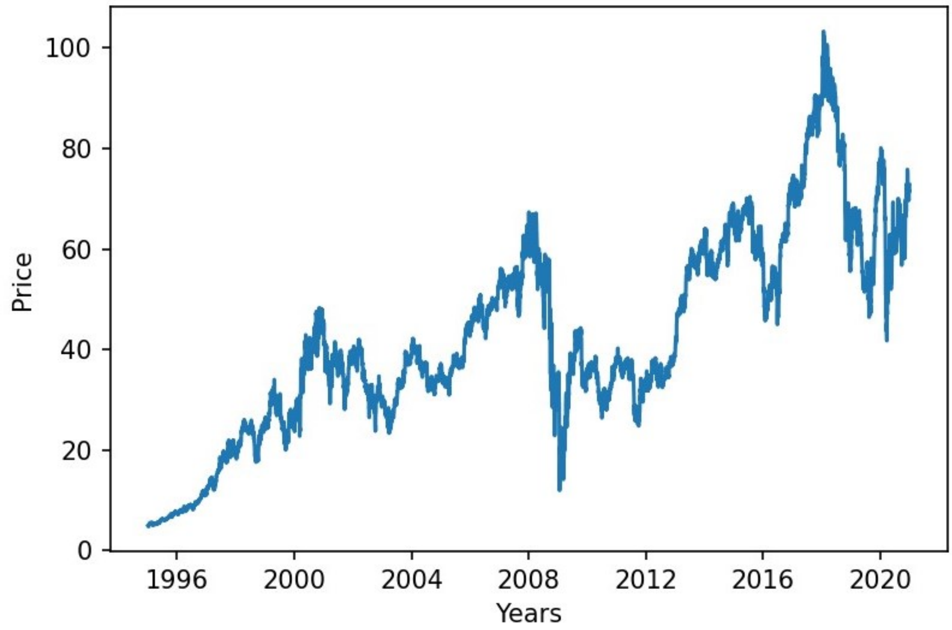

FIGURE 2.1: State Street Corp (STT) stock price graph between 1/1/1995-12/31/2020.

Several factors may affect stock price movement. Some of them can be quantified, e.g., demand and supply, media and news (company, market, industry or country specific) using sentiment analysis, earnings announcements, etc. There are also random influences that are difficult to measure (including investor psychology, reaction to various news - sometimes not directly related to the stock in question, and others), which manifest as noise [7]. Additionally, human psychology also affects the stock price movement. For example, the authors of [5] introduced the overreaction hypothesis and found that people tend to react in panic when dramatic news or events occur. This overreaction is just one of the few examples which makes stock price prediction challenging or even impossible.

There are many works in the current literature suggesting that changes in stock prices follow a random walk. The random walk hypothesis suggests that future stock price movements cannot be predicted based on past stock price changes or historical trends. As stated in [26], which advances the so-called "random walk" hypothesis, the market moves in an arbitrary manner, thus it cannot be modelled effectively. On the contrary, many recent articles question the Random Walk theory and suggests that stock market excess returns can be forecasted using historical information (see [13], [11], [34], [12]). Another hypothesis known as Efficient Market Hypothesis [10] holds that stock price includes all available information and that investors are rational and stocks are traded at a fair value. This means that risk is inevitable if one wishes to outperform the market.

Back in 1973, [3] developed the Black-Scholes model that is one of the most vital concepts for pricing options contracts. It is a differential equation that uses geometric Brownian motion, the affect of time and other quantitative components such as strike price of an option, the current stock price, the time to expiration, the risk-free rate, and the volatility. This method calculates a fair value for an option contract. For their contribution in modern finance, the authors later were awarded with the Nobel price in Economics.

Another study [19] suggests that stock prices are moving subject to trends. Stock price movements will tend to repeat themselves overtime, and as a result, historical information could be used to predict future stock prices. Similarly, [44] argue that security prices, as indicated by financial research, do not reflect all publicly available information. Consequently, investors can benefit by using stock screening rules and filters to minimize the search space and select the best assets out of a large pool of available stocks. The hypothesis that stock prices do not embed all publicly available information, is also supported by numerous other studies ([23], [14], [23], [37]) which also demonstrate the advantages and potential of stock screening techniques.

Overall, it is a well-accepted fact that predicting the prices (and profiting from their predicted movement) is a challenging task. This thesis aims to design a stock screener with the ultimate goal of identifying the best candidates for an investment portfolio after taking into consideration the following factors:

- Inventors' tendency to sell in cases of temporary price reductions.
- The "stocks move in trends" observation stated in the previous paragraph.

Because of the fact that hardware resources limited, very complicated and time-consuming machine learning models were not considered for the work underlying this thesis. Numerous machine learning approaches up to 1 day model training duration were examined(tree-based algorithms, XGB and LSTM neural networks). However, without an exceptional outcome. Instead, a more intuitive and expert-based approach was selected because it was found to produce the best outcome, at a reduced complexity and with a higher degree of explainability.

2.2 Stock allocation and portfolio optimization

The purpose of portfolio optimization or stock allocation is to select from a pool of stocks a portfolio that will lead to higher-than-market returns, taking into account the investor's risk appetite. In general, investors are more concerned about unexpected losses that they are about unexpected investment gains from their portfolio. A portfolio construction consists of a weight vector, indicating how much of the available capital should be invested in each stock or asset. Usually, it is represented with the vector w. Given a pool of stocks N, the vector w should follow these two constraints:

$$0 \leq w_i \leq \tag{2.1}$$

for i = 1,...,N and

$$\sum_{i=1}^{N} w_i = 1 \tag{2.2}$$

Given the selected stocks for the portfolio, the challenge is to optimize the weight w for achieving an investment goal. That can be maximizing the return, minimizing the risk, or a combination of both. Then the portfolio return is calculated as

$$w \cdot R \tag{2.3}$$

where R is an n-vector and stands for future return of assets.

In [29], Markowitz proposed a method for optimal trade-off among the expected return and risk. He constructed a mathematical framework that uses as input the expected return of the assets and the historical covariance matrix of these assets.

Then, by solving the formulated constrained quadratic optimization problem, the optimal weight vector is obtained a the given risk bound.

Over the years, many publications have been based on Markowitz's theory and tried to improve its shortcomings (out-of-sample performance) or adjust the framework to different problems. Well known risky measurements include, but are not limited to, semi-variance [28], semi-deviation [31], mean absolute difference [21], Value-at-Risk [8], Conditional-Value-at-Risk (CVaR) [36], Entropic Value-at-Risk (EVaR) [1], and Gini's Mean Difference [38]. We refer to [9] for a detailed review of risk measurements. Each risk measurement employed in the portfolio optimization frameworks constitutes a family of mean-risk models.

There are many attempts and various methods for solving the portfolio optimization and stock selection problem. The "key" to using Markowitz's work or any of its descendants is to somehow estimate r, the vector of asset returns. The most common approach is to predict the future behaviour of the stock, based on historical trends of the stock price, using technical analysis and qualitative information for the given stock. The primary assumption in technical analysis is that the stock price and market in general move subject to "momentum" and "trends" [20], [17]. This statement claims that the stock's performance will keep following its past behaviour for a time until the trend reverses. Numerous technical indicators can have been proposed for calculating this trend and movement, using mathematical formulas based on the asset's historical price. Most used indicators are trend indicators like 50-Day EMA and 200-Day EMA, mean reversion indicators like bollinger bands (BB), relative strength indicators like stochastics, momentum indicators like MACD, volume indicators like on-balance-volume (OBV) and more.

2.3 Literature review on portfolio selection strategies

The majority of stock selection frameworks are focused on forecasting the future price of the assets and then allocate the capital based on these prices and according to some other indicators, such as stock volatility, correlation, etc.

The authors of [18] tried to create a stock selection framework using two steps. First, they predicted the future trend of the stocks, either up or down. The authors used technical indicators as input features for the prediction and trained two machine learning algorithms for classification (xgb and logistic regression). As a dataset, then used 26 companies from S&P500 and a rolling window technique [6] for prediction and market simulation. Thus, the training period remains constant and moves forward in time as the price is predicted over a new week. In the second step, the capital allocation in stocks was optimized based on Gini's Mean Difference and Mean-Gini optimization model. The outcome dominated the S&P500 index in cumulative return using 20 years backtest. The overall cumulative return of the backtest was up to 26 times greater than the equal-weighted portfolio strategy using these 26 S&P500 stocks that consist of the backtest sample.

Similarly, in [41], the authors used technical indicators as input features for the classification problem. Then the stock selection was based on classifying/ranking the stock in N classes based on future excess returns. After the prediction, they buy the best 20 stocks from the first (best) class (and hold them) uniformly for 20 days. Then, using a rolling window technique, they tested their framework for five years. Based on a random forest model in the Chinese stock market, their portfolio selection strategy outperformed the CSI500 index, up to 5 times more (in terms of cumulative return) for the 5-year backtest period. The only downside was that an equal weight portfolio is not present in their in that work, and thus it was not possible to assess any survivor bias.

In [33], the authors used a combination of long- and short-term trends and technical indicators' momentum to identify stocks that are most likely to outperform the market index. Given 100 stocks of the SET100 index of the Thai stock market for the period 2011-2015, he used technical indicators as features for the model. Then, they performed a cluster analysis in x-20 day with technical indicators as features on that initial day. The clustering algorithm grouped those stocks in 10 clusters of 10 stocks each. The performance of these ten portfolios was assessed for the next 20 days (x-20, x). Then, the best stock was selected as a centroid. The 10 stocks closest to that centroid were used to construct an equal weight portfolio for the next 20 days. The main idea was that these stock trends would continue to occur based on the current market situation. The results show that the proposed framework can outperform the market index in the long run, generating 3.5 times higher cumulative returns compared to the market index.

While many studies use traditional machine learning models for the stock selection problem, Yang et al. [46] highlighted the power of the CI technique. A two-step framework for building a portfolio was proposed. In the first step, the future price was forecasted using ELM, a special case of a fast convergent single-hidden layer feedforward neural network. In the second step, the framework stored and ranked the stocks based on a differential evolution algorithm for optimizing the weights of the scoring function, which is based on factors like stock profitability, leverage, liquidity, efficiency, and growth, which are popularly used in existing stock selection models. The top 5% formed an equal weight portfolio until the next rebalance. This method returns in only a 2-year (2014-2016) backtest sample of A-share stocks of China market, 1.6 times more cumulative return than the market and the equal-weighted portfolio strategy.

In [30], the stock selection strategy was based on a Hidden Markov Model (HMM). The authors used HMMs to predict monthly regimes for four core macroeconomic factors: inflation, industrial production index, stock market index, and market volatility. The regimes are the two opposite states (high/low) of each variable. The predicted regime is then compared with similar historical behaviour. A composite score of each stock is calculated based on the scores that are a mixture of a stock factor under its behaviour in the previous similar macroeconomic stage. The stock factors are E/P, free cash flow/enterprise value, sales/enterprise value, long-term earnings per share growth, and long-term sales growth. The portfolio is then constructed based on this algorithm by choosing the top 50 ranking stocks to buy. The cumulative return of the out-of-sample tested period (2000 - 2015) is up to 6.5 times greater annual return compared to the S&P500 index.

Another popular stock selection technique is to invest in the worst-performing assets. The so-called mean-reversion principle argues that bad-performing assets will eventually return to their historical mean in the future [24]. Therefore, in [4], the authors used an anti-correlation technique for building a portfolio by taking advantage of price changes, constantly transferring the capital from the high-performing stock to the relatively low-performing stocks, and considering the correlation of the stocks. The results were promising and tested in a 5-year backtest, surpassing the 5x returns compared with the equal-weighted portfolio for S&P500. The downside was that mean reversion was prone to more significant unexpected losses.

Influenced by the principle of mean-reversion above, the authors of [22] proposed a novel kernel-based trend pattern tracking (KTPT) system for portfolio optimization that uses three stages for portfolio creation. Their work tried to start from a simple prediction of the stock price, based on historical inertia and maximum price, and then make that more robust by incorporating historical financial states and including a trend-reverting factor. Their method, tested and compared with many benchmarks and portfolio selection methods, could beat previous stock selection approaches. The returns are up to an 18-digit number of cumulative returns to a specific backtest sample. As we will discuss later, survivor bias, less-than-representative datasets, and risky strategies can lead to that return, however those returns are impossible to replicate moving forward. Overall, the best and comparable strategies in the literature that use S&P500 stocks and risk aversion methods are [18], and [30]. In 4, we will discuss a comparison of these strategies with the suggested strategy of this thesis. Of note, additional portfolio optimization strategies exist, but, as we will see in Chapter 4, are not comparable because they have selected different markets for their backtesting, or because they do not report cumulative results in their evaluation ([46], [33], [41]). The authors of [22] consider that it is state of the art by outperforming five other strategies using the same stock benchmark datasets [4].

Chapter 3

Methodology

This chapter describes the process of data used for evaluation and the main components of this thesis's proposed portfolio optimization framework.

3.1 Data Collection

3.1.1 S&P500 index, RSP index and S&P500 stocks

The S&P 500 Index is a market-capitalization index of the 500 largest publicly-traded companies in the United States. Stocks that participate in the index must to follow specific criteria. These include market capitalization, liquidity, and many more [40]. The index is weighted, allocating more to the top stocks and less to those on the lower end of the list. RSP is an ETF composed of the 500 companies included in the S&P500, investing equally in each stock. This uniform weighting increases the footprint of smaller S&P 500 stocks. Historically, there were periods when RSP outperformed S&P500. The equal-weighted RSP ETF makes the portfolio more robust to unexpected deviations of top-performing stocks.

This study will use S&P500/RSP as stock selection pool and performance benchmarks. The reasons for selecting this index are listed below:

- It includes many stocks with high volumes and high liquidity that are ideal for simulation. Usually, smaller (in) size stocks may not always have sufficient liquidity. Furthermore, any buy or sell decisions will not affect the stock price due to the high volume of these stocks. Other researchers use this or similar indexes like the NASDAQ index, A share (Chinese), etc.

- The weights that S&P500 applies to its component stocks are not publicly available. RSP, on the other hand, is handier to use and can be easily reproduced.

3.1.2 Survivor bias

The exclusion of failed stocks from a sample on which a portfolio strategy is assessed (because those stocks no longer exist) is known as survivor bias. Since only companies that were successful enough to survive until the time at which the strategy is being tested, the results of these studies are skewed towards over-performance. This is very common in the research literature [18], [22]. Without including the delisted stocks due to bankruptcy or pure performance, evaluating the stock selection strategy might lead to unreliable results which will not be achievable in the future.

For example, an investment agency wants to test a novel trading strategy in historical S&P500 companies by performing a backtest. Backtesting is a procedure of determining the appropriateness of a trading strategy by examining how it would perform in the real world using historical data. They find a list of S&P500 stocks as of today and downloads ten years of historical information for them. Much to their surprise, they obtained unexpected promising results using this dataset for the evaluation. They invest then in the real market, hoping that they will beat the S&P500 index. A couple of months or years later, they realize that their strategy out of sample do not perform as their backtest promised. Examining the backtest sample carefully, they realize that the effect of survivor bias is enormous. They had selected stocks that were the "survivors" of the past ten years, ignoring the stocks that went bankrupt, became delisted, or were part of an acquisition. In addition, these stocks are part of the S&P500 index as of today. This adds an extra bias to the sample since these were not only stocks that survived but also stocks that became so successful in this 10-year historical period that they became or kept being part of the top 500 stocks of the USA market. Figure 3.1 reveals the actual cumulative return of the RSP index vs. the equally-weighted cumulative return using historical stocks in the S&P500 index in 2020.

The dataset that includes historical information of the closing price for many stocks is called simulated dataset (SIM). This dataset should simulate the actual market in order to backtest a trading strategy on historical information. Creating a dataset that minimize the survivor bias effect can make the evaluation procedure more robust and trusted. An example of different simulation dataset are shown next.

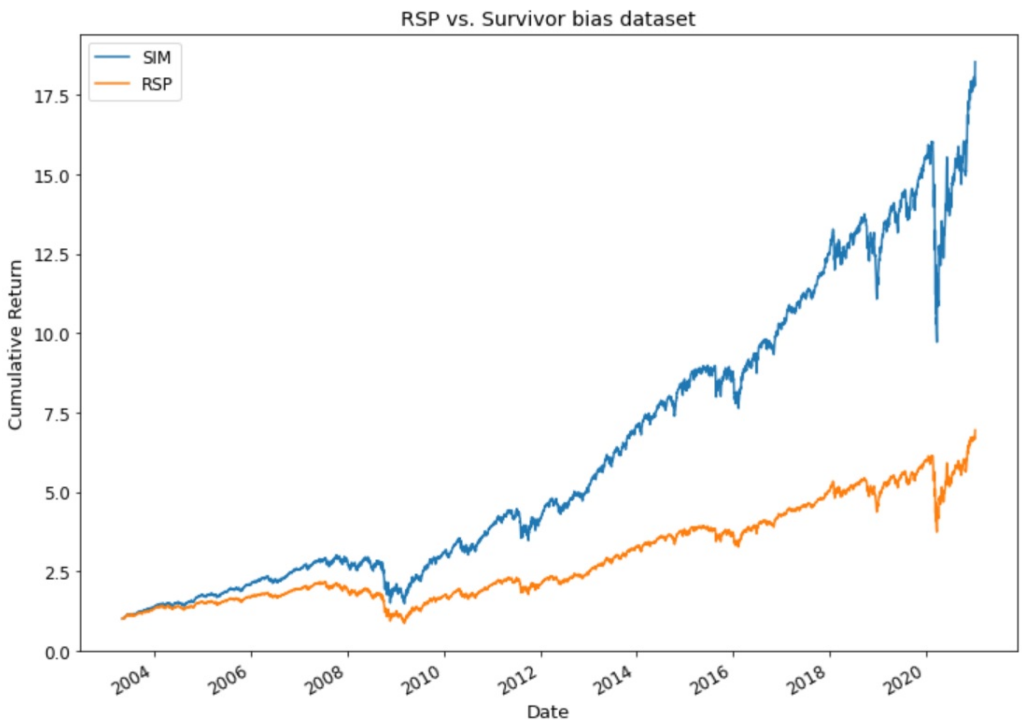

FIGURE 3.1: RSP vs stocks included in the S&P500 2021.

As depicted in Figure 3.1, the sample with survivor bias is up to 3 times greater than the actual returns of RSP. The simulation dataset (SIM) was created using 16 years of stock price information for stocks that are included in S&P500 at of the year 2021. Using the simulated dataset (blue), the overall performance becomes exagerated without including any portfolio optimization strategy. It also reveals that bad-performing stocks are not on the list to drive the average to the index (orange) line level. Subsequently, the results obtained by the equally-weighted survivors of the

SP500 index may be misleading. For more robust and trustworthy results, actions are needed to minimize the effect of survivor bias in the data.

3.1.3 Historical Information

To create S&P500 constituents and historical changes table, information obtained from Wikipedia [25]. A table, which includes the company, ticker name, date added, and date removed, was created using python and web scraping. When the date a stock was removed from the SP500 was unavailable, it was set to the end of the back-testing period, 31 December 2020. After excluding four tickers that were included and removed from S&P500 two times in their history (CBE, GAS, OI, AGN), the table with the components consists of 555 stocks for the period 2000 – 2020.

TABLE 3.1: Historical changes table (556 changes available).

Ticker	**Name**	**In**	**Out**
T	AT&T	30/11/1983	18/11/2005
SBL	Symbol Technologies	5/12/2000	10/1/2007
ABK	Ambac Financial	5/12/2000	10/6/2008
AYE	Allegheny Energy	5/12/2000	25/2/2011
ANR	Alpha Natural Resources	1/6/2011	2/10/2012
AMD	Advanced Micro Devices	1/1/1957	20/9/2013
SAI	SAIC	18/12/2009	20/9/2013
JDSU	JDS Uniphase	27/7/2000	23/12/2013
WPX	WPX Energy	31/12/2011	21/3/2014
CLF	Cliffs Natural Resources	18/12/2009	2/4/2014
COV	Covidien	28/2/2011	27/1/2015
PETM	PetSmart	10/10/2012	12/3/2015
...	...	...	...

The historical stock information on which this work is based was gathered from Yahoo! Finance [45]. The second source of historical data, The Quandl Wiki dataset, was used to add to the Yahoo! Finance source because the latter did not contain the delisted stocks. The only downside of the Quadl wiki dataset is that it is updated until March 2018.

TABLE 3.2: Stock historical dataset of STT stock example.

Date	Open	High	Low	Close	Adj Close	Volume
3/1/1995	7.28125	7.4375	7.21875	7.34375	4.92071	1569200
4/1/1995	7.34375	7.375	7.25	7.3125	4.899774	2515200
5/1/1995	7.3125	7.3125	7.1875	7.1875	4.816017	899200
6/1/1995	7.25	7.25	7.1875	7.1875	4.816017	994800
9/1/1995	7.1875	7.25	7.125	7.1875	4.816017	749200
10/1/1995	7.1875	7.1875	7.125	7.125	4.774138	891200
11/1/1995	7.1875	7.1875	7.09375	7.15625	4.795078	512000
12/1/1995	7.125	7.15625	7	7.09375	4.7532	1470000
...	...	...	...	...	...	...
21/12/2020	70.7	72.27	68.8	70.34	69.83493	5216400
22/12/2020	70.34	70.94	69.84	70.19	69.68601	5568400
23/12/2020	70.52	72.47	70.52	71.58	71.06603	2694400
24/12/2020	72.02	72.02	70.8	71.78	71.2646	665800
28/12/2020	72.37	72.89	71.45	72.12	71.60216	1738900
29/12/2020	72.52	72.81	71.73	71.94	71.42345	1168700
30/12/2020	71.76	72.92	71.76	72.42	71.9	909900
31/12/2020	71.93	72.91	71.47	72.78	72.78	951000

Date: Trading day.
High, Low: Highest and lowest prices at which a stock traded within a single day.
Open, Close: Prices at which a stock begins and ends the trading day.
Adj Close: Adjusted values incorporate changes resulting from corporate actions such as dividend payments, stock splits, or new share issuance.
Volume: Number of shares that were traded within a single day.

3.1.4 Data Manipulation

Data is downloaded using the table with the historical changes table 3.1. For each stock of the table, we download using yahoo Finance and Quadl WIKI dataset by setting as starting date the date when the stock was added to the S&P500 ("in" column of the table) minus three years back and end date the date when the stock went out of index ('out' date column of the table). Downloading historical values before the date added in S&P500 is necessary since at the date when a given stock is included, we need the information to forecast the future price, screen the stock based on historical price, etc. The minimum starting date is set as '02/01/1998' and the maximum date as '31/12/2020'. Out of 574 individual stocks, 556 were successfully downloaded, and 18 of them were not possible to be retrieved from either of the two data sources.

Next, the columns Open, Close, High, Low, Volume are scaled with the Adjusted Close value:

$$sf = \frac{Adjusted\ Close}{Close}$$

And for each variable in the dataset, the new adjusted to the close value computed as:

$$X_{scaled} = X \cdot sf$$

for X in { Close, Open, High, Low, Volume }.

3.1.5 Simulation dataset with reduced survivor bias

The final simulation sample is a concatenation of 545 files of S&P500 stocks[1], merged into one dataset keeping only the adjusted close price for each stock. The main dataset contains 5,787 trading days. For convenience reasons, dates are assigned to a day number. So, the first available date in the data, '02/01/1998', is set to day=1, and the last date, '31/12/2020', is assigned to day =5787.

TABLE 3.3: Survivor bias reduced simulation dataset format.

Day	**Date**	**AAPL**	**...**	**ABT**	**...**
0	2/1/1998	0.124763		8.843689	
1	5/1/1998	0.121884		8.892864	
2	6/1/1998	0.145397		8.745333	
3	7/1/1998	0.13436		8.802705	
4	8/1/1998	0.139638		8.917454	
5	9/1/1998	0.139638		8.819099	
6	12/1/1998	0.140118		9.122358	
7	13/1/1998	0.149715		9.190816	
8	14/1/1998	0.151635		9.215502	
5787	31/12/2020	132.2673	...	108.2397	

The creation of this dataset managed to reduce the effect of survivor bias and converge to the actual RSP index. A comparison is shown in the Figure 3.2:

[1] See full ticker list at Chapter 6

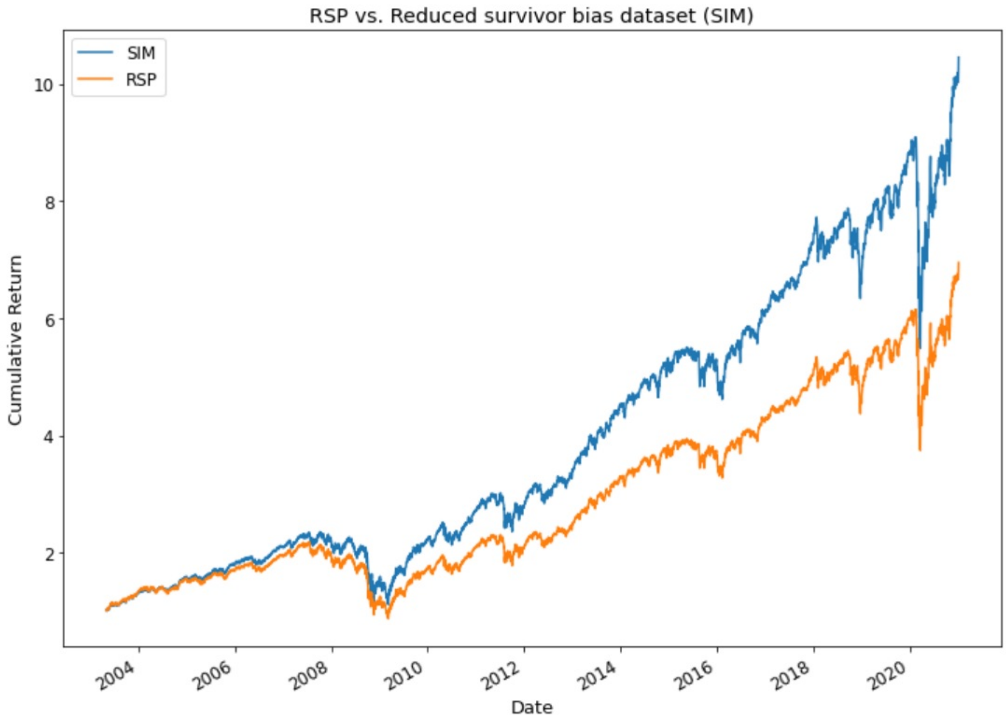

FIGURE 3.2: Simulated vs RSP for the period 2003 – 2020.

As shown in Figure 3.2, even by using this horizon (2003-2020), the simulated index is much closer to the real one when compared to Figure 3.1, where, for the same period, the cumulative return was up to 4 times greater than the actual. In the simulated dataset (SIM), the cumulative return is for this large horizon at 1000% compared to the actual, which is 700%. The closer we get to the end date and/or the smaller the period used, the better the approximation. In the Figure 3.3 below, the simulated dataset is very close to real RSP, with the effect of survivor bias being reduced to a minimum.

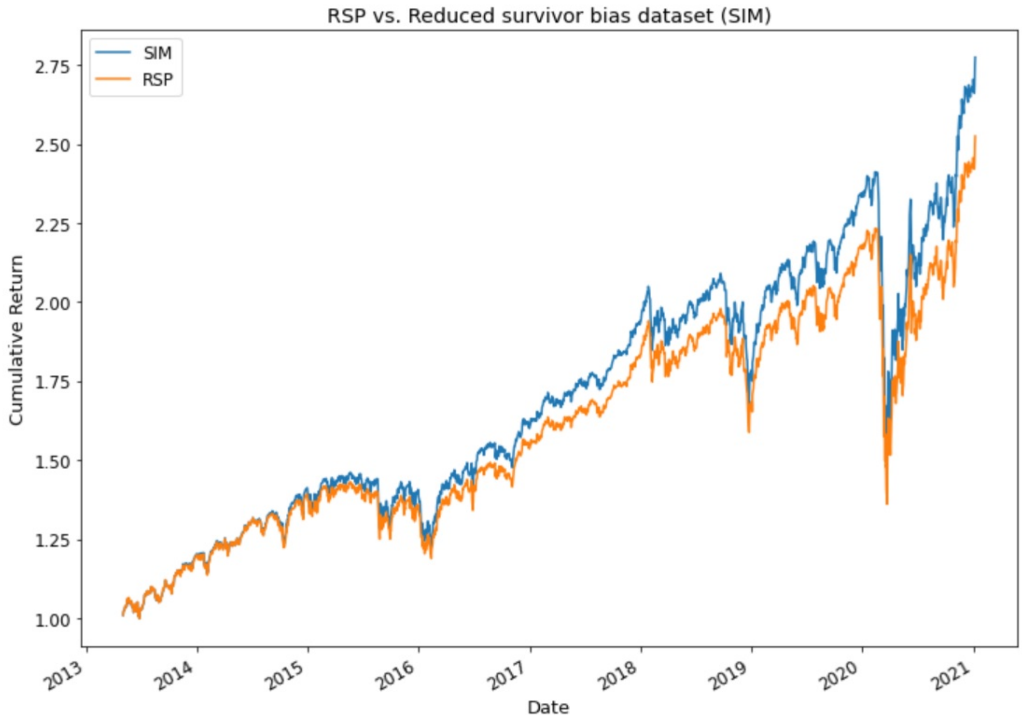

FIGURE 3.3: Simulated vs RSP for the period 2013 - 2020.

3.2 Screener

Stock screening is an essential step in the process of portfolio selection. The role of the stock screener is to minimize the search space in an extensive database of stocks. In the proposed framework, we describe a stock sceener that will choose from the pool of S&P500 stocks those that can potentially be the best candidates for the portfolio, according to quantitative criteria which we describe next.

3.2.1 Metrics used

The screener will be tested using three fundamental metrics as screening rules.

1. Historical Cumulative returns of a stock.
2. Volatility of a stock
3. Historical sharp ratio of a stock

Cumulative return (CR) is the total amount of change, it shows how much a portfolio has gained or lost in time. The cumulative return it's expressed as the ratio of the stock's price change over its initial price, indicating profit/loss compared to the portfolio's value at the beginning of the period. It is the raw mathematical return of the following calculation:

$$CR^i_{t,w} = \frac{p^i_t - p^i_{t-w}}{p^i_{t-w}} \tag{3.1}$$

where p^i_t is the stock price value of stock i at the day t, and w is the window of days we look back.

Volatility is a statistical measure of the fluctuation of returns for a given stock or portfolio for a specific period. It measures how much a stock price deviates from the mean value. In most cases, the higher the volatility, the riskier the asset. Volatility is measured either as the standard deviation or variance between returns from that same stock or portfolio. It is the raw mathematical return of the following calculation:

$$Vol_{t,w} = \sqrt{\frac{\sum_{i=t-w}^{t}(r_i - \bar{r})^2}{w}} \tag{3.2}$$

where t denotes the current day, r_i the daily in the given day i from day $i-1$, $\bar{r}$ is the mean value for the relevant period, and w denotes the window (number of days back).

The **Sharpe ratio** (SR) [39] is used to calculate the return-to-risk ratio of an investment by incorporating both the cumulative return and also volatility of the investment. The Sharpe ratio is defined as the average return earned over the risk-free rate per unit of volatility or total risk. Volatility is measured as the standard deviation of the excess returns and captures the price fluctuations of an asset or portfolio. Subtracting the risk-free rate from the cumulative return allows an investor to isolate the profits associated with risk-taking activities. The risk-free rate of return is the return on investment with no risk, e.g., buying a U.S. Treasury bond, meaning whatever investors could expect for taking no risk. In our setting, the risk-free rate is set to zero. Generally, the greater the value of the Sharpe ratio, the more attractive the risk-adjusted return is. The Sharpe ratio can be calculated by

$$SR_{t,w} = \frac{CR_{t,w}}{\sigma_{t,t-w}} \tag{3.3}$$

where $CR_{t,w}$ denotes the cumulative return an asset for the period (t-w, t) and $\sigma_{t-w,t}$ denotes the standard deviation (volatility) of the asset's excess return for the relevant period using daily returns r_i at the day i.

Next, for the creation of an effective screener based on these metrics' (i.e., CR, SR, Vol), experiments were conducted to answer the following questions:

1. How good is each of these metrics regarding selecting the most appropriate subset of stocks that will lead to high return given the risk?

2. What is the best value for the hyperparameter $w = t_2 - t_1$, i.e., how far into the past should we go when calculating the three metrics?

To answer these questions, we performed a backtest using the pool of RSP stocks to fine-tune the hyper-parameters of the screener. The process is summarized Algorithm 1.

Algorithm 1 Sreener Hyperparameter Optimization

Require: d_0: initial day,
M: Screening method,
$Data$: Daily historical stock data
O: Ranking order

```
procedure
    for w in 1..90 step=1 do
        while d <= d_0 + 1000 do
            for stock in Data do
                Calculate: M^{stock}_{d-w,d}
                Save: value in a list L.
            end for
            Calculate: L_O using the order method from O.
            Buy: A uniform portfolio (UP) with the first 50 stocks on the list L_O
            Set: d <- d + 5
            Calculate: C_w <- C_w * (Profit_UP)    ▷ Capital C_w is considered 1 at d_0 + 5
        end while
        Save C_w in a list F_M
    end for
end procedure
```

First, a 5-year historical period of 12/05/1998 till 03/01/2003 was selected using these RSP stocks of the dataset. Then each of these metrics separately used as a filter for the screener. From the initial day t, the algorithm calculates the values of CR, SR and Vol using a historical window w that can range from 1 to 90 days. Then the stocks are sorted in a both descending and ascending order using these 3 metrics. Now, the algorithm selects the top 50 stocks and forms an equal weight portfolio. This procedure is repeated every week (five market days). Every time, the portfolio is updated based on the top of the list of CR, SR and Vol assets. The results of the backtest are plotted and presented in the next two subsections.

3.2.2 Selecting using descending order

Stocks are ordered from highest to lowest based on the CR and SR metrics. This means selecting the portfolio's best current performers or winners using a relevant window w. For the volatility, descending order means that high volatility stocks are chosen first. The results are shown in the Figures 3.4, 3.5 and 3.6.

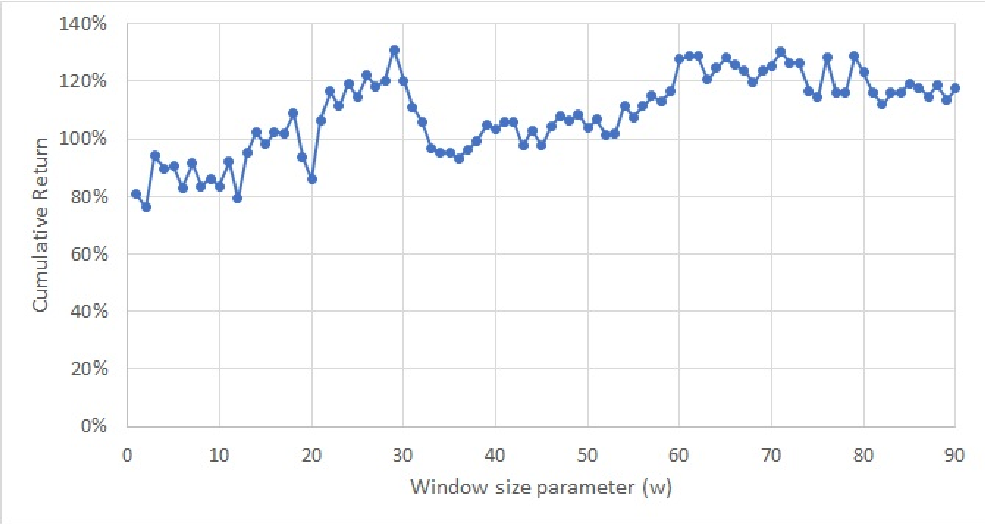

FIGURE 3.4: Performance of an equally-weighted portfolio using stocks selected by the screener. The screened used descending cumulative returns as the sole criterion for inclusion in the portfolio. The graph shows the portfolio's performance for a range of the parameter w.

Figure 3.4 shows the screener's performance when using the cumulative return as the criterion for selecting the "best" stocks. The x-axis shows the window size used for the calculation of the method. More specifically, the window size starts from 1 day back, which is just the return of the previous day, and can go up to 90 days, which gives us the cumulative return of the stock for the last 90 days. The outcome of this analysis using the information above is that recent information (returns) leads to poor portfolio performance. In other words, selecting stocks with an increasing trend in the past 1-12 days will lead to poor future returns. The reasoning is that this stock is already overpriced (at its peak), and constructing a portfolio with them will lead to a loss when they reverse their trend. The following graphs show that this portfolio strategy does not perform well at all.

Next, ranking based on the Sharpe ratio was tested. The results are shown in the following Figure 3.5.

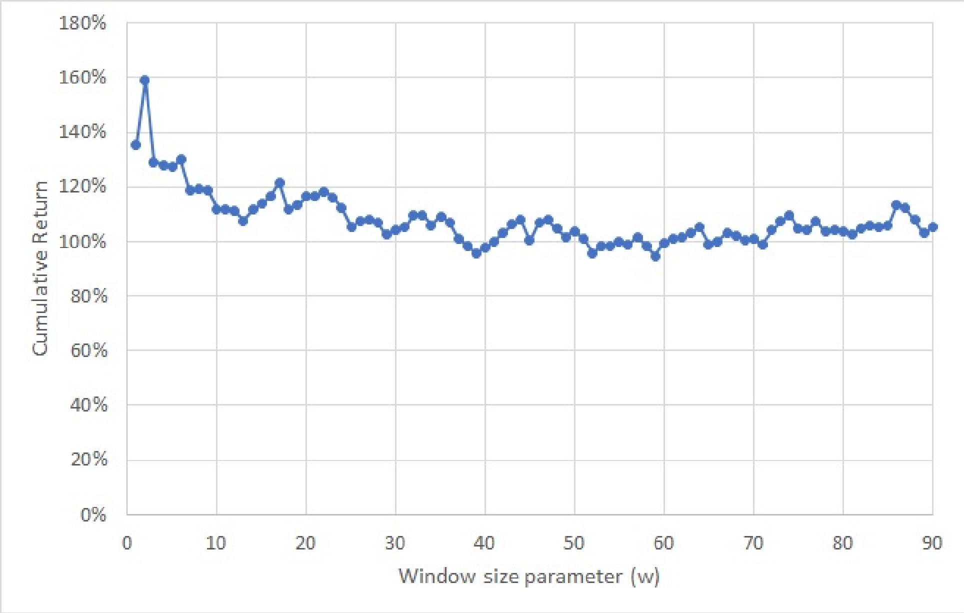

FIGURE 3.5: Performance of an equally-weighted portfolio using stocks selected by the screener. The screened used descending Sharpe ratio returns as the sole criterion for inclusion in the portfolio. The graph shows the portfolio's performance for a range of the parameter w.

Figure 3.5 shows the screener's performance when using the 50 stocks having the highest Sharpe ratios. Portfolio returns peak when using a window size of $w = 2$ and after that, the portfolio performance stabilizes close to 100%. The result is very poor, for similar reasons that are described above.

Finally, stock volatility was used as a screening criterion. The results are shown in the Figure 3.6

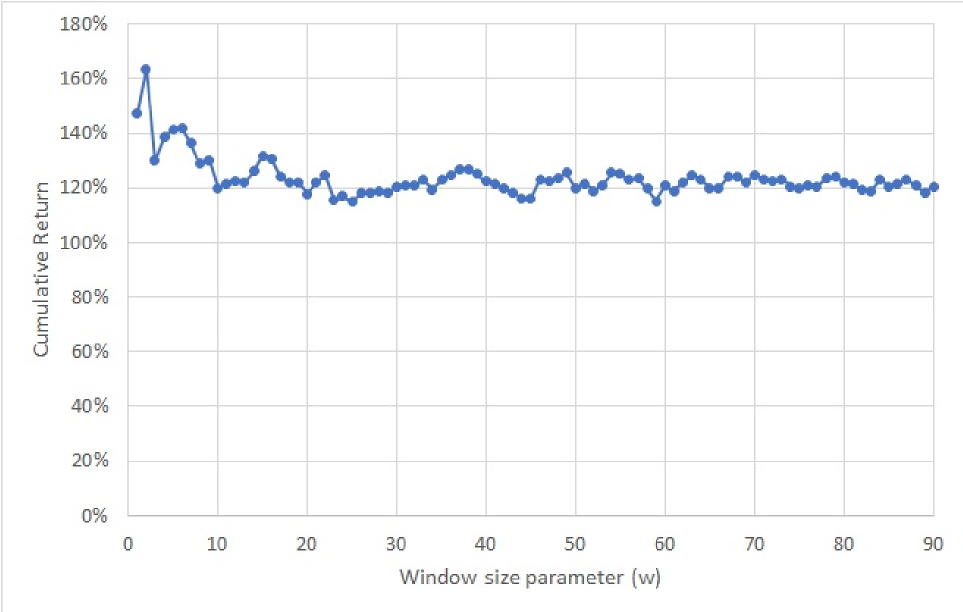

FIGURE 3.6: Performance of an equally-weighted portfolio using stocks selected by the screener. The screened used descending volatility returns as the sole criterion for inclusion in the portfolio. The graph shows the portfolio's performance for a range of the parameter w.

The graph in Figure 3.6 shows the screener's performance when using volatility to rank the stocks. Here the stocks selected are with high volatility. As depicted above, high volatility cannot be considered a screening method. After a window value of 10, portfolio returns are not affected. The peak using 1-10 windows is probably related to the increasing trend of stocks at that period.

3.2.3 Selecting using ascending order

Order by lowest to highest the assets using the CR and SR metrics; it's like selecting the worst current performers or losers for the portfolio using a relevant window size w.For the volatility, ascending order means that the lowest volatility stocks are chosen. The results are shown in the Figures 3.7, 3.8 and 3.9.

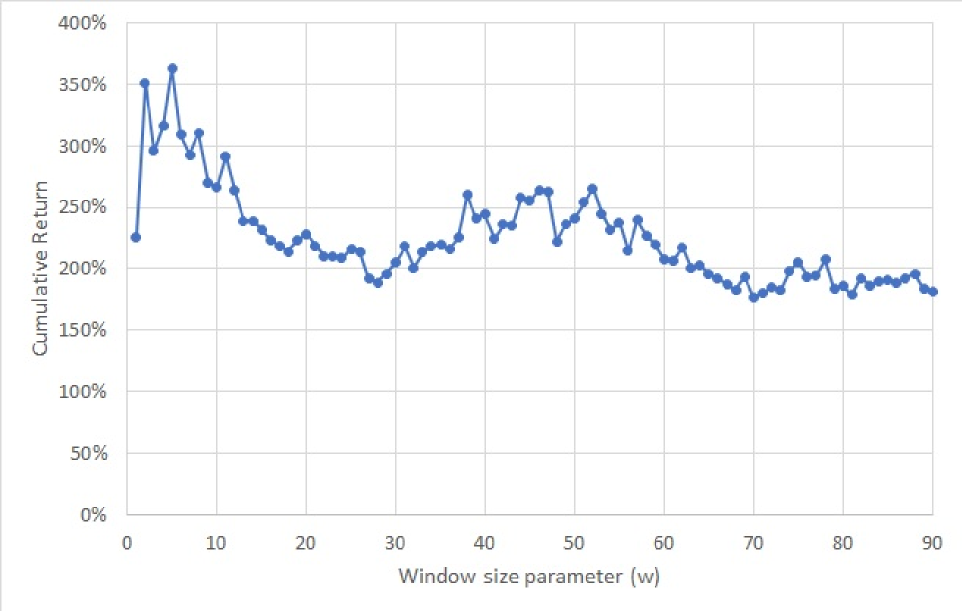

FIGURE 3.7: Performance of an equally-weighted portfolio using stocks selected by the screener. The screened used ascending cumulative returns as the sole criterion for inclusion in the portfolio. The graph shows the portfolio's performance for a range of the parameter w.

Figure 3.7 shows the screener's performance when using the cumulative return as a criterion and selecting the worst-performing stock for the portfolio. This strategy gives very high returns compared to the other strategies. The maximum return was observed using 2- and 5-day window sizes, with the five being slightly better, with a cumulative return at 360%. In general, the peak is using a 2-11 days window. After that, higher window numbers stabilize the returns at 200% on average. This Figure 3.7 shows that selecting short-term underperforming stocks is the best bet for maximizing future profit. Stocks tend to move based on trends. So, the best bet is to buy a stock "at its worst" period and expect the upcoming future to reverse its trend to overperforming [24], [4], [22]. In the literature, similar behavior is known as mean reversion, and a relevant strategy is to buy underperforming stocks below the historical average and expecting that they will reverse back to the mean value.

Next, ranking based on the Sharpe ratio was tested. The Figure 3.5 shows the result:

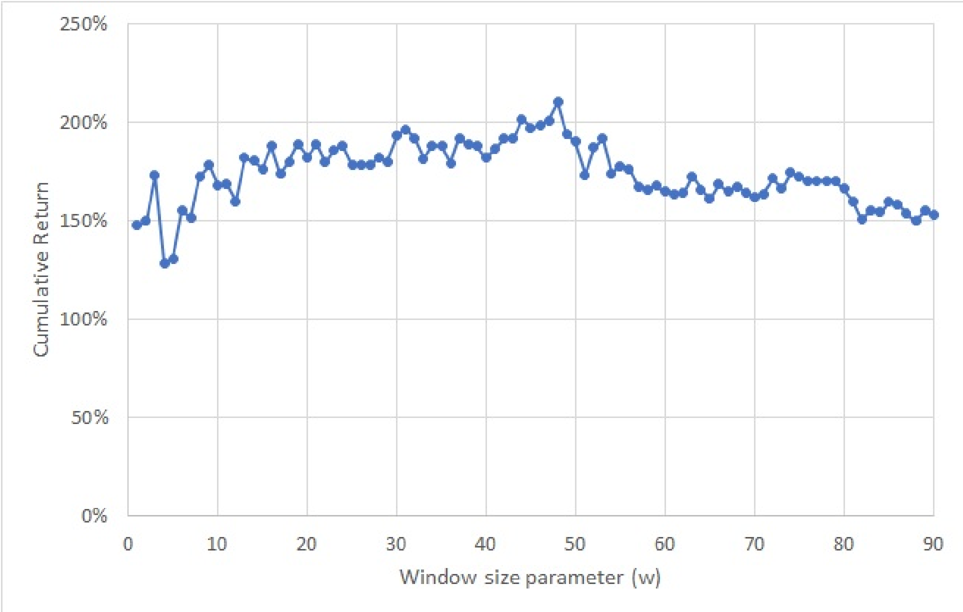

FIGURE 3.8: Performance of an equally-weighted portfolio using stocks selected by the screener. The screened used ascending Sharpe ratio returns as the sole criterion for inclusion in the portfolio. The graph shows the portfolio's performance for a range of the parameter w.

In Figure 3.8, Sharp ratio is used as a screening criterion, and stocks are selected using ascending order. No particular peak is observed. The maximum return is marked using 48 days window back, with values close to 210%, but the returns are similar for windows between 15 and 54.

Finally, stock volatility was used as a screening function. The results are shown in the Figure 3.9:

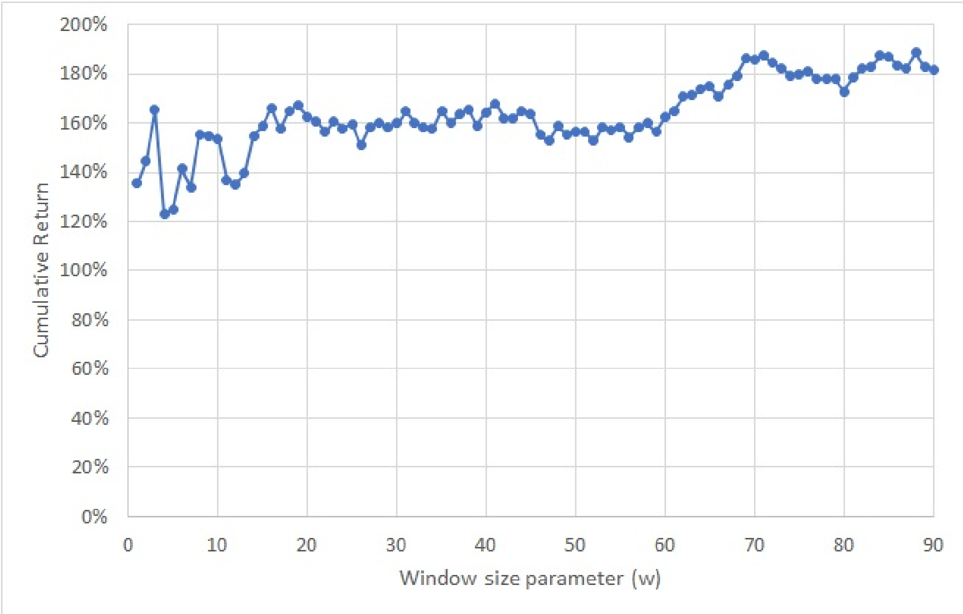

FIGURE 3.9: Performance of an equally-weighted portfolio using stocks selected by the screener. The screened used ascending volatility returns as the sole criterion for inclusion in the portfolio. The graph shows the portfolio's performance for a range of the parameter w.

Figure 3.9 shows the screener's performance when using the volatility criterion and selecting the least volatile assets. As expected, volatility itself does not significantly affect the portfolio's performance compared to cumulative return and Sharpe ratio when window size is small. However, when choosing stable stocks for an extended period (as shown above for the window size of 60 to 90), in the long run, when the market follows an increasing trend, low volatility stocks are suitable for safe investment and can promise higher returns compered to short-term low volatility stocks.

3.2.4 The KISS principle for stock price screening

Many screening techniques in the literature try to capture complex patterns by using numerous filters combining both fundamental and technical analysis. The screening in this work aims to minimize the search space of possible stocks that will be candidates for forming the portfolio for the next period. The KISS[2] principle states that most frameworks work best if they are kept simple instead of made complicated; subsequently, effortlessness ought to be a key objective, and unnecessary complexity should be dodged.

In stock price prediction, there is a lot of noise and the movement of the stock is the result of many factors that, for now, seem impossible to capture in their entirety. What remains for the investor is to select stocks that look the most promising. And the last observation that someone has for a given stock is its current price. Since stocks tend to move in trends, our best bet is that a stock that today shows a

[2]Keep It Simple Stupid

decreasing trend is likely to follow the opposite pattern a week later. This is a common scenario for large capitalization companies like S&P500, which are tested in this work.

On the other hand, for a stock that currently features a high increase in price in the recent week, it is difficult to state whether it will not be overvalued and start following the opposite trend, or for how long it will keep the upward direction. For those reasons, the first step for the screener is to select the stocks that show a downward trend in the short observable historical window. In our experiments, the best portfolio performance was observed when the selection criterion was the cumulative return, the value of w was set to w=5 days, and the worst performers were selected. The reasoning for this is to buy stocks at low prices today that are expected to perform better in the short-term future (five trading days ahead).

3.3 Portfolio Optimization

Putting all the wealth into one asset, i.e., the one with the best prediction best on some methodology or simple with the best historical performance is risky. Although an excellent future performance of the given stock can award the investor with tremendous profit, on the other hand, he puts at risk all of his capital in case the stock starts losing value. To avoid this situation, the rational investor should manage risk by distributing wealth to multiple assets. Therefore, in case of some companies start losing a large amount of their value, a good-performing stock in the portfolio can counterbalance that loss. The question that begins to rise in the market and financial work is how to optimize the distribution of the capital that will lead to high returns and, at the same time, keep the risk factor at a tolerable level.

3.3.1 Modern Portfolio Theory

Modern portfolio theory (MPT) is a portfolio selection strategy with the goal of maximizing overall returns while maintaining an acceptable level of risk. This simple but breakthrough idea was published in the Journal of Finance by the economist Harry Markowitz in his paper "Portfolio Selection" in 1952 [29]. For his contributions to portfolio theory, Markowitz was later awarded the Nobel Prize [42].

Diversification is an essential part of the MPT approach. Most investments are either high risk/return or low risk/return. Investors, according to Markowitz, may obtain the best outcomes by selecting an ideal balance of the two according to their particular risk tolerance. An investor might start with a certain amount of expected return and build a portfolio with the lowest possible risk that can provide that return. The screening step performs the filtering of stocks from the stock pool. Later, for these 20 candidates, portfolio weights are optimized based on the MPT framework by minimizing as possible the risk parameter. The optimization problem is formulated as follows:

$$\begin{aligned} \min_{v} \quad & v^T \Sigma v \\ \text{s.t.} \quad & \textstyle\sum_i^N v_i = 1 \\ & 0 \leq v_i \leq 1 \end{aligned} \tag{3.4}$$

where v is the weight distribution vector and Σ is the cross-covariance matrix of the N stock time series over a specified time window.

3.3.2 The covariance in Modern Portfolio Theory

Covariance in portfolio optimization is used to measure diversification in portfolio construction. It measures the directional relationship between two stocks. A positive value refers to movement in a similar direction, while a negative value signifies that the two stocks usually move in opposite directions. The goal is to choose assets that lower the variance of the combined portfolios returns, to a level lower than those of individual assets. This can reduce the volatility of the portfolio and thus the risk. Modern portfolio theory seeks to create an optimal mix of higher-volatility assets with lower-volatility assets. By diversifying the assets in a portfolio, investors can reduce risk and allow for a positive return.

The use of covariance does have a weakness. Covariance can only measure the directional relationship between two stocks. It cannot show the strength of the relationship between them. The correlation coefficient can add this dimension to measure that relationship strength. An additional drawback to using covariance is that the calculation is sensitive to higher volatility returns. More volatile stocks include returns that are far from the mean. These excess returns can have a disproportionate impact on the resulting covariance calculation. Huge single-day price moves can affect the covariance, leading to a wrong estimation measurement.

3.3.3 The impact of stock correlation

The correlation coefficient between two-time series defined in eq. 3.5 , takes on values between -1 and 1 and shows the association of each other.

$$Correlation = \rho = \frac{COV(X,Y)}{\sigma_x \sigma_y} \tag{3.5}$$

where:
$Covariance = cov(X,Y)$ = Covariance between stock X and Y
σ_x = Standard deviation of X
σ_y = Standard deviation of Y
$Cov(X,Y) = \frac{\sum(Price_X - Average_X) * (Price_Y - Average_Y)}{(Sample\ size - 1)}$
$\sigma_X = \sqrt{\frac{\sum(Price_X - Average_X)}{Sample\ size - 1}}$

In other words, when a correlation of two time series is close to +1, they move in the same way. When one goes up usually, the other also goes up. On the other hand, a negative and close to -1 correlation means that when one goes in one direction, the other tends to follow the opposite direction. A near-zero correlation means that the two time series are uncorrelated, and their relative movement is random.

A portfolio is considered robust when the lowest correlation possible between the selected stocks is achieved. This reduces the risk of loss because even if the price of one or more stocks tumbles, the remaining stocks will not follow the same trend. Instead, they are expected to perform better due to uncorrelated behaviour. Then, the loss would be counterbalanced. A high positive correlation close to 1 is not very different from the best stock portfolio[3] since all the stocks are expected to follow similar trends in the future, and thus, they will either go up or down. The reward

[3] Best stock portfolio is a portfolio strategy where one stock is selected only, based on same methodology.

can be very high, but the risk of loss is also high. As a conclusion, a neutral - near-zero correlated portfolio can help provide better diversification and reduce overall risk, since stocks will not be "dependent" with each other.

Markowitz formulated the return and risk trade-off idea as the constrained quadratic optimization problem defined in (3.4). As Σ, the correlation matrix will be used, instead of the covariance matrix, since the correlation matrix shows also the strength of correlation and not only the direction. The period used for calculating correlation is days [t-14, t-6] where t is the current trading day. The reason is explained below:

1. The period [t-5, t] is used for the screening. Since the worst-performing stocks are selected, this period is expected to be highly correlated for all selected assets since they have a downward trend (because the worst performing was selected) in that period.

2. The t-14 day was selected as the lower bound of the period interval based on a tune experiment. All the frameworks parameters remained fixed with the sigma's lower bound taking values from 10 to 29. The intervals tested had the form of $[t - t_s, t - 6]$ where t_s = 10, 12 ,..., 29. The performance was evaluated similarly with screener experiment by calculating the cumulative return at the end of the testing period. The training sample is the sample used for tuning the screener (first 5-year historical period of 12/05/1998 till 03/01/2003).The performance was evaluated similarly with screener experiment by calculating the cumulative return at the end of the testing period.

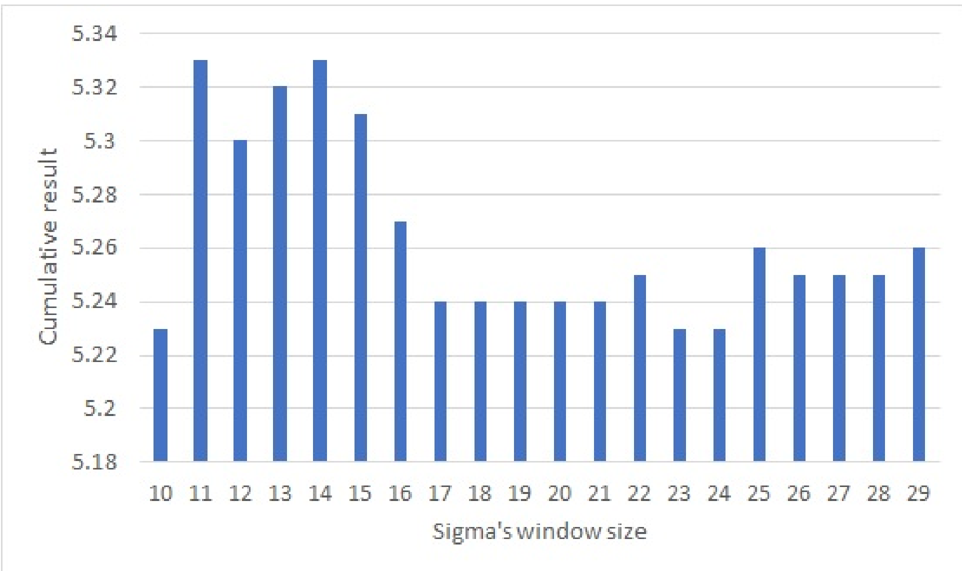

FIGURE 3.10: Performance of an equally-weighted portfolio by changing only the window size of sample for the calculation of the correlation matrix.

As depicted in the Figure 3.10, the best cumulative return is reached when t-12 and t-14 is used as beginning day of the sample. The t-14 is selecting since is gives bigger sampling period for the correlation matrix.
For the testing, the simulator, starts from the starting date of the testing dataset. The simulator feeds the screener with the available stocks and the screener return $N_{screener}$ in size stocks. Then, using the minimized volatility objective function the optimizer

returns a vector $v_{optimized}$ of $N_{screener}$ x 1. Finally, the weight vector v transformed to $v_{portfolio}$ with dimension N x 1 where the N is the total number of all available stocks in order to be the ruler for the buying strategy. The portfolio weight vector is defined as

$$v_{portfolio} = \{v_{optimized}(i)\ if\ s(i) \in SCRS\ else\ 0 \mid i \in [1..N]\}$$

where s(i) is the i-th stock and SCRS the set of stocks that were selected by the screener. In other words, all the stocks that are not part of the screener for the given trading day, allocation of the capital is forced to 0. The $v_{portfolio}$ follows summation to one rule since $\sum_{i=1}^{N} v_i = \sum_{i=1}^{N_{screener}} v_i = 0$.

Our setting used N_screener = 20 , the optimizer returns the optimal weights for each of the top20 stocks. The sum of the optimized weight vector summarizes to 1. For the remaining N-20 available stocks, the weight is forced to 0 since they are not considered for the portfolio at the current cycle. As a result, the portfolio is constructed using the final weight ($v_{portfolio}$) an Nx1 vector that sums to one and was the ruler on how to construct the current portfolio by buying stocks based on the weight of each stock times the capital to be invested.

3.3.4 The downturn trigger

When the price of stocks moves more slowly or drops and GDP (gross domestic product) diminishes, stagnates, or expands more slowly, an economic downturn occurs. An economic downturn is a typical element of the economy's cycle, which alternates between periods of expansion and contraction.

A downturn or recession is a contraction of the economic business cycle that occurs when there is a general decrease in economic activity. Then, there is an overall decrease in spending (an adverse demand shock). As a result, stock prices drop in value and can lead to significant losses. Recent downturn periods are the recession of 2008 and the covid-19 pandemic in Spring 2020. Our framework contains a simple downturn trigger that withdraws from the market until certain conditions are met to avoid these situations. In each 5-day buying and selling trading cycle, the previous cycle's profit is assessed. When the previous cycle was more than 10% loss (profit=-10%) of the total portfolio, our proposed approach withdraws from the market by not constructing a new portfolio for the current cycle.

Although the framework will not buy a new set of stocks for the current period, it will do it virtually. The new virtual profit will then be used for the downturn assessment in the next period. If the loss exceeds the 10%, a new set of stocks will construct the portfolio for the next period, and the framework will return to the market.

Algorithm 2 Trading Cycle

Require: today:int, window_size:int

Step 1: Select 20 stocks with the worst cumulative return in the period of (today-5, today).

Step 2: Calculate for these 20 stocks the correlation matrix using the period (today-16, today-6].

Step 3: Calculate the weight vector by solving the optimization problem.

Step 4: Assess the downturn trigger.

Step 5: Construct the portfolio of today by buying the stocks based on the weight vector.

State 6: Sell the portfolio using the prices of today + 5 trading day and calculate the profit.

Algorithm 3 Trading Framework

Require: initial_day, last_day

set: today = initial_day

set: total_capital = 1

while $today \leq last\_day$ **do**

  $Trading\ Cycle(today,\ window\_size)$

  $total\_capital = total\_capital * (1 + profit_{today})$

end while

Chapter 4

Results

4.1 Portfolio evaluation

The portfolio selection framework described in Chapter 3 and Algorithms 2 and 3 were tested using stocks from the S&P500 index from the year 2000 until the end of the year 2020. More details of the testing data can also be found in Chapter 3.

For assessing this thesis proposed strategy, the cumulative return (CR) was selected as the metric based on which to make portfolio evaluation and comparisons. The initial capital was set to 1 unit, to simplify the calculations. After the initial day, the profits of each trading week (5 days) were added to the total cumulative return as $c_t = c_{t-1} * (1 + r_t)$ where r is the portfolio return and c is capital, where $c_0 = 1$ at the initial day (03/01/2000). The Figure 4.1 presents the performance of the portfolio selection strategy by comparing the proposed framework with other strategies. These include:

1. **Framework portfolio with DT trigger**. This is the strategy for selecting stocks introduced in the current work using the Downturn trigger.

2. **Framework portfolio without DT trigger**. This is the strategy for selecting stocks introduced in the current thesis but without using the Downturn trigger as defined in Chapter 3.

3. **Screener only portfolio**. This variation creates a portfolio by buying uniformly the top 25 stocks of the screener.

4. **Uniform portfolio**. This portfolio construction selects all available stocks in our dataset using equal weights.

As performance metrics, for evaluating the performance of each portfolio construction strategy, the following measurements were considered:

1. **Simple return (R).** This is a simple portfolio return from a weekly portfolio construction.

$$R_\tau = \sum_{i=1}^{N} v_{i,\tau-1} \cdot r_{i,\tau}$$

where $v_{i,\tau-1}$ is the buying weight of stock i at time $\tau - 1$, $r_{i,\tau}$ is the return of stock i at time τ and τ refers to a trading day.

2. **Cumulative return (CR).** This is the aggregated capital gained. Initial invested capital is assumed to be 1.

$$CR_\tau = \prod_{t=\tau_{init}}^{\tau} R_t$$

3. **Volatility (Vol).** The variance of the weekly returns of the portfolio.

$$Vol_\tau = \sqrt{\frac{1}{|\tau - \tau_{init}|} \sum_{t=\tau_{init}}^{\tau} (R_t - \overline{R})^2}.$$

where $\overline{R}$ is the arithmetic average portfolio return and τ_{init} is the initial trading day.

4. **Sharpe Ratio (SR).** This the ratio of cumulative return divided by variance of weekly return. First defined by Sharpe at [39].

$$SR_\tau = \frac{CR_\tau - 1}{Vol_\tau \cdot \sqrt{|\tau - \tau_{init}|}}$$

5. **Annual Return (AR).** This is the return of the portfolio of the last 52 trading weeks.

$$AR_\tau = \frac{R_\tau - R_{\tau-52}}{R_{\tau-52}}$$

6. **Annual Volatility (AV).** This is the variance of the return of the last 52 trading weeks.

$$AV_\tau = \sqrt{\frac{1}{(52-1)} \sum_{t=\tau-52}^{\tau} (R_t - \overline{R})^2} \cdot \sqrt{52}.$$

7. **Drawdown (DD).** The decline in value of the portfolio from the maximum peak value to

$$DD_\tau = \frac{CR_\tau}{MAX(CR_t : t \in \{\tau_{init}, \tau_{init+1}, ..., \tau\})}$$

8. **Maximum Drawdown (Max DD).** This is the maximum drawdown drop in the portfolio from the beginning of the trading period.

$$maxDD_\tau = MIN(DD_t : t \in \{\tau_{init}, \tau_{init+1}, ..., \tau\})$$

Cumulative returns of the four strategies mentioned above are shown over time in the testing period[1] of 2000 - 2020 in the Figure 4.1. This period was selected in order to have a full 20-year sample and compare the results with different strategies from the literature.

[1]The results of out-of-time sample (2003-2020) exist in the appendix, Chapter 6.

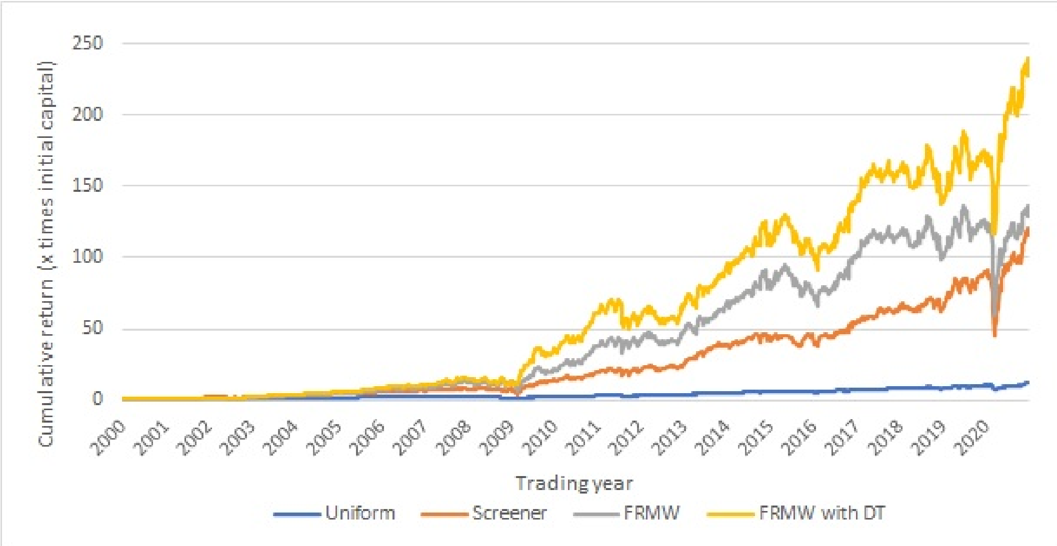

FIGURE 4.1: Weekly cumulative returns using 4 portfolio construction methods.

The results above show that Framework with Downturn trigger (FRMW with DT, yellow line) outperforms all the other strategies. This is the stategy that included stock screening, weight optimization and downturn trigger as described in Chapter 3. Compared with the uniform portfolio (Uniform, blue line), the FRMW with DT portfolio return was 20 times higher using cumulative return as performance measurement. The framework without the DT trigger (FRMW, grey line) performs worse than the inclusion of DT trigger. This indicates that the DT trigger positively impacts the overall performance and leads to better returns when applied as a safety/risk measure to the portfolio strategy. The Screener portfolio (Screener, orange line) performs 10 times better than the uniform portfolio. Nevertheless, it was only 2-3 times worse than the FRMW with DT strategy. This indicates that screening itself can create a robust stock selection. As a conclusion, the above graph shows that each component adds up to the overall performance. Screening itself outperforms uniform portfolio. Then, adding weight optimization to the screened stocks, it outperforms both uniform and screener portfolios and lastly, after the inclusion of the downturn trigger, the FRMW with DT strategy achieves the best cumulative return out of all.

TABLE 4.1: Final cumulative results as of last trading date using 20-year sample starting at 03/01/2000.

Date	Day	Uniform	Screener	FRMW without DT	FRMW with DT
18/12/2020	5781	1101.00%	12011.00%	13519.00%	23869.00%

In the Figure 4.1, an upward trend is shown after the year 2008. In the period before, the market was very stable. The framework also shows a horizontal trend. However, neither method drops significantly below the uniform portfolio. After 2008, when the market goes into a bullish phase, our approach takes the advantage and starts to outperform the index and the uniform portfolio significantly. The table

4.2 shows the performance of each portfolio construction strategy using key performance metrics like Cumulative return, average daily return, portfolio volatility, Sharpe Ratio, annual volatility, and max drawdown.

TABLE 4.2: Performance metrics for different strategies.

Metric/Strategy	Uniform	Screener	FRMW	FRMW with DT
CR	12.0100	121.1100	136.1900	239.6900
Volatility	2.68%	4.21%	4.78%	4.43%
SR	12.6592	87.9192	87.1180	166.0047
Average AR	13.45%	26.55%	30.20%	34.92%
Average AV	0.31%	0.59%	0.71%	0.64%
MAX_DD	-53.92%	-52.93%	-56.41%	-37.94%

Table 4.2, shows different portfolio measurement of these four portfolio constructions. Comparing cumulative return (CR) and average annual return (Average AR), FRMW with DT has by far the best performance. An interesting observation is that Screener and FRMW constructions have almost equivalent returns. This indicates that the inclusion of downturn trigger has a significant impact for avoiding unnecessary losses. This is confirmed also by the maximum drawdown results (maxDD) where FRMW with DT has the smallest value due to the DT trigger. Portfolio volatility and average AV is similar all portfolio strategies except the uniform, where is the half.

The cumulative return of the portfolio of the best strategy (FRMW with DT) is shown in the Figure 4.2 below. The comparison with the S&P500 index and uniform portfolio shows the dominant performance of the proposed strategy.

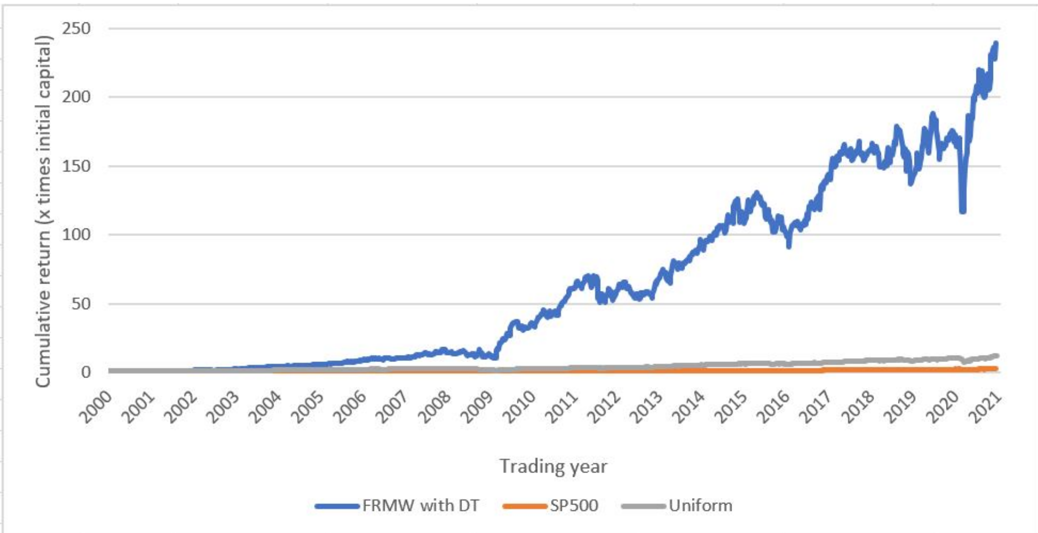

FIGURE 4.2: Suggested portfolio strategy (FRMW + DT) for the same period.

A more thorough look at drawdown and the performance of this framework's strategy are shown in Figure 4.3. The worst ten recession periods were tested based on the drawdown on the S&P500 and tested how the proposed framework performed in these situations.

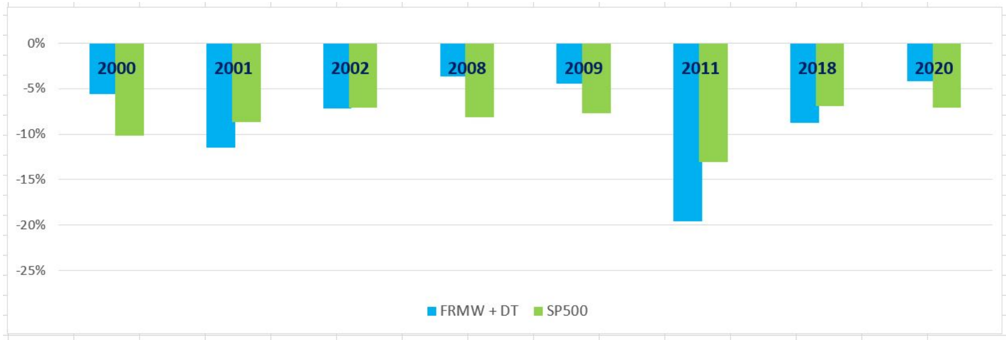

FIGURE 4.3: Drawdown comparison with S&P500 index.

As depicted in Figure 4.3, the FRMW+DT in blue has less drawdown in the most significant recessions of the last 20 years. For the big crisis of 2008-2009 and the coronavirus situation of 2020, the FRMW+DT outperformed the index significantly, producing high returns. It does not perform well under the other half of the cases. Overall, the FRMW+DT is similar and not worst than S&P500, where the risk is significantly lower in terms of drawdown.

4.2 Comparing with other strategies from the literature

This thesis attempts to shed light on the factors that contribute the most to good portfolio construction and propose a framework as investment advice for portfolio construction and risk/return optimization. While other studies and strategies use more sophisticated machine learning techniques and technical indicator analysis. We will see that - in some cases - our approach performs better.

In [18], the authors provide a portfolio construction strategy based on the classification of a stock's trend (either upwards or downwards trend) using a technical indicators as feature space. That approach was tested on 26 popular stocks from S&P500 from the year of 2000 until the end of 2019. The cumulative return are shown in Figure . The cumulative return reached in 20 years is up to 1500%. The result is three times better compared to the return of the equal-weight portfolio. To compare the performance of our own approach with that [18], a simulation of the same period was performed. Since the exact testing period is not explicit stated in [18] we performed a check by using different starting dates of 2000 using the S&P500 and S&P500 results documented in [18].

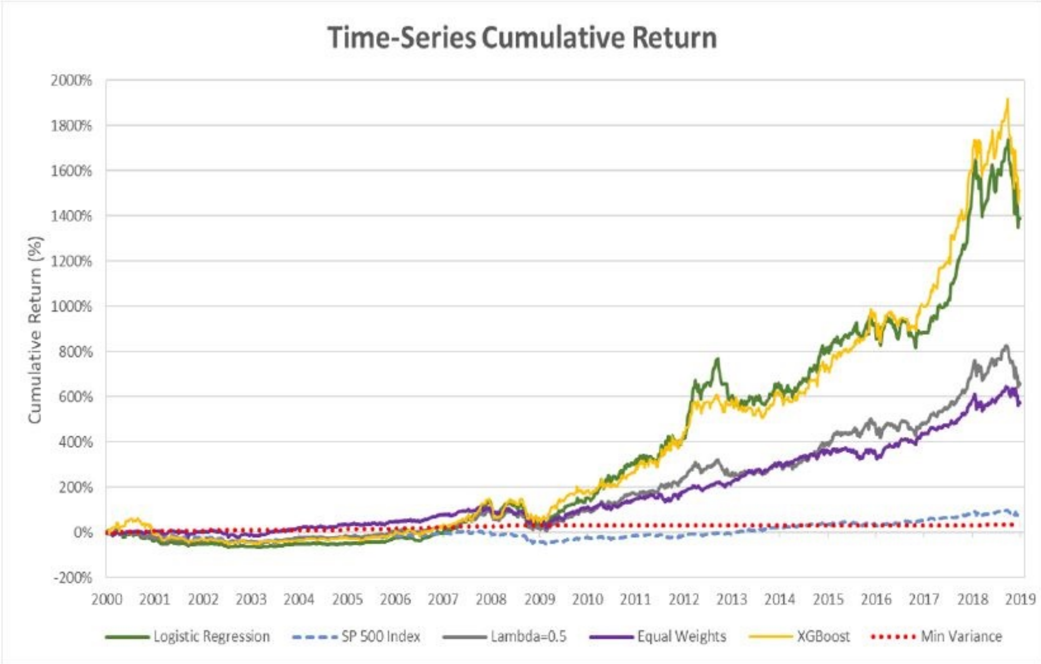

FIGURE 4.4: Time-Series Cumulative Returns comparison of different strategies from [18] [2].

Additionally, Table 4.3 shows very similar results between [18] and our trading simulation, indicating a "correct" approximation for the testing period used. The reproduced period starts at 19/07/2000 and ends at 20/12/2018.

TABLE 4.3: Simulation using the [18] testing period.

Metric	**S&P500([18])**	**S&P500 (Reproduced)**
Average Return	0.08%	0.09%
Standard Deviation	2.42%	2.43%
Sharpe Ratio	0.04	0.04
Cumulative Return	66.61%	68.00%
Annualized Return	2.72%	3.00%

Table 4.3 compares the best strategy based on cumulative return of [18] (BEST) with the best strategy of our proposal (FRMW +DT) along with the corresponding equal-weighted portfolios (EqW). The EqW portfolio of our proposed strategy is similar with [18] EqW portfolio. However, the cumulative results of our strategy is by far greater than the [18] best strategy. The Standard Deviations is slightly higher in our approach but compared with the return using the Sharpe Ratio our approach is superior.

Using our proposed strategy, and for the same period, the cumulative return was 13,503%, which is 18 times higher compared to the equal-weighted portfolio. Our FRMW+DT proposed strategy outperforms the [18] proposal for the same period in all performance metrics. Note that all performance metrics are as defined in the [18]. A comparison of the best strategy of each proposal along with an equal weight portfolio is presented in Table 4.4.

TABLE 4.4: Performance comparison with Jiang's proposal.

Metric	[18] BEST	FRMW+DT	[18] EqW	EqW
Average Return	0.33%	0.63%	0.22%	0.27%
Standard Deviation	3.68%	4.40%	2.29%	2.61%
Sharpe Ratio	0.0970	0.1439	0.0955	0.1022
Cumulative Return	1511.73%	13503.00%	574.16%	732.00%
Annualized Return	15.28%	32.33%	10.57%	12.84%

Figure 4.5 shows the cumulative return of FRMW+DT during the same test period used in [18]. This is the equivalent graph of this thesis framework for comparison with [18] result in Figure 4.4. The trend looks the same, but the overall cumulative results are on different scale, with our approach reaching more than 140x the initial invested capital.

FIGURE 4.5: Suggested portfolio strategy (FRMW + DT) for the testing period of [18].

In [30], their stock selection strategy was based on the Hidden Markov model (HMM). Using the HMM, the authors predicted monthly regimes for four core macroeconomic factors: inflation, industrial production index, stock market index, and market volatility. The regimes are the two opposite states of each variable (eg yes/no, up/down, high/low). The predicted regime is then compared with similar historical behaviour. Stock characteristics that performed well in similar historical periods are assigned with weights. A composite score of each stock is calculated based on the scores that are a mixture of stock factors following its behaviour in the previous similar macroeconomic stage. These factors include price/earnings ratio, free cash flow/enterprise value, sales/enterprise value, long-term earnings-per-share growth, and long-term sales growth. The portfolio is then constructed by selecting the 50 top-ranking stocks to buy and was updated every month. The Figure 4.6 shows the aggregated return of the out-of-sample tested period (December 1999 - December 2014). The Figure 4.6 shows on top the trading year on x-axis and the portfolio value on y-axis, with initial investment capital of 100$ (blue line) and the

S$P500 performance (red line). The bottom subgraph show the log excess cumulative return of their approach. The top subgraph was used for the comparison with our approach.

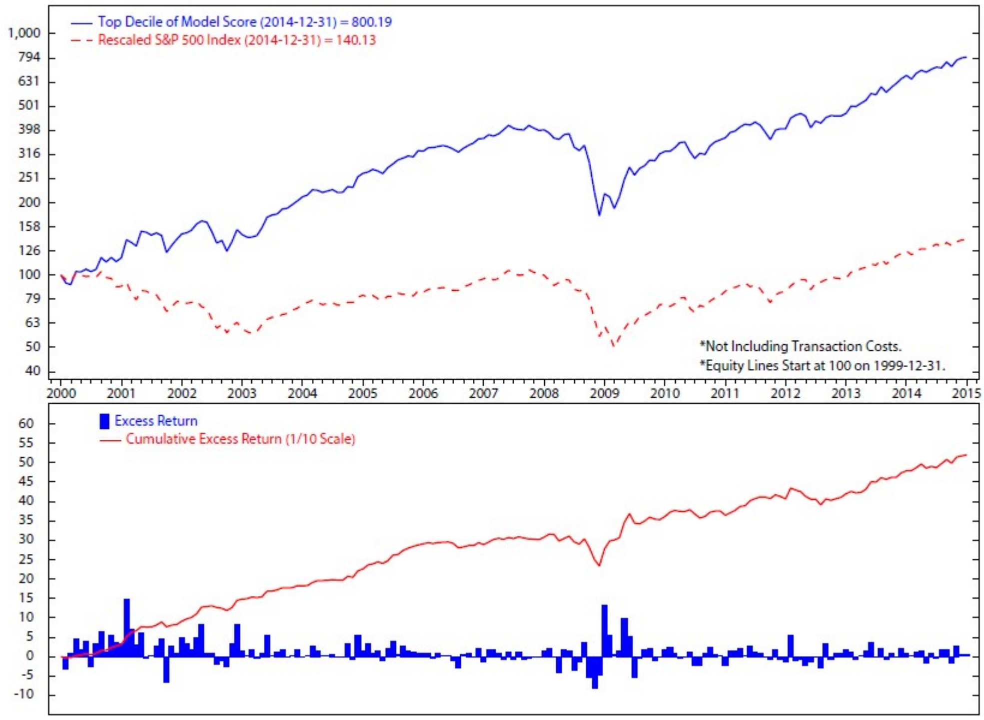

FIGURE 4.6: HMM based portfolio selection strategy [30] [3].

Figure 4.6 shows the approach suggested in [30] works quite well, and in the long rung of 15 years (December 1999 to December 2014), the return reaches 7.94x the initial capital invested (bottom graph). Compared with FRMW+DT, we notice better performance in the latter, with the excess return exceeding 80x the initial capital. The graphical performance of our proposed FRMW+DT approach during the same period is shown in the Figure 4.7.

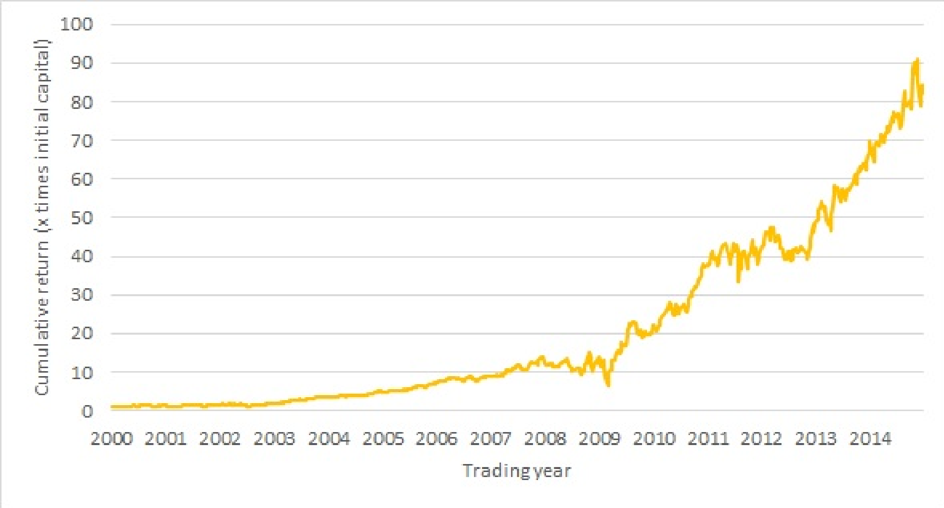

FIGURE 4.7: Suggested portfolio strategy (FRMW + DT) using testing period as of [30] until end of 2014.

The authors of [22] proposed a novel kernel-based trend pattern tracking (KTPT) system for portfolio optimization that uses three stages for portfolio creation. Their work starts from a simple prediction of the stock price based on historical inertia and the maximum p. Then, the authors adjusted the initial prediction by incorporating historical financial states and including a trend-reverting factor. Tested and compared with many benchmark datasets and portfolio selection methods, they showed that their method could beat previous stock selection approaches. In [22], there is a comparison with many other proposed stock selection strategies from the literature in their work. They used 5 datasets as benchmarks with historical close price for each stock. The comparison shows that KPTP outperforms all others and is considered the state-of-the-art strategy.

This thesis's proposed strategy could not be compared with these portfolio selection systems due to the following reasons and issues:

- The five benchmarks are small-sized datasets regarding the number of stocks included (n<20). This means that our screening method was not applicable.
- These strategies are not considering any risk management in their optimization strategy. The best performing approach is the KTPT. However, this approach selected on average for the portfolio construction 1-2 stocks on average. [22].
- All the stocks in the benchmark datasets suffer from survivor bias. This is indicated by the market returns (equal-weighted portfolio of the data), where the cumulative return is enormous (18.06 absolute cumulative return when the initial capital is 1). More importantly, all stocks in the datasets are survivors, with historical information covering the whole horizon. For some, this is up to 20 years (NYSE(N) dataset).

Chapter 5

Conclusions

5.1 Summary and conclusions

Our study aimed to design a simple, scalable, and efficient portfolio optimization prototype on S&P500 listed stocks. The framework was backtested for the period 2000 - 2020 (full 20 year period) and 2003 - 2020 (testing period). We extracted historical stock data to reduce the effect of survivor bias as much as possible by incorporating the insertion date and exit date for stocks from the index. After that, a screener was created for the pool of stocks, looking for candidates. Then, by performing a mean-risk aversion technique based on the correlation of stocks, we constructed a portfolio that will be sold five days after the purchase day. This process is repeated until the end date of the sample. The results are satisfying, with cumulative return exceeding the index and being significantly better than many other optimization strategies in the literature.

5.2 Research limitations

The first limitation of this study is the dataset for the backtesting. Although the sample was constructed to reduce as much as possible the survivor bias effect, it is still present in the data. The reason is that the full list of changes of S&P500, delisted stocks and the merges or renames are not all publicly available in open sources on the internet. The second restriction is hardware/computational power. More complex, sophisticated methods (complex tree-based algorithms or neural network architectures) are examined but without producing better results than this proposed strategy. Lastly, the stock market simulation was assumed to be a frictionless market, with no transaction costs and buy and sell options available anytime. There are many strategies implemented, with more realistic market simulations ([16], [15], [32], [2]).

5.3 Future extensions

This framework is simple, efficient, beats the market, and can run with minimum hardware requirements. The screener's parameters are tuned only at the beginning of the backtesting. The framework could perhaps perform better if these parameters (the window and the method) were adjusted weekly instead of fixed values. Moreover, adding well-known technical indicators for the stock screening may also increase the effectiveness of stock selection. Other risk-minimization techniques introduced in Section 2.2 could be examined, evaluated, and used in terms of optimizing portfolio weights. Although the downturn trigger is simple and effective, it would perform better if someone could incorporate rumors and news feed from the media, using NLP techniques and machine learning.

Chapter 6

Appendix

6.1 Out-of-time results

Cumulative returns of these strategies mentioned above are shown over time in the testing period of 2003 - 2020 in the Figure 6.1. This period is out-of-time sample. The results, although starting from 2003, dominate the index and the other compared strategies from literature even-though these evaluated from 2000). Comparison exist in 4.

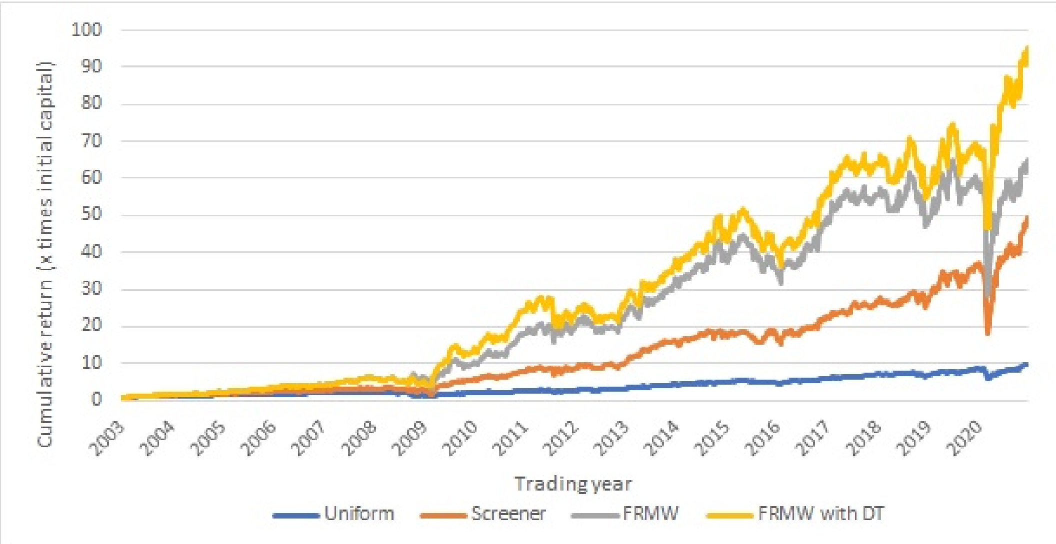

FIGURE 6.1: OOT weekly cumulative returns using 4 portfolio construction methods.

The results above show that Screener + Min Volatility + DT trigger (yellow line) can outperform all the other strategies. Compared with the uniform portfolio, the framework's return is up to 10 times using cumulative returns as a performance metric. The framework without the DT trigger (grey line) performs worse than the DT trigger included. This indicates that the DT trigger positively impacts the overall performance and leads to better returns when applied as a safety/risk measure to the portfolio strategy. The screener (orange line) performs up to 5 times better than the uniform portfolio. Nevertheless, it is only 2 times worst than the full framework strategy.

TABLE 6.1: Final results as of last trading date.

Date	Day	Uniform	Screener	FRMW without DT	FRMW with DT
18/12/2020	5781	869.00%	4843.00%	6381.00%	9418.00%

In the Figure 6.1, an upward trend is shown after the year 2008. In the period before, the market was very stable. The framework also shows a horizontal trend. However, both methods are not significantly dropping below the uniform portfolio. After 2008, when the market goes into a bullish phase, the framework takes the advantage and starts to outperform the index and the uniform portfolio significantly. The table below shows the performance of each portfolio construction strategy using key performance metrics like Cumulative return, average daily return, portfolio volatility, Sharpe Ratio, annual volatility, and max drawdown.

TABLE 6.2: Performance metrics for different strategies.

Metric/Strategy	Uniform	Screener	FRMW	FRMW+DT	S&P500
CR	8.6900	48.4300	63.8100	94.1800	3.0200
Return	1.36%	4.95%	5.21%	5.19%	1.01%
Volatility	0.0264	0.0413	0.0457	0.0416	0.0233
SR	10.9347	38.9845	46.4526	75.3167	4.3157
AR	14.81%	35.91%	7.19%	36.42%	16.86%
AV	0.0038	0.0078	0.0073	0.0040	0.0022
DD	0.0000	0.0000	-0.0011	0.0000	0.0000
MAX_DD	-54.01%	-53.12%	-56.39%	-37.94%	-56.80%

The cumulative return of the portfolio of the best strategy (FRMW_DT) is shown in the Figure 6.2 below. The comparison with the S&P500 index and uniform portfolio shows the dominant performance of the proposed strategy.

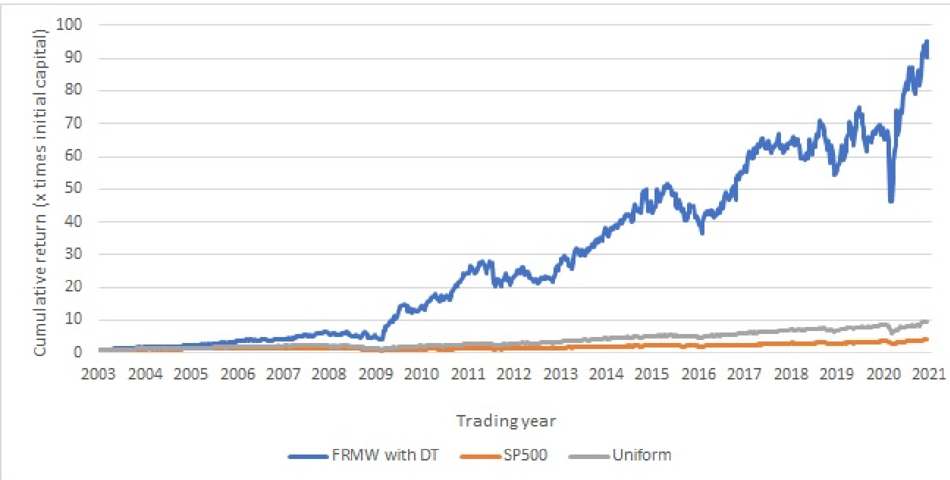

FIGURE 6.2: Suggested portfolio strategy (FRMW + DT) for OOT period.

6.2 Stock ticker list

Table 6.3 shows all 545 tickers of stocks that used for the backtest sample of this thesis.

TABLE 6.3: List of stock tickers used in dataset.

AAPL	ARG	CHRW	DLPH	FISV	HSY	LEN	MSI	PH	SNPS	URI
A	ARNC	CHTR	DLR	FITB	HUM	LH	MTB	PHM	SO	USB
AAL	ATO	CI	DLTR	FL	HWM	LHX	MTD	PKG	SPG	V
AAP	ATVI	CINF	DOV	FLT	IBM	LIN	MU	PKI	SPGI	VFC
ABBV	AVB	CL	DPZ	FMC	ICE	LKQ	NAVI	PLD	SRE	VIAC
ABC	AVGO	CLF	DRE	FOSL	IDXX	LLY	NCLH	PM	STE	VLO
ABMD	AVY	CLX	DRI	FOX	IEX	LMT	NDAQ	PNC	STT	VMC
ABT	AWK	CMA	DTE	FOXA	IFF	LNC	NEE	PNR	STX	VNO
ACN	AXP	CMCSA	DUK	FRC	ILMN	LNT	NEM	PNW	STZ	VRSK
ADBE	AYI	CMCSK	DVA	FRT	INCY	LO	NFLX	PPG	SWK	VRSN
ADI	AZO	CME	DVN	FTNT	INFO	LOW	NFX	PPL	SWKS	VRTX
ADM	BA	CMG	DWDP	FTV	INTC	LRCX	NI	PRGO	SYF	VTR
ADP	BAC	CMI	DXC	GD	INTU	LUK	NKE	PRU	SYK	VTRS
ADS	BAX	CMS	DXCM	GE	IP	LUMN	NKTR	PSA	SYY	VZ
ADSK	BBWI	CNC	EA	GGP	IPG	LUV	NLOK	PSX	T	WAB
ADT	BBY	CNP	EBAY	GILD	IPGP	LVLT	NLSN	PVH	TAP	WAT
AEE	BDX	COF	ECL	GIS	IQV	LVS	NOC	PWR	TDC	WBA
AEP	BEN	COG	ED	GL	IR	LW	NOV	PXD	TDG	WDC
AES	BF_B	COO	EFX	GLW	IRM	LYB	NOW	PYPL	TDY	WEC
AFL	BHF	COP	EIX	GM	ISRG	LYV	NRG	QCOM	TEL	WELL
AIG	BIIB	COST	EL	GMCR	IT	MA	NSC	QEP	TFC	WFC
AIZ	BIO	COTY	EMN	GME	ITW	MAA	NTAP	QRVO	TFX	WHR
AJG	BK	COV	EMR	GOOG	IVZ	MAC	NTRS	RCL	TGT	WLTW
AKAM	BKNG	CPB	ENDP	GOOGL	J	MAR	NUE	RE	TJX	WM
ALB	BKR	CPGX	EOG	GPC	JBHT	MAS	NVDA	REG	TMO	WMB
ALGN	BLK	CPRT	EQIX	GPN	JCI	MCD	NVR	REGN	TMUS	WMT
ALK	BLL	CRM	EQR	GPS	JDSU	MCHP	NWL	RF	TPR	WPX
ALL	BMY	CSCO	ES	GRMN	JEC	MCK	NWS	RHI	TRIP	WRB
ALLE	BR	CSRA	ESRX	GS	JKHY	MCO	NWSA	RIG	TROW	WRK
ALXN	BSX	CSX	ESS	GWW	JNJ	MDLZ	O	RJF	TRV	WST
AMAT	BWA	CTAS	ESV	HAL	JNPR	MDT	ODFL	RL	TSCO	WU
AMCR	BXLT	CTSH	ETN	HAS	JOY	MET	OKE	RMD	TSN	WY
AMD	BXP	CTVA	ETR	HBAN	JPM	MGM	OMC	ROK	TSO	WYNN
AME	C	CTXS	EVHC	HBI	K	MHK	ORCL	ROL	TT	XEC
AMG	CAG	CVC	EVRG	HCA	KEY	MJN	ORLY	ROP	TTWO	XEL
AMGN	CAH	CVS	EW	HD	KEYS	MKC	OTIS	ROST	TWTR	XLNX
AMP	CARR	CVX	EXC	HES	KHC	MKTX	OXY	RRC	TXN	XOM
AMT	CAT	CXO	EXPD	HFC	KIM	MLM	PAYC	RSG	TXT	XRAY
AMZN	CB	D	EXPE	HIG	KLAC	MMC	PAYX	RTX	TYC	XYL
ANET	CBOE	DAL	EXR	HII	KMB	MMM	PBCT	SBAC	TYL	YHOO
ANR	CBRE	DE	F	HLT	KMI	MNK	PCAR	SBUX	UA	YUM
ANSS	CCI	DFS	FANG	HOLX	KMX	MNST	PCLN	SCHW	UAA	ZBH
ANTM	CCL	DG	FAST	HON	KO	MO	PEAK	SEE	UAL	ZBRA
AON	CDNS	DGX	FB	HP	KORS	MOS	PEG	SHW	UDR	ZION
AOS	CDW	DHI	FBHS	HPE	KR	MPC	PEP	SIG	UHS	ZTS
APA	CE	DHR	FCX	HPQ	KRFT	MRK	PETM	SIVB	ULTA	
APD	CERN	DIS	FDX	HRL	KSU	MRO	PFE	SJM	UNH	
APH	CF	DISCA	FE	HRS	L	MS	PFG	SLB	UNM	
APTV	CFG	DISCK	FFIV	HSIC	LDOS	MSCI	PG	SLG	UNP	
ARE	CHD	DISH	FIS	HST	LEG	MSFT	PGR	SNA	UPS	

Bibliography

[1] Amir Ahmadi-Javid. "Entropic value-at-risk: A new coherent risk measure". In: *Journal of Optimization Theory and Applications* 155.3 (2012), pp. 1105–1123.

[2] Marianne Akian, Agnes Sulem, and Michael I Taksar. "Dynamic Optimization of Long-Term Growth Rate for a Portfolio with Transaction Costs and Logarithmic Utility". In: *Mathematical Finance* 11.2 (2001), pp. 153–188.

[3] Fischer Black and Myron Scholes. "The pricing of options and corporate liabilities". In: *World Scientific Reference on Contingent Claims Analysis in Corporate Finance: Volume 1: Foundations of CCA and Equity Valuation*. World Scientific, 2019, pp. 3–21.

[4] Allan Borodin, Ran El-Yaniv, and Vincent Gogan. "Can we learn to beat the best stock". In: *Advances in Neural Information Processing Systems* 16 (2003).

[5] Werner FM De Bondt and Richard Thaler. "Does the stock market overreact?" In: *The Journal of finance* 40.3 (1985), pp. 793–805.

[6] Victor DeMiguel et al. "A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms". In: *Management science* 55.5 (2009), pp. 798–812.

[7] Peter A Diamond. "What stock market returns to expect for the future". In: *Soc. Sec. Bull.* 63 (2000), p. 38.

[8] Darrell Duffie and Jun Pan. "An overview of value at risk". In: *Journal of derivatives* 4.3 (1997), pp. 7–49.

[9] Frank J Fabozzi, Dashan Huang, and Guofu Zhou. "Robust portfolios: contributions from operations research and finance". In: *Annals of operations research* 176.1 (2010), pp. 191–220.

[10] Eugene F Fama. "Efficient capital markets: A review of theory and empirical work". In: *The journal of Finance* 25.2 (1970), pp. 383–417.

[11] Eugene F Fama and Kenneth R French. "Size and book-to-market factors in earnings and returns". In: *The journal of finance* 50.1 (1995), pp. 131–155.

[12] Wayne E Ferson and Campbell R Harvey. "The risk and predictability of international equity returns". In: *Review of financial Studies* 6.3 (1993), pp. 527–566.

[13] Ramazan Gencay. "Non-linear prediction of security returns with moving average rules". In: *Journal of Forecasting* 15.3 (1996), pp. 165–174.

[14] Steven C Gold and Paul Lebowitz. "Computerized stock screening rules for portfolio selection". In: *Financial services review* 8.2 (1999), pp. 61–70.

[15] László Györfi and István Vajda. "Growth optimal investment with transaction costs". In: *International conference on algorithmic learning theory*. Springer. 2008, pp. 108–122.

[16] Garud N. Iyengar and Thomas M Cover. "Growth optimal investment in horse race markets with costs". In: *IEEE Transactions on Information Theory* 46.7 (2000), pp. 2675–2683.

[17] Narasimhan Jegadeesh and Sheridan Titman. "Returns to buying winners and selling losers: Implications for stock market efficiency". In: *The Journal of finance* 48.1 (1993), pp. 65–91.

[18] Zhenlong Jiang, Ran Ji, and Kuo-Chu Chang. "A machine learning integrated portfolio rebalance framework with risk-aversion adjustment". In: *Journal of Risk and Financial Management* 13.7 (2020), p. 155.

[19] C.D. Kirkpatrick, J. Dahlquist, and J.R. Dahlquist. *Technical Analysis: The Complete Resource for Financial Market Technicians*. Pearson Education, Incorporated, 2016. ISBN: 9780134137049. URL: https://books.google.gr/books?id=Ks1GrgEACAAJ.

[20] Charles D Kirkpatrick II and Julie A Dahlquist. *When You Can Use Technical Analysis for Investing: Whe Y Can Us Tec ePub_1*. Pearson Education, 2010.

[21] Hiroshi Konno and Hiroaki Yamazaki. "Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market". In: *Management science* 37.5 (1991), pp. 519–531.

[22] Zhao-Rong Lai et al. "A kernel-based trend pattern tracking system for portfolio optimization". In: *Data Mining and Knowledge Discovery* 32.6 (2018), pp. 1708–1734.

[23] Richard M Levich and Lee R Thomas III. "The significance of technical trading-rule profits in the foreign exchange market: a bootstrap approach". In: *Journal of international Money and Finance* 12.5 (1993), pp. 451–474.

[24] Bin Li et al. "Confidence weighted mean reversion strategy for on-line portfolio selection". In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings. 2011, pp. 434–442.

[25] *List of S&P 500 companies*. URL: https://en.wikipedia.org/w/index.php?title=List_of_S\%26P_500_companies&oldid=1042268881. (accessed: 08.09.2021).

[26] Burton. G. Malkiel. *A Random Walk Down Wall Street*. Norton, New York, 1973.

[27] *Market Index*. URL: https://www.forbes.com/advisor/investing/stock-market-index/. (accessed: 27.05.2022).

[28] H Markowitz. "The Semi-Variance". In: *Chapter IX, Portfolio Selection. Efficient Diversification of Investments* (1959), pp. 188–201.

[29] Harry Markowitz. "Portfolio Selection". In: *Portfolio Selection - Harry Markowitz.pdf - Portfolio Selection Author(s Harry Markowitz Source The Journal of Finance Mar 1952 Vol 7 No 1(Mar 1952 pp | Course Hero* (1952). URL: https://www.coursehero.com/file/119015085/Portfolio-Selection-Harry-Markowitzpdf/.

[30] Nguyet Nguyen and Dung Nguyen. "Global stock selection with hidden Markov model". In: *Risks* 9.1 (2020), p. 9.

[31] Włodzimierz Ogryczak and Andrzej Ruszczyński. "From stochastic dominance to mean-risk models: Semideviations as risk measures1This work was partially supported by the International Institute for Applied Systems Analysis, Laxenburg, Austria, and distributed on-line as the IIASA Report IR-97-027 (June 1997).1". In: *European Journal of Operational Research* 116.1 (1999), pp. 33–50. ISSN: 0377-2217. DOI: https://doi.org/10.1016/S0377-2217(98)00167-2. URL: https://www.sciencedirect.com/science/article/pii/S0377221798001672.

[32] Mihály Ormos and András Urbán. "Performance analysis of log-optimal portfolio strategies with transaction costs". In: *Quantitative Finance* 13.10 (2013), pp. 1587–1597.

[33] Ratchata Peachavanish. "Stock selection and trading based on cluster analysis of trend and momentum indicators". In: *Proceedings of the International Multi-Conference of Engineers and Computer Scientists*. Vol. 1. IMECS. 2016, pp. 317–321.

[34] M Hashem Pesaran and Allan Timmermann. "Predictability of stock returns: Robustness and economic significance". In: *The Journal of Finance* 50.4 (1995), pp. 1201–1228.

[35] Joseph D Piotroski. "Value investing: The use of historical financial statement information to separate winners from losers". In: *Journal of Accounting Research* (2000), pp. 1–41.

[36] R Tyrrell Rockafellar, Stanislav Uryasev, et al. "Optimization of conditional value-at-risk". In: *Journal of risk* 2 (2000), pp. 21–42.

[37] Pavel Sevastjanov and Ludmila Dymova. "Stock screening with use of multiple criteria decision making and optimization". In: *Omega* 37.3 (2009), pp. 659–671.

[38] Haim Shalit and Shlomo Yitzhaki. "Mean-Gini, portfolio theory, and the pricing of risky assets". In: *The journal of Finance* 39.5 (1984), pp. 1449–1468.

[39] William F. Sharpe. "The Sharpe Ratio". In: *The Journal of Portfolio Management* 21.1 (1994), 49–58. DOI: 10.3905/jpm.1994.409501.

[40] *SP500 overview*. URL: https://www.spglobal.com/spdji/en/indices/equity/sp-500/#overview. (accessed: 12.07.2022).

[41] Zheng Tan, Ziqin Yan, and Guangwei Zhu. "Stock selection with random forest: An exploitation of excess return in the Chinese stock market". In: *Heliyon* 5.8 (2019), e02310.

[42] *This Year's Laureates Are Pioneers In The Theory of Financial Economics and Corporate Finance*. URL: https://www.nobelprize.org/prizes/economic-sciences/1990/press-release/. (accessed: 29.12.2021).

[43] *U.S. Securities and Exchange Commission. 'Stocks'*. URL: https://www.investor.gov/introduction-investing/investing-basics/investment-products/stocks. (accessed: 18.08.2020).

[44] M Mark Walker and Gay B Hatfield. "Professional stock analysts' recommendations: Implications for individual investors". In: *Financial Services Review* 5.1 (1996), pp. 13–29.

[45] *Yahoo Finance*. URL: https://finance.yahoo.com/1. (accessed: 22.01.2022).

[46] Fengmei Yang et al. "A novel hybrid stock selection method with stock prediction". In: *Applied Soft Computing* 80 (2019), pp. 820–831.