ID Cover Page

Summary of WP Student Team

Analysis of Quantitative Investment Strategies

Group constitution:

Student Name	Program	Individual Title
Enrique Ferrari	Finance	Volatility Forecasting with GARCH models and Recurrent Neural Networks
Fynn Scherzler	Finance	Value and Momentum recently
Micha Sebastiaan Gerla	Finance	Value including intangibles
Gianluca Pecoraro	Finance	Residual Momentum
Hugo José Leitão Eusébio	Finance	LSTM Neural Networks

Work project carried out under the supervision of:
Advisor: Prof. Nicholas H. Hirschey

A Work Project, presented as part of the requirements for the Award of a Master's degree in Finance from the Nova School of Business and Economics.

Analysis of Quantitative Investment Strategies

Group part

Strategy comparison

Enrique Ferrari

Fynn Scherzler

Micha Sebastiaan Gerla

Gianluca Pecoraro

Hugo Eusébio

Work project carried out under the supervision of:

Prof. Nicholas H. Hirschey

16/12/2022

Abstract

This paper tests the combination of five different sub-strategies, resembling the performance of a multi-strategy hedge fund benchmarked against the popular buy-and-hold S&P 500 investing approach. The sub-strategies are: residual momentum, value including intangibles, value and momentum, volatility forecasting, and a long short-term memory strategy, the latter two being machine-learning-based, and all investing in the U.S. universe. The combined strategy's performance is analyzed by three weighting schemes: equal-weight, momentum, and mean-variance, resulting in a gamut of robustness and performance. The combined strategies reap diversification benefits, thereby giving investors a superior risk-reward trade-off compared to the buy-and-hold S&P 500 approach.

Keywords: Systematic trading strategy, Momentum, Value, Volatility Forecasting, Machine learning, Neural Networks, Quantitative trading strategy

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

Table of Contents

ABSTRACT **1**
LIST OF ABBREVIATIONS **3**
LIST OF FIGURES AND TABLES **4**
1. INTRODUCTION **5**
2. INDIVIDUAL STRATEGY PROFILES **7**
2.1 RESIDUAL MOMENTUM 7
2.2 LONG SHORT-TERM MEMORY 9
2.3 VOLATILITY FORECASTING 10
2.4 VALUE AND MOMENTUM 12
2.5 VALUE INCLUDING INTANGIBLES 13
3. DATA DESCRIPTIONS **15**
3.1 RESIDUAL MOMENTUM 15
3.2 LONG SHORT-TERM MEMORY 17
3.3 VOLATILITY FORECASTING 18
3.4 VALUE AND MOMENTUM 20
3.5 VALUE INCLUDING INTANGIBLES 21
3.6 LIMITATIONS 22
4. COMBINED STRATEGY ANALYSIS **24**
4.1 INDIVIDUAL STRATEGIES PERFORMANCE REVIEW 26
4.2 COMBINED STRATEGY PERFORMANCE 35
4.2.1 Equal weight *38*
4.2.2 Momentum optimization *39*
4.2.3 Mean-Variance optimization *42*
5. RESULTS **45**
6. CONCLUSION **48**
REFERENCES ..... **IV**

List of Abbreviations

ARCH	autoregressive conditional heteroskedasticity
B/M	Book-to-Market
CAPM	Capital Asset Pricing Model
ETF	Exchange Traded Fund
FANMAG	Reference to Facebook, Amazon, Netflix, Microsoft, Apple and, Google
FF3	Fama French 3-factor model
GICS	Global Industry Classification Code
HML	High-Minus-Low
iHML	High-Minus-Low Including Intangibles
LSTM	Long Short-Term Memory
S&P500	Standard & Poor 500
SMB	Small-Minus-Big
T-Bills	Treasury Bills
US	United States
VIX	CBOE Volatility Index
WRDS	Wharton Research Data Services

List of Figures and Tables

Figure 1: Strategies comparison 29

Figure 2: Correlation of individual best performers 31

Figure 3: Performance in-sample vs out-of-sample 32

Figure 4: Correlation in-sample vs out-of-sample 34

Figure 5: Combined strategies max Sharpe Ratio vs Max Return 35

Figure 6: Combined strategies In-sample vs. Out-of-sample 38

Figure 7: Yearly distribution of portfolios in Momentum Strategy 40

Figure 8: Efficient frontier replication based on Mean-Variance 43

Table 1: Metrics of individual best strategies 27

Table 2: Metrics combined strategy 37

1. Introduction

This paper tests the combination of five different sub-strategies, resembling the performance of a multi-strategy hedge fund. The intention is to find combined strategies that offer investors in the hedge fund superior returns consistently in all market environments. This is mainly achieved through the combination of different economic signals, through lowering the correlation between the sub-strategies, and thereby improving the combined strategy's risk-return profile. The performance of the combined strategy is evaluated with common performance measures and benchmarked against the popular S&P 500 buy-and-hold strategy.

The five strategy legs are the following: 1) residual value momentum, 2) Long Short-term memory network, 3) volatility forecasting, 4) value and momentum, and 5) value including intangible assets. The investment universe focuses on the U.S. equity market, and therefore the S&P 500 universe. Strategy 1, 2, 4, and 5 trade U.S. common equity (stocks), while strategy 3 trades assets that emulate trading strategies to take volatility positioning. The combined strategy is benefitting from diversification benefits as all strategies use different signals, resulting in potentially wide correlations and performances among the strategies. Strategy 2 and 3 are based on machine learning and aim to predict future returns and volatilities, respectively, while the remaining three use classic trading signals based on price performance and fundamentals. The in-sample period ranges from 01.01.2002 until 31.12.2011 and out-of-sample ranges from 01.01.2012 to 31.12.2021.

The analysis plots the individual as well as the combined strategies and then sorts through differences in performance among the sub-strategies in different market environments. There are three sub-strategies used. The first sub-strategy combines the five individual strategies with the highest cumulative return based on equal weighted returns. The second sub-strategy is constructed by the individual strategies that yield the highest combined Sharpe ratio. Lastly, the third strategy uses a momentum approach. For each period, the strategy/ invests in the best performing strategy of the previous time period. They are diverse sub-strategies, each by themselves expanding existing literature by unique adjustments and an extended time frame. Their combination furthermore forms the basis for a promising and interesting analysis, especially as all revolve in the same equity universe. The evaluation of the strategies and therefore the back-test is conducted with Python with data from Yahoo, CRSP or Compustat through API's.

The paper is constructed as follows. To start the analysis with a brief introduction to the economic motivation and construction behind the individual strategies' performance is explained. The next section describes the underlying data set and preparation of each individual strategy. After this, we compare the performance of the individual strategies, followed by the combined strategy under different weighting schemes. Finally, we summarize and discuss the results and limitations for practical implementation. The group members' student IDs are: 48296, 48297, 46267, 50915, 48992

2. Individual Strategy Profiles

This section introduces the five strategies and the economic motivation behind each, how they extend the existing literature, as well as a brief summary of their stand-alone performance. The five strategy legs are: 1) residual momentum, 2) long short-term memory network, 3) volatility forecasting, 4) value and momentum, and 5) value including intangible assets. As mentioned in the introduction, the common denominator of the strategies is U.S. equities and the time frame. Nonetheless, the strategies also differ in several ways apart from the trading signal. For instance, Strategies 2 and 3 are machine learning-based strategies, while the other strategies use classic signals based on price performance and fundamental data. Furthermore, strategy 2 trades volatility through ETFs or derivatives, while the other strategies focus on U.S. common equity. For a more in-depth analysis of the individual strategies, we refer to the individual papers.

2.1 Residual Momentum

The residual momentum strategy analyzed by Blitz Huij and Martens (2011) is derived from the conventional momentum strategy. Contrary to the latter which is based on total stock returns they investigate a momentum strategy on the basis of residual returns. The excess returns were estimated using the Fama and French three-factor model, reasoning that momentum has substantial time-varying exposures to the Fama and French three-factor-model. Grundy and Martin (2001) find evidence in their paper that the factor models can explain 95% of winner and loser return variability. Moreover, they argue that the main cause of the momentum phenomenon comes neither from industry nor cross-sectional differences in expected returns. They clearly distinguish between two return components, the “stock-specific-return” and returns related to the Fama French factors. Thus, they determine the winner and losers based on their “stock-specific-return” and compare those against total return winners and losers portfolios over the period of 1926 to 1995. In conclusion, they find the stock-specific-return strategy to be significantly more profitable than the total return strategy.

Economically there are multiple indicators for why a residual momentum strategy might be a good predictor for future returns. Apart from being a more profitable strategy compared to a total return momentum strategy, the residual momentum strategy appears to have further improvements according to Blitz, Huij and Martens (2011). Not only does the residual momentum have a higher long-run average Sharpe ratio than the conventional momentum strategy, but it seems to not have lost its profitability since the early 2000s. The latter confirms the findings of Grundy and Martin (2001) that the negative returns of the moment strategy are attributed to their time-varying exposures to the Fama and French Factors. Especially in 2008 and the first quarter of 2019, the total return momentum's negative market beta caused large losses, whereas residual momentum strategy was less negatively exposed (Blitz, Huij and Martens 2011). They argue in their paper that a residual momentum strategy can deliver not only positive returns during expansions but also in recessions as it is market neutral. Moreover, it is not critically dependent on a structural tilt towards small caps as it is nearly neutral to the Fama and French Size factor.

In line with Blitz, Huij and Martens (2011) argumentation evidence is found that a residual momentum strategy is able to perform well in times of expansion as well as in recessions. The best performing portfolio achieved cumulative returns of 1070% of the whole period of 20 years and outperformed the S&P 500 Index . However, most of the returns are attributed to the in-sample period and the strategy even fails to outperform the S&P 500 Index during the out-of-sample period.

2.2 Long Short-Term Memory

In an attempt to gain a competitive advantage in the market over their rivals, corporations are looking to integrate large data sets in the different decision-making processes. With the exponential increase in information needed to be curated and analyzed, it becomes vital to adopt systems capable of managing and processing said information. With this purpose in mind, companies are allocating large portions of their investments to the artificial intelligence field, with a particular focus in machine learning applications. Machine learning represents a powerful tool in the data analytics and forecasting space due to their ability of handling large sets of data and learning information over time from said data, leading to great results for organizations in terms of optimization and efficiency.

In the finance space, machine learning models are currently being employed in different areas across the industry, including stock market forecasting. Nowadays, multiple hedge funds incorporate predictive algorithms into their quantitative trading strategies, with several studies suggesting that deep learning algorithms have a great degree of prediction power. One report that supports the usefulness of machine learning models in terms of stock market forecasting can be found in detailed in Altay and Satman (2005). The authors implemented an artificial neural netowork model in order to forecast the movement of the ISE National 100 Index, an index tracking the performance of the Turkish stock market, using daily, weekly and montlhy data. They were able to achieve impressive results, in particular when using montlhly data, correctly prediting 78.3% of signs of index. A portfolio constructed on the basis of the model employed in this study would have yielded a far better performance when compared to a simple buy-and-hold strategy. To further enhance the literature in this space and analyze the potential of a machine learning based quantitative strategy, a Long Short-Term Memory Neural Network (LSTM) was implemented. These networks have been regarded as one of the best models to use for stock prediction due to their capacity of processing sequential data in addition to extracting and storing important information, while dropping the information that is not relevant.

The strategy was executed on the stocks that comprise the S&P 500 Index between the years of 2002 to 2021, with the 2002-2011 period serving as our training dataset and 2012-2021 representing our test set. Since the strategy is outlined as a classification problem, the target variable of the model acts as the trading signal, with the dependent variable only assuming two distinct values: 0 or 1. In order to take advantage of long and short positions, the target was first optimized to predict securities that would outperform the benchmark, the S&P 500 index, by at least 0.5% (the model predicted 1 if outperformance was identified, or 0 otherwise) and, afterward, the target was employed to predict the stocks that would underperform the benchmark by the same threshold utilized previously (the dependent variable output would be 1 if underperformance was predicted, or 0 otherwise).

With the predictions, two strategy portfolios were constructed: the Long Portfolio, which goes long on the stocks predicted by the model to outperform the benchmark, and the Long-Short Portfolio. The latter does also shorts the stocks predicted to underperform the S&P 500 Index, in addition of going long on the same securities of the Long Portfolio. At the end of each period, both portfolios are rebalanced to consider the new predictions of the model. Overall, the long portfolio generated the best performance, achieving a cumulative return during the out-of-sample period of 229.80%.

2.3 Volatility Forecasting

Financial Markets' consent for derivatives pricing is the Black-Scholes model. As explained by Sinclair (2008), this model allows to account for all possible, and by their probability weighted outcomes in price development of the underlying. The volatility that is reflected in the price of a derivative resembles the expectation of future volatility by the market. Known as the implied volatility, it is also called the "investor fear gauge". Naturally, the question arises how good of a prediction of the future the implied volatility is. Accordingly, if there is a way to estimate future volatility more accurately than what the markets' expectations are, an alpha can be generated. (Sinclair 2008)

Beside the implied volatility, there are two other pillars that research bases on to evaluate future volatilities: Time-series volatility models and neural networks. While Engle (1982) laid the cornerstone of time-series models with the autoregressive conditional heteroskedasticity (ARCH) model, the first neural network approaches only recently started to gain attention.

The strategy analyzed in the course of this paper attempts to accurately forecast volatility by making use of various time-series model, built on the foundation on the ARCH. Furthermore, also neural network approaches are implemented that range from simple linear regression models to recurrent neural networks. Before developing a successful trading strategy, these models are evaluated for their performance, where ultimately the best time-series model, as well as the best neural network are chosen to forecast volatilities. After accounting for the volatility risk premium, said forecast is then to be compared to the expectations of the market, the implied volatility. The volatility risk premium corresponds to a deviation from the implied volatility to the realized volatility. For further explanations about the volatility risk premium, please refer to the individual part.

If, on the one hand, the forecasted volatility lies above the adjusted implied volatility of the market, a signal that indicated long volatility position is created. On the other hand, when the forecasted volatility lies below that level, a signal to short volatility arises.

The volatility in this strategy is either traded through exchange traded funds, attempting to emulate the VIX, or through monthly options on the S&P 500. Please refer to the group part for an in depth description of the instruments used. While both methods, with both ways to forecast volatilities, yield attractive returns, they are very volatile. Due to a short track record, the VIX ETF strategy is only tested on the out-of-sample period of 2012-2022. However, in that time frame, the daily positions taken, result in a pleasant performance, that stands out with a large negative Beta regarding the S&P 500. When trading the volatility through options, where an index is used that tracks covered short straddles on the S&P 500, the performance is evidently lower. Here, monthly rebalancing and forecasting is applied. Due to the covered nature, however, investors are awarded with a lower volatility in performance. However, also here, low Betas make the performance very appealing.

Interestingly, both ways to trade volatility not only achieve profitable performances, but also demonstrate a neural network approach that is superior to the best time-series model from the ARCH-family.

2.4 Value and Momentum

This sub-strategy is a two-leg trading strategy that allocates portions of the portfolio based on value and momentum factors. Literature shows that both momentum and value strategies earn consistent and significant premia and are therefore widely discussed market anomalies (Jegadeesh and Titman 1993; Fama and French 1992). Asness (1997) finds that measures of momentum and value are negatively correlated across stocks, yet each is positively related to the cross-section of average stock returns. This begs the question of whether a portfolio combining the two could achieve an even higher Sharpe ratio by reaping diversification benefits.

The idea is that value strategies work and are strongest in low-momentum (loser) stocks and weakest among high-momentum (winner) stocks, while momentum strategies are strongest in these expensive (winner) stocks. Asness et al. (2013) confirm the superior performance of the combined strategy across diverse markets and asset classes with a fixed 50/50 portfolio in a time series up to 2011. However, with fixed weights, one captures the negative returns and large drawdowns that the momentum strategies achieve in market crashes (Barroso and Santa-Clara

2015), while value strategies tend to perform better in these times. Conversely, in longer-dated market bull runs momentum tends to perform better. This paper expands the existing literature by adjusting weights between the two based on market volatility to reflect this dynamic.

In conclusion, momentum and value continue their negative correlation in the recent past, and as a result, a combined strategy reaps diversification benefits, allowing for a better risk-return trade-off measured as a higher Sharpe ratio. Additionally, to the extended time frame, allocating overweight into momentum and underweight in value when market volatility is under a certain threshold and vice versa allows to enhance the strategies Sharpe ratio further. As assumed, this enables to limit negative returns in market turmoil, while kurtosis is largely unaffected or increases only slightly. It appears most benefits to limit kurtosis, skewness, minimum monthly return, and maximum drawdown are already realized in the equal-weight strategy, while the volatility-based allocation still enhances returns slightly. Testing out-of-sample the same improvements in Sharpe are found between equal-weight and volatility-based. However, the superior performance of the two strategy legs in the past, in general, stands out. Momentum L/S interestingly seems to have stopped working post-2008, while cumulative returns from the value L/S deteriorated continuously from there.

2.5 Value including intangibles

After the global financial crisis in 2007, value investing seems to have lost its edge to growth investing, experiencing its deepest and longest lasting drawdown since 1963 (Arnott et al. 2020). Value investing is defined as a portfolio going long on the stocks with the highest book-to-market (B/M) ratios and going short on the stocks with the lowest B/M ratios. As a result, there is a risen debate about the continued relevance of value investing. It even led to investors arguing that value investing is "dead". In the recent literature, there are some counter arguments against the "value is dead" argument. First, value stocks have tumbled in value relative to growth stocks. Second, the traditional B/M ratio fails to capture the value of intangible assets.

Therefore, it understates the book value of a firm. Academics found better results introducing a new factor (iHML) for value investing, which capitalizes intangibles into the book value (e.g. Li 2022). However, Berkin et. al (2022) concluded that the performance efficacy of the HML factor has decayed at a much faster rate in High Intangible Industries (High II) compared to Low Intangible Industries (Low II).

This strategy tests the performance of the traditional HML factor compared to the iHML factor. Both the HML factor and iHML factor are constructed in High II industries and the Low II indsutries. By doing this, we observe that the drawdown of value stocks is only based on the performance of the portfolio in the High II industries for both the HML factor and the iHML factor. The long-short strategies in the low II industries did not experience such a drawdown over the out-of-sample period. Furthermore, by constructing equal-weighted returns, the performance of value stocks increases compared to the performance based on value-weighted returns, which is commonly used in literature. Based on these findings, we propose an equal weighted long-only strategy in value stocks. In the out-of-sample period, all portfolios in both industries outperformed the SPX. Furthermore, the iHML factor outperformed the traditional HML factor in both industries. Nonetheless, the relative performance is not as significant in the in-sample period. The HML factor even slightly outperformed the iHML in high II industries.

3. Data descriptions

This chapter provides a brief overview of the data sourcing and the data preparation on all the individual strategies. The time frame of all the strategies is the same in order to perform a proper comparison. In addition, to perform the correct analysis in the combined strategy section, the value and momentum strategy, the LSTM and the value including intangibles strategies' monthly returns out of the individual parts have been converted into daily returns to match the return series of the remaining two strategies, which likely caused a slight smoothening of returns.

3.1 Residual momentum

The data for the residual momentum strategy has been retrieved from the Wharton Research Data Services WRDS web page and data ranges from 01.01.2002 to 31.12.2021. It's a combination of two data samples where the first data sample consisted of a historical composite of S&P 500 stocks with daily returns and the corresponding stock specific information such as ticker, permno, market capitalization, number of share outstanding and trading volume. For each individual year the actual composition of the S&P 500 was utilized and then merged to ensure having the correct compositions.

The second data sample is generated with an analytic tool called "Beta Suite by WRDS". Beta suite was launched in November 2016 by the WRDS research team and is a powerful web tool allowing individuals and researchers to compute stocks loading on various risk factors. It is designed to be flexible is capable of handling monthly, weekly, and daily regressions. (WRDS , 2016). Beta suite offers three regression models to calculate excess returns from which the CAPM and the Fama-French 3-factor model is used to determine excess returns. The estimation window applied on the regression is set to 252 days with a minimum window size of 126 days. The second dataset generated comprised of excess returns, estimated betas, idiosyncratic and total volatility for each individual stock.

The final dataset consists of both daily estimated parameters and daily actual market data. Datapoints were dropped that had less than 6-month of consecutive returns. This is important to ensure an appropriate comparison of cumulative returns for each stock for the analyzed period.

The residual momentum strategy applied in this analysis is based on previous research performed by (Blitz, Huij and Martens 2011). To contribution from a different angle on to the residual momentum research, not only Fama French 3-factor model excess returns are used to construct the investment signal but also CAPM excess returns. Moreover, within the constructed portfolios different weighting schemes are applied. The weighting schemes applied are simple equal weights, value weights based on market capitalization and a weighting based on the annualized return volatility of each stock. The return volatility was computed on a rolling basis of past 6-month returns to be in line with the signal estimation window of 6-month. Blitz, Huij and Martens (2011) in contrast computed only equal weighted portfolios to remain in line with previous research.

The first step to implement the strategy is to construct the signal. The signal indicates whether a certain stock will be bought or sold. In order to elaborate the signal cumulative excess returns are calculated over a window of 6 month or 126 days, assuming a business year has 252 business days. Jegadeesh and Titman (1993) argue that the ideal window of a momentum strategy lies within 3 to 12 months. Thus, 6-month are in line with the optimal time frame determined by Jegadeesh and Titman (1993). In contrast, Blitz, Huij and Martens (2011) use an estimation window of 12 month.

Based on their past 6-month cumulative excess returns, stocks are divided into decile or tercile. This resulted in stocks having performed best over the last six month being grouped to the top decile/tercile and stocks having performed worst over the last six month being grouped to the bottom decile/tercile. The portfolios get rebalanced daily and consequently the holding period for the portfolios is fixed to 1 day. This differentiated the analysis performed further from Blitz, Huij and Martens (2011)'s paper who focused part of their research on finding the ideal holding period and therefore compared holding periods of 1-month, 3-month, 6-month and 12-month.

3.2 Long Short-Term Memory

To make sure the model could make accurate and robust predictions, a diversified set of features was applied to the Long Short-Term Memory network. Contrasting with other studies in this area, financial indicators specific to each stock retrieved from their respective balance sheets and profit and loss statements (such as but not limited to, the company profit margin, dividend growth and return on assets) were incorporated into the model in an attempt to improve the predictive power of the algorithm. In addition to the latter, technical indicators computed from historical data (including the moving average convergence divergence, the relative strength index and the stochastic oscillator) and trend indicators in order to include information about the United States economy. The trend indictors used in the model comprise the business confidence index, which tries to incorporate future information about the US economy since its constructed upon surveys where firms are asked about their future expectations, the consumer confidence index, which is also based upon surveys but measures the optimism that consumers have about the overall state of the American economy, and the CBOE Volatility Index, which represents the stock market's expected volatility over the next thirty days. In addition, the monthly returns of the Standard & Poor's 500 Index were also fed into the model, acting as our benchmark.

Following the data collection process, several techniques were implemented in order to improve and refine the long short-term memory network. First, a MinMax scaler was applied to scale the dataset in the interest of retaining the original significance of the model features. Afterwards, the recursive feature elimination method was employed used to perform feature selection in order to guarantee the model made the predictions using the most consistent and relevant features to the problem in question. Overall, three financial indicators and two technical indicators were removed from the algorithm, with the final model including 16 features of the 21 that were originally tested.

On the whole, the long short-term network was much more prone in predicting securities that would outperform the benchmark, with an average monthly forecast of 21 stocks in this aspect. The maximum number of outperformers predicted in a single month by the model was 37 stocks, while the minimum number of outperformers forecasted was 11 stocks. In terms of underperformers, the average monthly prediction computed by the algorithm was 9 securities, with the maximum and minimum number of stocks predicted to underperform in a single month being 17 and 4 respectively.

3.3 Volatility Forecasting

The volatility forecasting is conducted on the S&P 500. Accordingly daily SPX closing prices from 2002 until 2021 are retrieved from Yahoo Finance. Furthermore, to develop the signal, as proxy for the at-the-money SPX implied volatility, closing prices of the VIX are retrieved. To trade the signal, closing prices of the VIXY, and the SVXY from their respective inception date are also downloaded from Yahoo Finance. When backtesting the signal through the use of options, a proxy index selling covered short straddles is used. This index, the CMBO, can also be retrieved for the full sample period of 20 years directly from the Chicago Board Exchange.

By using the historic volatilities of the S&P 500, the volatility forecasting is conducted with 16 different variations. On a 20-year time-window, using a train-test split with a shifting window to compare the forecast with the actual datapoint, the models are evaluated. The methodologies range from simple baseline models, over time-series models and neural networks. The underlying assumptions and model specifications can be found in the individual part. Finally, the time-series models, as well as the neural networks, each convey a model superior to the best primitive baseline approach. From each group, the best performing model according to the metric of Root Mean Squared Percentage Error is chosen to develop the trading signal. For the time-series models this is a TARCH (1,2,0), and for the neural networks a multivariate bi-direct LSTM model with two layers.

To develop the signal, again, these two are trained from scratch, but on the time frame of the in-sample period. While the two machine learning models have slightly different ways of having the data fed to them, the expectation is that they yield similar results, with the LSTM performing slightly superior. The forecasted one month realized volatility, which is the output received from the models, is then compared with the one month implied volatility in the form of the VIX. The implied volatility, however, contains an observable premium over the realized volatility. This is the so-called volatility risk premium. For further explanations please refer to the individual part. In this project, by factoring out the previous' day volatility risk premium of the difference between forecast and implied volatility, a unique way of dealing with this issue is attempted. Finally, the signal to buy or sell volatility arises, when the forecast lies above or below the adjusted implied volatility level, respectively.

This signal indicates whether on any given inspection time of observation, a long volatility or a short volatility position should be taken. By rebalancing futures, the VIXY tries to emulate the VIX. As explained in the individual part it is risky to short-sell this ETF. Therefore the SVXY, a short VIX ETF is chosen to short volatility. A long volatility position would then correspond to a position in the VIXY, whereas a short volatility position would correspond to a position in the SVXY.

Another way to trade volatility is through options. Theory shows that a straddle, where the investor buys a call and buys a put with the same strike is the best way to take a long volatility position with options. The CMBO is an index that tracks the performance of selling monthly puts and calls, while covering them with long SPX and Treasury positions. In spite of the fact that it is technically not a straddle, but a strangle with a strike gap of 2%, it is considered good enough of a proxy for the purpose of this back-test. Here, a long volatility signal would result in a short position in the CMBO, while a short volatility position results in a long position in the CMBO.

3.4 Value and Momentum

This strategy's data is based on a combination of CRSP and Compustat data sourced from WRDS, comprising monthly returns for companies in the U.S. equities universe comprising the CRSP stock codes 10 & 11, as it is standard in prior literature. The equities universe comprises ordinary common shares and excludes ADRs, REITs, companies incorporated outside the U.S., trusts, and closed-end funds. The investable equities universe is further restricted to exclude the lowest market capitalization decile firms, to ensure sufficient liquidity and trading volume in the stocks and spreads, as well as comparability to prior literature. Diverging from the remaining strategies, the in-sample period in the individual part encompasses the full 20 years from 01.01.2002 to 31.12.2021, as out-of-sample tests are conducted in the estimation windows of the most relevant prior literature. However, for the combined analysis, the return series is simply split into the two separate 10-year windows.

The value strategy leg relies on the common book-to-market value indicator, with annual data sourced from Compustat and lagged six months to ensure data availability at the time, as commonly done. This data is then merged with the current CRSP market capitalization each month, to arrive at the book-to-market ratio. The momentum strategy leg relies on 12-month raw cumulative return measures based on monthly price performance. For the volatility signal, daily S&P 500 returns are sourced from CRSP and rolling annualized volatility is computed based on the past 30-day returns. The volatility levels at each month's end are then merged with the monthly returns of the remaining securities. For portfolio construction, the equities universe will be split into deciles based on performance in terms of the value or momentum signal over a certain formation period J (1 month on value, 6 months in momentum in the base case), and the strategy will keep the long-short position for the holding period K (6 months for both legs in the base case). Therefore, in any given month t, the strategies hold a series of portfolio that are selected in the current month as well as in the previous K – 1 months, as done in the significant momentum literature by Jegadeesh and Titman (1993). The weighting between the two portfolios is equal-weight or fixed 70/30 weighting, depending on the volatility of the S&P 500.

3.5 Value including intangibles

The data for the strategy is retrieved from the WRDS database. The CRSP database is used for the monthly returns of U.S.-based common equities with share codes 10 and 11. The delisting returns (if applicable) are added to the total monthly return of the stocks. In order to split the data sample into High II and Low II industry classification, the four-digit Global Industry Classification Standard (GICS) is used. Accounting data is retrieved from the Compustat database.

Book Equity is calculated as SEQ + TXDITCS – PSTKRV. SEQ is the stockholders' equity, TXDITCS is the balance sheet deferred taxes and investment credit, and PSTKRV is the redemption value of preferred stock. If PSTKRV is not available or zero, the value of the preferred stock is computed based on the total preferred stock liquidating value (PSTKL). If PSTKL is also not available, the value of preferred stock is set to the total preferred stock capital. (PSTK). Otherwise, the preferred stock is set to zero.

The annual R&D expenses is measured as the data item XRD in Compustat. Missing values of XRD are set to zero. We measure annual SG&A as XSGA – XRD – RDIP. There are some exceptions for this computation. If XRD > XSGA and XRD < COGS, we measure SG&A as XSGA. If XSGA is missing or zero, we also set SG&A to zero. The Compustat standardization model includes annual R&D expenses in the XSGA data item, unless the company allocates this cost as COGS. Therefore, XRD is subtracted from XSGA, unless we allocate R&D to COGS.

In order to compute the BE/ME breakpoints for year t, we only use the NYSE stocks in accordance with Fama and French (1993). The breakpoints are separately calculated for the HML factor and the iHML factor. For the size breakpoints, data from the Kenneth French library is utilized. The industry groups are split after the computation of the breakpoints.

we exclude firms with total assets under $5 million in order avoid illiquid, unfrequently traded assets. Moreover, financial institutions are excluded from the sample given the deviating nature compared of the industry compared to other industries. Firms with negative book value are excluded from the sample as well.

3.6 Limitations

For all projects, general limitations are to be considered. First, no bid-ask spread is considered and all strategies are done without factoring in transaction costs. In addition, all strategies assume borrowing to be available and short-selling to be possible in case the asset allows to do so in normal market conditions. Finally, free market access and daily rebalancing are possible without any trading restrictions. The back-tests were done with these presumptions put in place.

With in- and out-of-sample tests, overfitting the data should be mostly ruled out, specially when considering the three non-machine-based-learning strategies use rather conventional signals, not further adopted to the time frame. A forward-looking bias should be avoided by lagging fundamental data. However, not all strategy legs account for transaction costs and short funding costs, which would soften returns further. Also, common issues with using day-close prices or month-close prices occur, not being able to fill orders at that exact price and quantity in practice. A difficulty with the combined strategy is the different rebalancing times, as the residual value uses daily rebalancing, while the remainder uses monthly returns in the combined part. The most critical implementation issue likely arises from the identical investment universe the five strategies use, as different strategy parts might be long and short the same security at the same time.

In terms of the machine learning-based strategies, some specific limitations should also be noted. Data collection and feature selection are crucial steps when building a predictive algorithm and an incorrect decision during this process can substantially alter the results achieved later in the model. It is not only important to collect enough data but also to choose the correct and relevant information to avoid typical machine learning problems such as an overfitting or underfitting model. Another critical procedure to an accurate algorithm is hyperparameter tuning. Unlike parameters, which are derived and learned from the dataset itself during the training process, hyperparameter values are specified by the user before the training stage in order to assert the values of model parameters that the algorithm ends up using. This means that they are very susceptible to human error and can, to a large extent, also dramatically change the final results one obtains.

4. Combined Strategy Analysis

In the following, first the individual sub-strategies' performance is analyzed, followed by the combined strategy, benchmarked in three distinct weighting schemes. The in-sample period ranges from 01.01.2002 until 31.12.2011 and out-of-sample ranges from 01.01.2012 to 31.12.2021. Out of the individual parts, the strategy portfolios with the highest cumulative returns were chosen, as opposed to the highest Sharpe ratio. The combined strategy is expected to reap diversification benefits from the combination of the different strategies, since their investment is based on distinct economic signals. One important thing to note about the combined strategy is that all sub-strategies revolve around the same equities' universe, so diversification benefits from the mere allocation into uncorrelated asset classes are not captured.

In the individual residual momentum strategy part, 36 different portfolios were compared with two factor models being applied, totaling 18 portfolios for each factor model (please refer to section Data and Methodology in the individual analysis for further elaboration). The 36 portfolios were compared based on their cumulative return, their Sharpe ratio, the significance of alpha generated and the tracking error. Overall, the long only volatility weighed portfolio generated the highest cumulative return in addition to the highest Sharpe ratio. The signal to construct portfolio was generated through CAPM excess returns.

In terms of the long short-term memory network strategy, the portfolio chosen to be included in the combined analysis ultimately relied on the long portfolio, since, out of the two portfolios tested in the individual strategy, the long portfolio performed considerably better during the full sample period. In addition to that, it is a more diversified portfolio in terms of the number of securities included, due the reasons mentioned in the data description section.

As for volatility forecasting, the long only (short volatility) CMBO strategy with the signal developed by the LSTM is used for the comparison. While the project of volatility forecasting technically also contains a better performing strategy when using the daily ETFs positioning on the VIX, this is not suitable, since it is only tested on the out-of-sample period. Therefore, to cover the whole time frame, the best performing strategy using optionality is chosen to be compared with the other strategies. Because this strategy consists of monthly covered short-straddle positions, the upside in any given month is limited. However, this should also result in a relatively macro environment neutral strategy. Due to the monthly rolling of the option position the return graphs can look very smooth relatively to the (daily) equity portfolios it is compared to.

For the value and momentum sub-strategy, the best performing L/S variant using top and bottom third of the US equities for both strategy legs is used, ranking securities by their performance among these factors at the end of each month, buying the top third and selling the bottom third. The portfolio is weighted 70/30 between them based on past 30-day annualized S&P 500 volatility signal, which is the main performance improvement developed in the individual part.

Regarding the value including intangibles strategy, the iHML long only strategy in the High II industries is used in the comparable analysis. The strategy invests each year in the top 30% in terms of book-to-market value. The measurement of the book-to-market value in this strategy, however, deviates from the traditional measure by adding intangibles to the book value. The iHML long only strategy in high II industries is the best performing strategy out of all strategies tested and compared in the individual report, overperforming them all in terms of cumulative returns and Sharpe ratio.

In the base scenario, the five sub-strategies are combined in an equal-weight portfolio. Afterwards, out of the individual sections, two variants of the sub-strategies are plotted against each other 1) the portfolio with the highest Sharpe ratio and 2) the portfolio where we use momentum as a signal. A notable drag to the efficient combination is the frequent rebalancing in some of the models. All strategies are code-based and can thereby be automated in their execution, which supports practical implementation. Since all individual strategies are implementable, as discussed in the individual parts, also the combination of them is. With the given weights, at any given inspection point, the portfolio of strategies would just be shifted accordingly. However, as all individual strategies use the same universe, one sub-strategy might be long a certain stock while at the same time another is short, posing a difficulty for practical implementation.

The historical price performance of the combined strategy and its sub-strategies are evaluated by their risk-reward profile expressed by the Sharpe ratio and other common performance measures such as the alpha generated, associated tracking error and information ratio. The analysis provides insights into the performance of the strategies and their variations over time, highlighting some important strong points and shortcomings. Furthermore, the strategies are also benchmarked against a buy-and-hold S&P 500 strategy in order to better understand how they perform relative to the overall market. A 60/40 stocks/bonds portfolio is not used in the comparison, as all strategies comprise equities and derivatives exclusively. Additionally, the respective correlations of the five individual strategies are also discussed in the analysis, in order to better understand how they interact and relate to each other.

4.1 Individual strategies performance review

The performance of the individual strategies is compared over the full 20-year horizon and then split into the in-sample and out-of-sample periods (see Table 1). This is crucial to gain insight into what strategy works best in which market environment, highlighting the potential diversification benefits of the combined strategy.

Table 1: Metrics of individual best strategies

2002-2021

Table 1 shows the main metrics for the performance of all strategies and respective signals for the full period of 2002-2021 with the S&P500 as benchmark.

Strategy	Cumulative Return	Annualized Sharpe Ratio	Annualized Alpha	Beta	Tracking Error	Information Ratio
Volatility forecasting	278.59%	1.01	1.54%	0.18	6.05%	0.26
Volatility forecasting IS	76.06%	0.73	1.47%	0.21	6.66%	0.22
Volatility forecasting OOS	115.03%	1.39	1.70%	0.14	5.34%	0.32
Value and momentum	45.50%	0.22	0.12%	-0.01	1.90%	0.06
Value and momentum IS	42.84%	0.48	0.27%	-0.01	2.29%	0.12
Value and momentum OOS	1.91%	0.03	0.01%	0.00	1.52%	0.01
LSTM Network	866.04%	0.99	1.05%	0.75	10.21%	0.11
LSTM Network IS	231.93%	1.31	1.45%	0.81	13,19%	0.10
LSTM Network OOS	229.80%	0.54	0.80%	0.70	9.37%	0.13
Value including intangibles	1151.15%	0.77	0.98%	1.10	4.68%	0.21
Value including intangibles IS	199.65%	0.62	0.92%	1.12	5..23%	0.18
Value including intangibles OOS	600.61%	1.20	1.04%	1.06	4.19%	0.25
Residual momentum	1073%	0.64	0.93%	1.09	11.42%	0.08
Residual momentum IS	347%	0.74	1.08%	1.20	15.47%	0.07
Residual momentum OOS	160%	0.56	0.83%	0.96	8.78%	0.10

During the full time frame, despite being only the third best strategy returns-wise during the in-sample period, the value including intangibles strategy outperforms the other individual strategies in the analysis with a cumulative return of 1151.15%. After picking up some heat at the beginning of the out-of-sample period, this strategy really shines after 2020, achieving impressive results after a considerable dip due to the coronavirus pandemic stock market crash.

At the other end of the spectrum and standing out as one of the two individual strategies to underperform the S&P 500 Index sits the value and momentum strategy, with a cumulative return during the 20-year period of 45.50%. Contrasting with the value including intangibles strategy, this strategy obtains considerably worse results during the out-of-sample period acquiring a cumulative return of 1.91%, a major difference when compared to the 42.84% earned in the in-sample period.

The residual momentum strategy is the best performer during the in-sample period, finishing 2011 with a cumulative return of 347%. Interestingly, this strategy sees a great boost in its returns in two distinct moments: the first occurs during the in-sample period, after the 2008 financial crises, and the second happens after the 2020 coronavirus pandemic, suggesting that this strategy performs particularly well during stock market recovery periods.

Like the residual momentum strategy, the Long Short-Term Network strategy also obtains good results during the full time frame, performing similarly during the in-sample and out-of-sample periods and achieving a cumulative return at the end of 2021 of 866.04%.

In terms of the volatility forecasting strategy, the good performance during the 2020 stock market crash stands out, with the strategy withstanding that specific time frame without any major dips. However, during the full sample period, the strategy achieves a cumulative return of 278.59%, underperforming the S&P 500 Index.

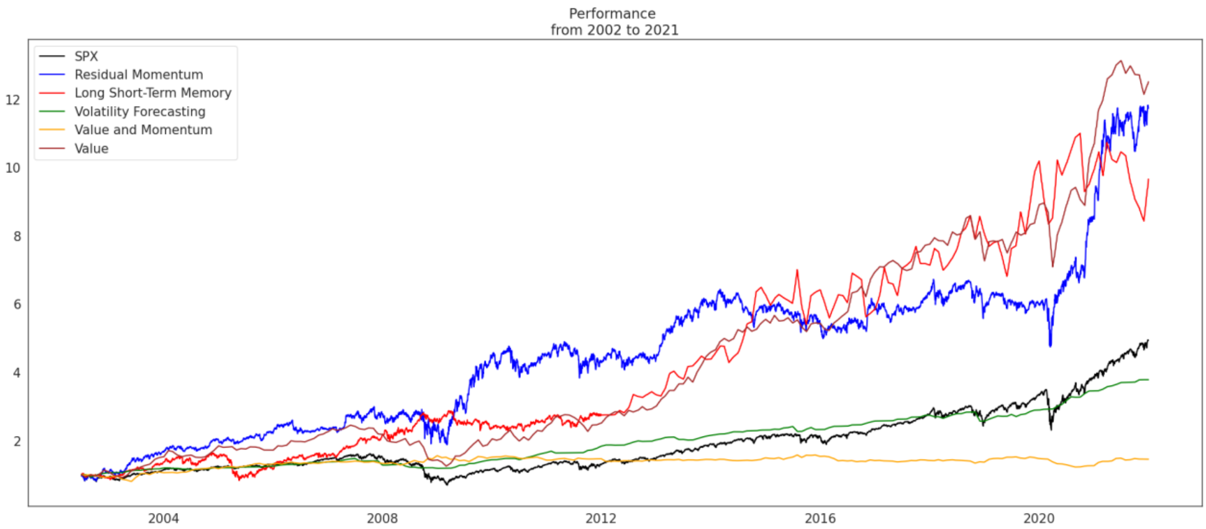
Regarding the risk profile of the individual strategies, during the full sample period, the volatility forecasting strategy performed best with a Sharpe ratio of 1.01. This is the only strategy able to achieve a Sharpe ratio above 1 during the entire time frame, suggesting that the other strategies, despite the impressive results obtained by some, are underlined by a great degree of risk and volatility. Since the volatility forecasting strategy essentially collects a steady premium by selling straddles, this makes sense. Therefore, less volatility in the performance is to be expected.

In this aspect, the value and momentum strategy performed the worst with a Sharpe ratio of 0.22, demonstrating that the strategy doesn't provide an adequate return rate for the risk that it possesses. For the other three strategies, despite obtaining good returns during the full sample period, the Sharpe ratios also suggest that they may take on excessive risk for the performance they end up achieving.

Figure 1: Strategies comparison

2002-2021

Figure 2 plots the performance of the best returns of all strategies from 2002 to 2021. As a benchmark the S&P 500 is plotted as well. While some strategies seem way more volatile than others, this is mainly due to the differences in daily and monthly returns.



In terms of excess returns relative to the market, all five strategies produce a positive alpha for the full sample, in-sample and out-of-sample periods, meaning that they all manage to generate excess returns relative to the market. Once again, the volatility forecasting strategy stands out from the rest with an annualized alpha of 1.54% for the entire time frame and, on the opposite side, the value and momentum strategy is the worst performer when it comes to this criterion, generating an annualized alpha of 0.12%. This can also be well observed in Figure 1 as the volatility forecasting strategy shows very low volatility throughout the whole period while standing out with a low beta of 0.18. This, accounting for the decent return, also explains the highest alpha.

Regarding systematic risk, the value and momentum strategy is the only in the analysis that generates a negative beta (-0.01 for the in-sample and full sample periods) suggesting that the strategy has an inverse relation with the market, meaning that it tends to increase in profitability when the overall market falls and vice versa. This may explain the overall underperformance of this particular strategy since the market, during the full sample period, generated positive returns for the investors. The value including intangibles and the residual momentum strategies both generated betas above 1 during certain time frames, suggesting that they were more volatile than the overall market in those specific periods.

In order to check if the chosen benchmark is appropriate and if the strategies are indeed loading up on additional risk factors, the tracking error is computed. Taking into consideration the general literature consensus, where an annualized tracking error above 6% suggests that the benchmark is not ideal for performance analysis, it's possible to conclude that for the long short-term memory strategy and the residual momentum strategy the benchmark used in the analysis may be inappropriate since both strategies generate tracking errors well above the threshold for all periods analyzed. The other three individual strategies all produce tracking errors below the 6% mark, apart from the volatility forecasting strategy, where a slight violation occurs during the in-sample and full sample periods (6.66% and 6.05% respectively).

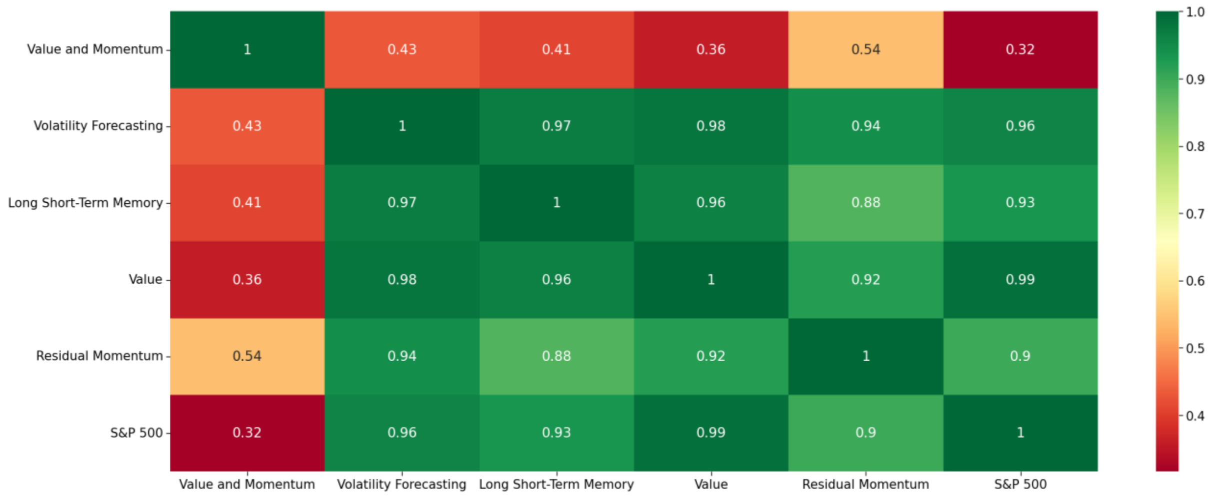
To conclude the individual strategies analysis, the information ratio (IR) is also calculated. This metric allows us to measure the strategies' risk-adjusted returns relative to the benchmark. We consider once again the literature consensus on the topic, which states that annualized information errors above 0.5 are considered adequate. In terms of this indicator, all strategies generate subpar results in all three periods. The highest information ratio is achieved by the volatility forecasting strategy during the out-of-sample period, with an IR 0.32, while the lowest information is produced by the value and momentum strategy, with an IR of 0.01.

In order to explore and understand potential diversification benefits, it's vital to analyze the correlation between the five individual strategies. Since all of them are built from different underlying economic drivers, the expected and desired result is to generate low correlation values between the strategies. When reviewing this metric for the sub-strategies over the 20-year period, it surprisingly stands out that most strategies approximately trade like a basket, with correlation values above the 0.9 mark. The one notable exception is the value and momentum L/S strategy, which has mostly correlation values ranging between the 0.3 and 0.4 thresholds with the remaining strategies. As a result, the combined strategy earns the most diversification benefits when the value and momentum L/S strategy is included, thereby lowering volatility and overall risk. However, due to the poor performance of this individual strategy, its inclusion will also most likely damps the returns of the combined strategy portfolio.

Figure 2: Correlation of individual best performers

2002-2021

Figure 3 shows correlation of the best return portfolios with each other, as well as with the benchmark index S&P 500. Most strategies seem to yield a market like performance.

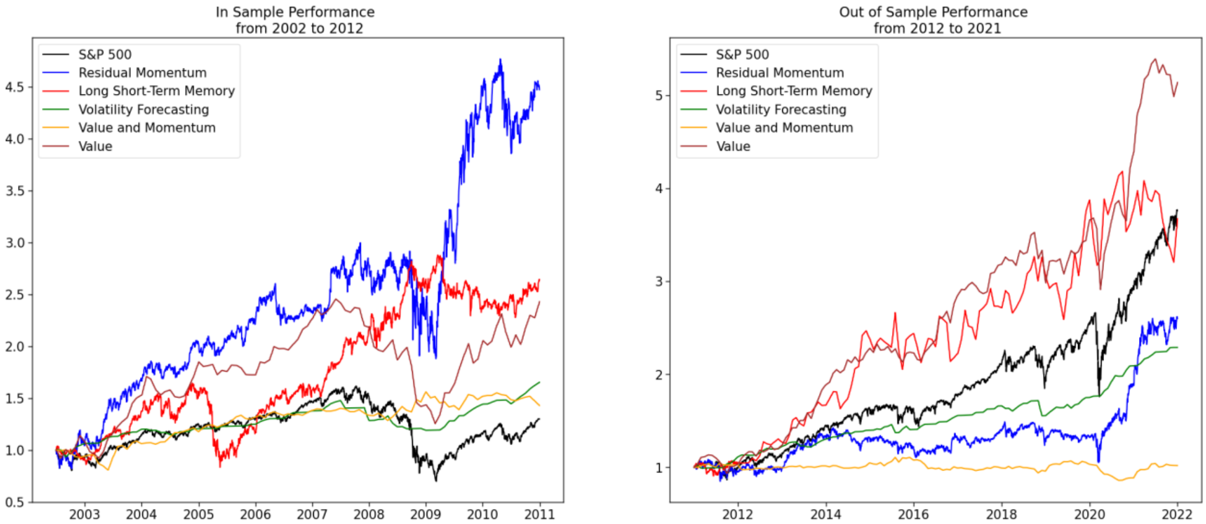


To further explain and understand the above findings and in the hope of getting a better grasp of the unexpected results obtained from the full sample period, it becomes necessary to look at the in-sample and the out-of-sample comparison.

Figure 3: Performance in-sample vs out-of-sample

2002-2021

Figure 4 shows in-sample and out-of-sample graphs of the performance of each strategy if 1 dollar was invested in each strategy at the beginning of each period.



Illustrating the in-and out-of-sample performance in Figure 3, there are notable differences in the performance of the same strategies in the two time periods. Beginning with the in-sample period, notably, all strategies outperformed the S&P 500 overall and especially in the great financial crisis of 2008 (see Figure 3). Overall, the residual momentum performance stands out, achieving the highest cumulative returns, followed by LSTM, with value catching up post the great financial crisis. Volatility forecasting and value and momentum remain rather stable with comparably low volatility, still outperforming the benchmark.

During the great financial crisis, in particular, the residual momentum and value strategy experienced larger drawdowns, however, achieved higher returns prior to the crisis, thereby remaining above the S&P's low. The volatility forecasting strategy remained rather muted during the crisis, while LSTM and value and momentum achieved positive returns during that time.

Conversely, during the out-of-sample period, the strategy comparison forms a different picture. In total cumulative return, the S&P 500 performance is only matched by value strategy, with LSTM matching and losing only in 2021. In contrast to the in-sample-period, residual momentum performs about flat until the stock market crisis due to the global corona pandemic in 2020. There, after a small dip, it rallies strongly with the overall market rebound in the S&P 500. Notably, the volatility forecasting strategy can avoid the 2020 stock market crash entirely, achieving rather consistent returns. It can be assumed that this strategy correctly predicts volatility in this time frame. Since this is the short volatility strategy, it apparently did not take a position during the crash, while benefitting from the high premium in the recovery. The value and momentum strategy flat lines in this period, while, however, being notably neutral to the 2020 crash, as the returns of value and momentum winners and losers seem to offset each other in the period, where especially large-cap securities performed well. This structurally strong performance might also explain the superior performance of the adjusted value strategy, reflecting intangible assets into the book to market equity.

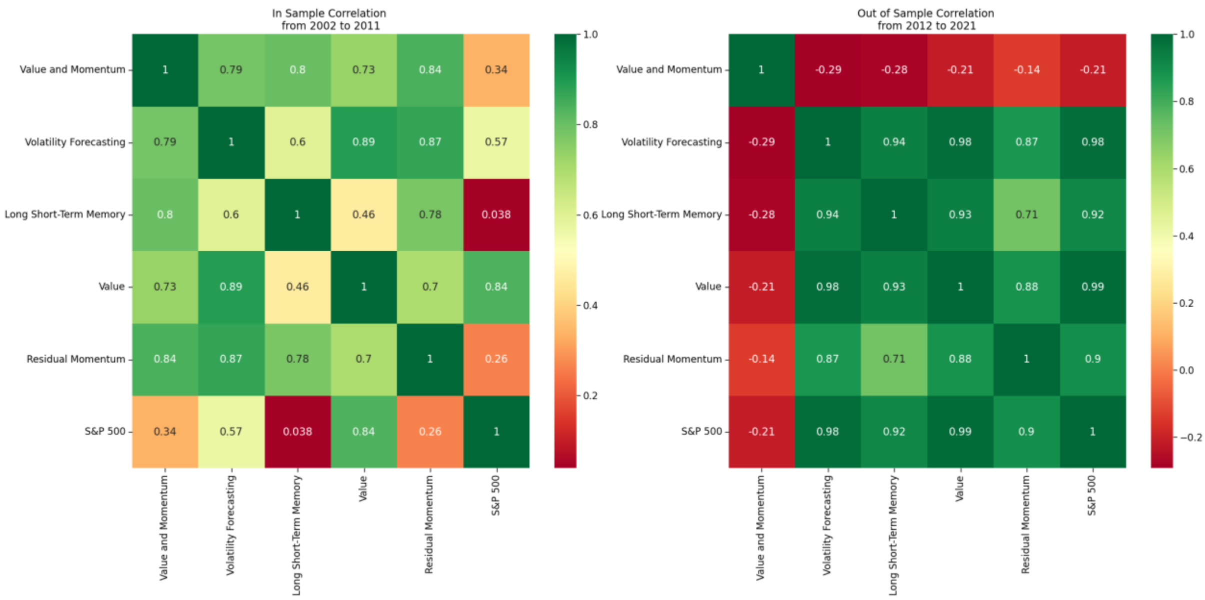
Comparing the two sample periods, the machine based-learning LSTM achieves sizeable, but volatile cumulative returns, mostly independent from the market and avoiding crashes. Residual momentum appears to perform better in sharp market rallies, while more muted in gradual uptrends. Value including intangibles has outperformed the market in both periods, exemplifying returns, while volatility forecasting and value and momentum L/S achieve the least volatile but small nominal returns.

On the left graph of Figure 3, it is now also apparent that the correlation seems to be way lower in the in-sample period, then during the out-of-sample period.

Figure 4: Correlation in-sample vs out-of-sample

2002-2021

Figure 4 shows in-sample and out-of-sample graph. As expected, the correlation with the market and therefore with each other is way higher in the Out-of-sample period.



Comparing the correlations between the strategies in- and out-of-sample period in Figure 4 forms an interesting picture. Evidently during the in-sample period the strategies experience comparably little correlation, while out-of-sample except for the value and momentum L/S the remaining strategies trade more like a basket with correlations above 0.9 for all – as it is the case for the full 20-years illustrated in Figure . This can be explained by noting that during the in-sample period the models develop their signal by fitting the data by trading with the market, while in the out-of-sample the models revert back to trading with the macroeconomic environment. Only the value and momentum strategy stand out as there is no actual training period for the signal, as it is simply rooted in price and fundamental data rather than model fitting. As a result, in the out-of-sample period the combined strategy earns most diversification benefits from the value and momentum L/S thereby lowering volatility, which, however, at the same time likely drags returns due to its poor performance.

In conclusion, we find different return patterns among the sub-strategies across the full 20-years, in particular with their performance during market crashes. When trying to combine the strategies, the goal should now be to profit from the stability of those strategies during market shocks, while also benefitting from the high returns during bull markets. The insights gained here inform the portfolio weighting of the combined strategy in the sub-sequent chapter.

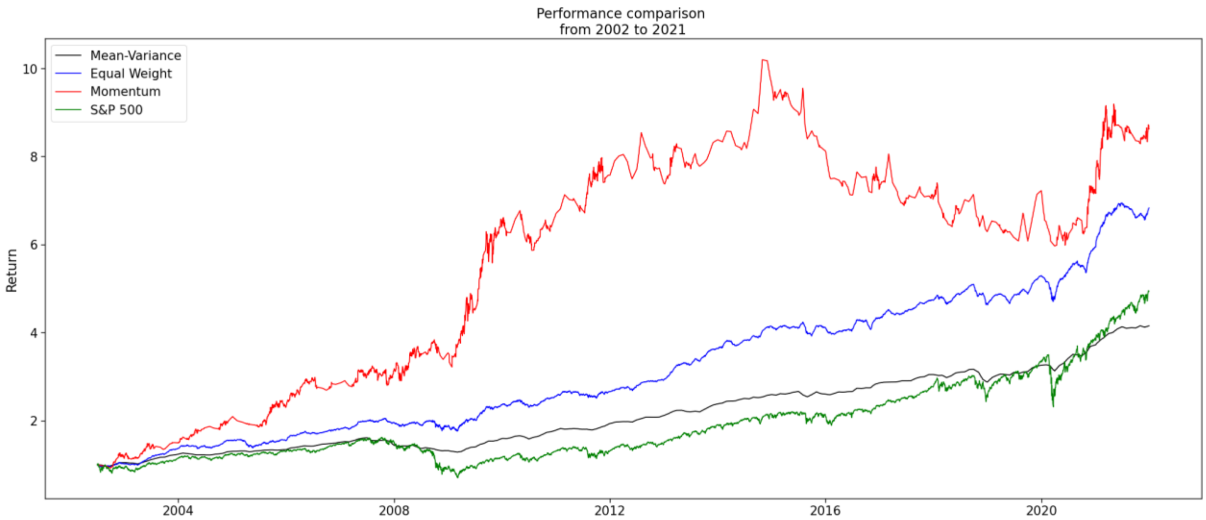
4.2 Combined strategy performance

In the following, three variants of the combined strategy are analyzed: 1) equal weighted returns, 2) mean-variance and 3) momentum. The in-sample period ranges from 2002 until 2012, the out-of-sample period from 2012-2021 and rebalancing happens daily. The value and momentum individual part deviates from this, as the in-sample comprises the full 2002-2021 and the out-of-sample compares to the most relevant prior literature, however, the return series was split here to match the group comparison.

Figure 5: Combined strategies max Sharpe Ratio vs Max Return

2002-2021

Figure 5 plots the performance of the portfolio that yields the best collective Sharpe Ratio, as well as the portfolio that has the maximized return. Furthermore, it shows the portfolio based on the momentum strategy.



The combined strategies are illustrated in Figure 5. The portfolio of the equal-weighted returns is constructed at the beginning of the time frame and is not rebalanced during the period.

The individual strategies chosen are the ones that yield the highest return comparing the strategies within each individual paper. The combined strategy based on equal-weighted returns invests equally distributed among all the individual investment strategies.

Contrary to the investment strategy based on equal-weighted returns, the mean-variance portfolio does not hold an equally distributed amount of all the individual strategies. The aim of this strategy is to maximize the Sharpe ratio, which is a measure of risk-adjusted returns. A large Sharpe ratio means that the portfolio has a relatively high return compared to its risk measured as volatility of the returns. The strategy is constructed based on a Monte Carlo simulation. A Monte Carlo simulation is a model that helps to explain the impact of risk and uncertainty in prediction and forecasting models by looping through iterations of various scenarios underlying assuming a normal distribution. It is used to predict the probability of a variety of outcomes if there is a presence of random variables.

The momentum strategy invests every day in one of the five portfolios meaning that the strategy rebalances daily. The rebalancing is based on the moving average over the last three months, where it is interesting to see the number of days per year the strategy invests in one of the five portfolios. It is attempted to create a combined strategy that uses the individual strengths of each strategy, as explained in the previous chapter. Ideally, this combination invests in the more stable strategies during market crashes, while it features the strongly performing strategies in bull markets in between.

The individual strategies are to be analyzed based on their individual performance over the whole time frame. Furthermore, again, the in-sample, as well as the out-of-sample period is looked at to potentially find additional learnings.

Table 2: Metrics combined strategy

2002-2021

Table 2 shows the main metrics for the combined strategies as well as the S&P 500. For all four, the in-sample, as well as the out-of-sample period is shown.

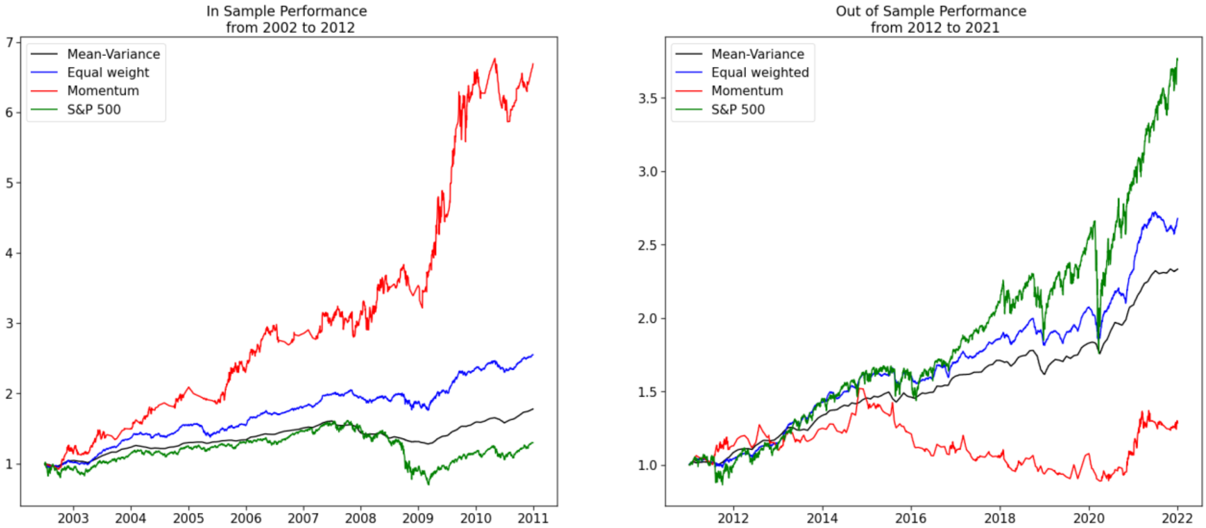
Portfolios	Cumulative Return	Annualized Sharpe Ratio	Annualized Alpha	Beta	Tracking Error	Information Ratio
Mean-Variance	315%	1.45	0.44%	0.03	1.19%	0.37
Mean-Variance IS	78%	1.32	0.42%	0.03	1.36%	0.31
Mean-Variance OOS	133%	1.56	0.46%	0.02	1.00%	0.44
Equal weight	583%	1.09	0.49%	0.22	3.94%	0.12
Equal weight IS	155%	1.05	0.63%	0.24	5.21%	0.12
Equal weight OOS	168%	1.13	0.40%	0.19	2.49%	0.16
Momentum	762%	0.89	0.65%	0.16	13.00%	0.05
Momentum IS	569%	1.15	1.40%	0.21	17.17%	0.08
Momentum OOS	28%	0.31	0.01%	0.09	8.28%	0.01
S&P 500	393%	0.52	0.00%	1.00	0.00%	n/a
S&P 500 IS	30%	0.25	0.00%	1.00	0.00%	n/a
S&P 500 OOS	276%	0.31	0.00%	1.00	0.00%	n/a

In the following, the individual combinations and their performance metrics shown in Table 2 are analyzed in more detail. The portfolio statistics are computed through an ex-post single factor regression. All alphas and betas are statistically significant for all convenient significance levels. At first glance, the differences in performance between the strategies in general, and between their performance throughout the in- and out-of-sample period is evident. The dissimilarities in metrics are confirmed when illustrating their performance in the following figures.

Figure 6: Combined strategies in-sample vs. out-of-sample

2002-2021

Figure 6 plots the in-sample, as well as the out-of-sample performance for the three combined portfolios. Furthermore, the S&P 500 is plotted to see how the benchmark performs in those two time frames.



4.2.1 Equal weight

The equal weighted strategy generates the second-highest total return of all the portfolios, achieving a cumulative return of 583% (see Table 2), with the performance being consistent in both the in-sample and out-of-sample periods. The Sharpe ratio of 1.71 for the full 20-year period is high in general, exceeding the benchmark S&P 500's 0.52 more than three-fold, providing investors with a superior risk-return adjusted profile. In the strategy comparison, this Sharpe ratio is only exceeded by the Max Sharpe strategy. A simple equal-weight allocation usually performs quite well in terms of risk-return trade-off by capturing diversification benefits of all the strategies in all market environments, without a dynamic weighting scheme and its underlying intuition distorting it.

What further stands out is the overall beta of 0.22, suggesting a relatively low sensitivity to market swings and showing the diversification benefits obtained in the combined portfolios. In addition to outperforming the S&P 500 by a wide margin, the equal-weighted portfolio also shows a steady growth over the years, despite suffering a significant downturn during the financial crisis. Nonetheless, the drop was significantly lower when compared to the dip of the S&P 500. Moreover, while the whole stock market experiences an exponential increase in its returns after the initial Covid-19 stock market crash, the equal weight portfolio, despite producing a low beta, also sees a large growth in its returns during this time frame.

In terms of excess returns relative to the market, the alpha generated by the equal weighted portfolios do not seem to suggest a significant difference when compared to the other strategies, with the values being in-line with the ones previously discussed. In the out-of-sample period, the portfolio produces an annualized alpha of 0.40%, while the alpha generated in the in-sample period sits at 0.63%.

As illustrated in Figure 6, there is a notable difference between the in- and out-of-sample performance. During the in-sample period, the equal-weight allocation outperforms both the S&P 500 Index as well as the Max Sharpe strategy, especially by achieving higher returns in the run-up to the great financial crisis 2008 and then by limiting drawdowns therein. The return series over time is an almost diagonal line to the up right with comparably low volatility, which is precisely what hedge funds seek. However, in the out-of-sample the equal-weight strategy cannot quite keep up with the steep rise of the S&P 500 until the coronavirus related market crash in early 2020 and the subsequent sharp rally. Still, the strategy outperforms the Max Sharpe and momentum counterparts, again with comparably low volatility returns.

4.2.2 Momentum optimization

As explained previously, the momentum strategy consists of a daily one-asset portfolio that shifts between the portfolios based on the past 30-day performance. Out of the three potential combinations, in absolute terms, this is the best-performing one.

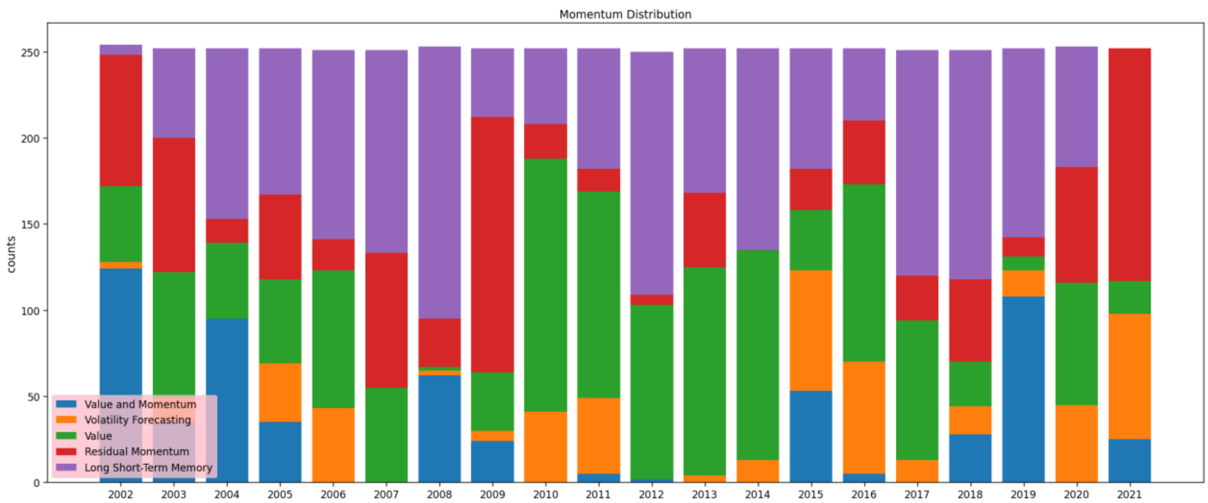
As observable in Table 2, the cumulative return of 762% of the momentum strategy outshines the other strategies. However, after looking at the in-sample and out-of-sample metrics, Figure 6 confirms the huge difference in the two windows. While momentum works extremely well during the in-sample period, it completely falls off in the out-of-sample period. Research has widely shown, that momentum as a factor in general declines strongly in performance as an investment factor after 2008. Baltas and Koswki (2019) argue that momentum strategies perform worse after 2008 due to increased correlation between markets and asset classes. Given the individual strategies focus on different asset classes the increased correlation between asset classes, this could be viable reason for the underperformance of the momentum strategy. While the in-sample period features a decent alpha of over 1.40%, no outperformance in the out-of-sample period against the SPX is recorded.

Obviously, certain strategies dominate the performance of the momentum strategy more strongly than others. Not only can this information help to understand the performance of the momentum combination, but can also show the respective strenghts of the individual strategies in certain market conditions.

Figure 7: Yearly distribution of portfolios in Momentum Strategy

2002-2021

Figure 7 shows how many days per year each portfolio is invested in. The Y-axis is the number of days, while the X-axis corresponds to the respective year.



As observable in Figure 7, the Long Short-Term Memory strategy seems to have a strong presence in most years. As the strategy is predicting stock returns, it is already expected to be outperforming other strategies in difficult market environments. The momentum strategy shows that especially during the two market shocks in the 20-year time frame from 2002 to 2021, this signal yields a strong performance.

The counterpart to this one would be the value strategy. Here, the macro environment has a big influence on performance. Accordingly, it is observable what has already been presupposed: During market crashes like the one 2008 or even 2020, fundamental strategies show an inferior performance. While it does not seem like it affects 2020 as much, this is mostly due to the fact that the crash here came in the early window of the year. The value strategy again is well featured during the recovery, explaining its presence in that year.

As for the value and momentum strategy, the overall performance is just not strong enough to have a heavy influence over the whole time frame. Because of the stable nature, this strategy is clearly also featured in calm markets, or whenever other strategies seem to to have a poor performance due to any macro economical reasons.

While it could be assumed that the volatility forecasting strategy would be heavily featured during market downturns, this is not the case. As discussed in Chapter 3, it doesn't perform bad, however, the overall performance seems to not be good enough to heavily influence the momentum strategy performance. Since the volatility strategy that is analyzed is a short-volatility strategy that sells covered straddles, extreme returns are unlikely, which caps the strategies' performance potential. Accordingly, whenever the market is performing well, almost no weight is in this signal.

Finally, a very clear pattern is observable for the residual momentum strategy. During the recovery phase after the two major events in the 20-year time frame, this strategy is heavily featured. This confirms the assumptions already made in previous chapters about this signal. No other strategy seems to capture the market recovery as well.

4.2.3 Mean-Variance optimization

Mean-Variance is often used in research papers as the standard model of portfolio construction even though investors rarely use it (Fischer and Statman 1997). The mean-variance optimization framework has been first introduced by Harry M. Markowitz in 1959 (Markowitz 1999). According to Markowitz (1999) an average investor primarily seeks to maximize returns. In his framework, Markowitz implicitly intended to maximize the previously not existing Sharpe ratio of his portfolio: The Sharpe ratio has been first implemented by Sharpe and Lintner in 1963 and is the expected portfolio returns divided by expected portfolio volatility. Thus, by iterating through multiple portfolio combinations a so-called efficient frontier is generated. The efficient frontier is a continuation of the mean-variance framework and reflects an optimal trade-off between return and volatility. Portfolios below the efficient frontier are not mean-variance efficient. Hence, the portfolio with maximized Sharpe ratio is located on the efficient frontier (Markowitz 1999).

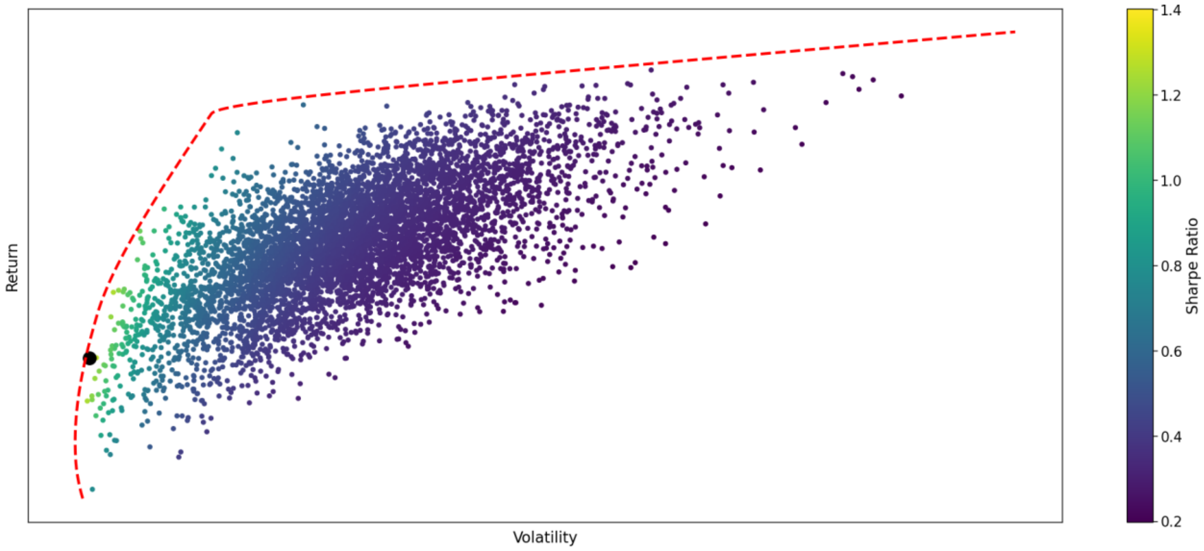
As aforementioned, Fischer and Statmann (1997) state that mean-variance is rather used on research-related topics than for investors' decisions on portfolio construction. Fischer and Statmann (1997) argue the reason for that being that mean-variance constraints overweight the optimization benefits. However, for the purpose of this analysis, it can make sense to create a mean-variance efficient portfolio. Additional learnings can be achieved by doing so, while a potential superior weighting scheme might be discovered.

The optimized portfolio weights are obtained by running a Monte Carlo simulation and iterating through different portfolio weightings to create 6000 different portfolios. The 6000 portfolios are visualized in Figure 7 and show the different Sharpe ratios for each portfolio, the efficient frontier, and the optimized portfolio as a black dot with the highest Sharpe ratio.

Figure 8: Efficient frontier replication based on Mean-Variance

2002-2021

Figure 8 plots the efficient frontier that is replicated by optimizing the Mean-Variance framework. The Monte-Carlo simulation iterates through different portfolio weights to seek the most efficient mean-variance portfolio which will result in having the highest Sharpe ratio. In this case the black dot reflects the portfolio with the highest Sharpe ratio.



The mean-variance efficient portfolio has the lowest cumulative return out of all combined strategies tested over the whole analyzed period of 20 years. The cumulative return achieved by this portfolio for the whole time frame is 315%, roughly 70% lower than the S&P 500 Index return. Comparing the in-sample and the out-of-sample periods, the mean-variance efficient portfolio performs better in the second half of the time frame, generating a cumulative return of 78% in the first 10 years and 133% in the later part of the analysis. This could be related to the low volatility and high growth environment after the financial crisis of 2008. Looking at the other portfolio statistics, the mean-variance portfolio provides the best results compared to the other strategies, obtaining a tracking error of 1.5% for all analyzed time periods, suggesting that nearly no additional risk has been taken to achieve the respective alpha, and it generates an information ratio for the full sample period of 0.37.

The weights in the mean-variance efficient portfolio in this analysis are fairly skewed towards strategies having higher Sharpe ratios such as volatility forecasting, value and momentum and value. This gives further implication on why returns achieved are lower compared to the momentum and equal-weighted portfolio. Especially the strategies volatility forecasting and value and momentum, which make up 80% of the portfolio have on average lower returns compared to the residual momentum and long short-term memory strategies.

In line to Fischer and Statmann (1997)'s comments, the mean-variance efficient portfolio performs worse against the other strategies in the long-run. As Fischer and Statmann (1997) state the optimization problem is a big limitation of a mean-variance portfolio. The primary goal of a mean-variance investor is to maximize the Sharpe ratio. This might limit downside risk as during the financial crisis in 2008 or during the pandemic in 2020, but at the same time add a low ceiling for potential returns.

5. Results

In conclusion, all combined strategies outperform the S&P 500 in terms of Sharpe ratio and thereby award investors with a superior risk-return profile. The superior risk-return arises through increased diversification that consequently led to lower volatility and by limiting larger drawdowns in market turmoil due to the diverse economic signals underlying the individual strategies. Also, the strategies achieve overall very low betas, ranging from 0.03 to 0.20, allowing investors to be largely market neutral and in any case capture some of the general uptrend the market experiences over the long term. An investor wanting to invest in this multi-strategy hedge fund can achieve its objective by applying different weighting schemes to achieve different risk/reward levels. However, limitations for practical implementation such as transaction costs would have to be considered. In particular, both the equal-weight and the mean-variance combined strategies exceed the risk-return trade-offs of all the individual sub-strategies, as expected from diversification, while only the more cyclical combined return momentum strategy slightly undershoots some of the individual Sharps for the full 20-years. The former two strategies achieve consistent returns with comparably low volatility, which is precisely what hedge funds seek. The Sharpe ratio of the equal weight and the mean-variance combination amounts to 1.09 and 1.45 respectively for full time frame, while the Sharpe of the return momentum combination amounts to lower 0.89.

Comparing the individual strategies in- and out-of-sample, the differences in performance between the two periods stands out especially compared to the S&P 500 benchmark. Interesting characteristics of the sub-strategies are derived as, for example, the LSTM strategy appears most independent of market swings, while residual momentum seems to capture the highest returns in market rallies. Furthermore, value including intangibles outperforms the market, experiencing similar swings, while volatility forecasting and value and momentum are the least volatile, avoiding larger drawdowns even in market crashes. In terms of excess returns relative to the market, all five strategies produce a positive alpha for the full sample, in-sample and out-of-sample periods, meaning that they all manage to generate excess returns relative to the market. Once again, the volatility forecasting strategy stands out from the rest with an annualized alpha of 1.54% for the entire time frame and, on the opposite side, the value and momentum strategy is the worst performer when it comes to this criterion, generating an annualized alpha of 0.12%. Regarding systematic risk, the value and momentum strategy is the only one in the analysis that generates a negative beta (-0.01 for the in-sample and full sample periods) suggesting that the strategy has an inverse relation with the market, meaning that it tends to increase in profitability when the overall market falls and vice versa.

For the combined strategies, the equal weighted strategy generates the second-highest total return of all the portfolios, achieving a cumulative return of 583%, with the performance being consistent in both the in-sample and out-of-sample periods. The Sharpe ratio of 1.09 for the full 20-year period is high in general, exceeding the benchmark S&P 500's 0.52 more than double, providing investors with a superior risk-return adjusted profile. In the strategy comparison, this Sharpe ratio is only exceeded by the mean-variance strategy.

The mean-variance framework is expected to return a maximized Sharpe ratio for a given risk/return profile. As aforementioned the mean-variance portfolio generated the highest Sharpe ratio out of all combined strategies. Opposed to all other strategies returns of the mean-variance efficient portfolio were not affected by the financial crisis in 2008 and the pandemic in 2020. Its resilience against market turmoil is determined by its low annualized volatility. The mean-variance portfolio holds 67% in the volatility forecasting strategy. Since the volatility forecasting strategy essentially collects a steady premium by selling straddles it is less dependent on the market environment. Volatility can actually favor the strategy as it might increase premiums and thus increase returns.

Momentum achieves the highest cumulative return, which is however highly dependent on the market environment, visible in the in- and out-of-sample analysis. While one key message is the clear emergence of patterns in the respective performance of the different underlying strategies, this is probably the hardest strategy to implement. The daily rebalancing of strategies would be a tough ask, mainly because of the volatility forecasting strategy that takes monthly option positions. Additionally, this way of combining the strategy seems to not be an attractive way of combining the portfolios going forward. The disappointing 28% cumulative return in the out-of-sample period is underperforming the benchmark, as well as all other approaches. Even if the past does not equal the future, it would be a hard sell to apply this way of combining the strategies going forward.

Conclusively it can be summarized that while both the mean-variance and the momentum combination of the underlying strategies yielded attractive results and interesting learnings, the equal weight performance seems to be the most attractive. Obviously, it has a lower Sharpe ratio than the mean-variance, but as Fischer and Statmann (1997) already found, the constraints in this way of combining the portfolios overweight the optimization benefits. Looking at our results, the most consistent way of combining the strategies is the equal weighted strategy. This strategy achieves to consistently outperform the S&P 500 over the in-sample and out-of-sample period, while having a low Beta and a Sharpe ratio of over 1. It shows great immunity to macro-economic conditions, and still aims to maximize returns in each period.

6. Conclusion

Overall, the findings provide a diverse comparison of strategies within the same investment universe and their performance in different market environments. The momentum strategy and the equal weighted portfolio were able to outperform the S&P 500 Index in terms of overall returns during the full sample period. Contrarily, the mean-variance efficient portfolio performed worse compared to the S&P 500 Index. However, the portfolio appears withstand large macroeconomic shocks such as the financial crisis in 2008 and the pandemic in 2020 and thus proved itself to be an excellent option for more risk adverse investors. The latter shows especially the highest risk adjusted return performance, with the mean-variance efficient portfolio obtaining the highest Sharpe ratio out of all the combined strategies.

The comparison of the in-sample period and the out-of-sample period shows further that none of the strategies is able to outperform the S&P 500 Index in the out-of-sample period. This considering that both the momentum portfolio and the equal weighted portfolio beat the S&P 500 with regard to return performance. The momentum strategy not performing well after the financial crisis in 2008 is, however, in line with previous research on momentum strategies. One of reason for the underperformance is increased correlation across markets and asset classes according to Baltas and Koswki (2019). Especially the latter appears to be a plausible explanation for the underperformance in the out-of-sample period given the individual strategies in this paper focus on different asset classes

This group analysis touched upon multiple known and well-studied portfolio frameworks. However, each framework has their limitations which opens the door to future research projects. In general, the individual strategies do not invest in a common asset class. Therefore, it could be interesting to align the asset class on the individual strategies to test the performance of the combined strategies for a common asset. At the other end of the spectrum it would also be interesting to test the strategies on five totally different asset classes to maximize diversification.

Moreover, to compute the combined strategies, the individual strategy that performed the best was chosen. However, extending the analysis by comparing only long-short strategies could especially be interesting for the performance on the mean-variance portfolio. As no transaction costs are assumed daily rebalancing is not a limiting factor. To improve the robustness of the result it would nevertheless make sense to incorporate a holding period of 1 to 6-month or even incorporate a comparison of multiple holding periods.

Furthermore, the estimation window for the combined momentum strategy implemented is fixed to 3 months as Jegadeesh and Titman (1993) argue that the ideal estimation window should be set to 3 to 12 months. Similarly, to the holding period, this analysis could be extended by comparing different estimation windows.

In general, this analysis provides on the one hand a brief overview of all individual strategies and on the other a combined strategy performance analysis. The individual strategies differ in the way they are constructed and in the way they perform. Nonetheless, the strategies appear to be highly correlated with each other. Especially in the out-of-sample period correlations are mostly above 0.9 and therefore in line with Baltas and Koswki (2019) statement of asset classes being larger correlated since 2008. At the same time the high correlation simplifies the implementation of a profitable equal weighted combined strategy without having to sacrifice potential upside returns.

Conclusively, this project helps to concisely demonstrate the hurdles that have to be overcome in the combination of different investment strategies. It demonstrates how hedge funds and asset managers these days must find a balance between risk and return, as the two worlds of modern and quantitative approaches collide with traditional, fundamental investing.

References

Asness, Clifford S. 1997. “The Interaction of Value and Momentum Strategies.” *Financial Analysts Journal,* 53(2): 29–36.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen. 2013. “Value and Momentum Everywhere.” *The Journal of Finance,* 68(3): 929–85.

Baltas, Nick, and Robert Kosowski. 2019. Demystifying Time-Series Momentum Strategies: Volatility Estimators, Trading Rules and Pairwise Correlations. London: SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2140091.

Barroso, Pedro, and Pedro Santa-Clara. 2015. “Momentum Has Its Moments.” *Journal of Financial Economics,* 116(1): 111–20.

Blitz, David, Joop Huij, and Martins Martens. 2011. "Residual momentum." *Journal of Empirical Finance,* 18(3): 506-521.

Dingli, Alexiei., and Karl S. Fournier. 2017. “Financial Time Series Forecasting-a machine learning approach.” *Machine Learning and Applications: An International Journal,* 4(1): 11-26.

Engle, Robert F. 1982. “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation.” *Econometrica,* 50(4): 987-1007.

Erdinç, Altay., and Hakan Satman. 2005. *Stock Market Forecasting: Artificial Neural Network and Linear Regression Comparison in an Emerging Market*. London: SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=893741.

Fama, Eugene F., and Kenneth R. French. 1992. “The Cross-Section of Expected Stock Returns.” *The Journal of Finance,* 47(2): 427–65.

Fischer, Kenneth L., and Meir Statman. 1997. “The Mean–Variance-Optimization Puzzle: Security Portfolios and Food Portfolios.” *Financial Analysts Journal,* 24(1): 41-50.

Grundy, Bruce D., and J. Spencer Martin. 2001. “Understanding the Nature of the Risks and the Source of the Rewards to Momentum Investing.” *The Review of Financial Studies,* 14(1): 29-78.

Jegadeesh, Narasimhan, and Sheridan Titman. 1993. “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency.” *The Journal of Finance,* 48(1): 65–91.

Markowitz, Harry M. 1999. “The Early History of Portfolio Theory: 1600–1960.” *Financial Analysts Journal,* 55(4): 5-11.

Prado, Marco L. 2018. “The 10 reasons most machine learning funds fail.” *The Journal of Portfolio Management Special Issue Dedicated to Stephen A. Ross,* 44(6): 120-133.

Sinclair, Euan. 2008. *Volatility Trading*. Hoboken: John Wiley & Sons.

WRDS. 2016. *Wharton University of Pennsilvania.* Accessed November 02, 2022. https://wrds-www.wharton.upenn.edu/pages/get-data/beta-suite-wrds/.

A Work Project, presented as part of the requirements for the Award of a Master's degree in Finance from the Nova School of Business and Economics.

Volatility Forecasting with GARCH models and Recurrent Neural Networks

Analysis of Quantitative Investment Strategies

Enrique Ferrari

48296

Work project carried out under the supervision of:

Prof. Nicholas H. Hirschey

16/12/2022

Abstract

The three main ways to estimate future volatilities include the implied volatility of option prices, time-series volatility models, and neural network models. This project investigates whether there are economically meaningful differences between those approaches. Seminal time-series models like the GARCH, as well as recurrent neural network models like the LSTM are investigated to forecast volatilities. An eventual informational advantage over the market's expectation of future volatility in the form of implied volatility is sought after. Through trading strategies involving options, as well as investment vehicles that emulate the VIX, it is attempted to trade volatility in a profitable way.

Keywords: volatility forecasting, implied volatility, realized volatility, volatility risk premium, Machine Learning, Neural Networks, GARCH, LSTM, GRU, VIX

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1. Introduction

Starting with the autoregressive conditional heteroskedasticity (ARCH) model, Engle (1982) laid the cornerstone for research to dynamically capture return variance. Bollerslev's (1986) generalized autoregressive conditional heteroskedasticity (GARCH) model added to the foundation of stochastic volatility models that are now the benchmark to build upon for volatility forecasting. Even though these models are still widely used, artificial neural network models start to rise in popularity and the questions as to what model best to use is still unsettled.

Three main approaches conclude the capturing of future volatility: Implied volatility (IV) from option prices, time-series volatility models, and neural network models. However, studies suggest that when assuming efficient option markets and accurate derivatives pricing, IV should contain all information that lies within other variables to explain future volatility (Jiang and Tian 2005). Accordingly, Harvey and Whaley (1992) show findings in which IV has a definite predictability of realized volatility (RV) in the future. To challenge this, the question that must be raised here is whether there are better, more accurate approaches to predict the RV rather than the IV which is the market consensus.

This project attempts to give insight into various seminal as well as new approaches to forecast volatility and build a successful trading strategy through that.

2. Theory

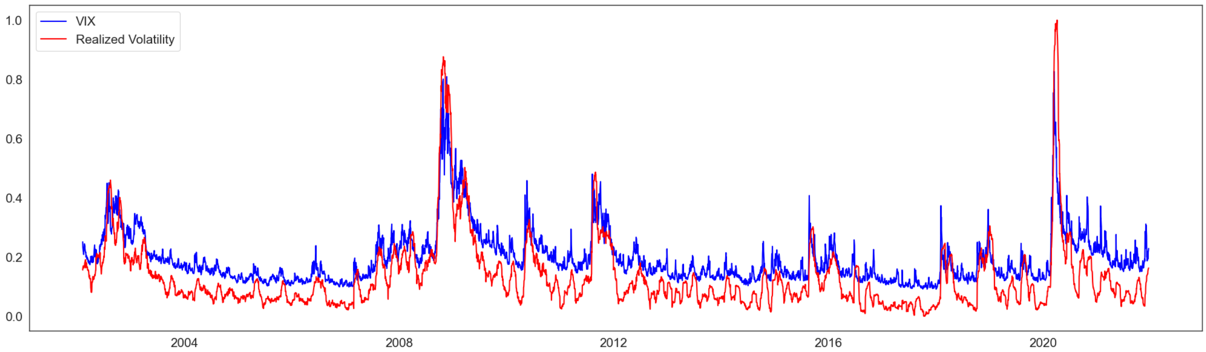
2.1 Realized Volatility vs Implied Volatility

The underlying consensus for derivatives pricing in which the markets think is the Black-Scholes model. Sinclair (2008) explains how this model factors in the multitude of likely - or unlikely - developments in the underlying, proxied by the volatility of price movement. Accordingly, the market holds a certain expectation for underlying's development. If there is a way to estimate future volatility more accurately, an alpha can be generated (Sinclair 2008).

However, it is observed that when trying to forecast volatility, the IV is materially above the modeled forecast (Guo 2000). This is due to the volatility risk premium that the market factors in. A trader that sells IV and is therefore on the short end of an option contract, is selling insurance. A multitude of potential outcomes can occur, that are not reflected in the past, which are to be accounted for. Furthermore, Sinclair (2008) elaborates that since market makers earn money by accumulating the bid/ask spread in the pricing of options, they purposely elevate the skew slightly in their favor. This naturally also shows on the short end when the market wants to sell insurance in heightened volatility levels.

Figure 1: Volatility Risk Premium

Figure 1 shows a proxy of the S&P 500 volatility risk premium by plotting VIX/100 against S&P 500 realized volatility between 2002 and 2021.



In Figure 1 it can be observed that markets tend to overestimate the volatility, here displayed as the Volatility Index (VIX), when markets are in a normal state and underestimate the volatility when the market is crashing. While Bakshi and Kapadia (2003) find that there is a conclusive existence of the volatility risk premium on a single stock level, Jackwerth and Rubinstein (1996) demonstrate that except for the 1987 crash period, S&P 500 at-the-money index options have an observable premium. Therefore, a model that purely forecasts volatility, based on historic volatilities, will be short volatility when futures are in contango and long when they are in backwardation. While research disagrees if and to what magnitude the volatility risk premium should be factored in when forecasting volatilities, there is further elaboration in Chapter 3 on how this is dealt with for the purpose of this project.

2.2 Volatility forecasting models

The market's volatility forecast, the IV, also known as the "investor fear gauge" is a widely researched field that marks the consensus of an investors' expectations of the future (Shaikh and Padhi 2015). However, alongside the inference of volatility levels from derivatives prices, there are also two other seminal modi operandi to forecast volatility: Time-series volatility models and neural network models.

While Harvey and Whaley (1992) demonstrate that IV holds predictive power of future RV, it is still debatable as to how objective of a forecast it really is (Jiang and Tian 2005) .

When looking at time-series models in discrete time settings, the GARCH is established as the standard model. As Rosenberg and Engle (2002) compare multiple variations of the GARCH model to forecast daily and monthly volatilities on index level, Engle and Sokalska (2012) use it for intraday forecasting of equity volatility levels.

Research evolving around neural network approaches to forecast volatilities is still in its infancy, as a multitude of possible methodologies are at hand. Sermpinis at al. (2013) for example exhibit the capabilities of higher-order neural networks in forecasting of future realized volatility on the FTSE 100 futures index.

Concurrently, this is enabling a great breeding ground to build upon and discover new findings. Whilst this project might not go as much in depth as its predecessors, it tries to capture a multitude of approaches to gain a holistic overview in the field of volatility forecasting.

3. Volatility forecasting

3.1 Dataset

In what follows, the dataset used to estimate the volatilities and conduct the back-test is being elaborated. Daily S&P 500 closing prices from 01.01.2002 until 31.12.2021 are directly retrieved from Yahoo Finance. Furthermore, closing prices of the VIX, the VIXY, and the SVXY from their respective inception date up until 31.12.2021 are also retrieved from Yahoo Finance. For the part of the backtest involving short strangles on the S&P 500, the CMBO index from the Chicago Board Exchange is used, which can be directly retrieved from there. For the comparison of the various forecasting methods, the whole timeframe is split into a 19-year training period and a test period consisting of the last 252 trading days. For the back-test the final models selected are trained again from scratch on the in-sample period of 2002 to 2011. They are then applied to the out-of-sample period ranging from 2012 to 2021.

3.2 Forecasting Models

In the following all models that are compared are presented, and the final selection of models that build the basis of the trading signal is conducted. The models are divided into three categories: Baseline models, GARCH models and neural network models.

A time shifting interval is used to determine how well the models perform. To predict the volatility of the future, the time-series is shifted back by one prediction interval to then compare the predictions with the actual volatilities. Therefore, the forecasted volatility is always compared with the actual volatility. When applying the models to develop the trading signal, the one month implied volatilities are looked at. Naturally, the one month (21 trading days) volatilities are forecasted.

To determine how good each method performs the two metrics of Root Mean Squared Percentage Error (RMSPE) and Root Mean Squared Error (RMSE) are used. While the RMSPE is prioritized due to the relative nature of the metric, those two are chosen because they punish large errors more than other metrics.

Table 1: Volatility Forecasting Models

Table 1 shows the Root Mean Squared Percentage Error and the Root Mean Squared Error of all inspected models on the test period. The Training Period consists of the 20 year timeframe except of the last 252 trading days that make up the test period. For all neural networks the standard model setup is: Loss = MSE, Optimizer = Adam, Metric = RMSPE.

Model	**RMSPE**	**RMSE**
1. Mean Baseline	1.073706	0.057519
2. Moving Average (14 days)	0.317809	0.021218
3. Random Walk	0.381630	0.027433
4. GARCH(1,1), Constant Mean, Normal Distribution	0.335344	0.033577
5. GJR-GARCH(1,1,1), Constant Mean, Skewed Distribution	0.442098	0.042082
6. TARCH(1,1,1), Constant Mean, Skewed Distribution	0.316358	0.032122
7. TARCH(1, 2, 0), Constant Mean, Skewed Distribution	**0.261749**	**0.023070**
8. Simple Linear Regression (n_past = 21)	0.387887	0.027928
9. LSTM 1 layer (layers= 1, units = 20, n_past = 21)	0.334205	0.025391

10. Bidirect LSTM (layers= 2, units = 32/16, n_past = 21)	0.515580	0.058730
11. Multivariate Bidirect LSTM (layers= 2, units = 32/16, n_past = 21)	**0.235793**	**0.019670**
12. Multivariate Bidirect LSTM (layers= 3, units = 64/32/16, n_past = 21)	0.269351	0.020519
13. Multivariate Bidirect LSTM (layers= 4, units = 128/64/32/16, n_past = 21)	0.274504	0.021291
14. GRU (layers= 1, units = 20, n_past = 21)	0.361871	0.028273
15. Multivariate Bidirect GRU (layers= 2, units = 32/16, n_past = 21)	0.288850	0.022405
16. Multivariate Bidirect GRU (layers= 3, units = 64/32/16, n_past = 21)	0.650380	0.040329

3.2.1 Baseline Models

Three primitive models are examined to forecast volatilities. The *mean baseline* model makes use of the assumption that volatility is mean reverting. It assumes that at every point in time the average volatility of the whole timeframe corresponds to the best estimate in future volatility. Meanwhile, the *moving average* uses the last two weeks' average volatility level as the forecasted volatility level. Finally, the *random walk* forecast just assumes that todays' volatility is the same as next periods volatility.

3.2.2 GARCH Models

Regarding time-series models, three different variations are tested that all root in Engle's (1982) ARCH model. While already Engle, Kane, and Noh (1994) show that the economic value of GARCH models is higher compared to implied volatility regression model, there might be enhanced variations. These models make use of the fact that volatilities come in clusters, resulting in the short-term volatility being close to the current levels of volatility, while the long-term volatility is reverting to its mean. These models are trained with an expanding window forecasting, where with every forecasted datapoint the window to train the model grows. The most widely used model, the *GARCH (1,1)* is a time-series model that can be defined as:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \tag{1}$$

with ε_{t-1}^2 being the model's residuals, σ_t^2 being the variance at time t. The factors ω, α and β then sum up to 1 with the former being the long-term variance.

Furthermore, an asymmetric GARCH model, known as the *GJR GARCH* (1,1,1) first introduced by Glosten, Jagannathan, and Runkle (1993) is used. Glosten, Jagannathan, and Runkle (1993) explain how the GJR GARCH allows for taking volatility clustering with seasonal patterns into account, where positive or negative innovations can have different impacts on the conditional volatility. While the basic GARCH model assumes equal impact for positive and negative shocks, the GJR GARCH penalizes negative shocks more by adding one parameter.

Finally, the threshold ARCH (*TARCH*) model is tested. First published by Rabemananjara and Zakoïan (1993), the authors explain how TARCH models use absolute, instead of squared values as inputs, while still maintaining the capability to capture asymmetric shocks with a dummy variable. This allows the model to forecast the standard deviation and still consider asymmetric behavior. Since this model yields the best performance, the TARCH is tuned for the best parameters of α, β and γ, resulting in the ultimate time-series model *TARCH(1,2,0)*

3.2.2 Neural Network Models

In contrast of the GARCH models, forecasting volatilities with neural networks (NN) has no state-of-the art approaches that must be tested. Different methods are examined and tuned, whereas here a sliding window forecasting approach is put in place. Unlike before, every newly forecasted volatility estimate makes the train-window shift, but it always has the same size.

The first model tested is a fully connected network that basically resembles a *simple linear regression*. When trying to forecast volatilities using NN, there is a wide range of possibilities. While Chen, Härdle and Jeong (2010), for example, suggest using support vector machine, an Artificial Neural Network, others point towards Recurrent Neural Networks. This class of NN seem to be well suited for the forecasting of time-series'. Liu (2019) suggests using Long Short-Term Memory (LSTM), which was first introduced by Hochreiter and Schmidhuber (1997). The model is described as receiving a new input at every time-step together with its own previous output resulting in this so-called memory. Due to the reception of new information the network constantly gets rid of some data and although it is capable of assigning data-points to long-term importance, information very far back could no longer be represented. This *basic LSTM* is then modified into a *Bidirect LSTM* approach that allows to train two LSTM's simultaneously with one of the models having a reversed time-series. While both approaches are not outperforming the best GARCH model, the introduction of *multivariate models* finally yields promising results. Here, alongside the volatility, the intra-day high-low spread, the open-close spread and the volume are considered. By adding further bidirectional layers - including dropout layers to prevent overfitting - the optimal 2-layered LSTM model is determined.

Finally, first introduced by Cho et al. (2014), the Gated Recurrent Unit (GRU), another Recurrent NN is tested to compare the results with the LSTM. GRU works like the LSTM, but being more efficient and less memory heavy. It should perform better on large datasets. Again, a simple *one-layered GRU* and *multivariate bidirectional GRUs* are tested, but none of them is able to outperform the best LSTM.

Conclusively, due to the performance reported in Table 1, the 2-layered multivariate bidirect LSTM and the TARCH(1,2,0) model are used for the signal as well as the trading strategy.

3.3 Trading Signal

By forecasting the volatility based on realized volatilities, both the GARCH and the NN methods essentially lead to the forecasted one month RV. For one, it is interesting to see whether the NN approach can outperform the time-series approach. But since the models are retrained and tested on different periods, it is not given that the LSTM model performs best again.

Sinclair (2008) argues that this predicted volatility should technically be compared with the implied volatility swap, a weighted average of a series of option prices. In practice, this is not really feasible, since there is only an insubstantial amount of strike prices with a wide bid-ask spread. Therefore, the forecasted volatilities are compared with the level of IV. As Sinclair (2008) explains that there are technically no listed at-the-money options, similar to Liu et al.

(2021), the CBOE Volatility Index (VIX) is assumed to be the optimal substitute for the implied volatility level. This paper compares the forecasted one month RV with the level of the VIX/100 on a daily basis. If the forecasted volatility is above the level of the VIX, the market is underestimating the one month volatility and therefore a long volatility position is taken. If, however, the level of the VIX is below the forecasted volatility level, the market appears to be overestimating the one month volatility and a short volatility position is taken. Since, as demonstrated in Chapter 2, the volatility risk premium would lead to short volatility positions in calm markets and long volatility positions during shocks, this project attempts to factor out the premium for the signal development. Because volatility trades in clusters and therefore also market conditions have a cluster-like behavior, the volatility risk premium behaves accordingly. Accordingly, on any given day of comparing the VIX with the forecasted volatility level, the volatility risk premium of the previous day is deducted. This leads to a fair comparison of the forecasted volatility level and the VIX. A basic assumption is made to account for the premium that does not accurately capture the behavior of the term structure or the development. This would go beyond the scope of this paper, but it can certainly be challenged in future examinations.

3.4 Backtest

This chapter gives an overview of how the signal is turned into a trading strategy. Literature suggests various ways of trading volatility. Sinclair (2020) explains how numerous option strategies allow to take positions in volatility. Furthermore, also the VIX can be traded by using options, futures or exchange traded funds (ETFs).

The VIX essentially replicates the one month implied volatility of S&P 500 options. Since there are practically no at-the-money options traded in the market, the VIX is constructed by weighted averaging of the options' volatility that are closest to the money.

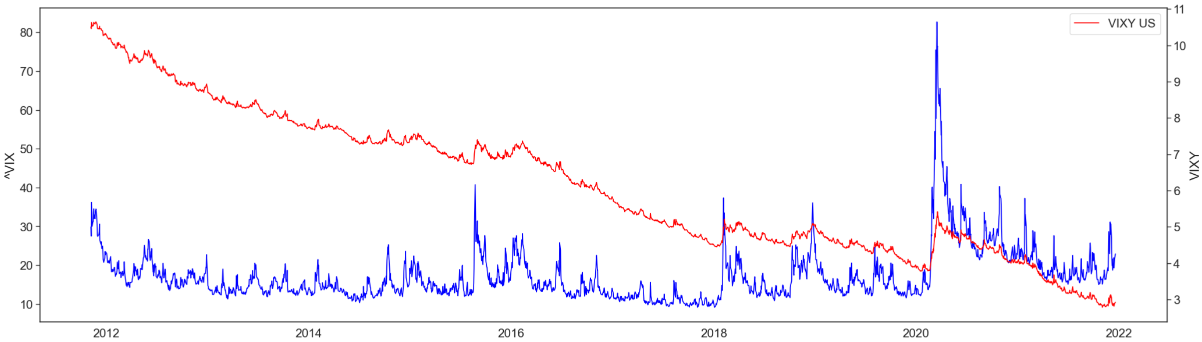
In this project the focus lies on two ways to trade volatility. One of them is the daily trading of the VIX using futures. The other method exploits the widely used strategy of using straddles to trade positional volatility. Long straddles are an option strategy that involves the buying of a put and a call option with the same strike price. For data and simplicity reasons both methods will make use of a proxy ETF and an index, respectively, that closely replicate the two ways of trading volatility.

3.4.1 VIXY and SVXY

When trading the VIX, and therefore taking daily positions, one can make use of ETFs. These ETFs try to replicate the VIX, by recreating its value through the balancing of futures positions. They essentially replicate what a trader would do when trying to emulate the VIX through futures. Unfortunately, due to markets usually being in contango (the longer dated future is higher), these products lose money simply by rolling into the next contract.

Figure 2: VIX vs. VIX ETFs

Figure 2 plots the VIX against the VIXY in log scale. Even though the VIX ETF should emulate the VIX, a big difference in price development can be observed from 2012 until the end of 2021.



However, while in Figure 2 it may look attractive to do so, it is not recommendable to simply short-sell these ETFs. Spikes during shocks in the market can result in high borrow rates, increases in margin requirements and per usual a recalling of the shares in such times.

Consequently, to go long volatility when the developed signal indicates, the VIXY index is bought. To take a short volatility position, a long position in the SVXY, a short VIX ETF, is taken. This ETF is essentially the counterpart to the VIXY that allows to take a short position in the VIX. There are several other VIX ETFs available in the market, but those two are chosen due to the extensive track record of over 10 years, which allows for a solid back-test. This approach is only tested on the out-of-sample period since there is no for both ETFs there is no prior data available.

Figure 3: VIXY/SVXY back-test

Figure 3 plots the development of a 1.00$ investment in the long only, short only, and long-short strategy, backtested using the LSTM and the TARCH signal with ETFs for the out-of-sample period of 2012 to 2021. This strategy has daily returns.

It can be observed that with both methods, the long-short strategy outperforms the long only, as well as the short only strategy. While the LSTM signal yields better results on the long-short strategy than the TARCH signal, both signals show a similar performance pattern overall. Evidently, as expected due to the volatility premium, short volatility seems to be a successful strategy. The charts show that taking long volatility positions only pays out in times of market shocks. In the period from 2011 to 2021, the only market shock came during Covid -19 in the first half of 2020. It is observable that only in that instance, the long-short strategy starts to outperform the short strategy. There seems to be potential for extreme returns in such times. However, when only taking long positions, portfolios like the long ones in Figure 3, suffer a lot during calm markets, resulting in a poor overall performance even with eventual market shocks.

3.4.2 S&P 500 Covered Combo Index

Ni, Pan, and Poteshman (2008) demonstrate that for trading volatility, the dominant option strategy is to trade straddles. Guo (2000) shows that forecasting based on GARCH in currency option markets can earn significant profits when using delta-neutral and straddle strategies. Bollen and Whaley (2004) furthermore explain that buying straddles is the most efficacious way when expecting volatility to increase, whereas selling straddles covers the other direction. Evidently, naked positions also lead to certain risks regarding the movement of the underlying.

While a move in IV would certainly change the value of the option strategy, there can also be other reasons. Therefore, it is recommended to ideally cover or even delta-neutralize the position to isolate the volatility exposure.

Due to limited data availability in historic at-the-money S&P 500 options, this project makes use of the CMBO. This index hypothetically tracks the performance of a portfolio that sells one month 2% out-of-the-money Call options and sells one month at-the-money Put options. To cover the short call position a long position in the S&P 500 is taken, while to collateralize the put one month T-bills are bought. Since all components that make up the index are tradeable in the market, the strategy is still implementable. In this strategy, there are monthly views taken. At each rolling date, when the VIX is above the forecasted one month volatility level, a long position in the index is taken. When the VIX is below the forecasted volatility level, a short position in the index is taken. To implement the latter, an investor would have to buy the strangle while shorting the money-market instrument and the S&P 500.

Figure 4: CMBO back-test

Figure 4 plots the long only, short only, and long short strategy, backtested using the LSTM and the TARCH signal with the CMBO for the period of 2002 to 2021 if 1 dollar is invested. It also shows the underlying index, the CMBO. This strategy has monthly returns and covers the in-sample and the out-of-sample period.

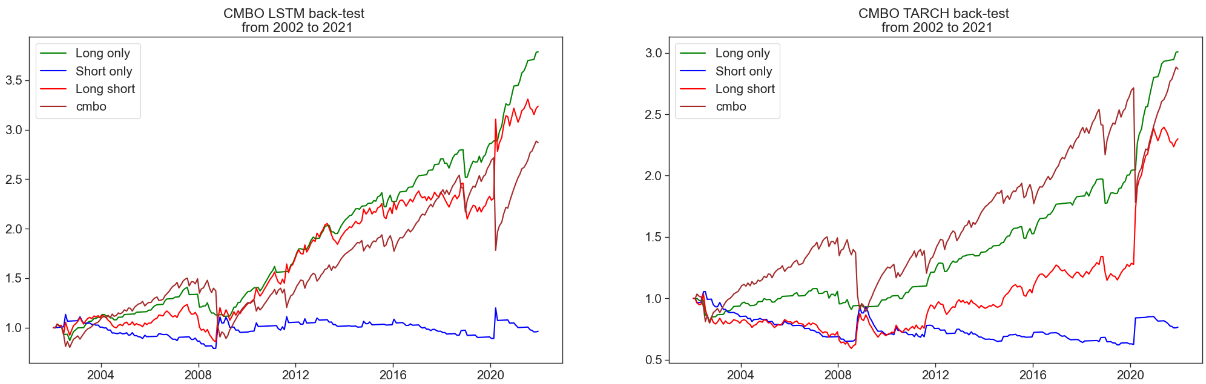


Figure 4 tells a similar story as Figure 3 already did. Again, the LSTM signal overall performs slightly better, even though the returns with the covered option strategy are considerably lower than with the ETFs. Here, a long only strategy corresponds to a long position in the CMBO index - a short volatility position. Therefore, as previously observed in Figure 3, shorting volatility is generally an effective strategy. Still, there is great potential for long volatility positions during market shocks. Assumably, due the covered nature of the option strategy, the performance spike in 2020 is not enough for the long-short strategy to outperform the short volatility (long) strategy. While during the in-sample period the LSTM roughly tracks the underlying, it manages to outperform the CMBO over 20 years in the long-only and long-short strategy. Obvious to see on the right graph, the in-sample period does not favor volatility trading as much as the out-of-sample period. Not only is the index itself performing worse, but the 2008 market shock is volatility-wise less extreme and therefore less profitable.

3.5 Limitations

While this project covers a wide range of methods, naturally some of them could not be covered in depth. No listed optionality for the past 20 years is used due to the data not being widely available and because at-the-money options are not actually traded in listed markets. Obviously, due to the inception date, the VIX ETF strategy has a limited timeframe of 10 years, only being applied to the out-of-sample period. In a more in-depth project, the back-test could be done using actual futures to replicate the index full timeframe. Finally, both strategies, the VIX ETFs and the CMBO, are unsuccessful at fully factoring out the volatility, but rather also take a view on the S&P 500 itself. It is beyond the scope of this project, but with the appropriate data, hedging ratios could be computed to optimally single out volatility.

4. Discussion

Even though both strategies to trade volatility seem to be profitable with either of the developed signals, they are also highly volatile strategies as Figure 3 and Figure 4 demonstrate.

Table 2: back-test metrics

Table 2 shows the main metrics for the performance of all strategies and respective signals for the full period of 2002-2021 with the S&P 500 as benchmark. The ETF strategy only includes the out-of-sample period from 2012-2021

Strategy	Cumulative Return	Annualized Sharpe Ratio	Annualized Alpha	Beta	Tracking Error	Information Ratio
VIX ETFs LSTM Long	129.14%	0.42	3.48%	-2.29	40.62%	0.09
VIX ETFs LSTM Short	612.26%	0.78	0.92%	0.69	29.23%	0.03
VIX ETFs LSTM Long-Short	1532.07%	0.75	4.4%	-1.6	57.83%	0.08

VIX ETFs TARCH Long	7.5%	0.28	3.01%	-2.29	40.87%	0.07
VIX ETFs TARCH Short	502.53%	0.73	0.81%	0.68	28.47%	0.03
VIX ETFs TARCH Long-Short	547.72%	0.60	3.81%	-1.6	57.56%	0.07
CMBO LSTM Long	278.59%	1.01	1.54%	0.18	6.05%	0.26
CMBO LSTM Short	-3.65%	0.04	1.57%	-0.55	7.83%	0.20
CMBO LSTM Long-Short	223.7%	0.48	2.94%	-0.37	12.81%	0.23
CMBO TARCH Long	200.85%	0.78	1.09%	0.23	6.18%	0.18
CMBO TARCH Short	-23.54%	-0.05	1.09%	-0.5	8.46%	0.13
CMBO TARCH Long-Short	129.97%	0.35	2.18%	-0.27	13.71%	0.16

Table 2 shows that when looking at the cumulative return, the VIX ETFs' long and long-short strategies, have a very attractive performance. Since these investment vehicles are highly volatile and the strategy consists of a one-asset portfolio, the Sharpe ratios are all below 1. They are yielding an attractive alpha over the S&P 500, but since they do not correlate with the index, the Information Ratio is completely off. Nonetheless, the Beta underlines what is expected: A crash in the market yields profitable returns for long volatility strategies. This results in negative Betas for all ETF strategies that profit from that. The profitable option strategies, however, have lower returns, but similar Sharpe ratios. The two short-only strategies (long volatility) yield negative returns, but the short and the long-short strategy have more promising metrics. Again, the strategies that allow a long volatility view, have negative Betas. It can be said that the long only (short volatility) option strategy using the LSTM signal is an appealing way to trade volatility. Not only does it have a very attractive 278.59% cumulative return, but also a Sharpe ratio of 1.01. Because of the negative Beta, paired with a decent performance and an acceptable Sharpe ratio of 0.5, also the LSTM long-short strategy can make a great addition to an equity portfolio in the S&P 500 universe. As observable in Appendix 1, both signals perform better in the out-of-sample period than during the in-sample period. Assumably, this is not due to a better prediction, but due to market environment, since it is unlikely that the model performs worse in the training period. The market shock in 2020 bolstered the performance so vigorously, that in the out-of-sample period the long-only option strategies for both signals even reach Sharpe ratios of over 1.3. While the impact of this occurrence on the overall performance is pleasing, it must be enjoyed with care. If the signals were to take the wrong direction in that market environment, the performance over the whole timeframe could be nullified.

5. Conclusion

Through the forecasting of volatilities using seminal methods, as well as state of the art approaches I realized that this is an important and still widely uncovered field of research. It feels like a breakthrough to have a NN approach outperform the best variation of the GARCH model, but there are a lot of puzzle pieces that must come together for this to perdure. Both ways of trading volatility and the signal development are implementable and tradeable strategies. In practice, they should be considered yield enhancement trades. This, especially in uncertain times, as part of a broader, more diversified portfolio. Comparing that many forecasting methods, and having them tested on various assets, while also attempting to factor in the volatility premium can prove valuable for further research. Furthermore, it is beyond doubt a unique way of shedding further light on the topic of volatility trading and forecasting. There are, however, still stones to be uncovered and ways to build on the findings done in this project. In a next step, the performance without the 2020 volatility spike could be analyzed. Moreover, it should be checked what would happen when factoring out the variance premium. Certainly, more short position would be taken, but since the short volatility strategies perform extremely well, this would maybe even augment the performance. On the other hand, however, this could also lead to the models not accurately capturing the volatility spike. Finally, it would be very interesting to see the performance when using the signals with daily delta adjusted option strategies that try to isolate the volatility exposure. Having raised these thoughts, I certainly believe that there are profitable volatility trading strategies. Yet, I assume that there are still models and ways to predict volatilities superior to the once covered in this project.

References

Bakshi, Gurdip, and Nikunj Kapadia. 2003. "Delta-Hedged Gains and the Negative Market Volatility Risk Premium." *Review of Financial Studies*, 16(2): 527-566.

Bollen, Nicolas P.B., and Robert E. Whaley. 2004. "Does Net Buying Pressure Affect the Shape of Implied Volatility Functions?. " *The Journal of Finance*, 59: 711-753.

Bollerslev, Tim. 1986. "Generalize Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics*, 31: 307-327.

Chen, Shiyi, Härdle, Wolfgang, and Kiho Jeong. 2010. "Forecasting volatility with support vector machine-based GARCH model." *Journal of Forecasting*, 29(4): 406-433.

Cho, Kyunghyun, Van Merriënboer, Bart, Gulcehre, Caglar, Bahdanau, Dzmitry, Bougares, Fethi, Schwenk, Holger, and Yoshua Bengio. 2014. "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation." Paper presented at the Conference on Empirical Methods in Natural Language Processing, Doha, Qatar.

Engle, Robert F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica,* 50(4): 987-1007.

Engle, Robert. F., and Magdalene E. Sokalska. 2012. "Forecasting intraday volatility in the us equity market. Multiplicative component GARCH." *Journal of Financial Econometrics*, 10(1): 54–83.

Glosten, R. Lawrence, Jagannathan, Ravi, David E. Runkle. 1993. "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks" *The Journal of Finance*, 48(5): 1779-1801.

Guo, Dajiang. 2000. "Dynamic Volatility Trading Strategies in the Currency Option Market." *Review of Derivatives Research* , 4(2): 133-154.

Harvey, Campbell R., and Robert E. Whaley. 1992. “Market volatility prediction and the efficiency of the S&P 100 index option market.” *Journal of Financial Economics*, 31(1), 43–73.

Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. “Long Short-Term Memory.” *Neural Computation,* 9(8): 1735–1780.

Jackwerth, Jens C., and Mark Rubinstein. 1996. “Recovering probability distributions from option prices.” *Journal of Finance,* 51: 1611–1631.

Jiang, George. J., and Yisong. S. Tian. 2005. “The model-free implied volatility and its information content.” *Review of Financial Studies*, 18(4): 1305–1342.

Liu, Dehong, Liang, Yucong, Zhang, Lili, Lung, Peter, and Rizwan Ullah. 2021. “Implied volatility forecast and option trading strategy.” *International Review of Economics & Finance*, 71: 943-954.

Liu, Yang. 2019. “Novel volatility forecasting using deep learning–Long Short Term Memory Recurrent Neural Networks.” *Expert Systems with Applications*, 132: 99-109.

Ni, Sophie. X., Pan, Jun and Allen M. Poteshman. 2008. “Volatility Information Trading in the Option Market.” *The Journal of Finance*, 63: 1059-1091.

Noh, Jaesun, Engle, Robert F., and Alex Kane. 1994. “Forecasting Volatility and Option Prices of the S&P 500 Index.” *The Journal of Derivatives*, 2(1): 17–30.

Rabemananjara, R., and Jean-Michel Zakoian. 1993. “Threshold Arch Models and Asymmetries in Volatility”, *Journal of Applied Econometrics,* 8: 31-49.

Rosenberg, Joshua V., and Robert F. Engle. 2002. "Empirical pricing kernels", *Journal of Financial Economics,* 64: 341–372.

Sermpinis, Georgios, Jason, Laws, and Christian L. Dunis. 2013. "Modelling and trading the realised volatility of the FTSE100 futures with higher order neural networks." *The European Journal of Finance*, 19(3): 165–179.

Shaikh, Imlak and Puja Padhi. 2015. "The implied volatility index: Is 'investor fear gauge' or 'forward-looking'?." *Borsa Istanbul Review,* 15(1): 44-52.

Sinclair, Euan. 2008. *Volatility Trading*. Hoboken: John Wiley & Sons.

Sinclair, Euan. 2020. *Positional Option Trading: An Advanced Guide*. Hoboken: John Wiley & Sons.

Appendix

Appendice 1: back-test metrics

Table 1 shows all metrics for the performance of all strategies and respective signals that are back-tested for the full period of 2002-2021. For the CMBO strategy, the in-sample (IS), as well as the out-of-sample (OOS) metrics are shown.

Strategy	Cumulative Return	Annualized Sharpe Ratio	Annualized Alpha	Beta	Tracking Error	Information Ratio
VIX LSTM long	129.14%	0.42	3.48%	-2.29	40.62%	0.09
VIX LSTM short	612.26%	0.78	0.92%	0.69	29.23%	0.03
VIX LSTM long-short	1532.07%	0.75	4.4%	-1.6	57.83%	0.08
VIX TARCH long	7.5%	0.28	3.01%	-2.29	40.87%	0.07
VIX TARCH short	502.53%	0.73	0.81%	0.68	28.47%	0.03
VIX TARCH long-short	547.72%	0.6	3.81%	-1.6	57.56%	0.07
CMBO LSTM long	278.59%	1.01	1.54%	0.18	6.05%	0.26
CMBO LSTM short	-3.65%	0.04	1.57%	-0.55	7.83%	0.2
CMBO LSTM long-short	223.7%	0.48	2.94%	-0.37	12.81%	0.23
IS CMBO LSTM long	76.06%	0.73	1.47%	0.21	6.66%	0.22
IS CMBO LSTM short	5.52%	0.1	0.78%	-0.54	7.84%	0.1
IS CMBO LSTM long-short	74.97%	0.42	2.07%	-0.33	13.59%	0.15
OOS CMBO LSTM long	115.03%	1.39	1.7%	0.14	5.34%	0.32
OOS CMBO LSTM short	-8.69%	-0.02	2.37%	-0.59	7.75%	0.31
OOS CMBO LSTM long-short	85.01%	0.53	3.9%	-0.45	11.9%	0.33
CMBO TARCH long	200.85%	0.78	1.09%	0.23	6.18%	0.18
CMBO TARCH short	-23.54%	-0.05	1.09%	-0.5	8.46%	0.13
CMBO TARCH long-short	129.97%	0.35	2.18%	-0.27	13.71%	0.16
IS CMBO TARCH long	26.09%	0.29	0.49%	0.22	6.79%	0.07
IS CMBO TARCH short	-24.55%	-0.17	-0.21%	-0.53	7.95%	-0.03
IS CMBO TARCH long-short	-4.88%	0.01	0.28%	-0.31	13.84%	0.02
OOS CMBO TARCH long	138.6%	1.36	1.63%	0.24	5.47%	0.3
OOS CMBO TARCH short	1.34%	0.06	2.27%	-0.5	7.88%	0.29
OOS CMBO TARCH long-short	141.76%	0.72	3.91%	-0.26	13.34%	0.29