Equity Sector Rebalancing via Machine Learning

Tiago Ramos

Dissertation written under the supervision of professor Dan Tran

Dissertation submitted in partial fulfilment of requirements for the MSc in Finance, at the Universidade Católica Portuguesa, 03/01/2023.

Abstract

In this dissertation the author will analyze whether supervised machine learning models namely Artificial Neural Networks, Support Vector Machines and Logistic Regressions can predict shifts in equity returns on a sector basis. Typically, in asset pricing linear factor models with a small number of variables are used. However, due to market efficiency, equity returns are highly influenced by unforecastable events making this task more challenging. Simple linear regressions also have difficulty incorporating a larger number of predictor variables, which the literature has accumulated over the decades, creating an opportunity for machine learning techniques.

The Machine Learning models will be used to forecast whether the excess return of each equity sector over a period of one month will be positive or negative. Then using the model's predictions capital will be allocated between the sectors and treasury bonds, building different portfolios namely an equal weighted, a value weighted portfolio. After all portfolios are built their performance will then be compared against the benchmark, namely the S&P500 index being back tested over a period of 25 years.

The portfolios built using the forecasts from the ML models lead to an increase in absolute and risk-adjusted returns beating the benchmark. The implemented strategies were shown to protect investors against larger market declines, showing the potential of Machine Learning as an investment tool.

Tittle: Equity Sector Rebalancing via Machine Learning

Author: Tiago Coutinho Ramos

Keywords: Sector Allocation, Market Timing, Machine Learning, Logistic Regressions, Support Vector Machines, Neural Networks

Resumo

Nesta dissertação o autor vai analisar se os modelos de machine learning nomeadamente Redes Neurais Artificiais, Máquina de Vetores de Suporte e Regressões Logísticas, conseguem prever mudanças nos retornos dos vários setores de mercado. Tipicamente, na definição do preço de ativos, são usados modelos de fatores lineares com um pequeno número de variáveis. Contudo, devido á eficiência do mercado, os retornos de ações são influenciados por eventos imprevisíveis aumentando a complexidade do problema. As regressões lineares simples têm dificuldade em incorporar um número vasto de variáveis, que a literatura Financeira veio a acumular ao longo das décadas, criando uma oportunidade para técnicas de machine learning.

Os modelos de Machine Learning serão utilizados para realizar previsões sobre se o retorno de excesso sobre o período de um mês será positivo ou negativo. Utilizando as previsões dos modelos, o capital será alocado entre os vários setores do mercado e obrigações de tesouraria, construindo diferentes portfolios. Estando os portfolios construídos a respetiva performance será avaliada e comparada contra o benchmark, nomeadamente o índice do S&P500, durante um período de 25 anos.

Os portfolios contruídos usando as previsões dos modelos de ML levaram a um aumento de retornos absolutos e ajustados ao risco batendo o benchmark. As estratégias teriam protegido investidores contra quedas acentuadas do mercado, mostrando o potencial de Machine Learning como ferramenta de investimento.

Título: Rebalanceamento de Setores de Investimento através de Machine Learning

Autor: Tiago Coutinho Ramos

Palavras-Chave: Alocação por Setor, Timing de Mercado, Machine Learning, Regressão Logística, Máquina de Vetores de Suporte, Redes Neurais Artificiais

Table of Contents

List of Tables

List of Figures

1. Introduction

In the beginning of 2022 heavy expectations existed regarding the re-opening of the economy on a global scale, with several restrictions put in place by Covid-19 expected to disappear consequently easing supply chain issues and inflation. However, the higher inflationary environment witnessed, which was first believed to be transitory, was feared to prolong longer than expected and desired. In order to keep inflation under the acceptable target level, interest rates have steadily increased globally, leading to shocks and higher volatility in assets prices from equity, bonds and even more noticeably in cryptocurrencies. Different equity sectors have also seen varying impacts to both their business fundamentals as well as their valuations.

In 2022 we have seen the Technology sector, the best performer over the past decade, greatly underperform the market. On the other hand, the energy sector, one of the worst performers over the same period, has witnessed one of its best years increasing in value by more than 50%. Historically, in periods of higher inflation value stocks have tended to outperform growth companies over the past century. These changes in regime are inherent to financial markets, and has been shown, when accounted for can increase our understanding and forecast ability of stock returns (Pesaran and Timmermann, 2002). The focus of this dissertation will be around developing trading strategies based on machine learning models, focusing on whether the excess returns in a period of 22 days are positive or negative, for each of the 10 stock market sectors, which compose the S&P500.

Every year the stock market creates massive amounts of data, which is hard to reliably interpret without proper computer-based tools and models. In finance one of the key topics helping us make sense of the data is asset pricing, which has the goal to understand and measure the expected stock returns in excess of the risk-free rate. Literature on stock return forecasting typically is divided into two methodologies. The first conducts regressions of aggregate portfolios return on several predictor variables linked to the economy, which we suspect explains broad market movements. The second approach is to perform a regression of individual stock returns on individual characteristics such as price-to-earnings or dividend yield. This dissertation will focus on the former, considering that the implementation of strategies using many securities is more costly due to transaction fees and will also be harder to implement for the average investor.

Risk premiums are known to be difficult to forecast, considering they are highly influenced by unforecastable news, creating an opportunity for Machine Learning that has essentially two great benefits for market timing. First, it allows for the usage of a greater number of predictor variables or features than in comparison with traditional return forecasting models, which typically collapse when the number of observations approximates to the number of independent variables. Secondly, machine learning techniques are also more flexible by allowing the usage of non-linear models, which better approximate the true relationship between the features and the dependent variables. These unique characteristics have increased interest in machine learning applications in finance in both a theoretical framework as well as in practical settings. In the paper by Giglio, Kelly and Xiu (2021) the authors give a in depth overview over recent findings in asset pricing using machine learning, which this thesis will use as a guide when building the models. Using the forecasts from the models, several portfolios were formed allocating capital between 10-year treasury bonds and the index of each respective equity sector.

This dissertation will address the following research questions:

Q1: Are there advantages of performing market timing strategies on a sector basis?

Q2: Do market timing strategies lead to similar performance gains across all sectors?

Q3: Can portfolios built using machine learning models be useful in protecting investors against periods of larger market declines?

In order to answer these questions this thesis will be organized as follows. Firstly, asset pricing literature will be overviewed alongside its challenges. Secondly, the data used to create the models, namely the chosen predictor variables will be described alongside the respective literature. Afterwards, the research methodology will be explained including the respective models used to perform the predictions. Finally, the results will be thoroughly presented alongside a conclusion and final remarks.

Machine learning models' literature in Asset pricing is still in its early stages, posing an interesting topic, which may lead to some new insight. The main models used will be neural networks and regression trees, which have been found to be the best performers (Gu, Kelly and Xiu, 2020). Other models, such as a logistic regression will also be used for comparison purposes.

2. Literature Review

Research into market timing strategies has been ongoing for decades, across several asset classes with its feasibility still being discussed. When mentioning market timing, it is typically referred to the practice of forecasting the direction of the market. In other words, predicting whether a certain asset is going to gain or lose value over a certain time period. The strategy itself has been experimented with and implemented in a variety of ways. One common case is for the portfolio managers to either take a long position in individual stocks or in a broad index fund when the market is expected to rise while investing in short-term bonds when it is expected to decline. In theory it seems straightforward, however Jensen (1968) found that not only do individual funds underperform a simple buy and hold strategy, but there were little signs they performed better than a simple coin toss. Decades later Malkiel (1995) arrived at a similar conclusion using data from all the funds in existence between 1971 and 1991.The author finds that mutual funds underperform the market both before and after accounting for management expenses pointing out investors would be better off investing in index funds. These early findings create a competition between passive and active management. Passive portfolio management as the goal of obtaining an adequate level of return for a certain level of risk with as minimal rebalancing as possible. On the other hand, a more active approach has the assumption that it is possible for the manager to exploit certain market inefficiencies.

The efficient market hypothesis (EMH) tells us that at any point in time prices should fully reflect all available information (Fama, 1970). This theory poses as an explanation to why actively managed funds have such difficulty beating the market, especially when considering transaction and management fees. However it was found that fund´s performance when evaluated using higher frequency trading data showed persistent positive returns (Puckett and Yan, 2011). They argue that institutional investors can take advantage of short-term information advantages when implementing their strategies.

Several arguments opposing the EMH can be found in field of behavioral finance. De BONDT and Thaler (1985) show that overreaction to unexpected news events, can affect decision making in the market, with portfolios formed of securities considered to be losers, outperformed portfolios with prior winners. Daniel, Hirshleifer and Subrahmanyam (1998) point out that, investors own psychological biases are responsible for several documented anomalies with investors tending to overreact to private information signals, while underreacting to public signals. In cases where market anomalies are identified they tend to not hold up in different time periods and weaken or even disappear after being diffused through academic literature (Schwert, 2003). This finding is corroborated by more recent papers such as Mclean and Pontiff (2016) who show that OOS performance declines post paper publication. These discoveries are consistent with the idea that when anomalies are commonly known, investors will quickly try to take advantage causing said anomalies to disappear increasing market efficiency.

An additional obstacle to forecasting returns is that financial data and their relationships can undergo periods of large unforeseen shifts overtime. These shifts are caused by a breakdown of market mechanisms typically observed during economic crisis, changes in fundamentals, country, and central bank policy among others. One example of such breaks is present in the paper by Ang and Bekaert (2004) who showed that in high volatility environments such as the ones observed during bear markets, can lead to uncorrelated securities and asset classes to suddenly become correlated. A possible explanation for these shifts is that complex interactions between asset prices combined with investors' varying expectations and risk-aversion can be jointly influenced in a way that could lead to multiple mis specified market equilibria (Branch and Evans, 2010). In the paper Pesaran and Timmermann (2002) the authors tackled the issue of regime change and model stability by incorporating a two-stage approach. In their solution a model would first be used to estimate the most recent break in the time series model and afterwards only data following the break would be used in the stock return forecast. They found that in comparison with traditional rolling and expanding windows the proportion of correctly predicted signs of stock returns improves when conditioning for breaks.

It can be argued that many of the anomalies due to mispricing are not found broadly in the market and are restricted to small subsets. Hou, Xue and Zhang (2015) discovered that of nearly 80 perceived market anomalies many were likely exaggerated by excessive weighting on microcaps. Furthermore, the paper argues that a factor model comprised of a market, size, investment, and profitability factors better captures several anomalies outperforming the Fama-French and Carhart models. Hence, we arrive at another challenge which comes down to understanding which variables can actually be useful in a model.

Another topic of discussion surrounds the potential predictors, which have been proposed over the years, such as financial ratios, macroeconomic and volatility measures among others. Campbell and Shiller (1988) show that price to earnings ratio can be a useful predictor of stock returns over a span of several years. Kothari and Shanken (1997) find evidence that dividend yields and book to market ratios are powerful predictors of returns over a yearly period. Macroeconomic variables such as yield spreads working as a proxy for risk premiums have also been explored as possible equity return predictors (Keim and Stambaugh, 1986).

Welch and Goyal (2008) evaluated the performance of several proposed variables and found that even though IS performance seemed promising, OOS all predictors fail to consistently beat a simple historical average. It was also shown that a model referred to as "kitchen sink", which included all individually tested predictors performed worst due to overfitting and overparameterization. A following paper Goyal, Welch and Zafirov (2021) reexamined the original predictors from the 2008 paper as well as 26 new variables. In the end, more than a decade after, the authors reiterated the original claim that the variables which could help us predict future equity premiums remain a mystery. In contrast Cenesizoglu and Timmermann (2012) further examined several predictors and models, arguing that many models which underperform the constant mean and variance benchmark, can still produce superior economic performance and be useful for investors. The authors defend forecasting performance should be more centered on economic measures instead of the overfocused statistical measures such as R^2 or root mean squared forecast errors. Kelly, Malamud and Zhou (2022) found that a timing strategy using machine learning models with all the predictors from Welch and Goyal (2008) lead to gains in Sharp ratio even with large and negative R^2. In the paper model parameterization was increased far beyond the number of training observations, increasing model performance even with minimal regularization. They defend the economic value of highly complex models whose performance can be understated with traditional statistical measures. The authors argue that smaller models should only be used if correctly specified, however they reiterate Box (1976) claim that models are never correctly specified and therefore higher complexity should overall be preferred.

The great majority of articles regarding machine learning applications in finance have been mostly published outside the main finance journals. Nevertheless, high value creation in real world applications outside of finance combined with growing interest and investment by larger institutions has led to a boom in machine learning research by finance academia. Recently, it was shown that neural networks and regression trees can help improve our understanding of asset prices, increasing predictive performance and beating simpler regression based models (Gu, Kelly and Xiu, 2020).In the paper Rapach, Strauss and Zhou (2013) global equity returns are predicted using lagged returns from several countries through a lasso regression. In Bianchi, Büchner and Tamoni (2021) machine learning lead to OOS increases in forecast accuracy of bond excess returns. Machine learning and more specifically neural networks, allow the functional form between returns and predictors to be left unspecified. They show that accounting for non-linearity is relevant, when building forecasting models of bond returns, overperforming traditional linear models.

Different applications of machine learning in finance continue to grow across different research subjects ranging from asset pricing, credit scoring, risk management among others. However, even with the growing literature, market timing using ML continues to be an underexplored subject with further research being needed.

3. Data

In order to conduct research on the value of machine learning models in forecasting market movements on a sector basis, data was retrieved from 1990 to 2022 covering a total of 32 years. Firstly, it was necessary to retrieve return data from the different sectors composing the S&P500. The S&P500 is commonly divided into 10 sectors ranging from consumer discretionary, energy, healthcare to technology. In order to track the performance of each sector respective indexes will be used, which are provided by the S&P Dow Jones Indices. The sector indexes are built using companies in the S&P500 and classifying them by their GICS. The Global Industry Classification Standard (GICS) is a method to classify companies to a sector which best characterizes it´s operations. The indexes built using this classification have a track record of more than 20 years providing sufficient data as well as a reliable proxy of each sector's performance.

The excess return for each sector was calculated over a 22-day period and was then labeled as a 1 (R_{t+22}-Rf_{t+22} > 33^{rd} percentile) or as a 0 (Rt_{+22}-Rf_{t+22} ≤ 33^{rd} percentile). The percentiles were calculated based on a rolling window of the previous 2 years of excess returns of the respective sector. Therefore, the retrieved data from 1990 to 1991 will be solely used to create the percentiles and won't be present in training or testing the models. The reason a classification based on moving percentiles was necessary, was to first accommodate for the inherent simplification of excess returns to a two-label classification or binary problem, which ignores the size of the returns. Secondly, financial markets and stock sectors go through changes over the years and therefore creating a stationary label could lead to inconsistencies in the training windows of the machine learning models. Several potential predictors across the literature were chosen, which can be classified into fundamental, macroeconomic, sentiment and technical measures. Further detail regarding the features can be found in the appendix.

3.1. Fundamental Features

Variables that represent the underlying performance of a company and help determine the fair value of financial assets in the market are referred to as fundamental indicators. The general usefulness of such indicators, in understanding and forecasting risk premiums, has long been investigated in finance literature. One of the chosen indicators is the Dividend yield, which has been found to be a forward-looking measure of stock returns and earnings (Lamont, 1998). The author argues dividend yields help measure time-varying components of returns, namely the inverse relationship between stock prices and expected returns. Lewellen (2004) overviewed several proposed financial ratios, finding strong evidence of dividend yields ability in predicting stock returns especially when compared to other ratios such as book to market.

The second chosen variable was the monthly updated Cape ratio, a valuation measure commonly known as the cyclically adjusted price to earnings ratio, which is calculated by dividing the price of the asset by the ten-year average of inflation adjusted earnings per share. The reason behind using a ten-year average is to avoid the inherent volatility in corporate earnings, which occurs through economic cycles. P/E ratios have been shown to be a powerful predictor of risk premiums, when forecasting lower frequency returns (Campbell and Shiller, 1988).

3.2. Macroeconomic Features

The chosen macroeconomic variables were the U.S unemployment rate, US unemployment initial claims, US Dollar Index, US money supply (M2), TED Spread, Term Spread, US prime rate, the WTI oil price and the Monthly CPI. Employment data was included as a potential predictor considering a rising unemployment environment has been shown to positively affect stock returns during periods of economic expansion, while negatively impacting the market in periods of economic contraction (Boyd, Hu and Jagannathan, 2005).The US dollar index was chosen since US monetary policy is a key driver of the global financial cycle, with changes in policy affecting global risk aversion and prices in risky assets (Miranda-Agrippino and Rey, 2020).Similarly, money supply changes in accordance with the monetary policy set out by the federal reserve board, which itself shifts depending on the state of the economy. Adjustments to the growth rate of money supply appear to have a bi-directional relationship with stock returns, with changes in money supply by FED policies having a direct impact in stock returns (Rogalski and Vinso, 1977).

Bond spreads have long been investigated as potential predictors of returns, with spreads believed to proxy changes in risk premiums (Keim and Stambaugh, 1986).Therefore, both the TED spread and the Term spread where chosen. The TED spread is simply the difference between short term US government debt and the interest rate on interbank loans. The Term spread is the difference in yield between US long term debt and short-term debt, being low near business cycle peaks and high during periods of economic stress (Fama and French, 1989).The prime rate also known as the prime lending rate is the reference interest rate charged by bank in loans and mortgages, being influenced by the FED funds rate reflecting changes in monetary policy.

Regarding the WTI oil price, large negative shocks to oil prices have been shown to influence the stock market positively when the US market is performing well, while in contrast positive shocks to oil prices have weak influence in the market (Sim and Zhou, 2015). In the paper by Phan, Sharma and Narayan (2015) the authors found the general usefulness of oil prices in forecasting stock returns is dependent on the frequency of the data. Furthermore, the paper finds evidence that the change in oil price is relatively more important for some sectors than others. Lastly, the monthly change in CPI was chosen considering, inflation appears to possess a negative relationship with stock returns in short-term time horizons and in contrast a positive impact in the long term (Boudoukh, Richardson and Whitelaw, 1994).

In terms of update frequency, the US prime rate, US money supply, the MoM CPI as well as the unemployment rate are updated monthly. The initial unemployment claims are updated weekly, and the remaining variables have a daily frequency. It’s important to consider that features with lower update frequency have varying release dates. The US bank prime rate is known at the end of each month and the US money supply is typically published at the third week of each month. The U.S unemployment rate is published by the Bureau of Labor Statistics on the first Friday of every month, while weekly U.S unemployment claims are published at the beginning of each week. Considering the varying publishing dates of the variables with lower update frequency and to avoid look-ahead bias a time lag was put in place for all variables except the ones, which are daily updated.

3.3. Technical Features

A technical indicator is derived from historical asset prices and is commonly used by investors to predict price trends. The first technical predictors chosen were variables, which measure volatility namely the CBOE Volatility index (VIX) as well as the volatility of companies in the financial sector over a 5-day and 22-day period. Gu, Kelly, and Xiu (2020) found that variables, which measure volatility are among the most powerful predictors. Additionally, Giglio, Kelly and Pruitt (2016) found that volatility in the financial equity sector is a more informative predictor of systemic risk than market volatility, after that analyzing over 19 proposed variables.

The last two remaining technical variables chosen were the S&P500 price minus it´s 100-day simple moving average and the return of the S&P500 over a 22-day period. Past returns over a 12 month period have been shown to persist into the future across asset classes from equity bonds and commodities (Moskowitz, Ooi and Pedersen, 2012). In the paper by Han, Yang and Zhou (2013) portfolios built using trading strategies, centered around moving averages, greatly outperformed a buy and hold strategy.

3.4. Market Sentiment Features

Another group of potential predictors are variables, which measure investor´s sentiment of the market. In the paper by Baker and Wurgler (2007) the authors argue it´s not only possible to measure investor sentiment, but such measures are useful in predicting stock returns. Market anomalies, such as stock overpricing have been associated to be stronger following periods of high levels of market sentiment (Stambaugh, Yu and Yuan, 2012). Another paper by Da, Engelberg and Gao (2015) showed that a sentiment index based on the volume of daily internet searches of worlds related to economic stress, such as recession or unemployment can predict market movements. From the sentiment measures chosen, four are among the indexes most watched by investment professionals which measure overall optimism and expectations of the economy. The chosen indexes are the US consumer confidence index, US national association of home builders housing index, US purchasing managers index and the US purchasing manager business barometer. The last two variables are based on surveys done to thousands of investors. The surveys measure overall bullishness and bearishness of the market, namely the advisors' sentiment bullish survey and advisors' sentiment bearish sentiment.

3.5. Data Analysis

In order to better understand the data used to build the different forecasting models several different data analysis methods were used. Firstly, several key statistical measures were calculated for both the 22-day excess returns of each sector as well as the respective percentiles, which were used to create the labels. As we can see from the following table the average excess return over a monthly period as well as the volatility vary significantly between sectors. When further analyzing the percentile distribution it becomes clearer the importance of creating a differentiated labelling process for the independent variables instead of labeling all sectors based on a singular criterion.

Table 3.1: Key Statistics for the 22-day sector excess returns

	Mean	SD	Maximum	75th Percentile	Median	25th Percentile	Minimum	Kurtosis	Skewness
CONS DISCRET	0,62%	6,75%	31,87%	4,50%	1,26%	-2,67%	-40,97%	6,05	-0,71
CONS STAPLES	0,39%	4,57%	21,90%	3,11%	0,79%	-1,79%	-32,43%	6,43	-0,85
ENERGY	0,58%	8,14%	40,57%	5,26%	0,88%	-3,84%	-64,16%	8,73	-0,67
FINANCIALS	0,38%	7,88%	54,53%	4,58%	0,94%	-3,23%	-51,16%	8,02	-0,50
HEALTH CARE	0,55%	5,28%	29,70%	3,84%	0,99%	-2,21%	-35,77%	5,73	-0,73
INDUSTRIALS	0,46%	6,73%	32,26%	4,44%	0,95%	-2,90%	-49,91%	7,06	-0,82
MATERIALS	0,41%	7,19%	31,34%	4,60%	1,03%	-3,41%	-44,52%	5,92	-0,63
TECHNOLOGY	0,73%	7,99%	39,70%	5,46%	1,51%	-3,46%	-38,76%	4,81	-0,46
COMM. SVS	0,16%	6,29%	36,69%	4,04%	0,63%	-3,22%	-36,47%	5,66	-0,44
UTILITIES	0,39%	5,42%	31,45%	3,42%	0,75%	-2,04%	-44,15%	8,16	-1,00
S&P 500	0,43%	5,74%	23,40%	3,85%	1,11%	-2,28%	-41,85%	6,96	-1,04

Table shows the key statistics of the 22-day excess returns for each sector and the benchmark index

It could be argued that simply introducing a constant threshold for each sector based on the respective historical distribution could be sufficient, therefore the same statistics were calculated on the percentiles used for the labelling. In the table we can see that not only are there disparities in the mean percentiles but also significant variation in their distribution and minimum and maximum values. As we can observe, the implementation of moving percentiles is relevant in order to better accommodate the inherent changes, the returns of each sector will go through with time.

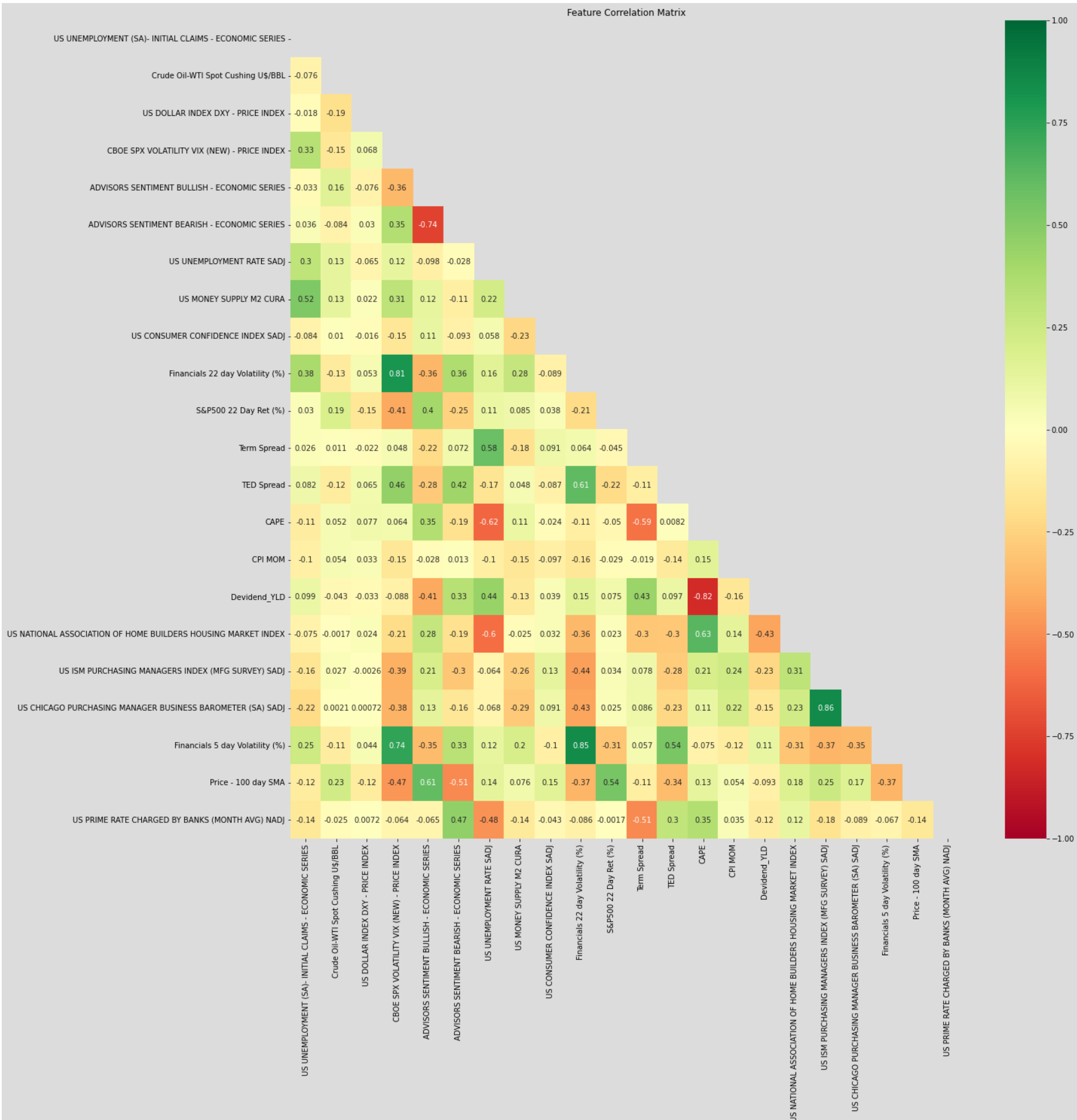
Table 3.2: Key Statistics for the 2-year rolling 33rd percentiles

	Mean	SD	Maximum	75th Percentile	Median	25th Percentile	Minimum	Kurtosis	Skewness
CONS DISCRET	-1,25%	1,83%	1,80%	0,02%	-0,80%	-1,81%	-6,66%	3,64	-1,17
CONS STAPLES	-0,95%	0,99%	0,70%	-0,30%	-0,72%	-1,49%	-3,52%	2,61	-0,76
ENERGY	-2,32%	1,84%	1,55%	-0,87%	-2,07%	-3,60%	-7,28%	2,58	-0,46
FINANCIALS	-1,93%	2,46%	1,65%	-0,37%	-1,29%	-2,72%	-10,76%	6,19	-1,66
HEALTH CARE	-1,08%	1,37%	1,45%	-0,13%	-0,96%	-1,77%	-4,58%	2,90	-0,57
INDUSTRIALS	-1,68%	1,87%	1,91%	-0,30%	-1,25%	-2,26%	-7,26%	3,66	-1,06
MATERIALS	-1,92%	1,71%	1,34%	-0,55%	-1,66%	-3,11%	-7,03%	2,96	-0,62
TECHNOLOGY	-1,85%	2,97%	2,15%	-0,14%	-1,14%	-2,28%	-13,37%	6,22	-1,77
COMM. SVS	-2,05%	2,43%	1,65%	-0,63%	-1,33%	-2,52%	-10,07%	4,67	-1,47
UTILITIES	-1,33%	1,73%	0,67%	-0,25%	-0,78%	-1,58%	-7,32%	5,30	-1,71
S&P 500	-1,11%	1,92%	1,76%	0,20%	-0,55%	-1,67%	-6,66%	3,96	-1,31

Table shows the key statistics of the 2-year rolling window used as a criterion when labelling the independent variables for the ML models.

Machine learning models, especially those whose coefficients are heavily penalized, are sensitive to highly correlated features. High correlation between features does not necessarily impact model performance, however in a small data environment, like the one we face in financial data the number of meaningful training observations included in the model´s training windows is relatively limited. Typically, in machine learning models as we increase the number of features and provide the model with more information the model's performance should increase up until a certain point. However, as we increase the number of features, while maintaining the size of our training sample constant, the model's performance will eventually start to decrease. When features are highly correlated between each other, the added information of the features will be limited. With that in mind it was necessary to plot a correlation matrix of the chosen predictors.

Figure 3.1: Feature Correlation Matrix

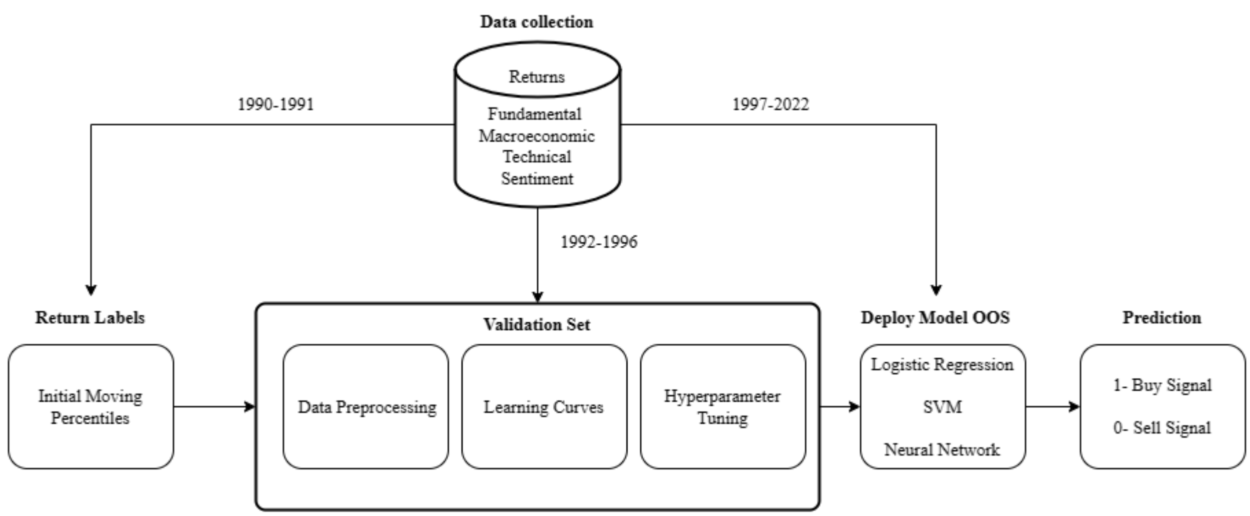


The correlation matrix was calculated based on the raw unscaled data, between the period of 1990 to 2022, in order to better approximate the true relationship between the predictors, without being affected by the scaling of the features. If two variables are perfectly correlated, then the model won´t receive additional information and therefore should be removed. The actual impact of highly correlated features does depend on the type of classifier, the number of features used and available observations. Overall, the features aren’t highly correlated, with the values ranging from -0.82 to 0.86 with a small portion of variables located on the tails of the range and therefore shouldn’t pose an issue when forming the models.

4. Methodology

In this section the methodology used to arrive at our results will be thoroughly explained. In the paper by Giglio, Kelly and Xiu (2021) the authors give an in depth overview of best methodology practices when using machine learning models, which this thesis will use as a guide when building the models. In order to build the ML models several open-source libraries were used such as pandas and scikit-learn. As mentioned previously, data from 1990 to 1991 will be used to build the initial percentiles and will be unavailable to construct the models. Additionally, data from 1992 to 1996 will be set aside for the validation sample, in order to further train and tune the models hyperparameters, without influencing the OOS results. Finally, data from 1997 to 2022 will be used to run the models OOS and implement the trading strategies.

Figure 4.1: Methodology Flowchart



4.1. Data Preprocessing

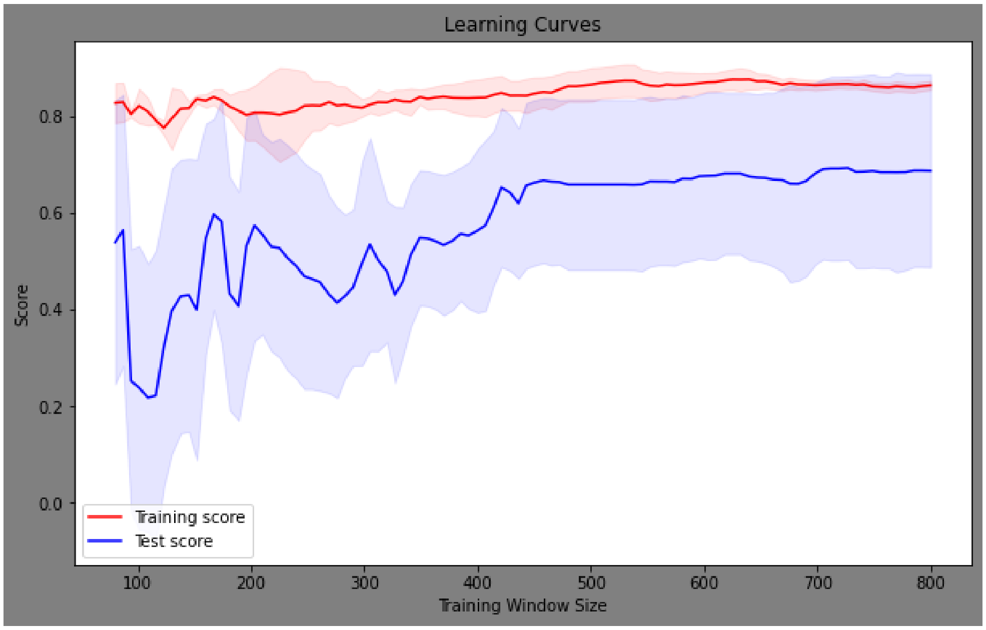
When dealing with datasets it is common to come across variables with distinct scales, which can introduce noise into the models by putting more weight into variables with a larger range. In this case implementing a preprocessing method such as normalization or standardization to rescale the data is one of the possible approaches. Typically, normalization works well when the distribution of the features doesn't follow a Gaussian distribution and can be useful when using ML algorithms such as Neural Networks which don't assume the data follows a particular distribution. In the case of normalization, the features will be rescaled so the variables values range between 0 and 1. On the other hand, standardization transforms continuous data, so it follows a mean of zero and a variance of one, making the features follow a normal distribution without a bounded range. In order to understand better what method works best with the data, both methods were thoroughly tested and compared across all machine learning models. Overall normalization was found to lead to better and more stables results and therefore was applied across all models.

Standardization: $x_{scaled} = \frac{x_i - \mu}{\sigma}$ **Normalization:** $x_{scaled} = \frac{x_i - x_{min}}{x_{max} - x_{min}}$

4.2. Training and Testing Sets

A supervised machine learning model uses information from a training dataset in order to arrive at the optimal combination of weights or parameters of the model, which is then used to make the predictions. Finance time series data adds an additional layer of complexity to the issue considering, as mentioned previously, the relationship between the predictors and the excess returns will change over time. If the training sets included older data, which no longer represents the current relationship of the variables, it will negatively affect the predictive performance. On the other hand, if the rolling window is too small the models will not have enough information to capture the relationship between the variables. Therefore, in order to understand the optimal size for the training window several learning curves were plotted for each sector and across all models. The learning curves were plotted using data from 1992 to 1996, which is the chosen validation period to avoid Look-Ahead Bias.

Figure 4.2: Learning Curves



A learning curve shows the training and testing performance of a model for varying number of training set samples. The performance measure chosen was the F1 score, a common evaluation metric for classification problems, which tends to work well when there is a disparity in the proportions of labels in a data set. Overall, across all models and sectors the gain in performance peaks after the training sets contain more than 500 observations, equivalent to two years of data. The performance of the models is unstable with smaller training sets and typically stabilizes after over two years of data is used. However, two sectors, namely the communication services and energy sector showed better results with a shorter training window. Therefore, for the communication services and energy sector the optimal training set size was a one-year training window, while the remaining sectors used two years of data.

4.3. Tuning Hyperparameters

In machine learning a model is defined by its parameters, which not only involves training the model with data, but also choosing the optimal hyperparameters. A hyperparameter is simply a parameter whose value controls the learning process the model takes and is chosen by the user. One of the hyperparameters is the regularization method, used to reduce the errors by fitting the model appropriately on the training set in order to avoid overfitting. If overfitting occurs the model will have a low accuracy score, because it will focus on features that may not represent the true relationship between the variables consequently introducing noise into the model. Regularization helps decrease the amount of noise by shrinking the coefficient estimates towards zero, reducing the weight on variables with low importance, decreasing the variance and overall complexity of the model.

Besides the penalization method another important choice when using machine learning techniques is the degree of penalization, which can be represented by λ ,typically referred to as the regularization parameter. As λ increases the value of the coefficients will shrink, further approximating 0 reducing the model´s variance. The increase in the penalization term will help avoid the issue of overfitting, benefiting the model up to a certain point. However, as the degree of penalization increases more bias is introduced, which could lead to underfitting, making the model lose its inherent properties. Underfitting occurs when too much bias is introduced into the model leading to the hypothesis function poorly mapping the trend of the data.

In the scikit-learn library three different regularization techniques are available. Firstly, the L1 penalty typically referred to as LASSO (least absolute shrinkage and selection operator) will perform the penalization based on the absolute value of the magnitude of the co-efficient.

$$\textbf{L1 Regularization: } \boldsymbol{Cost} = \boldsymbol{\lambda \sum_{j=1}^{m} |w_j|}$$

Another regularization technique is the L2 penalty commonly referred to as ridge regression, which performs the penalization based on the sum of the squared values of the coefficient weights.

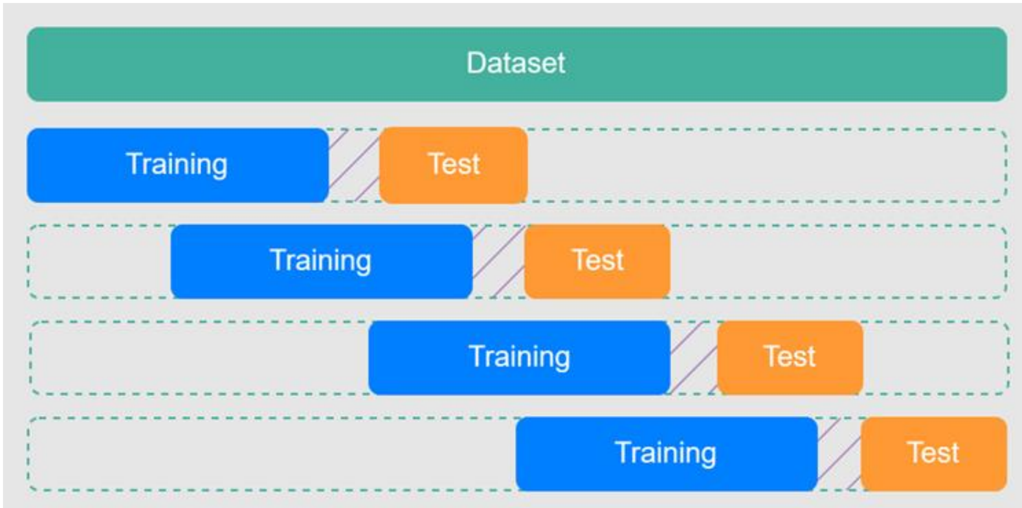
$$\textbf{L2 Regularization: Cost} = \boldsymbol{\lambda \sum_{j=1}^{m} w_j^2}$$

Finally, the last technique is a linear combination of the L1 and L2 penalty referred to as elastic net, which eliminates the need to choose between the two penalties. One significant difference between the three techniques is the elastic net and L1 penalty are not only able to shrink the coefficients, but also set them to zero consequently eliminating the predictors from the model.

$$\textbf{Elastic Net Regularization: Cost} = \boldsymbol{\lambda\left(\frac{1-a}{2} \sum_{j=1}^{m} w_j^2 + \mathbf{a} \sum_{j=1}^{m} |w_j|\right)}$$

In order to tune the hyperparameters, while avoiding look ahead bias a validation set was used, containing data from 1992 to 1996. Using the data from the validation set the combinations of hyperparameters are changed iteratively, with the best possible combination of each sector over that period being chosen, based on the achieved F1 score. Cross validation is a common tool used to assess and tune machine learning models.

Figure 4.3: Rolling Window Cross-Validation



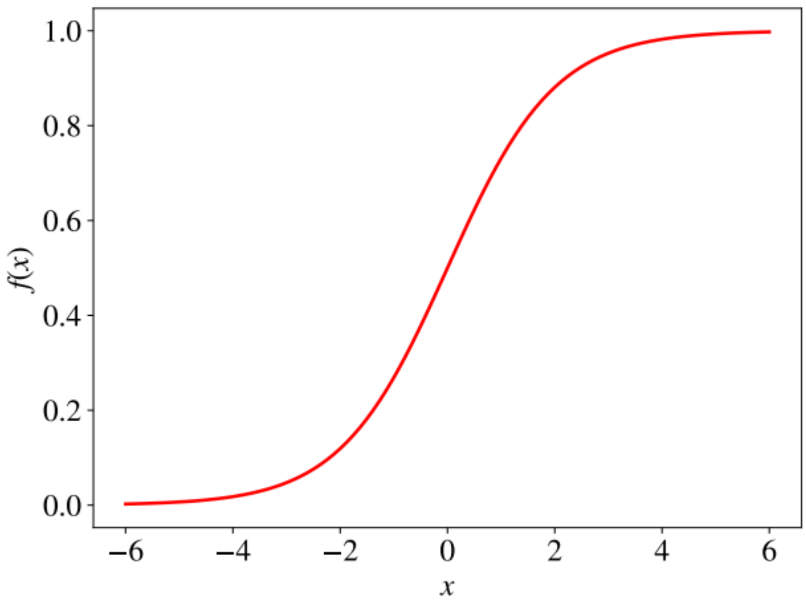
The validation set is split into several training and test sets corresponding to a rolling window with a 22-day break between training and testing sets

During the process the dataset is split into several training and testing subsets. Considering time series data is being used the validation set was split through rolling windows. Each window maintains its size, while being updated in a first in first out process, throughout the validation set. To avoid information spilling between the training and testing sets, commonly referred to as data leakage, a break of 22 days was introduced between the two. After the optimal hyperparameters are found they will then be used in the OOS period used to backtest the different strategies.

4.4. Logistic Regression

One of the most used supervised machine learning models is the logistic regression (LR). Despite the name, a logistic regression is still a linear model such as a linear regression, with the key difference being the output is discrete bounded between 0 and 1, instead of continuous. A LR is typically used in binary classification problems, which estimates the probability of a certain event occurring. After the probability is estimated a classification or prediction will be converted based on a certain threshold which is set to 50% by default.

Figure 4.4: Logistic Regression



(Burkov, 2019)

To construct the model, firstly a linear objective function is defined, which we think will be useful in making our predictions and approximates the true relationship between the predictors and the independent variables.

$$1.\ \boldsymbol{h_\theta(x) = \mathrm{g}(\theta^T x)}$$

The objective function will then be fed into a sigmoid function, which will transform any real number to values with an interval between 0 and 1. These transformed values will effectively become the estimated probabilities of a certain event occurring.

$$2.\ \boldsymbol{z = \theta^T x}$$

$$3.\ \boldsymbol{\mathrm{g}(z) = \frac{1}{1+e^{-z}}}$$

The probabilities will then be used to make the final classification with h_θ representing the probability that the output is equal to 1, which will be converted based on the previously mentioned threshold or decision boundary.

$$4.\ \boldsymbol{h_\theta(x) = P(y = 1|\,\mathrm{x}\,; \theta) = 1 - P(y = 0|\,\mathrm{x}\,; \theta)}$$

Or

$$5.\ p(y|\ \mathrm{x}\,;\boldsymbol{\theta}) = (h_{\theta}(x))^{y}(1 - h_{\theta}(x))^{1-y}$$

In order to fit the model or in other words obtain the optimal combination of weights (θ) the likelihood function present in step 5 will be maximized. However, calculating the maximum of the function in its current state might be difficult and therefore it is common to transform the function through logarithms so it can be minimized. To find the minimum the Scikit-Learn logistic regression package supplies several different solvers which can be used to arrive at the minimum, with Liblinear being found as the optimal choice.

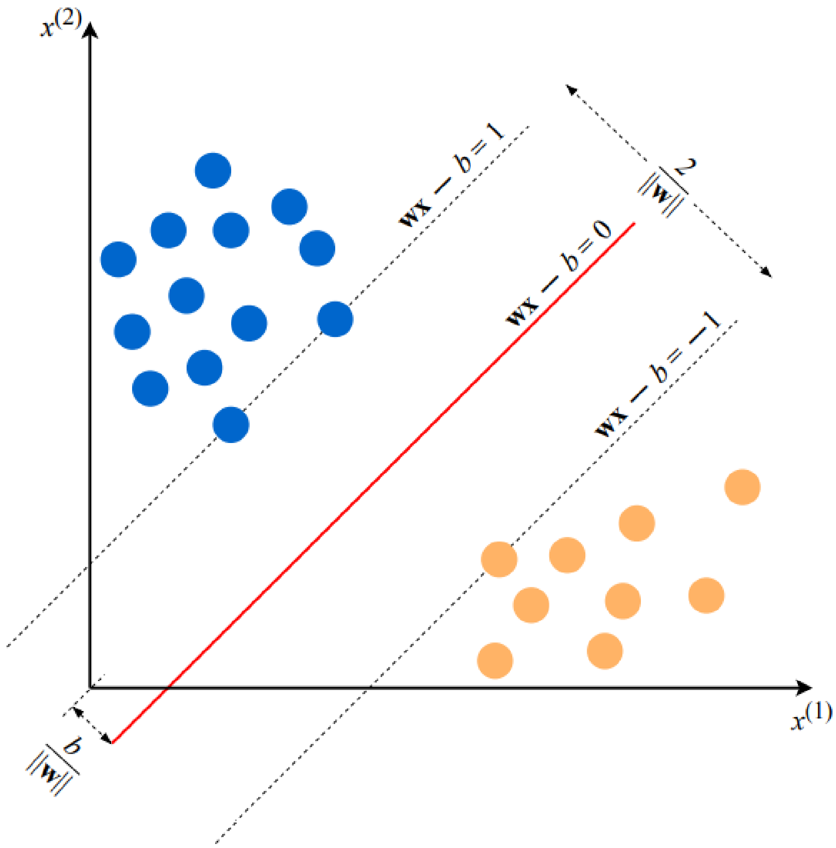
$$6.\ J(\boldsymbol{\theta}) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(h_{\theta}\left(x^{(i)}\right)\right) + \left(1 - y^{(i)}\right)\log\left(1 - h_{\theta}\left(x^{(i)}\right)\right)\right]$$

A LR provides several advantages over other ML models. Firstly, is one of the most transparent models by allowing the user to fully understand and track each step. Secondly, contrary to other ML models like SVM or Decision Trees, a logistic regression provides an estimated probability associated with each of the predicted labels. This probability allows for a better understanding of the model´s confidence on each prediction and if necessary, a more in-depth analysis. Lastly, a LR provides the coefficient weights providing the model´s overall importance to each predictor.

4.5. Support vector Machines (SVM)

Another supervised ML model is support vector machines which is used to tackle classification, regression, and outlier detection problems. Contrary to other ML models, SVM are still effective in situations where the number of dimensions or features is higher than the number of observations posing as a powerful tool in finance. The main disadvantage of using SVM over the previously explained LR is that the model doesn’t provide an estimated probability associated to each prediction. The way SVM works is by finding a hyper-plane that will be used to create a decision boundary between the different classes.

Figure 4.5: SVM model for two-dimensional feature vectors



(Burkov, 2019)

Each item in the data set is plotted in a N dimensional space, with N being the number of chosen features. Afterwards the optimal hyperplane that best separates the data is plotted and used to make the classifications. In a 2-dimensional space the hyper-plane is simply a line, however as the number of predictors increases it becomes gradually harder to visualize. The model defines the threshold as a midpoint between the edges of each identified class or cluster, maximizing de distance between the nearest observations or support vectors and the separating hyperplane.

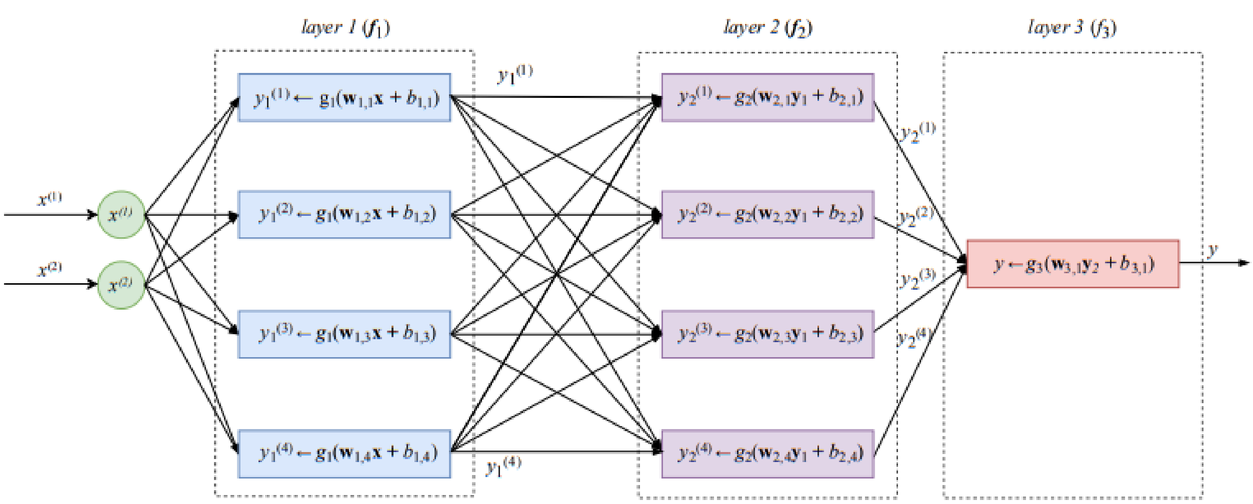
The SVM models can essentially be divided into Linear and Non-linear models. When the different classes can easily be separated by the hyperplane by using a straight line a linear SVM is used. However, when such a separation is not possible then kernel functions will transform the data, so the separation becomes possible. The kernel functions are a group of algorithms that use a linear classifier to solve a non-linear problem. Several different types of kernels exist and are applicable through the Scikit-Learn library namely the linear, RBF (non-linear), Sigmoid and Polynomial.

In order to choose the optimal hyperparameters for each ML technique, across all sectors a rolling window cross-validation was implemented. A validation sample set between 1992 to 1996 was used to create the training and testing windows allowing for the tuning of the hyperparameters to be out-of-sample. Several combinations of the different hyperparameters were tested iteratively for each sector, with the best combination being chosen based on the respective F1 score.

4.6. Artificial Neural Networks

The idea behind artificial neural networks (NN) originates from the human nervous system composed of connected neurons, being designed to learn and create abstracted features through multiple hidden layers. Each layer is made up of several interconnected nodes with each one having an associated weight, threshold, and an activation function. For this paper a feed-forward neural network called multilayer perceptrons will be used. This type of model is trained with a backpropagation algorithm, which propagates back the error from the output node to the hidden and input nodes. The final goal of backpropagation is to minimize the cost function by iteratively adjusting the neural networks weights. The size of each adjustment is defined by the gradients of the cost function relative to each of the parameters.

Figure 4.6: Multilayer Perceptron Structure



(Burkov, 2019)

The NN model begins with an input layer, which receives the raw values of the chosen features (x) that are going to be fed into the model as the first set of activations $a_{(1)}$.

$$\textbf{Input layer: } \boldsymbol{a}_{(\mathbf{1})} = \boldsymbol{x}$$

The raw inputs will then go through the hidden layers, where each additional layer will add further complexity in training the model. In each layer the inputs from the previous layer $\boldsymbol{a}_{(l)}$ are combined into a score $z_{(l+1)}$. Afterwards the scores are processed through an activation function g, whose transformed values $a_{(l+1)}$ will then be used as inputs in the following layer.

$$\textbf{Neuron value: } \boldsymbol{z}_{(l+\mathbf{1})} = \boldsymbol{\theta}_{(l)} \times \boldsymbol{a}_{(l)}$$

$$\textbf{Activation value: } \boldsymbol{a}_{(l+\mathbf{1})} = \mathbf{g}(\boldsymbol{z}_{(l+\mathbf{1})})$$

In the end the output layer receives and combines the data from the last hidden layer, producing the model´s prediction.

$$\textbf{Output layer } \boldsymbol{f}\textbf{ : } \hat{\boldsymbol{y}} = \boldsymbol{\theta}_{(\boldsymbol{f}-\mathbf{1})} \times \boldsymbol{a}_{(\boldsymbol{f}-\mathbf{1})}$$

Predictions from Neural Networks can be used in both regression and classification problems depending on whether a linear regression or a logistic regression is used in the output layer. Neural Networks tend to perform well in both linear and nonlinear problems posing as a powerful tool to solve real-world issues. A common criticism is these types of models are referred to as black boxes, considering they can be hard to interpret and don't provide enough information regarding the actual relationship between the features and the independent variables.

4.7. Portfolio Construction

In order to evaluate the model´s economic value and by extension its usefulness for investors it is necessary to implement the forecasts across several strategies. Therefore, a total of 13 different portfolios were built for each model using the different forecasts of the three machine learning techniques. Of the portfolios, 10 are simply sector portfolios, which invest in the respective sectors when the label is 1 and invest in 10-year treasury bonds when the label is 0. These portfolios were built in order to measure the change in performance in each sector from implementing a trading strategy using the model´s forecasts versus a simple buy and hold strategy. An additional portfolio was built using the same methodology as the sector portfolios but focusing on the S&P500 index. This portfolio was necessary in order to compare the economic gain of each of the ML models in implementing trading strategies on a sector basis versus of the overall market. The last 2 portfolios are equally and value weighted portfolios, which go long sectors forecasted to have a positive excess return and investing in bonds otherwise.

Lastly all the previously mentioned strategies were also applied using forecasts from a hybrid model, which combines the predictions from all three of the ML models. The hybrid model forecast was constructed using the forecasts of the logistic regression, SVM and neural network models. A positive signal of 1 would have been made, if at least two of the models also predicted a positive label. In all other cases the hybrid model would set the label to 0. Similarly, to traditional linear regression, ML techniques will also provide incorrect predictions. By combining the different model´s outputs into a single prediction it will hopefully decrease the chances of an incorrect prediction being made.

5. Results

In this section the results of the dissertation will be thoroughly explained with the following structure. Firstly, the different model's performance will be analyzed alongside the respective evaluation metrics. Afterwards, the relative importance of each feature will be measured across all the equity sectors. Lastly, the performance of the different portfolios constructed throughout the OOS period will be overviewed.

5.1. Model Performance

When using machine learning models' integrating performance metrics into the pipeline is fundamental in order to understand and quantify the progress made into solving a set problem. Algorithms solving classification problems, which this thesis is addressing, typically use four popular performance metrics namely accuracy, precision, recall and the F1 score. The first and one of the simplest metrics is accuracy, which tells us how often the algorithm predicts an outcome correctly. Secondly, precision measures the ability of the model to identify only the relevant data points (positive labels). Recall on the other hand, measures the capability of the model of identifying all relevant cases. The F1 Score is the harmonic means of recall and precision, therefore combining the two into a single metric.

Table 5.1: Model Performance

Model	Performance	CONSUMER DISCRET	CONSUMER STAPLES	ENERGY	FINANCIALS	HEALTH CARE	INDUSTRIALS	MATERIALS	INFO TECHNOLOGY	COMM. SVS	UTILITIES	S&P500
LR	Accuracy	65,21%	65,73%	65,06%	63,81%	64,60%	62,97%	64,72%	65,49%	64,24%	64,15%	63,99%
	Precision	65,97%	66,41%	67,85%	65,49%	65,61%	64,88%	66,23%	66,23%	66,86%	65,97%	65,06%
	Recall	97,91%	98,12%	90,96%	94,54%	97,57%	93,86%	95,49%	97,41%	89,70%	93,83%	96,48%
	F1	78,83%	79,21%	77,73%	77,38%	78,46%	76,73%	78,21%	78,85%	76,62%	77,47%	77,72%
SVM	Accuracy	64,06%	65,39%	65,00%	58,69%	65,28%	62,83%	65,73%	64,36%	64,17%	63,87%	62,42%
	Precision	65,55%	66,77%	67,70%	66,87%	65,81%	64,64%	66,98%	66,28%	66,63%	65,52%	65,08%
	Recall	96,26%	95,50%	91,38%	73,10%	98,78%	94,55%	95,31%	93,69%	90,43%	94,95%	91,22%
	F1	77,99%	78,59%	77,78%	69,85%	78,99%	76,79%	78,67%	77,63%	76,72%	77,54%	75,96%
NN	Accuracy	64,25%	65,27%	65,07%	57,41%	65,01%	62,71%	65,79%	64,55%	63,05%	63,76%	61,90%
	Precision	65,61%	66,76%	67,72%	66,83%	65,71%	64,58%	66,99%	66,62%	66,16%	65,54%	64,90%
	Recall	96,60%	95,19%	91,49%	69,34%	98,40%	94,46%	95,47%	92,81%	88,90%	94,56%	90,32%
	F1	78,14%	78,48%	77,83%	68,07%	78,80%	76,71%	78,73%	77,57%	75,86%	77,42%	75,53%
Hybrid	Accuracy	63,63%	65,10%	63,85%	56,19%	64,51%	62,02%	64,84%	64,51%	61,93%	63,88%	61,68%
	Precision	65,51%	67,00%	67,64%	67,15%	65,58%	64,60%	67,11%	67,00%	66,40%	65,87%	65,06%
	Recall	95,07%	93,71%	88,34%	64,73%	97,41%	92,03%	92,12%	91,12%	84,45%	93,43%	88,82%
	F1	77,57%	78,13%	76,61%	65,92%	78,39%	75,91%	77,65%	77,22%	74,35%	77,26%	75,11%

Accuracy = (TP + TN)/(TP+FP+TN+FN) | Precision = TP/(TP+FP) | Recall = TP/(TP+FN) |
F1 Score = 2×(Recall× Precision)/ (Recall+Precision)

Across all sectors the different models possessed an accuracy of more than 55%, with the lowest accuracy being registered in the financial sector. Overall, the models were able to reasonably predict the direction of the market over a 22-day period across all 10 sectors. Furthermore, both the precision and the recall scores are within an acceptable range, with the recall being typically higher than precision. It is important to keep in mind that finance data is known for its low signal to noise ratio and inherent difficulty in forecasting returns.

In general accuracy combined with the other statistical measures gives a positive picture of the different model's performance. It is important to consider that due to the nature of the problem being simplified to a two-label classification, namely whether the excess return is going to be positive or negative, it comes with certain setbacks. The inherent simplification ignores the actual size of the returns and therefore an excess return of 0.5% could have the same label as a return of 5%. Therefore, even a model which presents superior performance based on statistical measures may still provide lower economic value, underperform a buy, and hold strategy, if it wrongly predicts returns positioned in the tails of the distribution. Additionally, even if the model increases portfolio performance it could be the case the additional costs originated from transactions fees may erase potential economic gains from the model.

5.2. Feature Importance

In order to understand the relevance of the chosen features in forecasting returns, the relative importance of each variable was ranked. Understanding the importance of each variable towards the obtained results is relevant, not only because ML models criticized as black boxes, but also because of the nature and risk evolving trading strategies. Several methods exist to evaluate the importance of a variable to a model. However, the usage of rolling windows over a 22-year period combined with the importance of the variables changing over time increases the tasks difficulty.

Considering these challenges, the chosen metric bases the ranking of the variables on the value of the coefficients of the model. For each feature, the mean of the absolute value of the coefficients obtained during the OOS period was calculated. Afterwards, for each index the features were evaluated based on the calculated averages. This approach is possible, since the obtained coefficients already take into account the normalization of the features, making the average values comparable. The downside of this method is that of the three models only the logistic regression generates coefficients similarly to a linear regression and therefore the evaluation of the features will be restricted to the LR model.

Table 5.2: Feature Ranking

Features	CONSUMER DISCRET	CONSUMER STAPLES	ENERGY	FINANCIALS	HEALTH CARE	INDUSTRIALS	MATERIALS	INFO TECHNOLOGY	COMM. SVS	UTILITIES	S&P500	Global
Devidend Yield	1	1	1	1	1	1	1	1	1	1	1	1
US Unemployment Rate	2	4	3	3	2	2	2	2	5	2	3	2
Advisors Sentiment Bearish Index	3	3	2	2	3	3	3	3	2	6	2	3
Financials 22 day Volatility (%)	4	6	6	5	5	4	5	7	3	13	6	4
US Money Supply M2	5	8	4	9	6	5	8	4	4	21	5	5
US Unemployment Initial Claims	6	11	7	4	8	6	4	5	6	11	4	6
CBOE SPX Volatility VIX	7	14	9	8	10	7	10	10	7	4	7	7
WTI Crude Oil	11	10	21	6	4	8	9	6	14	10	8	8
US Dollar Index DXY	8	5	5	11	11	18	11	11	11	7	9	9
CPI MOM	20	2	8	7	19	9	12	19	10	3	18	10
US National Association of home builders housing market Index	14	13	12	14	7	12	17	17	9	5	13	11
Advisors Sentiment Bullish Index	19	9	16	15	9	11	13	16	13	9	14	12
US Consumer Confidence Index	13	12	10	16	12	13	7	14	16	15	15	13
US Prime Rate	16	15	22	10	13	10	6	12	22	8	10	14
TED Spread	9	21	13	12	18	16	15	8	12	20	11	15
Price - 100 day SMA	10	7	11	19	16	14	16	18	15	18	16	16
Term Spread	12	17	19	13	14	19	20	9	21	14	12	17
Financials 5-day Volatility (%)	18	16	20	17	15	15	19	21	19	16	17	18
CAPE Ratio	17	19	14	22	22	17	14	13	17	19	19	19
US Chicago Purchasing Manager Business Index	21	20	15	18	17	21	21	20	8	17	21	20
S&P500 22 Day Ret (%)	15	18	18	20	21	20	18	15	18	22	20	21
US ISM Purchasing Managers Index	22	22	17	21	20	22	22	22	20	12	22	22

It is important to keep in mind, since the absolute values are used this method only evaluates the weight the predictor has in the model and not the influence or relationship it may have with the returns. The top five features across all the sectors remain relatively the same, with the model giving more importance to the Dividend Yield, US Unemployment rate and Bearish sentiment index.

5.3. Portfolio Performance

As mentioned previously, simplifying returns into a binary classification problem could lead to the typical statistical valuation metrics misrepresenting the actual performance the different models have in predicting market movements. In the paper by Kelly, Malamud and Zhou (2022) the authors find that machine learning models, with large and negative R2,still led to increases sharp ratio. Therefore, to evaluate the economic value of the models several portfolios were built and backtested in the OOS period, from 1997 to 2022.

Table 5.3: Portfolios Performance Overview

Portfolios	Annualized Return	Annualized Volatility	Sharpe Ratio	Minimum Return	Maximum Return	Var
S&P 500 (Benchmark)	9,58%	19,34%	0,27	-11,98%	11,58%	-2,0%
LR Equal Long	10,28%	15,68%	0,38	-11,54%	8,53%	-1,6%
LR Weighted Long	10,14%	16,95%	0,35	-11,83%	9,34%	-1,8%
LR S&P500	9,79%	17,47%	0,32	-11,98%	9,39%	-1,8%
SVM Equal Long	10,73%	15,19%	0,42	-11,54%	9,54%	-1,6%
SVM Weighted Long	10,61%	16,10%	0,39	-11,83%	9,34%	-1,7%
SVM S&P500	12,90%	16,97%	0,51	-11,98%	9,39%	-1,8%
NN Equal Long	10,26%	15,21%	0,39	-11,54%	8,85%	-1,6%
NN Weighted Long	10,00%	15,99%	0,36	-11,83%	9,34%	-1,7%
NN S&P500	10,73%	17,19%	0,37	-11,98%	9,39%	-1,8%
Hy Equal Long	11,16%	14,27%	0,48	-11,54%	8,53%	-1,5%
Hy Weighted Long	11,07%	15,10%	0,45	-11,83%	9,34%	-1,6%
Hy S&P500	11,13%	16,73%	0,41	-11,98%	9,39%	-1,7%

Across all the models, the different portfolios outperformed the benchmark both in terms of absolute and risk adjusted returns. When comparing the different trading strategies, the answer to the first research question becomes clearer. The different strategies lead to both an increase in performance as well as a decrease in volatility, leading to higher sharp ratios and a decrease in the daily value at risk (Var). In order to evaluate the benefits of performing timing strategies on a sector basis, the ML models were also applied to predict the returns of the S&P500 index. The portfolios built using the S&P500 forecasts underperformed all the sector-based portfolios across the different models except for the SVM.

Table 5.4: Hybrid Portfolios Performance

'ortfolios	Annualized Return	Annualized Volatility	Sharpe Ratio	Minimum Return	Maximum Return	Var
S&P 500 (Benchmark)	9,6%	19,3%	0,27	-12,0%	11,6%	-2,0%
Hy Equal Long	11,2%	14,3%	0,48	-11,5%	8,5%	-1,5%
Hy Weighted Long	11,1%	15,1%	0,45	-11,8%	9,3%	-1,6%
Hy S&P500	11,1%	16,7%	0,41	-12,0%	9,4%	-1,7%
CONSUMER DISCRETIONARY	9,3%	20,6%	0,24	-12,1%	13,1%	-2,1%
Δ (Hy)	-2,4%	-1,3%	-0,09	0,0%	0,0%	0,1%
CONSUMER STAPLES	10,5%	14,5%	0,43	-9,2%	8,5%	-1,5%
Δ (Hy)	1,6%	-1,0%	0,13	0,0%	-0,7%	0,1%
ENERGY	10,7%	23,6%	0,27	-20,0%	14,2%	-2,4%
Δ (Hy)	-0,5%	-4,0%	0,02	0,0%	-4,3%	0,4%
FINANCIALS	9,5%	18,3%	0,28	-14,0%	13,2%	-1,9%
Δ (Hy)	-0,4%	-10,9%	0,09	3,0%	-5,5%	1,1%
HEALTH CARE	10,2%	18,3%	0,32	-10,0%	12,4%	-1,9%
Δ (Hy)	-0,9%	-0,4%	-0,04	0,0%	0,0%	0,0%
INDUSTRIALS	8,3%	19,1%	0,21	-11,4%	12,8%	-2,0%
Δ (Hy)	-1,3%	-2,2%	-0,04	0,0%	0,0%	0,2%
MATERIALS	12,2%	20,7%	0,38	-11,4%	11,6%	-2,1%
Δ (Hy)	2,9%	-3,0%	0,17	0,7%	-1,6%	0,3%
TECHNOLOGY	19,3%	24,2%	0,62	-13,9%	12,0%	-2,5%
Δ (Hy)	5,8%	-3,4%	0,29	0,0%	-5,5%	0,4%
COMM. SVS	10,2%	18,2%	0,33	-10,4%	9,2%	-1,9%
Δ (Hy)	3,7%	-4,1%	0,22	0,0%	-4,6%	0,4%
UTILITIES	11,5%	16,7%	0,43	-11,5%	13,1%	-1,7%
Δ (Hy)	2,3%	-2,5%	0,18	0,0%	-0,4%	0,3%

Note: Δ (Hy) referrers to change in performance relative to original index

The portfolios based on the best performing ML algorithm, namely the Hybrid model, are overviewed in table 5.4. The gain in performance is not consistent across the different sectors, with cases such as the Materials, Technology, Communication Services and Utilities sector, which doubled in sharp ratio. In contrast, the models' predictions of the other sectors either lead to similar performance to the original index or even worsened such as the consumer discretionary or health care sectors. While the models might present similar performance using statistical metrics, when evaluated based on the economic value gained in the portfolios larger disparities become clearer. While the financial sector possessed the lowest accuracy and F1 score, the forecast would have still benefitted an investor increasing the sector's sharp ratio. Therefore, the sectors benefited unevenly from the usage of the different ML models answering the second research question.

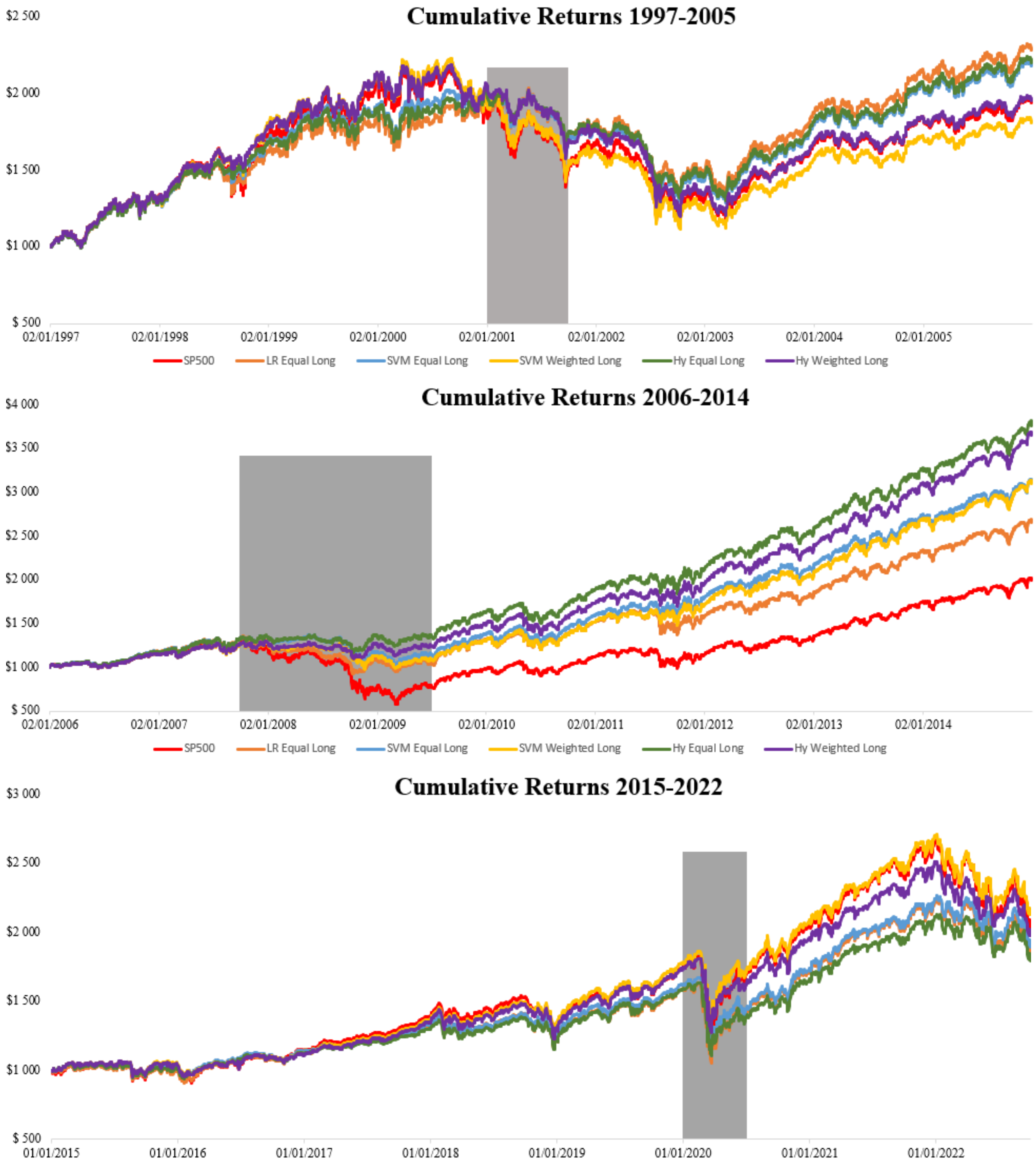
Table 5.5: Yearly Portfolio Returns

Yearly Returns	S&P 500	LR Equal Long	LR Weighted Long	SVM Equal Long	SVM Weighted Long	NN Equal Long	NN Weighted Long	Hy Equal Long	Hy Weighted Long
1997	33,4%	31,8%	33,4%	31,4%	32,6%	31,2%	32,4%	31,4%	32,7%
1998	28,6%	23,2%	25,1%	28,1%	35,4%	28,3%	34,9%	27,9%	34,2%
1999	21,0%	13,5%	20,3%	14,5%	19,1%	13,9%	20,8%	13,5%	20,2%
2000	-9,1%	7,7%	-2,0%	5,6%	-8,3%	5,1%	-8,8%	6,6%	-3,5%
2001	-11,9%	-9,3%	-10,4%	-13,5%	-17,5%	-15,4%	-19,2%	-11,9%	-14,8%
2002	-22,1%	-18,1%	-21,5%	-18,7%	-22,8%	-19,2%	-23,6%	-18,7%	-23,7%
2003	28,7%	27,4%	28,6%	26,0%	26,1%	25,5%	26,5%	25,0%	25,4%
2004	10,9%	14,5%	10,9%	13,7%	9,4%	14,2%	10,8%	13,6%	9,7%
2005	4,9%	6,0%	4,9%	6,2%	5,2%	6,0%	5,0%	6,3%	5,6%
2006	15,8%	18,3%	15,7%	18,1%	14,8%	18,1%	14,9%	17,8%	14,8%
2007	5,5%	9,7%	5,4%	14,2%	11,1%	14,6%	11,8%	14,9%	11,9%
2008	-37,0%	-16,4%	-19,5%	-12,6%	-11,2%	-13,7%	-13,2%	0,4%	-1,9%
2009	26,5%	20,4%	20,4%	16,2%	15,0%	13,4%	12,2%	19,3%	17,9%
2010	15,1%	16,3%	15,1%	16,4%	16,6%	15,5%	15,1%	16,3%	16,6%
2011	2,1%	3,9%	2,1%	12,3%	12,3%	7,2%	6,9%	12,6%	12,4%
2012	16,0%	15,2%	16,0%	18,9%	19,8%	19,8%	20,8%	19,5%	20,4%
2013	32,4%	29,0%	32,4%	29,1%	32,4%	29,5%	32,8%	29,5%	32,8%
2014	13,7%	13,1%	14,4%	13,1%	14,6%	14,1%	16,5%	14,7%	17,2%
2015	1,4%	-1,1%	1,2%	2,3%	4,6%	1,9%	4,5%	0,8%	3,6%
2016	12,0%	14,7%	11,6%	10,4%	6,1%	10,7%	6,2%	10,9%	6,8%
2017	21,8%	16,1%	21,9%	17,5%	22,5%	16,4%	21,2%	16,0%	21,3%
2018	-4,4%	-6,6%	-4,5%	-5,0%	-0,9%	-5,0%	-1,5%	-5,4%	-2,2%
2019	31,5%	28,9%	31,7%	28,6%	31,5%	28,5%	31,5%	29,2%	31,9%
2020	18,4%	8,7%	18,1%	7,6%	18,0%	8,8%	15,1%	4,7%	13,5%
2021	28,7%	28,8%	28,3%	28,6%	28,6%	27,1%	26,9%	27,1%	26,9%
2022	-23,9%	-16,8%	-23,3%	-15,4%	-20,9%	-16,0%	-22,0%	-14,9%	-20,7%

Green cells highlight the strategies which beat the S&P500(benchmark) in a given year. Grey Cells highlight years with larger market declines.

In order to answer the third and last research question, namely whether the ML based portfolios are able to protect investors against market declines, the yearly returns of the different strategies will be overviewed. During the period of 2000 to 2002, commonly associated with the burst of the dotcom bubble, the different strategies were able to not only decrease losses, but also in certain cases achieve a positive return. It is important to note the portfolios built using predictions from the logistic regression, consistently beat the benchmark as well as all the other strategies over the three-year period.

Figure 5.1: Top Portfolios cumulative returns of $1000



Highlighted grey areas represent periods when the US Economy was facing a Recession.

Similarly, during the 2008 Financial crisis all models would have been useful in protecting investors against the large market decline. During the year, the best strategy was the equal weighted allocation based on the forecasts from the hybrid model leading to a return of 0.4%, remarkably better than the -37% decline of the S&P500 index. More recently in 2022 all the models proved useful, with the best strategy being the equal weighted sector allocation based on forecasts from the hybrid model, achieving a return of -14.9% versus the -23.9% return of the benchmark. During the 26-year period, the models not only reduced the exposure of the investor to sharp stock market declines but would have also been useful outside recessionary periods.

6. Conclusion

This dissertation focused on predicting market movements on a sector basis, showing that not only can ML models increase Sharpe Ratios, but also the performance gained varied significantly across sectors. Three different research questions were outlined. Firstly, regarding the advantages of performing market timing strategies on a sector basis. The second question focused on understanding whether disparities existed in the performance gained. Lastly, the third question focused on the general usefulness of ML models in protecting investors against market declines.

The first and second question can be answered jointly, all the equal and value weighted portfolios outperformed the buy and hold strategy, beating the S&P500 both in absolute and risk adjusted returns. While performance gains varied, with some sectors greatly benefiting from the models' predictions, while others either maintained or worsened in performance. However, when building the equal and value weighted portfolios the best performing sectors, such as Technology or communication services, compensated for the model's underperformance in the remaining sectors. When observing the yearly performance of the different portfolios the answer to the third question becomes clear. All the models would have been useful in protecting investors against large declines in the stock market. During more volatility years, when the S&P500 declined substantially such as the 2008 financial crisis, the models would have not only reduced the drawdown in the portfolios, but in certain cases lead to a positive yearly return.

The different results mentioned through the thesis possess some limitations. In this paper a kitchen sink model was used composed of several variables, which don´t provide individual information over each sector. It could be argued that if sector specific variables were chosen the results could change significantly. Additionally in many cases the chosen potential predictors based on finance papers were published years after the backtesting period started. Consequently, the results could be positively influenced by including variables shown to have worked during the OOS period. Similarly, the study was performed based on historical data and therefore there is no guarantee both the usefulness of the chosen predictors, as well as the results from the portfolios will persist in the future. Furthermore, the hyperparameter tuning of the models was performed using a validation period set between 1992 and 1996 and remained fixed throughout the OOS period. Therefore, a different approach, which could potentially improve results, would be to tune the hyperparameters periodically during the backtesting period.

With the obtained findings several further questions arise to be studied in the future. In this thesis forecasts were performed based on a time series approach, however would similar disparities in results between sectors appear using a cross-sectional methodology. In the dissertation each sector was represented based on a single index, limiting the amount of data available regarding the underlying securities included in each sector. Therefore, a future avenue could be performing a similar study applying the forecasting models on individual stocks and applying strategies based on the company's sector.

Bibliography

Ang, A. and Bekaert, G. (2004) 'How Regimes Affect Asset Allocation', Financial Analysts Journal, 60(2), pp. 86–99. Available at: https://doi.org/10.2469/faj.v60.n2.2612.

Baker, M. and Wurgler, J. (2007) 'Investor Sentiment in the Stock Market', Journal of Economic Perspectives, 21(2), pp. 129–152. Available at: https://doi.org/10.1257/jep.21.2.129.

Bianchi, D., Büchner, M. and Tamoni, A. (2021) 'Bond Risk Premiums with Machine Learning', The Review of Financial Studies, 34(2), pp. 1046–1089. Available at: https://doi.org/10.1093/rfs/hhaa062.

Boudoukh, J., Richardson, M. and Whitelaw, R.F. (1994) 'Industry Returns and the Fisher Effect', The Journal of Finance, 49(5), pp. 1595–1615. Available at: https://doi.org/10.1111/j.1540-6261.1994.tb04774.x.

Box, G.E.P. (1976) 'Science and Statistics', Journal of the American Statistical Association, 71(356), pp. 791–799. Available at: https://doi.org/10.1080/01621459.1976.10480949.

Boyd, J.H., Hu, J. and Jagannathan, R. (2005) 'The Stock Market's Reaction to Unemployment News: Why Bad News Is Usually Good for Stocks', The Journal of Finance, 60(2), pp. 649–672. Available at: https://doi.org/10.1111/j.1540-6261.2005.00742.x.

Branch, W.A. and Evans, G.W. (2010) 'Asset Return Dynamics and Learning', The Review of Financial Studies, 23(4), pp. 1651–1680. Available at: https://doi.org/10.1093/rfs/hhp112.

Burkov, A. (2019) The Hundred-Page Machine Learning Book. Available at: https://hal.archives-ouvertes.fr/hal-02271698 (Accessed: 14 December 2022).

Campbell, J.Y. and Shiller, R.J. (1988) 'Stock Prices, Earnings, and Expected Dividends', The Journal of Finance, 43(3), pp. 661–676. Available at: https://doi.org/10.1111/j.1540-6261.1988.tb04598.x.

Cenesizoglu, T. and Timmermann, A. (2012) 'Do return prediction models add economic value?', Journal of Banking & Finance, 36(11), pp. 2974–2987. Available at: https://doi.org/10.1016/j.jbankfin.2012.06.008.

Da, Z., Engelberg, J. and Gao, P. (2015) 'The Sum of All FEARS Investor Sentiment and Asset Prices', The Review of Financial Studies, 28(1), pp. 1–32. Available at: https://doi.org/10.1093/rfs/hhu072.

Daniel, K., Hirshleifer, D. and Subrahmanyam, A. (1998) 'Investor Psychology and Security Market Under- and Overreactions', The Journal of Finance, 53(6), pp. 1839–1885. Available at: https://doi.org/10.1111/0022-1082.00077.

De BONDT, W.F.M. and Thaler, R. (1985) 'Does the Stock Market Overreact?', The Journal of Finance, 40(3), pp. 793–805. Available at: https://doi.org/10.1111/j.1540-6261.1985.tb05004.x.

Fama, E.F. (1970) 'Efficient Capital Markets: A Review of Theory and Empirical Work', The Journal of Finance, 25(2), pp. 383–417. Available at: https://doi.org/10.2307/2325486.

Fama, E.F. and French, K.R. (1989) 'Business conditions and expected returns on stocks and bonds', Journal of Financial Economics, 25(1), pp. 23–49. Available at: https://doi.org/10.1016/0304-405X(89)90095-0.

Giglio, S., Kelly, B. and Pruitt, S. (2016) 'Systemic risk and the macroeconomy: An empirical evaluation', Journal of Financial Economics, 119(3), pp. 457–471. Available at: https://doi.org/10.1016/j.jfineco.2016.01.010.

Giglio, S., Kelly, B.T. and Xiu, D. (2021) 'Factor Models, Machine Learning, and Asset Pricing'. Rochester, NY. Available at: https://doi.org/10.2139/ssrn.3943284.

Goyal, A., Welch, I. and Zafirov, A. (2021) 'A Comprehensive Look at the Empirical Performance of Equity Premium Prediction II'. Rochester, NY. Available at: https://doi.org/10.2139/ssrn.3929119.

Gu, S., Kelly, B. and Xiu, D. (2020) 'Empirical Asset Pricing via Machine Learning', The Review of Financial Studies, 33(5), pp. 2223–2273. Available at: https://doi.org/10.1093/rfs/hhaa009.

Han, Y., Yang, K. and Zhou, G. (2013) 'A New Anomaly: The Cross-Sectional Profitability of Technical Analysis', Journal of Financial and Quantitative Analysis, 48(5), pp. 1433–1461. Available at: https://doi.org/10.1017/S0022109013000586.

Hou, K., Xue, C. and Zhang, L. (2015) 'Digesting Anomalies: An Investment Approach', The Review of Financial Studies, 28(3), pp. 650–705. Available at: https://doi.org/10.1093/rfs/hhu068.

Jensen, M.C. (1968) 'The Performance of Mutual Funds in the Period 1945-1964', The Journal of Finance, 23(2), pp. 389–416. Available at: https://doi.org/10.2307/2325404.

Keim, D.B. and Stambaugh, R.F. (1986) 'Predicting returns in the stock and bond markets', Journal of Financial Economics, 17(2), pp. 357–390. Available at: https://doi.org/10.1016/0304-405X(86)90070-X.

Kelly, B.T., Malamud, S. and Zhou, K. (2022) 'The Virtue of Complexity in Return Prediction'. Rochester, NY. Available at: https://papers.ssrn.com/abstract=4121369 (Accessed: 23 September 2022).

Kothari, S.P. and Shanken, J. (1997) 'Book-to-market, dividend yield, and expected market returns: A time-series analysis', Journal of Financial Economics, 44(2), pp. 169–203. Available at: https://doi.org/10.1016/S0304-405X(97)00002-0.

Lamont, O. (1998) 'Earnings and Expected Returns', The Journal of Finance, 53(5), pp. 1563–1587. Available at: https://doi.org/10.1111/0022-1082.00065.

Lewellen, J. (2004) 'Predicting returns with financial ratios', Journal of Financial Economics, 74(2), pp. 209–235. Available at: https://doi.org/10.1016/j.jfineco.2002.11.002.

Malkiel, B.G. (1995) 'Returns from Investing in Equity Mutual Funds 1971 to 1991', The Journal of Finance, 50(2), pp. 549–572. Available at: https://doi.org/10.2307/2329419.

Mclean, R.D. and Pontiff, J. (2016) 'Does Academic Research Destroy Stock Return Predictability?', The Journal of Finance, 71(1), pp. 5–32. Available at: https://doi.org/10.1111/jofi.12365.

Miranda-Agrippino, S. and Rey, H. (2020) 'U.S. Monetary Policy and the Global Financial Cycle', The Review of Economic Studies, 87(6), pp. 2754–2776. Available at: https://doi.org/10.1093/restud/rdaa019.

Moskowitz, T.J., Ooi, Y.H. and Pedersen, L.H. (2012) 'Time series momentum', Journal of Financial Economics, 104(2), pp. 228–250. Available at: https://doi.org/10.1016/j.jfineco.2011.11.003.

Pesaran, M.H. and Timmermann, A. (2002) 'Market timing and return prediction under model instability', Journal of Empirical Finance, 9(5), pp. 495–510. Available at: https://doi.org/10.1016/S0927-5398(02)00007-5.

Phan, D.H.B., Sharma, S.S. and Narayan, P.K. (2015) 'Stock return forecasting: Some new evidence', International Review of Financial Analysis, 40, pp. 38–51. Available at: https://doi.org/10.1016/j.irfa.2015.05.002.

Puckett, A. and Yan, X. (sterling) (2011) 'The Interim Trading Skills of Institutional Investors', The Journal of Finance, 66(2), pp. 601–633. Available at: https://doi.org/10.1111/j.1540-6261.2010.01643.x.

Rapach, D.E., Strauss, J.K. and Zhou, G. (2013) 'International Stock Return Predictability: What Is the Role of the United States?', The Journal of Finance, 68(4), pp. 1633–1662. Available at: https://doi.org/10.1111/jofi.12041.

Rogalski, R.J. and Vinso, J.D. (1977) 'Stock Returns, Money Supply and the Direction of Causality', The Journal of Finance, 32(4), pp. 1017–1030. Available at: https://doi.org/10.2307/2326509.

Schwert, G.W. (2003) 'Chapter 15 Anomalies and market efficiency', in Handbook of the Economics of Finance. Elsevier (Financial Markets and Asset Pricing), pp. 939–974. Available at: https://doi.org/10.1016/S1574-0102(03)01024-0.

Sim, N. and Zhou, H. (2015) 'Oil prices, US stock return, and the dependence between their quantiles', Journal of Banking & Finance, 55, pp. 1–8. Available at: https://doi.org/10.1016/j.jbankfin.2015.01.013.

Stambaugh, R.F., Yu, J. and Yuan, Y. (2012) 'The short of it: Investor sentiment and anomalies', Journal of Financial Economics, 104(2), pp. 288–302. Available at: https://doi.org/10.1016/j.jfineco.2011.12.001.

Welch, I. and Goyal, A. (2008) 'A Comprehensive Look at The Empirical Performance of Equity Premium Prediction', The Review of Financial Studies, 21(4), pp. 1455–1508. Available at: https://doi.org/10.1093/rfs/hhm014.

Appendix

Table A1: Description of the chosen features

Name	Type	Update Frequency	Description	Source
CAPE Ratio	Fundamental	Monthly	Cyclically adjusted price-to-earnings ratio	http://www.econ.yale.edu/~shiller/data.htm
Dividend Yield	Fundamental	Monthly	Financial ratio which measures the annual value of dividends relative to market value	http://www.econ.yale.edu/~shiller/data.htm
US Unemployment Initial Claims	Macro	Weekly	Weekly number of Americans filing for unemployment benefits	U.S. Department of Labor
WTI Crude Oil	Macro	Daily	WTI oil price percentage change over the last 22 trading days (t-22)	Refinitiv Eikon Datastream
US Dollar Index DXY	Macro	Daily	Change of the value of the US dollar over a 22 day period against a basket of major currencies (t-22)	ICE Data Services
US Unemployment Rate	Macro	Monthly	Percentage of unemployed actively seeking employment within the past four weeks	Bureau of Labor Statistics
US Money Supply M2	Macro	Monthly	Monthly percentage change of the US money supply	Federal Reserve
Term Spread	Macro	Daily	Difference in yield between US long term debt and short-term debt	Refinitiv Eikon Datastream
TED Spread	Macro	Daily	Difference in yield between short term US Government debt and the interast rate on interbank loans	Refinitiv Eikon Datastream
CPI MOM	Macro	Monthly	Monthly percentage change of the CPI (Consumer Price Index)	http://www.econ.yale.edu/~shiller/data.htm
US Prime Rate	Macro	Monthly	Reference interest rate charged by bank in loans and mortgages (monthly average)	Federal Reserve
Advisors Sentiment Bearish Index	Sentiment	Weekly	Percentage of surveyed investors who have a bearish outlook (survey)	Refinitiv Eikon Datastream
US Consumer Confidence Index	Sentiment	Monthly	Indicator which measures the degree of optimism of consumers regarding current and expected economic conditions.	The Conference Board
US National Association of home builders housing market Index	Sentiment	Monthly	Index measuring home builders sentiment on the level of current and future single-family home sales	NAHB - National Association of Home Builders, United States
US ISM Purchasing Managers Index	Sentiment	Monthly	Index indicative of the demand for products by measuring the amount of ordering activity at factories	ISM - Institute for Supply Management
US Chicago Purchasing Manager Business Index	Sentiment	Monthly	Index based on a monthly survey of supply chain managers measuring economic health of the manufacturing sector	MNI Indicators
Advisors Sentiment Bullish Index	Sentiment	Weekly	Percentage of surveyed investors who have a bullish outlook (survey)	Refinitiv Eikon Datastream
Financials 22 day Volatility (%)	Technical	Daily	Volatility of the Financial sector over the last 22 trading days	Refinitiv Eikon Datastream
S&P500 22 Day Ret (%)	Technical	Daily	Return of the S&P 500 over the last 22 trading days	Refinitiv Eikon Datastream
Financials 5-day Volatility (%)	Technical	Daily	Volatility of the Financial sector over the last 5 trading days	Refinitiv Eikon Datastream
Price - 100 day SMA	Technical	Daily	Spot price of the S&P500 minus its 100-day moving average	Refinitiv Eikon Datastream
CBOE SPX Volatility VIX	Technical	Daily	Implied volatility of the S&P500 derived from option prices	Chicago Board Options Exchange

Table A2: Logistic Regression Portfolios Performance

	Annualized Return	Annualized Volatility	Sharpe Ratio	Minimum Return	Maximum Return	Var
S&P 500 (Benchmark)	9,6%	19,3%	0,27	-12,0%	11,6%	-2,0%
LR Equal Long	10,3%	15,7%	0,38	-11,5%	8,5%	-1,6%
LR Weighted Long	10,1%	17,0%	0,35	-11,8%	9,3%	-1,8%
LR S&P500	9,8%	17,5%	0,32	-12,0%	9,4%	-1,8%
CONSUMER DISCRETIONARY	10,5%	21,2%	0,30	-12,1%	13,1%	-2,2%
Δ (LR)	-1,2%	-0,8%	-0,04	0,0%	0,0%	0,1%
CONSUMER STAPLES	9,4%	15,2%	0,34	-9,2%	9,2%	-1,6%
Δ (LR)	0,5%	-0,2%	0,04	0,0%	0,0%	0,0%
ENERGY	10,5%	24,1%	0,26	-20,0%	14,2%	-2,5%
Δ (LR)	-0,7%	-3,6%	0,01	0,0%	-4,3%	0,4%
FINANCIALS	8,8%	22,9%	0,19	-14,0%	13,2%	-2,4%
Δ (LR)	-1,1%	-6,3%	0,01	3,0%	-5,5%	0,7%
HEALTH CARE	10,4%	18,4%	0,33	-10,0%	12,4%	-1,9%
Δ (LR)	-0,7%	-0,4%	-0,03	0,0%	0,0%	0,0%
INDUSTRIALS	8,6%	19,3%	0,22	-11,4%	12,8%	-2,0%
Δ (LR)	-1,0%	-1,9%	-0,03	0,0%	0,0%	0,2%
MATERIALS	10,0%	21,5%	0,27	-11,4%	11,6%	-2,2%
Δ (LR)	0,7%	-2,2%	0,05	0,7%	-1,6%	0,2%
TECHNOLOGY	15,8%	26,1%	0,44	-13,9%	12,1%	-2,7%
Δ (LR)	2,3%	-1,5%	0,11	0,0%	-5,3%	0,2%
COMM. SVS	7,1%	19,1%	0,15	-10,4%	9,2%	-2,0%
Δ (LR)	0,6%	-3,2%	0,05	0,0%	-4,6%	0,3%
UTILITIES	11,8%	16,7%	0,45	-11,5%	13,1%	-1,7%
Δ (LR)	2,6%	-2,5%	0,20	0,0%	-0,4%	0,3%

Note: Δ (LR) referrers to change in performance relative to original index.

Table A3: Support Vector Machines Portfolios Performance

Portfolios	Annualized Return	Annualized Volatility	Sharpe Ratio	Minimum Return	Maximum Return	Var
S&P 500 (Benchmark)	9,6%	19,3%	0,27	-12,0%	11,6%	-2,0%
SVM Equal Long	10,7%	15,2%	0,42	-11,5%	9,5%	-1,6%
SVM Weighted Long	10,6%	16,1%	0,39	-11,8%	9,3%	-1,7%
SVM S&P500	12,9%	17,0%	0,51	-12,0%	9,4%	-1,8%
CONSUMER DISCRETIONARY	8,5%	20,9%	0,20	-12,1%	13,1%	-2,2%
Δ (SVM)	-3,2%	-1,1%	-0,13	0,0%	0,0%	0,1%
CONSUMER STAPLES	9,7%	14,9%	0,37	-9,2%	9,2%	-1,5%
Δ (SVM)	0,8%	-0,6%	0,06	0,0%	0,0%	0,1%
ENERGY	11,8%	24,4%	0,31	-20,0%	16,3%	-2,5%
Δ (SVM)	0,6%	-3,3%	0,06	0,0%	-2,2%	0,3%
FINANCIALS	10,2%	21,7%	0,27	-17,0%	18,8%	-2,2%
Δ (SVM)	0,3%	-7,6%	0,08	0,0%	0,0%	0,8%
HEALTH CARE	10,8%	18,6%	0,35	-10,0%	12,4%	-1,9%
Δ (SVM)	-0,3%	-0,1%	-0,01	0,0%	0,0%	0,0%
INDUSTRIALS	8,1%	19,6%	0,19	-11,4%	12,8%	-2,0%
Δ (SVM)	-1,5%	-1,6%	-0,05	0,0%	0,0%	0,2%
MATERIALS	10,5%	21,8%	0,29	-12,1%	13,3%	-2,3%
Δ (SVM)	1,2%	-1,9%	0,07	0,0%	0,0%	0,2%
TECHNOLOGY	17,6%	26,0%	0,51	-13,9%	17,4%	-2,7%
Δ (SVM)	4,1%	-1,6%	0,18	0,0%	0,0%	0,2%
COMM. SVS	9,3%	20,0%	0,25	-10,4%	13,8%	-2,1%
Δ (SVM)	2,8%	-2,3%	0,15	0,0%	0,0%	0,2%
UTILITIES	10,9%	17,0%	0,39	-11,5%	13,1%	-1,8%
Δ (SVM)	1,7%	-2,2%	0,13	0,0%	-0,4%	0,2%

Note: Δ (SVM) referrers to change in performance relative to original index

Table A4: Neural Networks Portfolios Performance

Portfolios	Annualized Return	Annualized Volatility	Sharpe Ratio	Minimum Return	Maximum Return	Var
S&P 500 (Benchmark)	9,6%	19,3%	0,27	-12,0%	11,6%	-2,0%
NN Equal Long	10,3%	15,2%	0,39	-11,5%	8,8%	-1,6%
NN Weighted Long	10,0%	16,0%	0,36	-11,8%	9,3%	-1,7%
NN S&P500	10,7%	17,2%	0,37	-12,0%	9,4%	-1,8%
CONSUMER DISCRETIONARY	8,3%	21,0%	0,19	-12,1%	13,1%	-2,2%
Δ (NN)	-3,4%	-1,0%	-0,15	0,0%	0,0%	0,1%
CONSUMER STAPLES	9,5%	14,6%	0,36	-9,2%	8,5%	-1,5%
Δ (NN)	0,6%	-0,8%	0,06	0,0%	-0,7%	0,1%
ENERGY	11,9%	25,2%	0,30	-20,0%	18,5%	-2,6%
Δ (NN)	0,7%	-2,5%	0,05	0,0%	0,0%	0,3%
FINANCIALS	9,4%	21,9%	0,23	-17,0%	18,8%	-2,3%
Δ (NN)	-0,4%	-7,4%	0,04	0,0%	0,0%	0,8%
HEALTH CARE	10,6%	18,6%	0,34	-10,0%	12,4%	-1,9%
Δ (NN)	-0,4%	-0,1%	-0,02	0,0%	0,0%	0,0%
INDUSTRIALS	7,4%	19,6%	0,16	-11,4%	12,8%	-2,0%
Δ (NN)	-2,1%	-1,7%	-0,09	0,0%	0,0%	0,2%
MATERIALS	10,4%	22,6%	0,27	-12,1%	13,3%	-2,3%
Δ (NN)	1,1%	-1,1%	0,06	0,0%	0,0%	0,1%
TECHNOLOGY	15,6%	25,4%	0,44	-13,9%	17,4%	-2,6%
Δ (NN)	2,1%	-2,1%	0,11	0,0%	0,0%	0,2%
COMM. SVS	8,9%	19,8%	0,23	-10,4%	13,3%	-2,1%
Δ (NN)	2,3%	-2,5%	0,13	0,0%	-0,5%	0,3%
UTILITIES	10,5%	16,9%	0,37	-11,5%	13,1%	-1,7%
Δ (NN)	1,4%	-2,3%	0,12	0,0%	-0,4%	0,2%

Note: Δ (NN) referrers to change in performance relative to original index