FORECASTED SHARPE RATIO RANKING STRATEGY

Master Thesis

Copenhagen Business School
Programme: Msc Finance & Accounting (CM FIR)
Hand in date: 15th May 2018

Student: Benjamin Callesen

Student: Christian Buhl Møller

Supervisor: Søren Ulrik Plesner

Characters: *189,113*
Pages: *97*

Abstract

The purpose of this thesis is to understand if a portfolio constructed on the basis of ranking stocks by their predicted Sharpe ratio can produce a mean excess returns, a mean volatility, or a mean Sharpe ratio that is superior and different from an equally weighted portfolio of the same stock universe with statistical significance. The relevant stock universe is the current constituents of the Dow Jones Industrial Average, excluding 3 of the stocks, resulting in sample data consisting of 27 stocks over the period 28-03-1991 to 26-03-2018.

The study uses four different implementations of linear regressions to predict the Sharpe ratio as the dependent variable and the lagged 12-month forward, the lagged 12-month trailing earnings per share, the lagged 3-month excess return, the lagged 6-month excess return, the lagged 9-month excess return, and the lagged 12-month excess returns as independent variables for all 27 stocks. The portfolio findings are based on holding the 20 stocks with the highest predicted Sharpe ratio with a holding period and rebalancing frequency of 3-months.

The portfolios are tested against each other and the equal weight portfolio to determine both the internal hierarchy of the model portfolios and the performance relative to the simpler portfolio. The models are found to have some differentiating traits regarding their ability to minimize squared errors and maximize the number of significant Sharpe ratio estimates in relation to the actual observed Sharpe ratios of the different stocks. However, the linear relations between the dependent variables and the independent variables are found to be weak, which means the predicted Sharpe ratio is most often not statistically significant, and as a result, the wrong stocks are excluded or included based on the Sharpe ratio criteria. Consequently, the result is that the mean excess return, mean volatility, and mean Sharpe ratio produced by the four model portfolios, are not found to be different from the equal weight portfolio with statistical significance. Although it is shown that if the ranking procedure is based on ex post Sharpe ratios, a portfolio with statistically significant and superior excess return and Sharpe ratio can be constructed.

Contents

Introduction

Assets managers are always chasing higher risk adjusted performance in order to deliver as value much as possible to their clients. Whenever capital is invested into an asset or a trading strategy, there is always a risk of incurring losses. If investors single-mindedly try to maximize the expected return, then they may take on unacceptable levels of risk. According to asset pricing models, such as the capital asset pricing model (CAPM), risk is not taken for free and investors expect compensation in line with the risk incurred. In a perfect capital market, it can be more attractive to apply leverage to assets with lower expected returns, but higher risk adjusted returns. This is exemplified by the betting against beta strategy. The Sharpe ratio gives the amount of excess return for each unit of volatility (standard deviation) associated with an asset or portfolio. Investors are interested in obtaining as high a Sharpe ratio as possible as this indicates a more optimal trade-off between risk and return.

With the rise in popularity of low cost passive index investing strategies, we are interested to see if asset managers would be able to improve upon the indexing strategies risk adjusted returns by implementing an equity ranking model which excludes the stocks producing the worst risk adjusted returns from the portfolio.

Problem definition

We have been inspired to write our thesis largely by the following three papers:

Time-Varying Sharpe Ratios and Market Timing (Tang & Whitelaw, 2018) researches the relation between Sharpe ratio estimates and business cycles to take advantage of a market timing strategy. The paper finds that there is a substantial time-variation in the estimated Sharpe ratios that coincides with the phase of the business cycle. The relation is such that Sharpe ratios are low at the peak and high at the trough of the cycle.

Anomalous Returns in a Neural Network Equity-Ranking Predictor (Satinover & Sornette, 2008) rank orders a fixed universe of about 1,500 stocks by their predicted price change over the next quarter, and investments are made based on the predicted ranking. The inputs they have used to forecast the predicted price change are the ten prior quarterly percentage changes in price and earnings for each equity with quarterly frequency.

Time Series Momentum (Moskowitz, Ooi, & Pedersen, 2011) researches the relation between expected returns and past returns of commodities, stocks, currencies, and bond futures and forwards. They find there is significant prediction power in the past 1-, 3-, and 12-month return across all assets on the future return upwards of one year forward in time. They differentiate between time series momentum as looking at a single individual asset, and cross-sectional momentum as looking at an individual asset compared to peers.

Reading these papers inspired us to an idea of a combination of these different methodologies and approaches. We draw upon the following from each paper:

1. (Tang & Whitelaw, 2018): We wish to use the Sharpe ratio as our main portfolio selection tool, because the performance measure weighs returns according to risk. We also intend to implement the models using both a rolling and expanding forecast method.
2. (Satinover & Sornette, 2008): Investing based on the ranking of a predicted performance measure, where the predictions are based on prior quarter's earnings and return performance.
3. (Moskowitz, Ooi, & Pedersen, 2011): Prediction estimators for each equity will include the lagged return of the given equity, where the lagged return is at maximum the past 12-months return to take advantage of the positive autocorrelation of the time series momentum effect.

As such we are interested in researching how this specific combination of methodologies can be combined and how it will compare to a buy and hold strategy of all the stocks included in the selection universe.

All of the above has given occasion for us to examine the following problem statement:

How does a trading strategy that ranks and picks between stocks based on their forecasted Sharpe ratio compare to an equally weighted portfolio of all the stocks in the selection universe?

Sub questions

To forecast the Sharpe ratio we have decided to use a linear ordinary least squares regression on a quarterly basis. Given that there are a large multitude of ways to implement such a model we focus on using either a rolling or expanding time window regression model, which naturally gives rise to the question:

Which is the better way to implement a linear ordinary least square forecasting model, when choosing between a rolling or expanding time window?

Since the Sharpe ratio can be deconstructed into its formula components, we wonder what the better approach will be when forecasting; either directly forecasting the Sharpe ratio or indirectly forecasting the Sharpe ratio through the two components (excess return and volatility). Consequently, we want compare the two different methods, therefore, we ask:

Which is the better way to implement a Sharpe ratio forecasting model, when choosing between a component or direct approach?

Delimitation

The thesis uses the 30 constituents of the Dow Jones Industrial Average, excluding three, to end up with a total investable universe of 27 stocks.

We assume for all intents and purposes that the coefficients of the regressions are unbiased and consistent. Due to the sheer number of regressions required to implement the models, a total of 13,608 quarterly regressions, we do not test for all of the ordinary least squares assumptions such as a mean error term of 0 and heteroscedasticity, normality among the variables, linear relations between independent and dependent variables, multicollinearity, and autocorrelation. Furthermore, we limit ourselves to not look at implementation costs related to taxes, trading and financing. Given the fact that we are analyzing large cap US companies, a large share of the total returns will be in the form of dividends. As such, our total returns do reflect this and assume that all dividends are reinvested. However, there is made no distinction between the return of the dividends that are invested during the holding period and the rest of the investment gains.

As a result of our delimitations as well as the small sample size, we are aware of the fact that our results are not generally applicable.

Literature Review

The literature review consists of a selected range of academic and scientific research studies in the form of published papers by internationally renowned academics. The material is used as a basis to give insights to the financial subject of the thesis and the theories behind those subjects. Furthermore, we draw upon their expertise when designing the methodology, while not replicating their work, such that we have a more explorative approach to the subject.

Time series momentum is one of the most studied and researched phenomena in finance. In short, it is the relation between a security's past performance and its current performance. The subject has been covered extensively, and in 1993 Jegadeesh and Titman wrote about the momentum and mean reversion effect in their paper *Returns to Buying Winners and Selling Losers: Implications for Stock Market*. Momentum and mean reversion strategies were shown on average to generate cumulative returns of 9.5% over a 12 month period, after which more than half of the returns were lost over the following 24 months.

Time Series Momentum Investing

In the finance literature, there is a distinction between momentum and time series momentum investment strategies. Momentum is cross sectional, focuses on the relative performance of securities, and finds that the securities, which have been outperforming their peers over the past three to twelve months, will continue to do so on average over the following month. Time series momentum, on the other hand, looks at the individual security's past performance.

Mosokowitz, Ooi, and Pedersen (2011) documents their findings regarding time series momentum. They find that the past twelve months excess return for a financial instrument has a consistent predictability power for its future excess return across equity indices. The results are also found to be consistent across different asset classes; more specifically, futures, and forward contracts on equity indices, as well as currencies, commodities, and sovereign bonds.

Time series momentum is based around the hypothesis that the trend of the return on the financial asset in the last period will continue in the next period and then partially reverse over the long term. Thus, investment strategies that focus on the time series momentum exploit the positive autocorrelation that individual securities exhibit.

Investment strategy

The strategy is divided into two phases; the look-back phase and the holding phase. Moskowitz, Ooi, and Pedersen (2011) find the optimal look-back and holding period is twelve months or less, which supports the fact that momentum is a trend-following strategy in the short to medium run, whereas the strategy faces mean reversion in the long run. This theory is consistent with the theory of initial under reaction and delayed over reaction to new information, which can produce such return patterns, and will be elaborated on in the behavioral finance section. The findings of the time series momentum are a breach of the random walk assumption incorporated in the efficient market hypothesis, because the momentum strategy has consistent predictability power of the next period's return. The strategy can, in theory, exploit both the momentum effect on the short to medium run, and mean reversion in the long term, by inverting the investment before moving into the long-term period. Jegadeesh and Titman (1993) found the momentum effect to last over the next 12 months when using a 6-months look back period with a 24 month mean reversal.

Time Series Momentum Robustness

The concept of an investment strategy focused on the positive autocorrelation in the individual financial securities dates back to Jegadeesh and Titman (1993), who analyzed the US equity market with data from 1965 to 1989. They ranked the stocks according to their past performance and constructed a portfolio long the stocks in the top 10 percentile while shorting the bottom 10 percentile. With a look-back and holding period varying from three to twelve months, they found evidence of positive autocorrelation present for the individual stocks. In *Profitability of Momentum Strategies: An Evaluation of Alternative Explanations* Jegadeesh and Titman (1999) retest their model using eight additional years of data and are able to arrive at the same conclusions. The momentum profits continue to be predictable in the 1990's, which adds to the robustness of the theory and shows the results were not necessarily a product of targeted data mining. The findings are consistent with the conclusions of several momentum studies following the original paper by Jegadeesh and Titman (1993) such as Moskowitz, Ooi, and Pedersen (2011), which finds a momentum premium across financial future instruments on equity indexes, currencies, commodities, and sovereign bonds, and even shows that there is significant autocorrelation across the different asset classes' time series momentum factors.

Mean Reversal

Mean reversion is the correction of a security's price to be in line with its fundamental value. Both Jegadeesh and Titman (1993) and Moskowitz, Ooi, and Pedersen (2011) find that financial securities with time series momentum premiums eventually revert or correct themselves. This translates to the fact that securities exhibit both positive autocorrelation and negative autocorrelation with past performance. Trading on the reverse signals of the momentum predictions would therefore also yield abnormal returns in some holding periods. Satinover and Sornette (2008) find that their model is able to segregate their sample stocks into those that will rise and fall relative to one another. However, for many quarters, the model not only fails to segregate correctly, but rather inverts the predictions. Buying past losers and selling past winners ends up a profitable trading strategy in some holding periods, because even when model predictions are different from the actual ex post rankings, current winners are likely to become future losers at some point, while the reverse is true for current losers.

Explaining the Time Series Momentum Premium

Many different research papers have found evidence of a time series momentum premium, and Moskowitz, Ooi, and Pedersen (2011) argue that the tendencies that the momentum strategy exhibits match predictions of behavioral finance theory.

Rational and risk based theories

Classic rational finance theory proclaims that any excess return is a compensation for a given level of risk. Given this, one would assume that the momentum portfolio of recent winners outperforms its peers that were recent losers, because they have a relatively higher risk associated with their returns. Jegadeesh and Titman (1993) use the capital asset pricing model in their momentum study and emphasize that momentum premium is not explained by compensation for market risk, which would be in line with what the capital asset pricing model argues for. They prove that the securities that are recent losers have a greater systematic risk relative to their winning peers.

If macroeconomic factors, on the other hand, could explain the momentum premium, there should exist a positive correlation between the return of the momentum strategies and swings in the economic cycles. Proxy measures for the economic cycles could be the Baa-Aaa spread, the commercial paper-Treasury spread, the one-year Treasury yield, and the dividend yield, which Tang and Whitelaw (2018) use in a market timing strategy. However, in *A Century of Evidence on Trend-Following Investing,* Hurst, Ooi, and Pedersen (2017) find that the momentum strategy exhibits strong diversification properties because it achieves abnormal returns in both booming and distressed markets. In their analysis of the past distressed markets since 1893, they find that the momentum model would only have failed to deliver anomalous returns during the 1937 recession and the 1987 market crash. When comparing returns in recession versus boom, low and high inflation, war versus peace, and bull versus bear, they find only marginal differences at a 95% confidence interval. They conclude that the momentum investment strategy has performed well in each decade for more than a century with significant and robust out of sample evidence across both markets and asset classes.

Behavioral finance

Finance scholars and professionals lean towards behavioral finance as the explaining factor for the momentum premium due to the lack of correlation with rational and risk-based theories. In *Do Industries Explain Momentum?* Moskowitz and Grinblatt (1999) find a strong effect between momentum and industry components of stock returns using data from 1963 to 1995 on stocks from NYSE, AMEX, and Nasdaq. They find that buying stocks from past winning industries and shorting stocks from past losing industries is a very profitable strategy. The industry momentum effect is also found to provide a considerable explanation power for the anomalous returns from the time series momentum of individual securities.

The finding is in contrast to Hurst, Ooi, and Pedersen's (2017) statement saying that the momentum strategy has strong diversification properties and therefore performs well in all economic cycles. Moskowitz and Grinblatt (1999) instead argue that momentum strategies are associated with a much higher risk, given that winners and losers are clustered together in separate industries. As such, the momentum strategy is heavily invested in certain industries while shorting other industries, and as a consequence, the strategy is highly vulnerable to shocks to the specific individual industry clusters.

Moskowitz and Grinblatt (1999) believe the explanation for the industry momentum effect is found in behavioral finance theories. Investors focus on the attractive industries and neglect the less attractive industries, which causes the prices to move accordingly, and even persistently, over a period. Investors may also be overconfident and exhibit biased self-attribution when assessing value in certain industries. This effect could be especially prevalent when investors are asked to assess new or changing industries, such as Internet stocks in the 90s, leading to mispricing of industries compared to the fundamental value. Another explanation is a conservatism bias, where investors underreact to new information, causing the price to climb upwards over a longer period instead of jumping immediately. In combination with representativeness bias, investors may extrapolate the information for a given industry too broadly, leading to mispricing and reversals in the long run, as observed in the momentum strategies.

Moskowitz, Ooi, and Pedersen (2011) mention a wide range of behavioral finance theory papers and studies that can explain the momentum premium as an initial under reaction, followed by a delayed over reaction, and eventual reversion. However, their findings also challenge behavioral finance as an explanation to the momentum premium, because the markets and assets included in the research have very different types of investors, while the return patterns are consistent across the asset classes. They find the correlation of time series momentum returns between asset classes are higher than positions within the same asset class. This implies that the momentum effect is able to extrapolate an explaining component from a given asset across all asset classes, which is not present in the individual asset itself.

Earnings momentum

As there is no current established and broadly accepted theory that can explain the anomalous returns of the momentum strategy. Chan, Jegadeesh, and Lakonishok argue in *Momentum Strategies* (1996) that there is a high likelihood of not working out of sample and it simply being a statistical fluke. However, given the plethora of empirical studies that have found significant momentum returns, it seems unlikely that it is a statistical fluke, and more likely that the explanatory factor or object has not been identified or cannot be observed. This belief is also strengthened by the fact that Jegadeesh and Titman (1999) are able to repeatedly find evidence of the momentum effect producing abnormal profits when expanding their data set from their 1993 analysis.

Chan, Jegadeesh, and Lakonishok (1996) research parallel analyses looking at both earnings and price momentum. They find that earnings are an explanatory factor for the changes in price of a given asset and explore the relation between time series momentum in stock prices and investors under reaction to new earnings information. Everything else equal, one would assume that a company producing higher earnings would also deliver higher returns relative to its poorer performing peers.

An earnings momentum strategy would exploit the under reaction to information in the market in the short term, while the price momentum strategy exploits the markets under reaction to a broader range of information, such as long term profitability, either increasing or decreasing. Under these assumptions, both price and earnings momentum could be independently successful, given that reported earnings are based on accounting principles. This could cause a discrepancy between the future economic earnings and reported earnings of the company. For example, if the asset prices reflect wider information than is available in the accounting numbers, such as industry specific trends in profitability, then the asset might have a positive price momentum factor, even with weakened reported earnings. This is also reflected in the research of

Chan, Jegadeesh, and Lakonishok (1996), which finds that the price momentum effect is stronger and persists for a longer time than the earnings momentum effect.

The effects of the earnings momentum are grouped around the announcement date of the asset's earnings. On average, shocks from either good or bad information continue to affect the market prices over the next two following earnings announcements. Bad news regarding stock prices spreads more slowly. When stock analysts have to adjust their forecasts to reflect negative information, they risk their own standing with the stock company's management, and are therefore more reluctant to share the bad news (Chan, Jegadeesh, & Lakonishok, 1996).

Forecasting Sharpe ratio

Tang and Whitelaw (2018) use a range of financial variables to forecast equity excess returns, volatility, and Sharpe ratio using the value weighted CRSP index over the period May 1953 to December 2010. The value-weighted CRSP index is an index of the S&P 500, S&P 500 Composite and the NASDAQ Composite (CRSP, 2018). The Sharpe ratio is estimated by forecasting the individual components that make up the Sharpe ratio; excess return and volatility, and by directly forecasting Sharpe ratio with their chosen estimators in a linear setup. They construct a linear model that has statistically significant estimators when predicting excess return and volatility. They find a significant linear relation between the excess return as the dependent variable, and the dividend yield and the 1-year Treasury rate as the significant independent variables. For volatility, they find that the lagged volatility, the Baa-Aaa spread, the dividend yield, and the commercial paper Treasury spread to be significant estimators. Their results reveal that the model is significantly better at predicting volatility compared to excess returns with a R^2 of 2.75% for excess returns and 54.37% for volatility over their full sample period. When they forecast Sharpe ratio directly, they find that the lagged volatility, an adjusted dividend yield, and the 1-year Treasury rate to be significant estimators. Their model ends up explaining 3.71% of the variation in the Sharpe ratio over the full sample period.

When Tang and Whitelaw (2018) predict the Sharpe ratio out of sample using a rolling and expanding model, their R^2 is in a lower range of 0.12% to 1.21%. Furthermore, they are not able to produce significant estimators for all of their regressions, which could indicate that the Sharpe ratio is not well suited to being predicted using a linear model.

Trading strategy signals and implementation

Moskowitz, Ooi, and Pedersen (2011) document their time series momentum findings through a combination of portfolios containing a long short strategy that is constructed to finance itself and neutralize the market exposure. The long positions consist of the past winners, and the short positions of the past losers. In their research they regress the excess returns of their individual instruments on the instrument's lagged return:

$$\frac{r_t^s}{\sigma_{t-1}^s} = \alpha + \frac{\beta_h(r_{t-h}^s)}{\sigma_{t-h-1}^s} + \epsilon_t^s$$

Formula 1.1.1: Excess returns scaled by ex ante volatility (Moskowitz, Ooi, & Pedersen, 2011)

The excess returns are scaled by their ex ante volatility to be able to compare the excess returns across assets. The lags take on the values of *h* = 1 to 60 months. They find consistent and significant positive return continuation on the first 12 months of lagged excess returns, followed by a reversal effect in the following years, which is most significant in the second year.

Their trading signals are not based on the forecasted excess returns of the regressed model, but instead of varying look back and holding periods. They construct portfolios where all securities are either part of a long or short investment given their past *k* months returns (long if positive and short if negative), where *k* = 1 to 48 months. The holding period *h* = 1 to 48 as well. The results reveal that the optimal look back and holding period is less than or equal to 12 months in total. In comparison, Jegadeesh and Titman (1993) have overlapping investment periods when their holding period is more than 1 month, because they still add new investments to their portfolio every month.

Satinover and Sornette (2008) rank their fixed universe of around 1,500 stocks by predicted change in price over the next quarter and construct long portfolios of the best 10, 20, …, 100 performing stocks, short portfolios of the worst 10, 20, …, 100, and combination portfolios to hedge against market risk. The best and the worst stocks are then included in a larger number of the portfolios, and as such are weighted more when combining the return from all the portfolios. All of their portfolios are held for a quarter, because their input is based around information with a quarterly frequency.

Methodology

The methodology section explains how we conduct the research into the topic of using Sharpe ratio as a ranking tool for a selected universe of stocks, and how we evaluate the performance of such a strategy when compared to an equally weighted portfolio of the stocks in the selection universe. The research takes on an explorative approach to the subject, as we try to find, and understand, the link between using a forecasted Sharpe ratio and future performance.

Strategy

The literature review has covered a range of different independent studies of times series momentum and earnings showed prediction power for future excess returns. The studies, like that of Tang and Whitelaw (2018), have found estimators that are significant when forecasting excess returns, volatility, and Sharpe ratio. Others, such as Moskowitz, Ooi, and Pedersen (2011), have found that past excess returns are consistent estimators of future excess returns, but are unable to consistently prove a significant linear relation. Most of the literature reviewed used historical price or earnings momentum to decide whether a long or short position should be taken in a given financial security without forecasting the actual excess return. Instead, their intent has been to use the past excess return as a signal in a trading strategy. We wish to see if these variables, lagged excess return and earnings, have prediction power in a linear model when forecasting the next periods excess return, volatility, and Sharpe ratio.

The motivation to predict the Sharpe ratio is to take risk into consideration, such that the ranking is done based on risk adjusted returns. The Sharpe Ratio is used as a predictor for which stocks will underperform compared to their peers in the next period.

A total of four different methods of forecasting Sharpe ratio are used before comparing and testing them to see which one has the most accurate predictions, consistent significant results, and if any produce better performance than a simple, equal weighted portfolio of all the stocks in our selection universe. The methods are split into forecasting the components of the Sharpe ratio individually in the form of the excess return and volatility, and a forecast directly of the Sharpe ratio. Both of these are done as both an expanding and a rolling twenty period forecast, such that we have the following 6 types of regressions shown in figure 1.2.1:

1. Separate forecasts of excess return and volatility and calculation of the next period's Sharpe ratio for each of the 27 stocks with a rolling regression, using a moving time window of 20 quarters.
2. Forecast of the Sharpe ratio for each of the 27 stocks with a rolling regression, also using a moving time window of 20 quarters.

3. Separate forecasts of excess return and volatility and calculation of the next period's Sharpe ratio for each of the 27 stocks with an expanding regression, using an initial window of 20 quarters.
4. Forecast of the Sharpe ratio for each of the 27 stocks with an expanding regression, also using an initial window of 20 quarters.

The quarters defined in this thesis follow the calendar year quarters. For easier referencing, the names for these models are chosen according to their method. This is done such that the models forecasting Sharpe ratio are prefaced with *direct*, while the ones forecasting both excess return and volatility are prefaced with *component*. The latter part of the name is either rolling or expanding, depending on the time window used for the forecasting model.

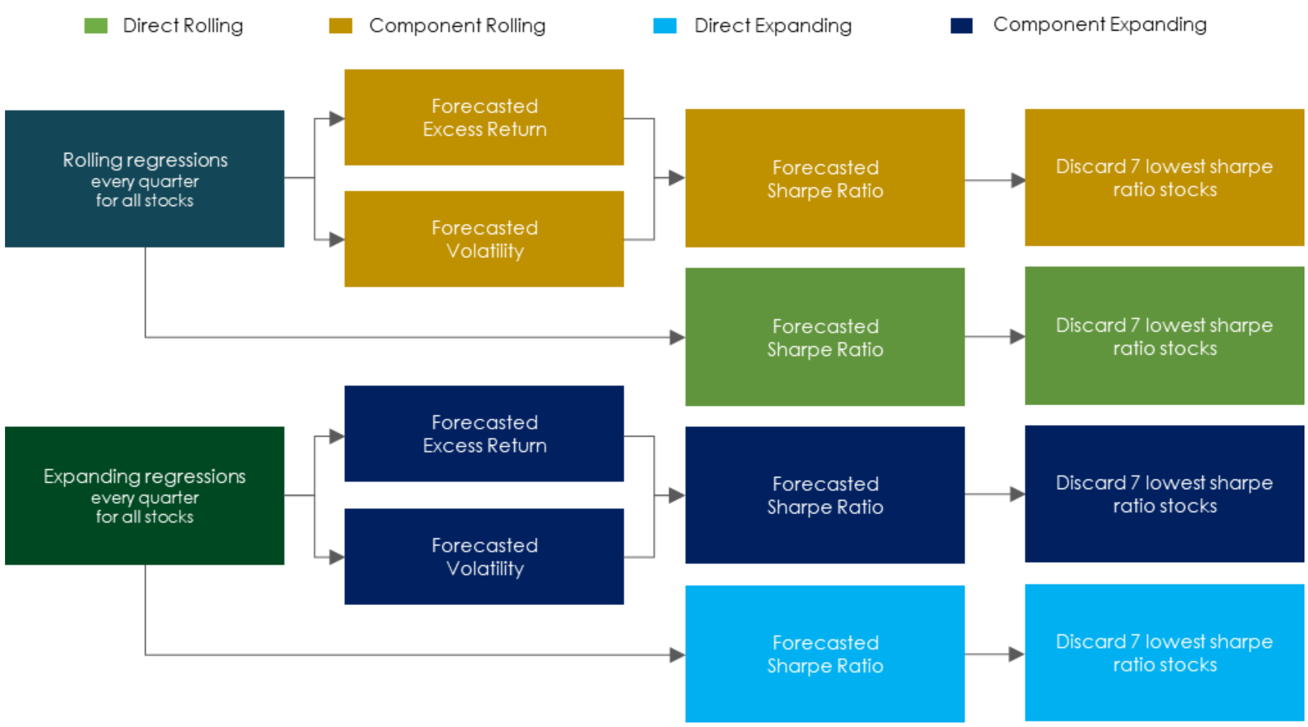

Figure 1.2.1: Illustration of the four different models to forecast Sharpe ratio and choose which stocks to exclude in the investment period

With a rolling twenty-period forecast, the last 20 quarterly observations are used as input in the regression. Thus, the regression contains information about what has happened in the last 5 years. With the expanding method, the initial regression is also only the last 20 quarterly observations, but when a new observation is added, the oldest one is not removed. Therefore, eventually the model ends up with all 84 quarterly observations included in the regression when forecasting the last period in the sample data. The hypothesis regarding implementing the different model methodologies is that the rolling models will identify the trends more easily, while the expanding model will be less error-prone in times of extreme or rapidly changing volatility.

The forecasting models only forecast one period ahead, which means that they compute a single regression for all stocks in all quarters from 1997 Q2 to 2018 Q1 with all of the four different model setups. We are aware of the fact that if we forecast more than one period ahead, the model is very likely to move out of synchronization with the actual observations in the market, because the error terms would be autocorrelated. In order to avoid these autocorrelated error terms as much as possible, only the following quarter is always forecasted.

Portfolio selection and size

The data set contains 27 of the 30 constituents of the current Dow Jones Industrial Average, since 3 stocks are excluded due to reasons explained in the data description section. The thesis intends to exclude the 7 stocks with the lowest Sharpe ratio in all investment periods. The logic behind this approach is that the model does not have to successfully distinguish between the best performing stocks, but rather, has to identify the worst performing ones. This makes it relatively easier for a linear model, because the realized positive return outliers can take any positive value, while the realized negative return outliers are limited at -100%. Therefore, a return would mean that the stock is worthless and the company bankrupt.

The rationale for this selection approach is that the estimators can identify the stocks that have performed poorly in the look-back phase and use this to forecast the ones that perform poorly in the next holding period as well. Depending on the model's ability to correctly identify trends, shorting the stocks with the poorest performance in the look-back period could lead to overall higher returns. However, this is not explored in the thesis.

When determining the ideal number of stocks that should be included in the portfolio from the 27 stocks in total, financial theory regarding diversification and portfolio risk is applied. The reason to exclude 7 stocks is that this corresponds to having a portfolio of 20 individual stocks. The decision is founded in financial theory regarding realizing a diversification benefit by diversifying idiosyncratic risk: risk that relates to individual stocks, which can be diversified by holding multiple stocks. It can be shown that when the stocks in a portfolio are equally weighted – that is, each of N stocks are allocated $\frac{1}{N}$ of the portfolio – then as N approaches 20, the majority of the gain from diversification is realized already and diversifying further only brings very marginal gains (Bodie, Kane, & Marcus, 2014). This is also the reason why the weight between the highest ranked 20 stocks is equal, and not relative, compared to the forecasted Sharpe ratio.

Look-back and holding period

A look-back and holding period of one to twelve months has, through all the literature covered, been found to exhibit positive correlated excess return with mean reversal in the form of negative correlated excess return beyond that period. The momentum effect of the changes in earnings is supposed to have the largest effect in the short run, and momentum effect of the changes to the price of the stock has a longer-lasting effect, per the reviewed literature.

The look-back period for the change in the price uses the past 3-, 6-, 9- and 12-months excess returns and the change in earnings happens when quarterly earnings are announced.

Portfolio rebalancing

When deciding upon the frequency and timing of rebalancing, there are many implications, such as trading activity and computational requirements. As mentioned in the section above, the rebalancing was selected to coincide with the new release of quarterly earnings.

Figure 1.2.2: Illustration of first investment period and first rebalancing

As illustrated in figure 1.2.2, the portfolios are rebalanced every quarter. It is assumed that the stocks realize their respective returns based on the closing price of the last day in the holding period's quarter. It is also assumed that it is possible to buy stocks for the next holding period based on the closing price of the last day of the current quarter. Thus, the assumption is that the closing price of the last day in the current quarter is equal to the opening price of the first day in the following quarter.

Trading costs, tax related costs, or financing costs will not be accounted for, as this is not the purpose of the thesis. As a consequence, there is no difference for holding a stock between multiple holding periods and completely selling and rebuying the stock again to rebalance the portfolio.

The Dow Equal Weight portfolio is constructed such that all 27 stocks have an equal weight at the beginning of all investment periods. This means that the portfolio is also rebalanced each quarter, such that the portfolio weights are always equal at the start of every holding period. As a result, the portfolio is not equally weighted between the stocks at the end of the holding period, because of the changes to the respective price of the stock between each investment window.

The perfect Sharpe ratio rank portfolio is constructed based on ex post knowledge of Sharpe ratios. These are used to rank and select the 20 best performing stocks for the next holding period, which is then rebalanced every quarter.

Testing for significance and accuracy

To test if the explanatory variables used to forecast in the models have any significant explanation power, an ordinary least squares linear regression is computed for each of the 27 stocks for all of the three different dependent variables; excess return, volatility, and Sharpe ratio, over the full sample period. The independent variables are lagged by one quarter. That is to say that it is consistent with the method used for forecasting. The results from these regressions, and therefore the information regarding whether the chosen estimators are significant or not, would not be possible to know ex ante for implementing the trading strategy in 1997. However, it is still important to know for the purpose of evaluating the model for future use outside the tested strategy time period.

In general, statistical tests are necessary in order to verify if the findings have any statistical significance beyond just showing if there is a significant linear relation between the estimators and the dependent variables when forecasting.

The models and constructed portfolios are tested on three different levels:

1. The first level of tests evaluate the model's ability to accurately forecast Sharpe ratio. The Diebold Mariano test is used to test if there is a statistical significant difference in the squared errors of the models. The model with the best accuracy does not necessarily have to produce statistical significant estimators, but is simply the model with the least amount of squared errors between the four competing forecasting models.
2. The second level of tests computes how well the forecasted Sharpe ratio relates to the actual Sharpe ratio. If the models are not able to consistently predict Sharpe ratio, then as a consequence they will not consistently rank the correct stocks to exclude. It is then harder to generalize and accept the results as being a product of a valid model.
3. The third level of tests includes ANOVA and two population hypothesis testing of the portfolios realized excess return, volatility and Sharpe ratio. The performances of all four portfolios are benchmarked against each other, as well as the equal weighted portfolio. The ANOVA test is used first to see if any of the portfolio results are statistically different. If any difference is found, a two population test will be used to evaluate which of the portfolio pairs produce the performance that is significantly different.

Lastly, the optimal portfolio, which is always able to pick the correct stocks with the highest Sharpe ratios in the following holding period, is constructed. This portfolio is included to see if there is any significant performance advantage to gain with the perfect model. The portfolio results are included in a new ANOVA test against all the other five portfolios.

Data description

In general, it is advantageous to have as many observations as possible to better achieve statistically significant results. The description of the three-month US Treasury Bill middle rate DataStream reveals that observations earlier than 1985 are not exact and are merely estimates (DataStream). To be on the safe side, the data is chosen from the early 1990's. More specifically, the data collection starts from the second quarter of 1991 to avoid unnecessary estimation errors in the raw data.

In this thesis, the daily total returns, quarterly twelve-month trailing earnings per share, and quarterly twelve-month forecasted earnings per share are collected for all the individual companies that make up the constituents of the Dow Jones Industrial Average index as of 28-03-2018. The data has been sourced through the DataStream Excel plugin available at Copenhagen Business School and has been collected for the period 28-03-1991 to 26-03-2018, which corresponds to 27 years of data. The following constituents are excluded due to limited amount of observations being available in the chosen time frame:

- Goldman Sachs Group Inc. (NYSE: GS), Company IPO May 1999
- Visa Inc. Cl A (NYSE: V), Company IPO March 2008
- DowDuPont Inc. (NYSE: DWDP), Merger between Dow Chemical and DuPont August 2017

The 3-month US Treasury Bill middle rate has also been sourced from DataStream for the same period. The 3-month US Treasury Bill rate is chosen because it corresponds with the selected holding period and all of the sample companies are US based.

Daily total returns are collected in order to have a larger sample size when estimating the quarterly volatility. The earnings per share measures are collected on a quarterly basis, as companies announce quarterly updates regarding their earnings, and these affect both the historical twelve-month trailing measure and the consensus twelve-month forward estimate.

The assumption in choosing these variables as estimators is that the literature shows that they have predictive power regarding the expected excess returns. The rationale is that these variables have proved their ability to predict future stock returns in prior research when used as trading signals. The historical and forward earnings per share estimates are used as a proxy for changes to the expected future profitability and expected cash flow from the company to its shareholders. The daily total returns are used to capture the time series momentum effect for the given stock.

Daily Total Return Index

The total return index data is based on the price of a given stock and the value of dividends paid, which are added to the price on the ex-date of the payment such that:

Method 1 (using ex-dividend date):

$$Total\ return\ index_t = Total\ return\ index_{t-1} * \frac{P_t}{P_{t-1}}$$

Formula 1.1.15: Return index before dividends

Method 2 (when t = ex-date of the dividend payment):

$$Total\ return\ index_t = Total\ return\ index_{t-1} * \frac{P_t + D_t}{P_{t-1}}$$

Formula 1.1.16: Return index post dividends

This assumes that the dividends paid are reinvested into the stock. The reinvested dividends would not realize the same gross return, because the reinvestment would be made at a different price than the initial investment price. However, we disregard this as mentioned in the delimitation section.

Total return index is used instead of the historical stock prices to calculate quarterly returns because larger and mature companies tend to pay out a high proportion of their earnings to shareholders as dividends. Therefore, the historical stock prices would miss a lot of returns over the span of the sample period if this was not corrected for.

Quarterly EPS Measures

The twelve-month forward earnings per share estimate is the consensus earnings forecast for the next twelve-month. The twelve-month trailing earnings per share is based on historical quarterly earnings publicly available from the individual companies.

Daily Three-Month US Treasury Rates

The three months US Treasury rates is given as the midpoint between the offered bid and ask rates. The three-month rate is used because the investment holding period is 3 months and as such to find the equivalent risk-free rate and thus the excess return the three-month US Treasury rate is used as a proxy.

Calculations

The purpose of this section is to describe how the calculations are performed. First the section describes how the computations are used to arrive at the regression input data. It then turns to the forecasting models employed and describes their specific construction. Lastly it covers the measures calculated for evaluation of the final portfolio results.

Daily and quarterly/3-month total returns

The daily total returns are calculated for each of the 27 stocks by using their total return index:

$$total\ return_{t,\ t-1} = \frac{total\ return\ index_t}{total\ return\ index_{t-1}} - 1$$

Formula 1.1.2: Total return for one period

The daily returns are used for computing the volatility over a given quarter. The daily volatility is estimated and then converted to a quarterly volatility. As discussed later, it is not necessary to compute the daily excess returns as the excess volatility is not required for calculating the Sharpe ratio, given the risk-free rate is constant.

The quarterly total return is found by using the same method as with daily total return, but instead using the last total return index observation in the quarter, and the observation of the last day from the previous quarter. This is because it is assumed that positions are entered at closing prices from the last day of the quarter such that the return reflects holding the stock over the entire quarter.

3-, 6-, 9-, and 12-month excess returns

Initially the 6-, 9- and 12-month total returns are found by using the same method as with the 3-month / quarterly total return; by taking the total return index observation the last day in the final quarter and dividing with the observation from the last day in the quarter before then subtracting 1. To get to the excess return the risk-free rate that is calculated for each holding period is subtracted according to the section below:

$$r_t^{excess} = r_t^{total} - r_t^f$$

Formula 1.1.3: Excess return for period t

The excess returns are calculated for all stocks and used in the regressions as lagged variables and the 3-month excess return is also used as a dependent variable.

Risk free rate

The Treasury bill middle rate from Datastream is annualized, so for the purposes of the analysis the annualized rate is converted to a quarterly rate in the following way:

$$r_{quarterly}^{T-bill} = \left(1 + r_{annual}^{T-bill}\right)^{\frac{1}{4}} - 1$$

Formula 1.1.4: Converting the annualized 3-month T-bill rate to a quarterly 3-month rate

Which is then used in the calculations for excess returns.

For the 3-month excess return the closing 3-month Treasury bill middle rate is used from the period before, as this indicates the expected rate which will be realized over the following quarter. For all the other periods a risk-free rate is calculated by assuming that the 3-month Treasury bill returns would be compounded in quarterly intervals. For example, the 9-month excess return would compound the risk-free rate for the 3 quarters that the total return corresponds to.

Volatility

The quarterly volatility is found by taking the sample standard deviation for the daily total returns in each quarter, this is then transformed from daily to quarterly by using the following formula:

$$\sigma_{quarterly} = \sigma_{daily} * \sqrt{trading\ days\ in\ quarter}$$

Formula 1.1.5: Converting the daily volatility to quarterly volatility using the observed trading days

For each quarter the actual number of trading days in the given quarter are counted, these corresponds to the number of observations for a given quarter.

Change in forward and trailing 12-month EPS

The percentage change in the forward and trailing 12-month EPS was both found for each quarter using the following formulas:

$$\delta\ \%\ forward\ EPS_{t,\ t-1} = \frac{forward\ EPS_t}{forward\ EPS_{t-1}} - 1$$

Formula 1.1.6: The percentage change in the forward consensus earnings per share

$$\delta\ \%\ trailing\ EPS_{t,\ t-1} = \frac{trailing\ EPS_t}{trailing\ EPS_{t-1}} - 1$$

Formula 1.1.7: The percentage change in the trailing earnings per share

The earnings change once every quarter, when new earnings are announced by the companies.

Forecasting

By using the approach described for the rolling and expanding window regressions, regressions that can make predictions for each of the 84 quarters from 1997 Q2 – 2018 Q1 for all 27 stocks are computed.

The regressions are computed every quarter to get the most updated estimators, which will then be used for forecasting. For each quarter the regression coefficients are multiplied by the respective historical data from the previous period and the intercept is added to find the expected excess return, volatility or Sharpe ratio for the following quarter.

OLS forecast models

The calculations above are used to compute all the data needed to proceed with computing the coefficients using ordinary least squares linear regressions. The model setup for forecasting the next period's excess return for each stock is as follows:

$$\hat{r}_t^{excess} = \hat{\alpha} + \hat{\beta}_0 \delta\%EPS_{t-1}^{forward} + \hat{\beta}_1 \delta\%EPS_{t-1}^{trailing} + \hat{\beta}_2 r_{t-1}^{3m} + \hat{\beta}_3 r_{t-1}^{6m} + \hat{\beta}_4 r_{t-1}^{9m} + \hat{\beta}_5 r_{t-1}^{12m}$$

Formula 1.1.8: The ordinary least squares regression used in the two component models

The model setup for forecasting the next period's volatility for each stock is as follows:

$$\hat{\sigma}_t^{excess} = \hat{\alpha} + \hat{\beta}_0 \delta\%EPS_{t-1}^{forward} + \hat{\beta}_1 \delta\%EPS_{t-1}^{trailing} + \hat{\beta}_2 r_{t-1}^{3m} + \hat{\beta}_3 r_{t-1}^{6m} + \hat{\beta}_4 r_{t-1}^{9m} + \hat{\beta}_5 r_{t-1}^{12m}$$

Formula 1.1.9: The ordinary least squares regression used in the two component models

The model setup for forecasting the next period's Sharpe ratio for each stock is as follows:

$$\widehat{SR}_t = \hat{\alpha} + \hat{\beta}_0 \delta\%EPS_{t-1}^{forward} + \hat{\beta}_1 \delta\%EPS_{t-1}^{trailing} + \hat{\beta}_2 r_{t-1}^{3m} + \hat{\beta}_3 r_{t-1}^{6m} + \hat{\beta}_4 r_{t-1}^{9m} + \hat{\beta}_5 r_{t-1}^{12m}$$

Formula 1.1.10: The ordinary least squares regression used in the two direct models

To forecast out of sample the ordinary least squares estimators are computed by using estimated lagged explanatory variables to predict the following period's independent variables.

Rolling and expanding window regressions

The estimation of the forecasting models for excess return, volatility and Sharpe ratio are implemented in two separate ways, by using a rolling and an expanding time window for the regressions.

Using a rolling basis for the regression means that a constant sample size of historical observations are used as input in the regression. Specifically it was chosen to include 20 quarters of observations per regression. The first and last observation of the regression sample is moved every single quarter to include one new observation and exclude the oldest one.

Below the original observations are included, where the variables refer to a point in time:

$$[t, t+1, \dots, N]$$

Next quarter the included variables are now shifted:

$$[t+1, t+2, \dots, N+1]$$

And the same is repeated for next quarter and for all future quarters:

$$[t+2, t+3, \dots, N+2]$$

Using a rolling forecast basis for the linear regressions, can be appropriate when the purpose is to make sure that the model is not attributing old observations too much importance over new observations, as new observations may more accurately reflect the current market environment. This is of course at the potential expense of losing longer term trends in the data. As mentioned earlier the model is using 20 quarters of data. The initial regression uses the data from 1991 Q2 to 1997 Q1 and is used to make the first out of sample prediction for the holding period 1997 Q2. In each new quarter the regression is computed again using the new time window which is shifted forward by a quarter.

The expanding basis for regression keeps including more observations for every single quarter that progresses:

$$[t, t+1, \dots, N]$$

$$[t, t+2, \dots, N+1]$$

$$[t, t+2, \dots, N+2]$$

The expanding model has the benefit of being able to include longer term trends or any other insights the older observations could provide, although as mentioned with the rolling basis regression, the older observations might have become irrelevant to current market conditions. As with the rolling window regression the model initially includes 20 quarters of observations, with the first regression including observations from 1991 Q2 to 1997 Q1 to make the first out of sample prediction for the holding period 1997 Q2. In each new quarter the regression is computed again but this time a new quarter of observations is included, thereby expanding the number of observations in the model.

Sharpe ratio

Expected excess returns are divided by the expected volatility to compute the expected Sharpe ratio:

$$\widehat{SR} = \frac{\hat{r}^{excess}}{\hat{\sigma}^{excess}}$$

Formula 1.1.11: Forecasted Sharpe ratio for the Component models

It is assumed that the risk-free rate is constant over the holding period, which simplifies the calculation of the Sharpe ratio, due to the excess return variance being equal the total return variance. It can be shown that:

$$\sqrt{var[R - R_f]} = \sqrt{var[R]}$$

Formula 1.1.12: Total return variance equals excess variance

Realized Sharpe ratios are calculated using the realized volatilities and excess returns.

Ranks

Once all the forecasted Sharpe ratios are computed, it is then possible to rank them. They are ranked in a descending order for each period in all the 84 forecasted quarters. The 7 stocks with the lowest expected Sharpe ratios are excluded from the portfolios, in other words the portfolios consist of 20 stocks.

Portfolio analysis

The approach described below is used for both the Dow Equally Weight portfolio and the four model portfolios, where the only difference is the exclusion of the 7 stocks with the worst predicted Sharpe ratios for each individual model.

The portfolio are constructed by weighting the stocks equally, such that the quarterly return for the portfolio equals the average quarterly return for the selected stocks.

Excess return is found by subtracting the treasury bill rate at the beginning of the period from the realized gross return.

The volatility for each quarter is found by computing the standard deviation of the daily returns for the portfolio during the quarter and transforming it to a quarterly measure. The daily return is calculated by assuming rebalancing on a daily basis, in effect taking the average of the daily returns for the stocks included in the portfolio for the quarter.

Sharpe ratio is computed as the realized excess return divided by the realized volatility.

The high water mark and draw down is also computed for each portfolio. The high water mark is the highest cumulative return and the draw down is the percentage deviation of the current cumulative return from the high water mark.

In order to evaluate the relative performance of the portfolios a range of tests are used. The methodology regarding the Diebold Mariano test is described in-depth in the next section since it is required calculated without the use of statistical software.

Diebold Mariano

In order to test if there is a significant difference in the accuracy between the rolling, expanding, component, and direct forecast models, a Diebold Mariano test statistic is used.

The test statistic for each individual stock is computed, and compared for the four forecast models, that is; Component Expanding, Component Rolling, Direct Expanding, and Direct Rolling. The test compares the models in pairs.

Initially the error terms for the two models are calculated as the difference between the forecasted Sharpe ratio and the actual Sharpe ratio:

$$\epsilon_t^{SR} = Forecasted_t^{SR} - Actual_t^{SR}$$

Formula 1.1.13: The error terms for the models

The error terms are used in a squared loss function, such that they indicate a distance from the actual value, regardless of it being positive or negative, while also punishing observations that are further away from the actual value progressively more:

$$L(\epsilon_t) = (\epsilon_t)^2$$

Formula 1.1.14: The squared error loss function for the models

Then d is defined as the difference between the error terms for each period:

$$d_t = L(\epsilon_t^1) - L(\epsilon_t^2)$$

Formula 1.1.15: The difference between the squared error loss function

The mean and sample variance of the d values are then computed and used in the test statistic:

$$\bar{d} = \frac{1}{T}\sum_{t=1}^{T} d_t$$

Formula 1.1.16: Average squared error loss function difference

Sample variance:

$$\sigma_{\bar{d}}^2 = \frac{1}{T-1}\sum_{t=1}^{T}(d_t - \bar{d})^2$$

Formula 1.1.17: Sample variance of the squared error loss function difference

The Diebold Mariano test statistic simplifies to the following formula using the model setup outlined in this thesis, the reason for this is discussed in the theory section:

$$DM\ test = \frac{\bar{d}}{\sqrt{\frac{\sigma_d^2}{T}}} \rightarrow N(0,1)$$

Formula 1.1.18: The Diebold-Mariano test statistic

The p value, that is the probability of observing a more extreme value, can then be calculated using this standard normal test statistic. The two-sided test is defined as:

$$Probability\ value = 2 * P(X > |DM\ test|)$$

Formula 1.1.19: The Diebold-Mariano two-sided test p-value

Bias

Since the strategy uses the constituents of the Dow Jones Industrial Average as of 2018, there is a selection bias typically called survivorship bias in terms of picking the winners that end up being included in the index. This information would not have been available over the period that the portfolios are simulated and leads to selecting the winners, which inherently guarantees good return results. However, since the purpose is to outperform an equally weighted portfolio of these same winner stocks and not providing a high total return, it is a relative target, which does not compromise the analysis and the results.

Data mining is another type of bias where data is picked due to its characteristics such that the results are overfitted. This would lead to results and conclusions that are not going to be consistent because the data does not reflect the choice that would have been made ex ante. Our selection of data is based around having as many of the current Dow Jones Industrial Index constituents included, while only including data, for which every company has observations for every period. It can be discussed if this is an objective choice with regards to data mining, however again since the hypothesis and problem statement is a relative target made up from the same data, this is not a major issue. The sample period includes multiple economic cycles, such that the models are exposed to markets with very favorable and not favorable conditions.

Theory

The theory section explains the theoretical applications and assumptions of the different models, calculations, and tests used in the methodology of the thesis.

Ordinary least squares regression

Linear ordinary least squares regression is a statistical model that describes the relationship between a dependent variable and one or multiple explanatory variables, known as the independent variables, as a linear function. For a simple linear regression with one explanatory variable the population function would be:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, i = 1, \ldots, N$$

Formula 2.1.1: Simple linear regression

Where the subscript, i, refers to the index of the specific observation. The value of Y depends on the value X in a linear connection. The error term ϵ is defined as the difference between the value of an observation and the average population of the population.

The expected value of Y when X equals 0 is defined by the intercept β_0. The term β_1 indicates the effect of a change by 1 in the explanatory variable X on Y. The terms β_0, β_1 are found mathematically by solving a minimization problem, such that predicted values provide the best possible linear fit with the observed values. The model can also be extended to multiple explanatory variables:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \epsilon_i, i = 1, \ldots, N$$

Formula 2.1.2: Multiple linear regression

Where k refers to the k'th term.

The multiple regression model coefficients, $\beta_0, \beta_1, \ldots, \beta_k$, for the input variables are interpreted as the change in Y when an X term changes by 1 while holding all other inputs constant.

To find the ordinary least squares estimators for the coefficients, the estimation is made by minimizing the predicted squared error terms.

The error terms can be summed, SSE, to get a sense of how accurately the model predicts:

$$SSE = \sum_{i=1}^{N} (\hat{\epsilon}_i)^2$$

$$= \sum_{i=1}^{N} \left(Y_i - \hat{Y}_i\right)^2$$

$$= \sum_{i=1}^{N} \left(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki}\right)^2$$

Formula 2.1.3: Sum of the squared errors

The coefficients of the function are chosen such that the sum of the squared error terms, SSE, are minimized for the observations with which the regression is computed:

$$Min \sum_{i=1}^{N} \left(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki}\right)^2$$

Formula 2.1.4: Minimization objective of the ordinary least squares regression

The coefficients are referred to as the OLS estimators and they are found by using calculus and linear algebra. Once estimated, these values define the ordinary least squares regression line, which can be used to make predictions (Stock & Watson, 2012):

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \cdots + \hat{\beta}_k X_{ki}$$

Formula 2.1.5: Multiple linear regression

OLS assumptions for multiple regression model

Under the OLS assumptions the estimators $\hat{\beta}_0 + \hat{\beta}_1 + \cdots + \hat{\beta}_k$ are unbiased, consistent and normally distributed in large samples. Below the OLS four assumptions are listed.

Assumption 1: The conditional distribution of the error terms given the input variables $X_{1i} \ldots X_{ki}$ has a mean of zero. This makes it such that on average the computed Y values will be on the regression line, as the ones lying above the line will balance the ones below.

Assumption 2: The input variables $X_{1i} \dots X_{ki}$, are independently and identically distributed random variables. This assumption is true when random sampling is used to obtain the data.

Assumption 3: It is unlikely that there are large outliers. This is because the OLS estimation is sensitive to large outliers when computing the estimators. Since the squared error terms are used, larger differences between the actual and estimated Y value contribute much more than smaller differences.

Assumption 4: There exists no perfect multicollinearity. Whenever there is perfect multicollinearity between the explanatory variables it is not possible to compute the OLS estimators as it ultimately leads to division by 0. Perfect multicollinearity is present whenever one of the explanatory variables is described by a perfect linear function of the one of other explanatory variables.

Measures of fit

Various summary statistics regarding the regression is usually computed by software packages.

To test the null hypothesis that the OLS estimator is zero against the alternative hypothesis that it is different from zero, a p-value is computed. If the p-value is small, the null hypothesis is rejected as the estimator's true value is significantly different from zero. The specific p-value required for rejecting the null hypothesis is dependent on the chosen significance level. As noted by Ruppert (2011):

"It is important to keep in mind that the p-value only tells us if there is a linear relationship" (Ruppert, 2011)

An estimate of the standard deviation of the error term is called the standard error of the regression (SER). It measures how Y is distributed around the regression line. For a multiple regression model, the standard error of regression can be computed as:

$$Standard\ error\ of\ regression\ (SER) = s_{\hat{\epsilon}}\ where\ s_{\hat{\epsilon}}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}\hat{\epsilon}_i^2 = \frac{SSR}{n-k-1}$$

Formula 2.1.6: Standard error of regression

R squared is an expression of the portion of the sample variance explained by the explanatory variables:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS}$$

Formula 2.1.7: R squared

Where:

$$Explained\ sum\ of\ squares\ (ESS) = \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2$$

$$Total\ sum\ of\ squares\ (TSS) = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

A potential problem with the R squared measure is that it never decreases and potentially increases whenever a new explanatory variable is added to the regression, which means adding more variables will improve R squared. This is because the ordinary least squares minimization method will find the same solution as before the new variable was introduced if the new variable does not contribute anything, or a better solution if possible.

The measure adjusted R squared takes this into account and does not increase by simply adding more explanatory variables:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1} * \frac{SSR}{TSS} = 1 - \frac{s_{\hat{\epsilon}}^2}{s_Y^2}$$

Formula 2.1.8: Adjusted R squared

Hypothesis testing

To test if the different OLS linear regression models produce statistically different results, a two-population hypothesis test can be used (Newbold, Carlson, & Thorne, 2013). The statistical interpretation of a two-sided null hypothesis is expressed as the two means being:

$$H_0 : \bar{x}_1 = \bar{x}_2$$

With the alternative hypothesis is expressed as the two means being unequal:

$$H_1 : \bar{x}_1 \neq \bar{x}_2$$

The hypotheses must be expressed such that the two outcomes cover all possible outcomes, otherwise it is possible to end up with an inconclusive analysis. The use of a one-sided or two-sided hypothesis depends on the intent. The one-sided test can be used to test if the forecast model produces results that are statistically lower or higher than the baseline model, whereas the two-sided test can be used to test if there is a statistical absolute difference in the results between two of the forecasts.

The statistical results cannot be interpreted as absolute proof of one of the results being correct, however the test can conclude that one of the two alternatives have a statistically small probability of being correct. Because of this the terms reject and fail to reject are used about the null hypothesis when the tests are conducted. The method is the fundamental basis of decision making in scientific research.

Type I and Type II errors and confidence level

Using the terminology reject and fail to reject is more correct given that when the test fails to reject the null hypothesis it either means that it is correct or that the test and input data are not sufficient to reject it and as a result the test is wrong. Given the sample mean is likely not the true population mean, there is a statistical possibility of coming to a false conclusion.

The two different types of errors the test can commit are classified as Type I and Type II errors. Type I error is rejecting the null hypothesis, when it is in fact true. Type II error is when the test fails to reject a false null hypothesis.

Type I & Type II matrix		
Decisions on the null hypothesis	Null hypothesis is true	Null hypothesis is false
Fail to reject the null hypothesis	Correct decision Probability = 1 – α	**Type II error** **Probability = β**
Reject the null hypothesis	**Type I error** **Probability = α**	Correct decision Probability = 1 - β

Figure 1.2.3: Type I and Type II error matrix (Newbold, Carlson, & Thorne, 2013)

As Newbold, Carlson and Thorne (2013) writes, the decision rule is chosen such that the probability of a type I error is small, which translates to picking a small alpha also called the significance level of the test. The probability of committing a type II error is found as β, whereas the inverse (1 – β) is referred to as the power of the test. The significance level of the test is chosen and as such the probability of committing a type I error and the probability of committing a type II error is a function of the chosen significance level. The lower the significance level is the lower the probability of committing a type I error is while the opposite effect is true for a type II error, which will increase in probability at lower significance level.

Test for equal variance

The test for equal population variances between two independent samples is conducted to decide, which test to use, when testing for equal means. The test uses a F distribution:

$$F = \frac{s_x^2/\sigma_x^2}{s_y^2/\sigma_y^2}$$

Formula 2.1.9: F-test for test of equal variance (Newbold, Carlson, & Thorne, 2013)

Where s_x^2 is the sample variance for a random sample of n_x observations from a normally distributed population with the variance σ_x^2. The s_y^2 is the sample variance for a random sample of n_y observations again from a normally distributed population with the variance σ_y^2. The distribution relates the population and sample variance for a normally distributed population. Using hypothesis tests depending on the F distribution are very dependent on the assumption of normality and as such the results will not be interpretable if the assumption is broken.

To test for equal variance the F-test is defined as:

$$F = \frac{s_x^2}{s_y^2}$$

Formula 2.1.10: F-test for test of equal variance assuming equal population variance (Newbold, Carlson, & Thorne, 2013)

Given that if the population variances are equal they cancel each other in formula 2.1.10 above. Thus, the test looks at the relative difference between the sample variance of x against the sample variance of y and if these statistically significantly differ from each other the population variances are not equal.

The null hypothesis and alternative hypothesis are as follows:

$$H_0: \sigma_x^2 = \sigma_y^2$$

$$H_1: \sigma_x^2 \neq \sigma_y^2$$

The decision rule to reject the null hypothesis is as follows:

$$F = \frac{s_x^2}{s_y^2} > F_{n_x-1, n_y-1, \frac{\alpha}{2}}$$

Formula 2.1.11: F-test critical value (Newbold, Carlson, & Thorne, 2013)

Where s_x^2 is the larger of the two sample variances, such that only the upper cutoff points are needed to test the hypothesis of equality in variances.

$F_{n_x-1, n_y-1, \frac{\alpha}{2}}$ is the number for which $P\left(F_{n_x-1, n_y-1} > F_{n_x-1, n_y-1, \alpha}\right) = \alpha$.

Analysis of Variance

Where the two-population hypothesis test can test the difference between two populations, the analysis of variance test can test more than two at once. The method specifically is used to test for the difference in means between K populations, where all of them are assumed to have the same variance. The framework of the one-way analysis of variance is a hypothesis test, where the null hypothesis is that all means of the K populations are the same:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_K$$
$$H_1: \mu_i \neq \mu_j; For\ at\ least\ one\ pair\ of\ \mu_i, \mu_j$$

The test for equal means in the populations is based on two different measures of variability exhibited by the different samples.

- SSW is the sum of the within groups variability, which is the sum of all the squared errors within the individual samples.

$$SSW = SS_1 + SS_2 + \cdots + SS_K; where\ SS_1 = \sum_{j=1}^{n_1} \left(x_{1j} - \bar{x}_1\right)^2$$

Formula 2.1.12: Sum of the within groups variability (Newbold, Carlson, & Thorne, 2013)

- SSG is the sum of the between group sum of squares, where the squared error between the mean of each individual sample and the overall mean for all sample observations is calculated. The overall mean is calculated as

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{K} \sum_{j=1}^{n_i} x_{ij}}{n}$$

Formula 2.1.13: The overall mean of the ANOVA samples (Newbold, Carlson, & Thorne, 2013)

The squared error component for each sample mean is weighted with its corresponding number of observations.

$$SSG = \sum_{i=1}^{K} n_i(\bar{x}_i - \bar{\bar{x}})^2$$

Formula 2.1.14: Sum of the between groups variability (Newbold, Carlson, & Thorne, 2013)

The test of the equal means in the populations is based on the assumption that the K populations have a common variance, because then SSW and SSG can each be used as the basis for an unbiased estimator of the population variance. To obtain the estimates the two sum of squares have to be divided by the relevant number of degrees of freedom.

The MSW within groups mean square is an unbiased estimator of the variance of the population:

$$MSW = \frac{SSW}{n - K}$$

Formula 2.1.15: Within groups mean square (Newbold, Carlson, & Thorne, 2013)

And MSG between groups mean square is another unbiased estimator of the population variance:

$$MSG = \frac{SSG}{K - 1}$$

Formula 2.1.16: Between groups mean square (Newbold, Carlson, & Thorne, 2013)

If the population means are not equal, then the between groups mean square is not an unbiased estimate of the population variance. If the null hypothesis is true, then the two different estimates are both unbiased estimates of the population variance, as such it would be expected that they are close to one another in terms of their value. The difference between the two estimates is thus the basis for the test on the null hypothesis and is formulated as an F test:

$$F = \frac{MSG}{MSW}$$

Formula 2.1.17: ANOVA F-test (Newbold, Carlson, & Thorne, 2013)

The decision rule regarding the test is based on the chosen significance level α.

$$Reject\ H_0\ if \frac{MSG}{MSW} > F_{K-1,n-K,\alpha}$$

Where $F_{K-1,n-K,\alpha}$ is the number for which:

$$P\left(F_{K-1,n-K} > F_{K-1,n-K,\alpha}\right) = \alpha$$

Formula 2.1.18: ANOVA F-test critical value (Newbold, Carlson, & Thorne, 2013)

Assumptions for the test are equal variance in the populations as mentioned and that the populations are normally distributed.

Diebold-Mariano

The Diebold-Mariano test is used to test if two different forecasting models have equal predictive accuracy. If $y^1_{t+h|t}$ and $y^2_{t+h|t}$ denote two different forecasts of y_{t+h} then the forecasted errors from the two models are calculated as:

$$\epsilon^1_{t+h|t} = y_{t+h} - y^1_{t+h|t}$$

$$\epsilon^2_{t+h|t} = y_{t+h} - y^2_{t+h|t}$$

Formula 2.1.19: Forecast model error terms

Here *h* denotes the number of steps or periods forecasted ahead from t, and t is the relevant starting period. In this thesis *h* will be 1 as the models are forecasting the following quarter at all times. The errors quantify how far the model prediction is from realized observations.

The errors are then used in a loss function. This is to get a sense of the cost or impact of the errors. A popular loss function is the squared loss function, which increases the size of the loss progressively as the errors diverge from zero either in the positive or negative direction:

$$L\left(\epsilon^i_{t+h|t}\right) = \left(\epsilon^i_{t+h|t}\right)^2$$

Formula 2.1.20: Squared loss function

There are other loss functions available, another popular option is the absolute loss function.

To determine whether if one of the forecasting models relatively outperforms the other in terms of predictive accuracy, a null hypothesis can be tested. The hypothesis tests if there is a significant difference in the expected loss function of the errors between the two different forecasting models, so the hypothesis specification is as follows:

$$d_{t+h|t} = L\left(\epsilon^1_{t+h|t}\right) - L\left(\epsilon^2_{t+h|t}\right)$$

$$H_0: E\left[d_{t+h|t}\right] = 0$$

$$H_1: E\left[d_{t+h|t}\right] \neq 0$$

The test is based on the difference in losses of the two models and that the two models are equal in accuracy.

$$DM\ test = \frac{\bar{d}}{\sqrt{\frac{2\pi\hat{f}_d(0)}{T}}} \rightarrow N(0,1)$$

$$\hat{f}_d(0) = \frac{1}{2\pi} \sum_{\tau=-(T-1)}^{T-1} l(\frac{\tau}{h-1})\,\hat{\gamma}(\tau)$$

$$l\left(\frac{\tau}{h-1}\right) = \begin{cases} 1\ for\ \left|\frac{\tau}{h-1}\right| \leq 1 \\ 0\ otherwise \end{cases}$$

$$\hat{\gamma}(\tau) = \frac{1}{T} \sum_{t=|\tau|+1}^{T} (d_t - \bar{d})(d_{t-|\tau|} - \bar{d})$$

Formula 2.1.21: Diebold-Mariano test and components

It can then be shown that:

$$l\left(\frac{\tau}{h-1}\right) = 0\ when\ |\tau| > h-1$$

When $h = 1$, that is when only forecasting the following period, the Diebold Mariano test does not need a spectral density adjustment for the autocorrelation between the error terms. The only time to account for the γ value is when $\tau = 0$ since all other values will result in the absolute value of τ being larger than $h - 1 = 1 - 1 = 0$ which then results in $l\left(\frac{\tau}{h-1}\right) = 0$.

A simplification of the second formula found with formula 2.1.21 under the $h = 1$ assumption:

$$\hat{f}_d(0) = \frac{1}{2\pi}\hat{\gamma}(0)$$

Formula 2.1.22: Simplified component of Diebold-Mariano test

When substituting, $\tau = 0$, it can be shown that $\hat{\gamma}(0)$ is just the variance of d:

$$\hat{\gamma}(0) = \frac{1}{T} \sum_{t=|0|+1}^{T} (d_t - \bar{d})(d_{t-|0|} - \bar{d})$$

$$\hat{\gamma}(0) = \frac{1}{T}\sum_{t=1}^{T}(d_t - \bar{d})(d_t - \bar{d}) = \frac{1}{T}\sum_{t=1}^{T}(d_t - \bar{d})^2 = \sigma_d^2$$

This finding is used to simplify the first formula with formula 2.1.21 for the Diebold Mariano test (Diebold & Mariano, 1995):

$$DM\ test = \frac{\bar{d}}{\sqrt{\frac{\frac{2\pi}{2\pi}\hat{\gamma}(0)}{T}}} = \frac{\bar{d}}{\sqrt{\frac{\frac{2\pi}{2\pi}\sigma_{\bar{d}}^2}{T}}} = \frac{\bar{d}}{\sqrt{\frac{\sigma_{\bar{d}}^2}{T}}}$$

Formula 2.1.23: Simplified Diebold-Mariano test

Which is used later in the analysis to compute the Diebold Mariano test statistic.

Performance metrics

In this section some of the measures relevant to portfolio evaluation are examined. This is such that the theoretical framework of the metrics are introduced already, when used to evaluate the portfolios of the different forecasting models. The first section is about returns which are the most essential metric, followed by volatility which describes the variation of the returns from the mean. The third introduced metric; Sharpe ratio relates the return to the risk taken. The high water mark is the highest cumulative return achieved at any given point, and from the high water mark the draw down can be calculated to get a sense of the downward risk faced by the strategy. Lastly the risk metrics; value at risk and expected shortfall are introduced.

Returns

There are many ways to compute returns; when choosing the level of expenses included such as gross, net, excess, or cumulative, and there are choices regarding the assumption about when returns are compounded. What the returns all have in common is that they communicate how much value has been gained or lost over a certain period (Bodie, Marcus & Kane, 2014). When evaluating performance, it is essential to examine the economic value gained or lost such that realistic expectations for future performance can be grounded in data before determining on forward looking investment actions.

- Gross returns are the returns earned on an investment in a period. This includes changes in price and potential dividends.
- Net returns are the realized returns after all expenses are paid, in other words gross returns with trading and financing costs subtracted.
- Excess returns are adjusted for the opportunity cost of investing in a risk-free asset. It should be positive for a trading strategy to be attractive, such that the trading strategy is expected to outperform a risk-free asset.
- Cumulative return looks at the total return earned over a given period, which provides insight into the total expected gains or losses. A variation of this is called the high water mark.
- The arithmetic average of returns provides an estimate for the returns of future holding periods. The measure only concerns itself with the distribution of returns and not with the effects of compounding.

All returns can also be given either in nominal or real terms, which are adjusted for inflation or purchasing power. All the numbers in this thesis are nominal and need no adjustments. Note that it is assumed that there are no trading costs in this thesis and in that case, gross and net returns are equal. Since the model is rebalanced quarterly it has discretely compounded returns.

Volatility

Volatility, also known as standard deviation, expresses how much the returns are expected to vary, specifically away from the mean. It is considered a very central measure of risk within finance, since it provides insight into the range of expected outcomes. It is the square root of the variance, which in turn is the expected squared difference from the mean (adjusted for degrees of freedom):

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(r_i - \bar{r})^2}$$

Formula 2.1.24: Sample variance

If returns are assumed to be normally distributed, then 90% of the possible returns are within a range of plus or minus 1.645 standard deviations from the mean. The mean and standard deviation are the only characteristics that need to be known about the return distribution. Since the standard deviation does not differentiate between positive and negative deviations from the mean, the later sections on draw down, value at risk, and expected shortfall are about measures that are concerned with losses only. (Bodie, Marcus & Kane, 2014)

Sharpe ratio

The Sharpe ratio is a measure for the risk adjusted returns. The excess returns are compared to the risk of the position to provide a measure of excess return per unit of risk. The risk is the standard deviation of the excess return. The Sharpe ratio is calculated as:

$$SR_t = \frac{r_t^{excess}}{\sigma_t^{excess}}$$

Formula 2.1.25: Sharpe ratio

Considering rational investors expect to be compensated for the risk they take, the Sharpe ratio is a valuable measure to compare different strategies.

High water mark

The high water mark is the highest cumulative return that has been achieved in the past and is calculated as:

$$HWM_t = \max_{s \leq t} P_s$$

Formula 2.1.26: High water mark

Where P_t is the cumulative return at time t.

Portfolio managers use the measure to see the peak cumulative returns of their investments and it is used to calculate the draw down for the investment (Pedersen, 2015).

Draw down

The drawdown measures the cumulative loss since the portfolio's losses started. The percentage drawdown from the high water mark is given by:

$$Drawdown = \frac{HWM_t - P_t}{HWM_t}$$

Formula 2.1.27: Draw down

The drawdown lets the portfolio manager see how the portfolio is currently performing compared to the high water mark. The drawdown reveals the magnitude of the realized losses that the portfolio suffers. The most common use is such that investors are not paying a performance fee on returns that are below the current high water mark, such that when a portfolio drops in value the investor will not pay performance fees before the portfolio breaches its prior peak (Pedersen, 2015).

Value at Risk

The value at risk (VaR) is the loss that corresponds to a certain percentile of the return distribution. The percentile of the return distribution is called α and the level of α can be chosen by the user. The value at risk is the value at which α% of the possible values are below and is calculated as:

$$VaR(\alpha) = \mu - Z(\alpha)\sigma$$

Formula 2.1.28: Value at risk for chosen significance level

The most common level to measure at is 5%, meaning that 95% of returns will be above the estimated value at risk and 5% will be below. VaR can be interpreted as the best return possible out of the worst 5% scenarios or as the worst loss with 95% certainty (Bodie, Marcus & Kane, 2014).

Expected Shortfall

Whereas the value at risk takes the most optimistic return of the worst case scenarios, expected shortfall looks at the average return conditional on only looking at the worst α% values in the distribution. Expected shortfall is a more realistic measure of worst case expectations because the worst α% values may be significantly worse than the exact α% value (Bodie, Marcus & Kane, 2014)

Analysis

The analysis is broken into two main sections; the first section concerns the forecasting models, the second section concerns the results from implementing the trading strategies. All results in the analysis section are specified on a quarterly basis, unless otherwise stated.

The first section examines whether the explanatory variables used by the models across all 27 stocks are statistically significant and to what degree the models can explain the variation in the dependent variable for the following period. This is done using the full data set available to explore if the chosen estimators provide a general explanatory power that could potentially extend to the analysis of the four shorter period forecast models used for the implementation of the trading strategies. To gauge how correct the forecasted Sharpe ratios are, a single variable ordinary least squares regression is performed with the realized Sharpe ratio as the dependent variable and the forecasted Sharpe ratio as the independent variable for each of the four models. The relative accuracy, as measured by the squared error terms, are then examined and tested to offer insights into which specific model offers the most accurate forecast.

The second section starts with a description of the results obtained from implementing the trading strategies. This is focused on the key portfolio parameters; excess return, volatility, Sharpe ratio, as well as additional risk statistics. To get a sense of the overall accuracy of the final strategy regardless of the forecast accuracy, the number of misclassifications made when ranking the stocks for portfolio selection is examined. Lastly the key portfolio parameters are tested for statistical differences from the equally weighted Dow benchmark portfolio and each other. This is done using ANOVA and hypothesis testing.

Choosing the significance level

When selecting the appropriate significance level, several things needs to be taken into account. First, considering the thesis takes an explorative approach with regards to the problem statement the significance level needs to be chosen such that the tests do not fail to reject the null hypothesis if the alternative hypothesis is true and commit a type II error. On the other hand, the tests should also not falsely reject the null hypothesis, when it is in fact true and commit a type I error. Thus, weighting the consequences of the different types of errors, it was chosen to use a significance level of 10%, such that the tests may be rejecting the null hypothesis more often than a more conservative level of significance level would do. From this decision it follows that the tests are also more likely to find statistically significant relations at the cost of accepting more type I errors.

Forecasting models

Before constructing the trading strategies and resulting portfolios, the validity of the forecasting models is tested. In the following section it is tested if the explanatory variables for each forecasting model have had statistically significant prediction power over the period of the full data set. The ordinary least squares assumption regarding multicolinearity is also examined. To test the relative predictive power of the models the Diebold Mariano test is used to compare the squared error terms.

Testing chosen coefficients for statistical significance

To see if the chosen coefficients have any statistical significant explanatory power on the three different dependent variables; Sharpe ratio, excess returns, and volatility, multivariate regressions on all 27 stocks are computed with the following regression where all r are the excess return:

$$\widehat{SR}_t = \hat{\alpha} + \hat{\beta}_0 \delta EPS_t^{forward} + \hat{\beta}_1 \delta EPS_t^{trailing} + \hat{\beta}_2 r_t^{3m} + \hat{\beta}_3 r_t^{6m} + \hat{\beta}_4 r_t^{9m} + \hat{\beta}_5 r_t^{12m}$$

Formula 3.1.1: Regression of selected coefficients on the actual Sharpe ratio at time t

$$\hat{r}_t^{excess} = \hat{\alpha} + \hat{\beta}_0 \delta EPS_t^{forward} + \hat{\beta}_1 \delta EPS_t^{trailing} + \hat{\beta}_2 r_t^{3m} + \hat{\beta}_3 r_t^{6m} + \hat{\beta}_4 r_t^{9m} + \hat{\beta}_5 r_t^{12m}$$

Formula 3.1.2: Regression of selected coefficients on the excess return at time t

$$\hat{\sigma}_t^{excess} = \hat{\alpha} + \hat{\beta}_0 \delta EPS_t^{forward} + \hat{\beta}_1 \delta EPS_t^{trailing} + \hat{\beta}_2 r_t^{3m} + \hat{\beta}_3 r_t^{6m} + \hat{\beta}_4 r_t^{9m} + \hat{\beta}_5 r_t^{12m}$$

Formula 3.1.3: Regression of selected coefficients on the volatility at time t

The regressions are computed for all stocks on the full data set available, such that it can be analyzed if the last periods change in the 12-months forward and trailing EPS, and the 3-, 6-, 9- and 12-month lagged excess returns have any prediction power on the current periods performance in the form of Sharpe ratio, excess returns, and volatility. The regressions include 104 observations, equal to the 104 quarters from the second quarter of 1992 to the first quarter of 2018.

Summary statistics	Sharpe ratio	Excess Return	Volatility
Average Adjusted R^2	-0.003	0.012	0.079
# F-significance value below 0.05	0	3	13
# F-significance value below 0.10	1	4	13

Table 3.3.1: Summary statistics of all 27 stock regressions

Examining the summary statistics for the regression on all the 27 stocks in table 3.3.1 the average adjusted R^2 values indicate that the chosen coefficients are not very good at predicting the following period's Sharpe ratio, excess return or volatility. The coefficients even have a negative explanatory power on the Sharpe ratio on average across the 27 stocks. For the excess returns, the coefficients are on average able to explain 1.2% of the variation of the realized excess returns, which is still not good considering the model is used to forecast. However, given the unpredictable nature of the analyzed data, it cannot be expected to achieve an impressive adjusted R^2, because of the implications if everyone could easily create a model with a very significant explanation power of future returns.

The p-value of the F-test indicates if the regression as a whole is statistically significant or not. When looking at the Sharpe ratio only 1 regression is statistically significant at a significance level of 10%. The volatility regressions have the best significance rate, where 13 of the 27 stock regressions are significant at even the 5% level.

	Sharpe ratio # p-value below		Excess Return # p-value below		Volatility # p-value below	
	0.05	0.10	0.05	0.10	0.05	0.10
Intercept	25	27	18	21	27	27
Delta 12MTH FORWARD EPS	0	4	2	4	0	2
Delta 12MTH TRAILING EPS	2	3	1	2	5	6
3 month	0	3	2	4	0	2
6 month	0	0	1	4	0	0
9 month	1	3	1	2	0	1
12 month	0	2	2	3	1	2

Table 3.3.2: Summary statistics of all 27 stock regressions for estimator significance

Examining the p-values of the individual coefficients in table 3.3.2 reveals the same message as looking at the F-significance level. None of the coefficients are close to being significant for the majority of the regressions, which means that their explanatory power in the model generally is poor. The statistics are only observable ex post and no test of the ordinary least squares assumptions were performed before computing the models, besides using the literature to find estimators with a possibility of having significant explanatory power. This was done to avoid selection and overfitting bias, when choosing the data set and explanatory variables. Given the poor significance results, the models and the independent variables are analyzed in depth to understand the causes.

Multicollinearity complications

The four lagged excess return variables are by definition expected to carry some semblance of multicollinearity, given that the lagged 3-month excess return is an explicit part of both the lagged 6-, 9- and 12-month excess returns. The data does not fully overlap and as such it would be assumed that the multicollinearity is not perfect and as such is not a modelling issue with regards to the model assumptions.

Average correlation between variables	3-month excess return	6-month excess return	9-month excess return	12-month excess return	12-month trailing EPS	12-month forward EPS
3-month excess return	1.00					
6-month excess return	0.67	1.00				
9-month excess return	0.54	0.78	1.00			
12-month excess return	0.47	0.66	0.83	1.00		
12-month trailing EPS	0.12	0.22	0.30	0.35	1.00	
12-month forward EPS	0.27	0.41	0.47	0.46	0.58	1.00

Table 3.3.3: Average correlation matrix of regression explanatory variables for all 27 stocks

The correlations in table 3.3.3 between the independent variables indicates that on average there is a strong positive correlation. The correlation is strongest between the sequential time lagged variables, which coincidentally is also the pairs that have the most overlapping number of observations. The highest correlated pair on average for all the stocks is the 9-month and 12-month excess return. It could potentially be an indication that the model may have an equal or better fit when excluding one of these, which could be tested. However, doing so could be considered overfitting since the correlations in table 3.3.3 include all observations. The imperfect multicollinearity does not affect the assumptions and the theory behind the ordinary least squares regression, according to Newbold, Carlson, and Thorne (2013) there are derivative consequences for the variable estimators in the regression due to the highly correlated variables such as:

- The ordinary least squares estimators are unbiased but are affected by a larger sampling variance and covariance that makes it difficult to precisely estimate the partial effects (Watson & Stock, 2011).

However, the partial effect does not have to be precise, given that the important result is the full model Sharpe ratio estimate and not necessarily how the 9-month return affects the Sharpe ratio individually. The partial effects are used to explain why and how the model is able to predict the way it does. Knowing that the estimators' partial effects are not precise, makes it hard to know, which estimator effects are the cause of the model making incorrect estimations, given that the information is disguised by the imperfect multicollinearity.

- The larger than usual variance contributes to larger confidence intervals and as a result there is a larger possibility of the estimators to be statistically insignificant due to the fact that the confidence level for a given alpha level is more likely to contain 0 as a value. This in turn makes it more likely to find insignificant relations.

Drawing conclusions from such small number of significant variables is difficult and especially considering the miniscule differences in how many times each coefficient is significant. The 6-month coefficient is never significant at a 10% level when predicting the Sharpe ratio. This indicates that either the variation in Sharpe ratio that the 6-month coefficient can explain is already fully explained by the other lagged coefficients or it contains no information that can be used in a linear function to explain the Sharpe ratio. For both Sharpe ratio and excess return there is only a marginal difference in how many times the EPS and lagged excess return coefficients are significant. As such it cannot be shown that the change in earnings or the change in excess return is better suited for predicting the following quarter.

The insignificant estimators are expected, given that there is a very strong positive correlation between most of the variables, which increases the variances and as such makes it more likely for the individual estimator to be insignificant. Chan, Jegadeesh, and Lakonishok (1996) found that the earnings effect is more short term minded, given that it exploits the under reaction to information in the market on the short term after the earnings announcement. The changes in return on the other hand exploits the markets under reaction to a broader range of information like changes to the expected long-term profitability not visible in the accounting numbers yet. This finding is consistent with that of Moskowitz, Ooi, and Pedersen (2011) who find that:

"Our finding of positive time series momentum that partially reverse over the long-term may be consistent with initial under-reaction and delayed over-reaction, which theories of sentiment suggest can produce these return patterns."

Due to the imperfect multicollinearity, the partial effect of each estimator is not as accurate as could be due to increased estimator volatility.

Earnings per share as estimators

According to the current SEC regulations, amended in 2005, public companies with a float of more than \$700m have 40 days to publicly announce their quarterly earnings pending from the end date of a given quarter (SEC, 2018). Furthermore, the companies may choose to start their fiscal year in any month. For example, Apple Inc. chooses their fiscal year such that the year ends on the last Saturday in September (Apple, 2018). All of these company specific fiscal years may even have been different throughout the sample period. The models intent was to use the release of new quarterly earnings as a proxy for short term economic changes in the company, which the market has not fully reacted to yet.

Sharpe ratio significant EPS coefficient (α = 10%)						
Stock	3M	BOEING	GENERAL ELECTRIC	JOHNSON & JOHNSON	MICROSOFT	TRAVELERS COS.
EPS trailing coefficient p-value	0.22	0.96	0.06	0.19	0.06	0.06
EPS forward coefficient p-value	0.05	0.09	0.16	0.02	0.42	0.34
Fiscal year matches trading window	Yes	Yes	Yes	Yes	Yes	Yes

Excess return significant EPS coefficient (α = 10%)				
Stock	3M	GENERAL ELECTRIC	INTERNATIONAL BUS.MCHS.	TRAVELERS COS.
EPS trailing coefficient p-value	0.06	0.02	0.06	0.00
EPS forward coefficient p-value	0.02	0.09	0.69	0.12
Fiscal year matches trading window	Yes	Yes	Yes	Yes

Volatility significant EPS coefficient (α = 10%)								
Stock	AMERICAN EXPRESS	JOHNSON & JOHNSON	JP MORGAN CHASE & CO.	MERCK & COMPANY	PFIZER	PROCTER & GAMBLE	TRAVELERS COS.	WALMART
EPS trailing coefficient p-value	0.06	0.98	0.11	0.56	0.18	0.74	0.06	0.68
EPS forward coefficient p-value	0.44	0.00	0.05	0.03	0.03	0.03	0.70	0.05
Fiscal year matches trading window	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No

Table 3.3.4: Stocks where EPS estimators are significant when estimating the Sharpe ratio, excess return, and volatility

As such it is explored if there is a link between which of the regressions have significant EPS coefficient and which of the stocks have their fiscal year not align with the quarterly rebalancing. For the regressions of Sharpe ratio, excess return and volatility in table 3.3.4, the stocks which have significant earnings coefficients all have fiscal years that align with the quarterly rebalancing except for Walmart. It can even be seen that three of the stocks; 3M, General Electric and Travelers COS. all have significant earnings coefficients for both the Sharpe ratio and excess return regressions.

Fiscal year starts	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
# Stocks	19	2	0	0	0	1	2	1	0	2	0	0
Fits with quarterly trading	Yes			Yes			Yes			Yes		
Stocks	3M	HOME DEPOT				NIKE B	MICROSOFT	CISCO SYSTEMS		APPLE		
	AMERICAN EXPRESS	WALMART					PROCTER & GAMBLE			WALT DISNEY		
	BOEING											
	CATERPILLAR											
	CHEVRON											
	COCA COLA											
	EXXON MOBIL											
	GENERAL ELECTRIC											
	INTERNATIONAL BUS.MCHS.											
	INTEL											
	JOHNSON & JOHNSON											
	JP MORGAN CHASE & CO.											
	MCDONALDS											
	MERCK & COMPANY											
	PFIZER											
	TRAVELERS COS.											
	UNITED TECHNOLOGIES											
	UNITEDHEALTH GROUP											
	VERIZON COMMUNICATIONS											

Table 3.3.5: When the stocks begin their fiscal year and potential overlap with the calendar year quarters

However, when looking at the individual fiscal years for all 27 companies in table 3.3.5 it is found that 23 of the 27 companies have a fiscal year that aligns their quarterly earnings with the quarterly rebalancing. As such there is not enough evidence to be able to conclude that the earnings coefficients are significant for stocks that report on the basis of the calendar year or starting in months that align with the calendar year quarters.

Referencing to the SEC regulations the sample companies have 40 days to disclose their quarterly earnings. After the quarter has been concluded, it is examined how many days the 27 companies had used before releasing their latest 10-Q form, which is the official SEC form for quarterly earnings. On average the companies use 23.6 days to inform the public about their newest financial standing with the longest time used being 37 days from Walt Disney and the shortest being 12 days from JP Morgan Chase & Co, which can be seen in table 8.3.1 in the appendix.

In the models the 12-months trailing earnings per share and 12-months forward consensus earnings per share are used with a quarterly frequency downloaded from DataStream. However, given that quarterly earnings are not released at the very start of the following quarter the downloaded data does not reflect this when retrieved with a quarterly frequency. This means that DataStream gives the newest earnings per share update and consensus for each quarter, but the update does not happen in real time at the beginning of the quarter.

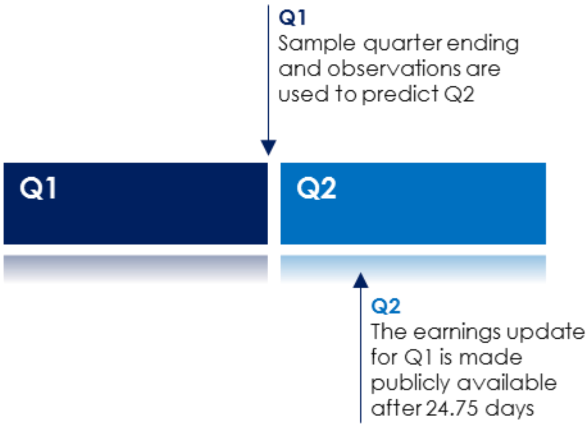

Figure 3.2.1: Quarter earnings are released on average 23.6 days into the next quarter

As such the models use observations from the future to predict the future when analyzing the stocks that have fiscal years which follow the calendar year quarters. Given how the trading strategy and modelling is implemented, referencing figure 3.2.1, the strategy is not feasible for a real life implementation.

The following section examines the ordinary least squares estimators for the four Sharpe ratio forecasting models; Direct Rolling, Direct Expanding, Component Rolling, and Component Expanding.

Partial effect of estimators in the four forecasting models

A more in-depth analysis of the partial effect of the estimators can possibly shed light on model limitations or complications that can be adjusted for a potentially better fit. To aggregate the data from the 6 different regressions, which are computed individually on each of the 27 stocks and for every 84 quarter period, the median output is analyzed. The median is chosen over the average due to the limited number of significant estimators and to neutralize the effect of outliers, such that the analysis is as representative for the full sample as possible.

Median beta coefficient	Excess return		Volatility		Sharpe ratio	
	Component Expanding	Component Rolling	Component Expanding	Component Rolling	Direct Expanding	Direct Rolling
Intercept	0.04	0.15	0.04	0.14	0.33	0.11
Delta 12MTH F EPS	0.17	-0.05	0.21	0.02	1.54	0.09
Delta 12MTH T EPS	-0.19	0.09	-0.44	0.04	-1.64	0.06
3 month	-0.05	-0.05	-0.06	-0.04	-0.42	0.03
6 month	-0.04	-0.01	-0.06	0.00	-0.13	0.04
9 month	-0.02	-0.01	0.02	-0.03	-0.08	0.04
12 month	0.03	0.00	-0.08	0.00	0.16	0.27

Table 3.3.6: Median value of the regression estimators' beta across all 27 stocks

Table 3.3.6 lists the calculated median estimators' beta value, however because of the imperfect multicollinearity found in table 3.3.3 the exact beta value is most likely not interpretable in any decisive way. Consequently, the interpretation is instead centered around the estimator having a positive or negative effect on the dependent variable. While Tang and Whitelaw (2018) also forecast excess return, volatility and Sharpe ratio, they did so with a different set of estimators and for stock indices. As such there are no parallels to be drawn regarding the partial effect of the chosen estimators. Moskowitz, Ooi, and Pedersen (2011) found in their research, that the h-month lagged excess return of the equity indices included in their analysis are positively correlated with the excess return. They find this to be consistently true for h equal to 1 and up to 14. The results for excess return and Sharpe ratio in table 3.3.6 are the comparable results. Because Sharpe ratio is just a volatility scaled excess return performance measure, the positive or negative partial effect of the estimators can be interpreted the same as the excess return. The estimators of the lagged 3-, 6-, and 9-month excess return are negative for the expanding models with the lagged 12-month excess return having a positive beta. Consequently, the finding is almost inverted in comparison showing a negative correlation between excess returns for h equal to 3- to 9-months and a positive return for h equal to-12 months. The two rolling models exhibit conflicting effects of the lagged excess return. After more closely examining the partial effect there seems to be no consistent pattern or findings that are equal to the literature, which needs to be seen in the context that the broad full sample estimators are not significant referencing table 3.3.1 and 3.3.2.

To examine which of the estimators have the most effect on the forecasted variables the absolute numerical partial forecasted value is indexed.

Indexed (100%) median koefficient effect on forecast	Excess return		Volatility		Sharpe ratio		Average
	Component Expanding	Component Rolling	Component Expanding	Component Rolling	Direct Expanding	Direct Rolling	Across all 4 models
Intercept	47.6%	49.7%	90.4%	90.9%	56.2%	57.6%	65.40%
Delta 12MTH F EPS	9.0%	16.4%	3.0%	2.3%	6.0%	7.5%	7.37%
Delta 12MTH T EPS	9.8%	14.7%	2.7%	3.2%	9.0%	17.2%	9.42%
3 month	5.3%	0.0%	0.8%	0.4%	3.8%	1.5%	1.97%
6 month	8.4%	6.6%	0.9%	0.8%	7.3%	3.6%	4.59%
9 month	10.2%	6.0%	1.1%	1.0%	8.7%	4.4%	5.23%
12 month	9.7%	6.7%	1.1%	1.4%	9.0%	8.2%	6.01%

Table 3.3.7: Indexed median effect on the dependent variable of the regression by coefficients across all 27 stocks

The result from table 3.3.7, that shows the indexed median effect of the estimators on the forecast, indicates that the regression suffers from omitted variable bias. Referencing table 3.3.2 the intercept is found to be consistently statistically significant across both the three forecasted variables and almost all of the stocks. Consequently, when the intercepts are significant and predict 47.6% to 49.7% of the excess return, 90.4% to 90.9% of the volatility and 56.2% to 57.6% of the Sharpe ratio, the models likely have omitted variables that are significant in predicting a linear relation.

Test for accuracy of models

Since four different implementations, of essentially the same trading strategy, are available it is necessary to be able to determine which is objectively better at forecasting, such that it can be determined which model is most attractive to implement in a real-life scenario. Regression statistics only provide information about the individual regression, which can be compared, but this does not provide a definitive statistical answer. The Diebold Mariano test is used to evaluate the relative attractiveness of different forecast models and in this section, it is implemented to compare the four different forecast models. Among its strengths is that the error terms do not need to be normally distributed, they can have a mean different from zero and are allowed to be serially correlated. The specific method for computation is discussed in the methodology section, while the formulas used are presented in the theory section.

# p-values below 0.10	Direct Rolling	Direct Expanding	Component Rolling	Component Expanding
Direct Rolling				
Direct Expanding	23			
Component Rolling	16	25		
Component Expanding	22	3	24	
Average squared errors	170	98	121,582	96
Median squared errors	152	95	842	89

Table 3.3.8: Summary statistics for Diebold-Mariano test on the four different models

The output available in table 3.3.8 is the product of the Diebold-Mariano test of the four different models. The test reveals that for most of the pairs of the models there is a statistically significant difference in their respective accuracy relative to one another. This is tested in their ability to predict the actual Sharpe ratio and comparing their squared errors.

The difference in squared errors is almost always significant across all the model pairs except for the two expanding models against each other, where the difference in squared errors is only significant for 3 out of the 27 stocks. This indicates that they are able to attain the same level of accuracy with one model possibly being marginally better.

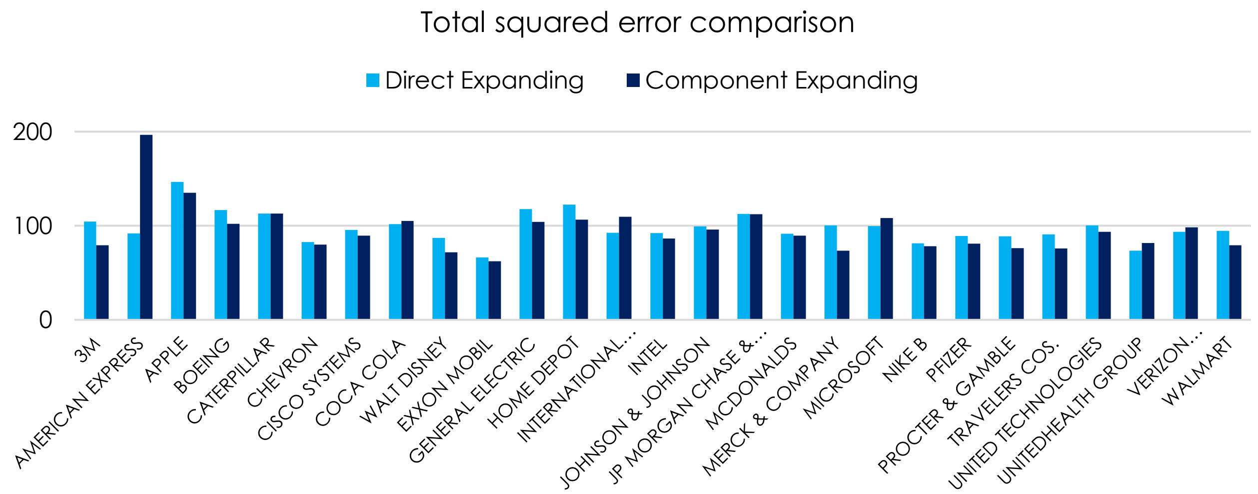

Figure 3.2.2: Squared errors of the Direct Expanding and Component Expanding model for all 27 stocks

In figure 3.2.2 it can also be visually seen that the squared errors of the two models are very similar. When comparing the total squared errors that each of the expanding regressions provide across the 27 stocks they generally track each other quite well and exhibit stable error across the stocks with few outliers. As such one would expect the two models to both have roughly the same accuracy when predicting Sharpe ratio. Examining the number of significant p-values, average, and median squared errors across all four models it is also clear, that the two expanding models are more accurate than the two rolling models. Examining the actual squared errors of both models for the three stocks, where the test finds a significant difference, it can be seen in table 8.3.2 in the appendix that for all 3 stocks; 3M, Boeing and Walt Disney the Component Expanding model has lower squared errors. Consequently, the Component Expanding is marginally the most accurate model ahead of the Direct Expanding model. Going forward the assumption is that the two expanding models are the most accurate and have the most prediction power on the actual Sharpe ratio.

The same effect to a much lesser extent is seen between the two rolling models in table 3.3.8, where 16 of the 27 stocks show a significant difference in squared errors. As such it looks as if the models that share the same regression window more strongly exhibit traits that makes them more alike across the stocks compared to the models that uses the same method of calculating the forecasted Sharpe ratio.

An in depth look at the component rolling model reveals squared errors that both have an average and median value much above its' peer models. The combination of both predicting the excess return and volatility in a rolling regression and the model estimators focusing on recent observations means that the component rolling model is very susceptible to drastic and rapid changes in the market. The most extreme effect this has on the model is for the American Express stock, where the model predicts the stock to have a Sharpe ratio of negative (1,765.40) in the second quarter of 2009 during the financial crisis. This happens because the linear model predicts an excess return of negative (54.55%) with an extremely low derived volatility of 0.03%. The squared error between the actual Sharpe ratio of 1.50 and the estimated negative (1,765.40) is 3,250,746.58, which is a very big outlier with the second largest squared error being 5,964. The main underlying reason for this outlier is the fact that the component models are estimating both components of the Sharpe ratio; excess return and volatility. This can lead to periods where excess return and volatility are very much out of synchronization, such that while forecasted excess return exhibits large amounts of volatility the volatility is forecasted to be low and stable. This leads to the scenario above, where an extreme negative return paired with an extremely low volatility yields an extremely negative Sharpe ratio forecast.

The same effect is also seen in the component expanding model, where the American Express stock provides the largest squared errors with a value of 196.71 compared to the average of 96. As such it seems that providing the model with more observations does not fully remove the extreme prediction errors but smoothens them out.

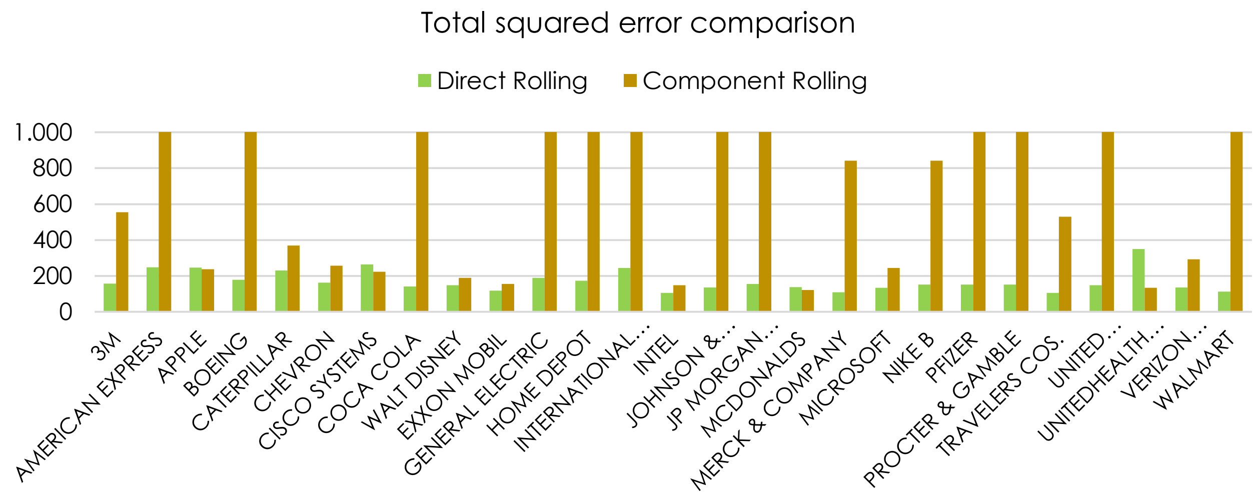

Figure 3.2.3: Squared errors of the Direct Rolling and Component Rolling model for all 27 stocks

In figure 3.2.3 the Component Rolling model consistently produces large squared errors, such that 12 of the 27 stocks are not visible while on the same scale as the Direct Rolling model. Where the three other models operate with squared errors in the hundreds, the component rolling model also has squared errors in the hundreds, but 11 of the 27 stocks have squared errors that are in the thousands and 1 that goes into the millions. The component rolling model thus has the most unstable and the most extreme squared errors and as such it is expected to have the worst prediction power on realized Sharpe ratio.

The squared errors for the Direct Rolling model in figure 3.2.3 are also like the two expanding models relatively smooth across the 27 different stocks with a maximum of 351 and minimum of 106. The error spread is still much larger than for the two expanding models, but much smaller than the error spread for the component rolling model. It has on average 1.74 and 1.78 times as high squared errors when compared to the direct expanding and component expanding model respectively. The total squared errors for each stock is also relatively stable albeit to a lesser extent than the two expanding models. However, the stable squared errors can indicate that the three models do have some consistency in prediction ability across the 27 stocks, which can help validate the choice of estimators. On the other hand, the stable squared errors can be interpreted as the models not being able to reflect the differences in the return patterns of the different stocks. Running the analysis ex post has the advantage that the stocks realized excess returns over the sample period is known. To further confirm or deny if the models have any prediction power on the actual Sharpe ratio this is tested.

Test for predictive power of forecasted Sharpe ratio

In this section the forecasted Sharpe ratios ability to predict the variability of the realized Sharpe ratio is tested. If the models do not have significant prediction power on the realized Sharpe ratio, then it is not possible to rank the stocks correctly according to the strategy. As a result of this the performance of the portfolios, that have been created on the basis of the model output is not attributable to the strategy but instead random.

The test is computed as a regression where each of the four models forecasted Sharpe ratio is individually regressed on the actual realized Sharpe ratio for every forecasted quarter. The single independent variable ordinary least squares regression is computed for all 27 stocks with the following regression from Q2 1997 to Q1 2018:

$$SR_t^{realized} = \alpha + SR_t^{predicted}$$

Formula 3.1.4: Regression with realized Sharpe ratio at time t as the dependent variable and predicted Sharpe ratio at time t as the independent variable

Given the findings from the Diebold-Mariano tests, the expanding models is expected to have the most prediction power on the basis that they produce the smallest and most stable squared errors across the 27 stocks. While it is expected that the component rolling model will have the least prediction power due to its extreme squared errors.

	Direct Rolling # p-value below	Direct Expanding # p-value below	Component Rolling # p-value below	Component Expanding # p-value below
	0.10	0.10	0.10	0.10
Intercept	27	24	27	23
Forecasted sharpe ratio	2	1	5	6
Average R2	0.013	0.007	0.013	0.017
Median R2	0.008	0.002	0.004	0.008
%-significant estimator	7%	4%	19%	22%

Table 3.3.9: Summary statistics of forecasted Sharpe ratio used as estimator for realized Sharpe ratio

The findings in table 3.3.9 from regressing the models individually predicted Sharpe ratio on the actual realized Sharpe ratio provides a mixed message compared to the possible conclusions from the Diebold-Mariano tests. First, it is seen that most of the predicted Sharpe ratios are not significant estimators of the actual realized Sharpe ratio. This is also reflected in the average and median values of the regressions R^2 statistics. The median R^2 statistic indicates that 0.2% to 0.8% of the variation in the actual Sharpe ratio is explained by the predicted Sharpe ratios. As such the output of the models is just as insignificant to predict with as the explanatory variables for the models. With the low amount of explanatory power, it is likely that the elusive stock return pattern is too complicated for a linear ordinary least squares model to reproduce this pattern reliably, and with validity. There is however consistency in the fact that the most accurate model according to the Diebold-Mariano test, the Component Expanding, also has the most amount of significant Sharpe ratio predictions on the actual Sharpe ratios and the highest average R^2 statistic.

Portfolio results

This section will examine the results achieved by the model portfolios based on the four forecasting models. First the results will be described, secondly the results are tested for statistically significant differences.

Portfolio analysis

Initially the quarterly total return data was used to calculate the quarterly return, but it was then discovered that the data did not properly line up with the actual daily data. The error was such that the quarterly total return index for a given quarter would not correspond to the last daily total return index observation. Therefore, it was decided to use daily observations as these were more accurate than the erroneous quarterly total return index.

Returns

The cumulative excess return index for all portfolios were computed to get a sense of the returns achieved by the strategies over the entire investment period. The values are indexed starting at 1. As can be seen on figure 4.2.1, the best performing model is the Direct Rolling forecast model which initially outperforms all the other portfolios from Q2 2000 and then keeps the lead through the rest of the investment period.

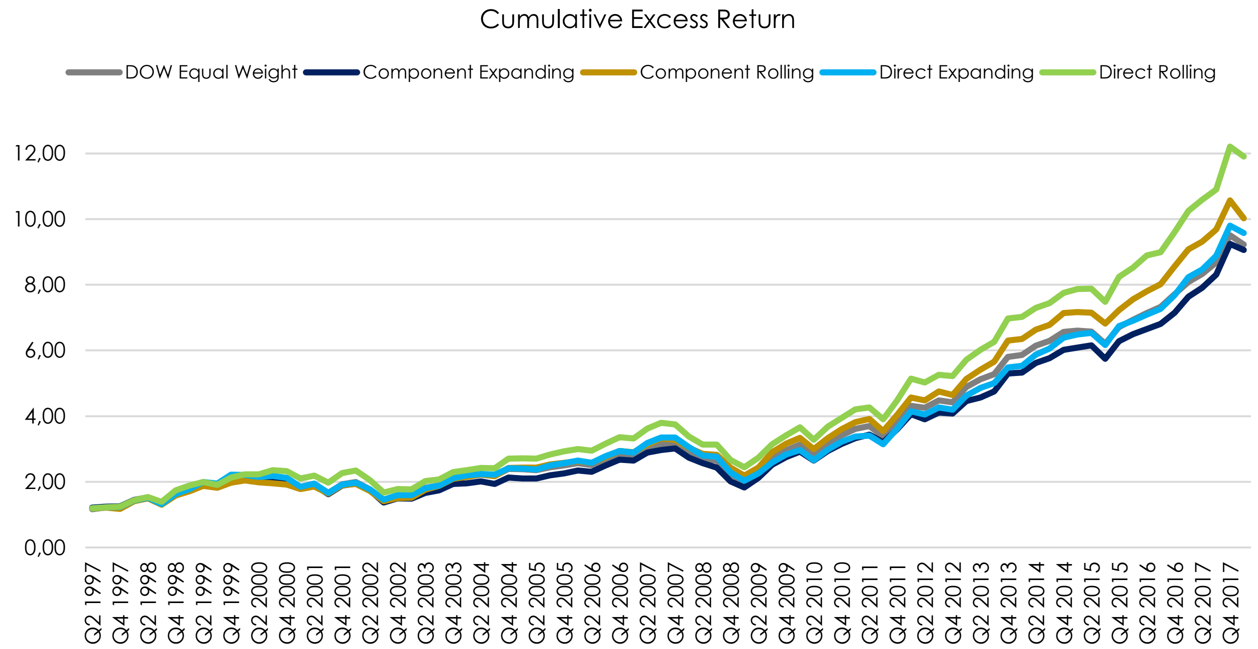

Figure 4.2.1: The cumulative excess return realized by the five different portfolios over the period starting 1997 Q2 and ending 2018 Q1

According to the tables 4.3.1 and 4.3.2, the Component Expanding forecast model yields the lowest cumulative excess return of 9.06, which is equivalent to a quarterly compounded excess return of 2.66%. In second last place is the equally weighted Dow benchmark portfolio with a cumulative excess return of 9.22 corresponding to 2.68% as a quarterly compounded excess return. In third place is the Direct Expanding forecast model, with a cumulative excess return of 9.58 or 2.73% as a quarterly compounded excess return. In second place is the Component Rolling forecast model with a cumulative excess return of 10.02 and a quarterly compounded excess return of 2.78%. The highest cumulative excess return of 11.91 belongs to the Direct Rolling forecast model, which computes to a quarterly compounded excess return of 2.99%.

To summarize, only the Component Expanding portfolio underperforms the equally weighted Dow benchmark portfolio in terms of excess return, implying that the three other outperform. Also, the two highest performing portfolios both employed a rolling window regression method.

1997 Q2 to 2018 Q1	Direct Rolling	Direct Expanding	Component Rolling	Component Expanding	DOW Equal Weight
Cumulative excess return	11.91	9.58	10.02	9.06	9.22

Table 4.3.1: The cumulative excess return realized by the five different portfolios over the period starting 1997 Q2 and ending 2018 Q1

Quarterly compounded excess return				
Direct Rolling	Direct Expanding	Component Rolling	Component Expanding	DOW Equal Weight
2.99%	2.73%	2.78%	2.66%	2.68%

Table 4.3.2: The quarterly compounded excess return realized by the five different portfolios over the period starting 1997 Q2 and ending 2018 Q1

In the appendix figure 8.2.9 it is shown that the Direct Rolling forecast model has been able to outperform the equally weighted Dow benchmark portfolio from Q2 2000 and during the remaining investment period up until Q1 2018. The shape of both the lines are very similar, but it seems that the Direct Rolling forecast model is a bit steeper for certain holding periods. There is a large difference between the ending cumulative excess return values of the two portfolios.

In the appendix figure 8.2.10 equally weighted Dow benchmark portfolio and the Component Rolling forecast model are computed. They are very similar and almost identical until around Q2 2012, when the Component

Rolling portfolio starts outperforming the equally weighted Dow benchmark portfolio for the rest of the investment period until Q1 2018. There is a sizeable gap between the ending cumulative excess return values.

The appendix figure 8.2.11 shows the equally weighted Dow benchmark portfolio and the Component Expanding forecast model. They are quite similar, but the Component Expanding portfolio is outperformed from around Q1 2004 and is below the equally weighted Dow benchmark portfolio for the rest of the investment period. The ending values for cumulative excess return are quite close to each other.

In the appendix figure 8.2.12 are the equally weighted Dow benchmark portfolio and the Direct Expanding forecast model. They are extremely similar, tracking each other most of the time with periods of minor under- and overperformance. The ending values of cumulative excess return are very similar.

The cumulative return, high water mark, and draw down for each strategy is graphed in the appendix 8.2.13 to 8.2.16. They are all very similar, all with the biggest spikes in draw down around 2002 and 2008 to 2009 coinciding with the end of the dot.com bubble and the 2008 financial crisis. As described in figure 4.2.1 for cumulative excess return while the shapes overall are similar, the absolute level does differ. As the model portfolios consists of 20 out of 27 of the stocks included in the equally weighted Dow benchmark portfolio, it would not be expected to see very large differences, since they mostly consist of the same stocks. It is also worth noting that the longest draw down an investor would experience was 20 quarters from 2000 till 2004.

As seen in figure 4.2.2 for the equally weighted Dow benchmark portfolio, overall, there is a positive upward sloping trend for the cumulative excess return, ending significantly higher than the starting point at 9.22. There are multiple times with big spikes in the draw down amount, the most significant being periods in 2001, 2002, 2008, and 2009. The largest draw down of 37% is around Q4 2008.

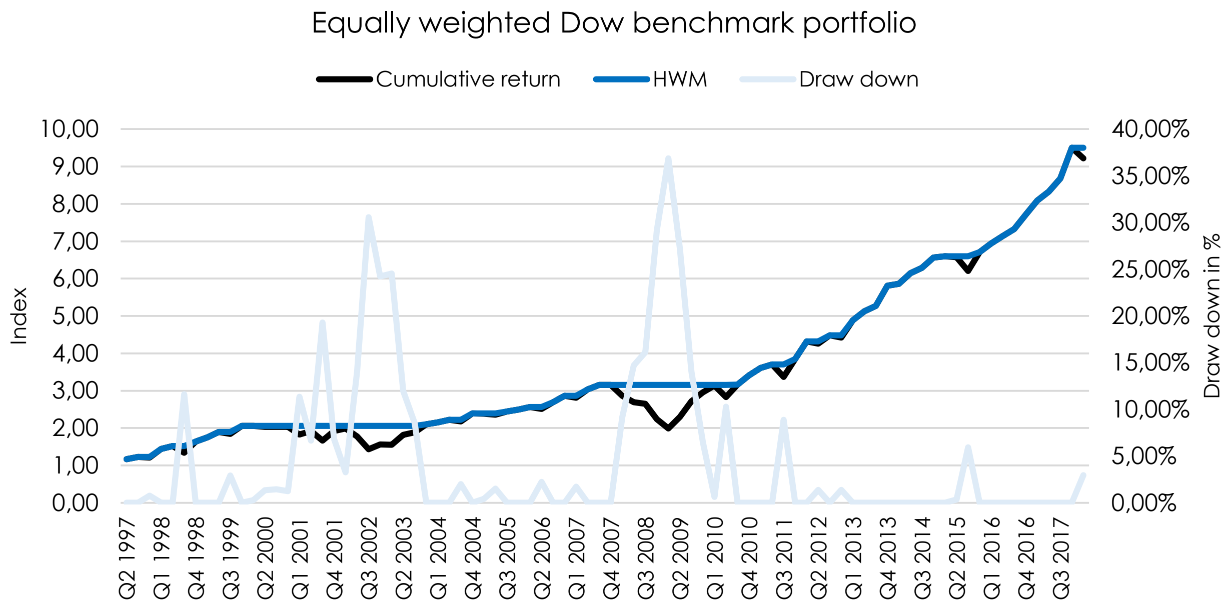

Figure 4.2.2: The highwater mark and drawdown for the equally weighted Dow benchmark portfolio over the period starting 1997 Q2 and ending 2018 Q1

The Component Expanding portfolio, as seen in appendix figure 8.2.13, achieves a positive upward sloping trend for the cumulative excess return, resulting in a significantly higher ending value at 9.06. 2001, 2002, 2008, and 2009 are the years with large draw downs with the largest being 39% in Q2 2009.

The Component Rolling portfolio in appendix figure 8.2.14 ends with a cumulative excess return of 10.57 with a positive trending slope for the entire period. The most significant spikes in the draw downs of the portfolio are in 2001, 2002, 2008, and 2009. Q1 2009 is responsible for the largest draw down of 34%.

As can be seen in the appendix figure 8.2.15, the Direct Expanding portfolio has a positive upward sloping trend for the cumulative excess return, ending significantly higher than the starting point at 9.80. There are multiple times with big spikes in the draw down amount, the most significant being 2001, 2002, 2008, and 2009. The largest draw down of 39% is around Q1 2009.

The Direct Rolling portfolio, located in appendix figure 8.2.16, is the portfolio that ends up with the highest cumulative excess return for the investment period which is 12.21. Just like the other portfolios the portfolio sees that largest draw downs in 2001, 2002, 2008, and 2009. Again Q1 2009 is responsible for the largest draw drown of around 36%.

To better understand what kind of excess returns are achieved for any given quarter, the distribution of realized quarterly excess returns are graphed in figure 4.2.3. They are compared to the equally weighted Dow benchmark portfolio. The distribution of the realized quarterly excess returns for each portfolio is discussed below. They are all similar in terms of their distribution with minor differences.

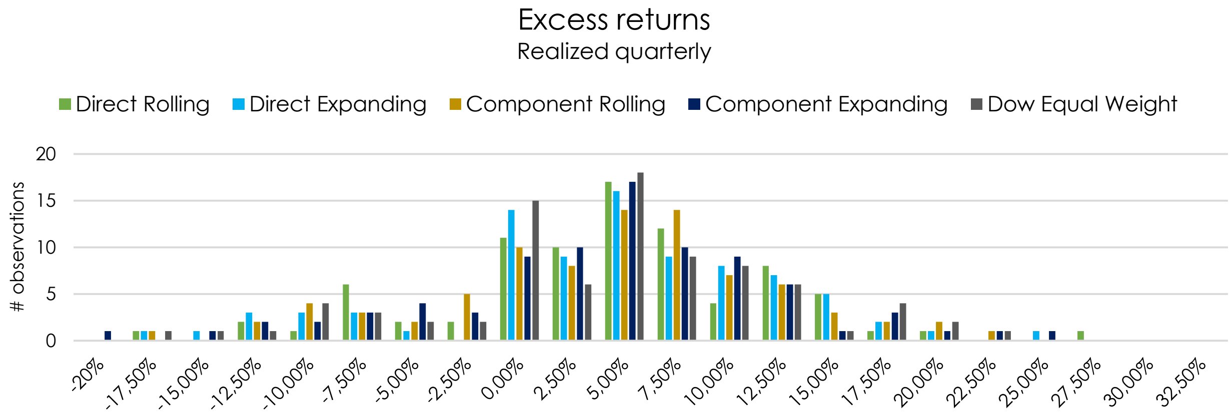

Figure 4.2.3: The distribution of realized quarterly excess returns for all five portfolios over the period starting 1997 Q2 and ending 2018 Q1

Shown in the appendix figure 8.2.17 are the distribution of quarterly excess returns for the Component Expanding portfolio and the equally weighted Dow benchmark portfolio. They look very similar and have a shape that resembles a normal distribution with some smoothness imperfections. The component expanding portfolio has both the highest and lowest realized quarterly return when compared to the equally weighted Dow benchmark portfolio in isolation.

The appendix figure 8.2.18 shows the quarterly excess return of the Component Rolling portfolio and the equally weighted Dow benchmark portfolio. Both seem to follow a normal distribution with few imperfections. They both share an observation in both the top and bottom quarterly excess return bin.

Seen in appendix figure 8.2.19 is the distributions of realized quarterly excess returns of the Direct Expanding portfolio and the equally weighted Dow benchmark portfolio. Once again both seem to follow a normal distribution although the direct expanding portfolio lacks smoothness as it has two large spikes of observations and zero observations around -2.5%. The Direct Expanding portfolio has the largest excess return and they both share the most negative bin.

Looking at the distribution of realized quarterly excess returns for the Direct Rolling portfolio, found in appendix figure 8.2.20, it shows that it is quite similar to the equally weighted Dow benchmark portfolio. Both look similar to a normal distribution, although this portfolio has problems with the smoothness of the curve. The Direct Rolling portfolio achieves a higher maximum excess return than the equally weighted Dow benchmark portfolio.

From looking at table 4.3.3 of summary statistics below, it is seen that the mean excess returns are close to each other, although the Direct Rolling is around 1.1 times higher than the other portfolios. Standard deviations are within a small range from 7.77% for the Direct Rolling portfolio to 8.15% for the Component Expanding portfolio. The standard deviations here should not be given as much weight as the summary statistics for volatility in a later section. This is because the volatility section is based on daily returns, which are then converted to quarterly, such that they better correspond to the actual risk experienced over the holding period. The largest realized excess return is achieved by the Direct Rolling portfolio with 25.23%, which is quite a bit higher than the lowest max of 20.96% for the Component Rolling portfolio. The 75th percentile observations vary between the ranges of 7.12% for the Direct Rolling portfolio to 8.84% for the Direct Expanding portfolio. The quarterly median excess returns range from 2.97% for the Direct Expanding portfolio to 3.42% for the Direct Rolling portfolio. Quarterly 25th percentile excess returns are in the range -1.63% for the Component Rolling portfolio to -0.67% for the Direct Rolling portfolio. The minimum quarterly excess returns ranged from -21.09% for the Component Expanding portfolio to -17.62% for the Direct Expanding portfolio making it quite a wide band.

Excess return	Direct Rolling	Direct Expanding	Component Rolling	Component Expanding	DOW Equal Weight
Mean	3.27%	3.04%	3.07%	2.97%	2.99%
Standard deviation	7.77%	8.13%	7.83%	8.15%	7.96%
Max	25.23%	24.43%	20.96%	23.69%	22.09%
75th percentile	7.12%	8.84%	7.57%	8.20%	7.94%
Median	3.42%	2.97%	3.41%	3.17%	3.16%
25th percentile	-0.67%	-1.23%	-1.63%	-0.84%	-1.06%
Min	-18.30%	-17.62%	-18.11%	-21.09%	-19.20%

Table 4.3.3: Summary statistics for the realized quarterly excess returns for all five portfolios over the period starting 1997 Q2 and ending 2018 Q1

Volatility

In this section the volatility realized by the four different model portfolios and the equally weighed Dow benchmark portfolio are presented.

As discussed in the methodology section a quarterly volatility measure that is based on the daily volatilities over a given quarter is employed. This is to better reflect the actual volatility during the holding period.

In general, as seen from figure 4.2.4 below, it seems that all the portfolios follow a lognormal distribution and are extremely similar to the equally weighted Dow benchmark portfolio. From a purely mathematical perspective since variance will always be a positive number, volatilities cannot be negative and as such they share a starting point of 0 with the lognormal distribution.

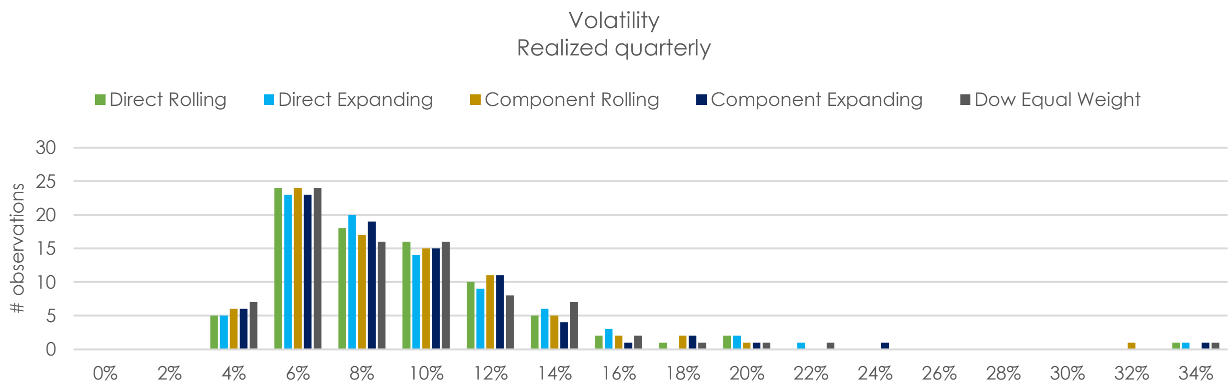

Figure 4.2.4: The distribution of realized quarterly volatility for all five portfolios over the period starting 1997 Q2 and ending 2018 Q1

In the following section the graphs of the quarterly volatility of the four different portfolios are compared to the equally weighted Dow benchmark portfolio.

Found in the appendix figure 8.2.21, the Component Expanding portfolios volatility seem to follow a lognormal distribution with a slim right tail. Most observations seem to be the same for Component Expanding portfolio compared to the equally weighted Dow benchmark portfolio.

As seen in the appendix figure 8.2.22, the realized quarterly volatilities for the Component Rolling portfolio seems to follow a lognormal distribution that is left skewed with a long thin right tail. Even with most of the observations being the same as the Dow portfolio, the highest volatility observation of 34% is realized by the Dow portfolio with the Component Rolling portfolio realizing 32% as its highest.

According to figure 8.2.23 in the appendix, the Direct Expanding portfolio seems to follow a lognormal distribution like the equally weighted Dow benchmark portfolio. The distributions are extremely similar with both sharing observations for the highest and lowest values.

As seen in the appendix figure 8.2.24, the Direct Rolling portfolio like the other 3 forecast portfolios looks like a lognormal distribution on top of the Dow observations. It also shares both the minimum and maximum values of realized quarterly volatility with the equally weighted Dow benchmark portfolio.

In table 4.3.4 summary statistics for the volatility of the different portfolios is seen. It can be seen that all of the four portfolios based on ranking the forecasted Sharpe ratios have around the same mean volatility and they are in a somewhat tight range from around 8.19% for the Direct Rolling portfolio to 8.43% for the Direct Expanding portfolio. The standard deviations of all portfolios are very close from 4.37% for the Component Rolling portfolio to 4.53% for the Component Expanding portfolio. The maximum volatilities range from 31.87% for the Component Rolling portfolio to 33.09% for the Component Expanding portfolio. The same is true for the median values where the Dow portfolio is in the range of the four portfolios spanning from 7.12% for the Component Expanding portfolio to 7.39% for the Component Rolling portfolio. For the 25th percentile it can be seen once again that the value for the equally weighted Dow benchmark portfolio is within the range of the four portfolios, the range of the portfolios are from 5.32% for the Direct Expanding and equally weighted Dow benchmark portfolio to 5.48% for the Component Rolling portfolio. The minimum volatility observations are similar with a maximum difference of 0.34% points. The portfolios range from 2.55% for the Direct Rolling portfolio to 2.89% for the Component Expanding portfolio.

It is found that the equally weighted Dow benchmark portfolio when compared to the four other portfolios has very similar mean volatility, maximum volatility, 75th, 50th, 25th percentile and minimum volatility.

Volatility	Direct Rolling	Direct Expanding	Component Rolling	Component Expanding	DOW Equal Weight
Mean	8.19%	8.43%	8.32%	8.29%	8.20%
Standard deviation	4.39%	4.52%	4.37%	4.53%	4.46%
Max	32.99%	32.39%	31.87%	33.09%	32.18%
75th percentile	9.98%	10.21%	10.28%	10.00%	9.93%
Median	7.14%	7.32%	7.39%	7.12%	7.13%
25th percentile	5.37%	5.32%	5.48%	5.37%	5.32%
Min	2.55%	2.61%	2.87%	2.89%	2.84%

Table 4.3.4: Summary statistics for the realized quarterly volatility for all five portfolios over the period starting 1997 Q2 and ending 2018 Q1

Sharpe ratio

In the following section the Sharpe ratio results from the five portfolios are examined. In general, as can be seen in figure 4.2.5, the distribution of the Sharpe ratios are quite similar for the different portfolios.

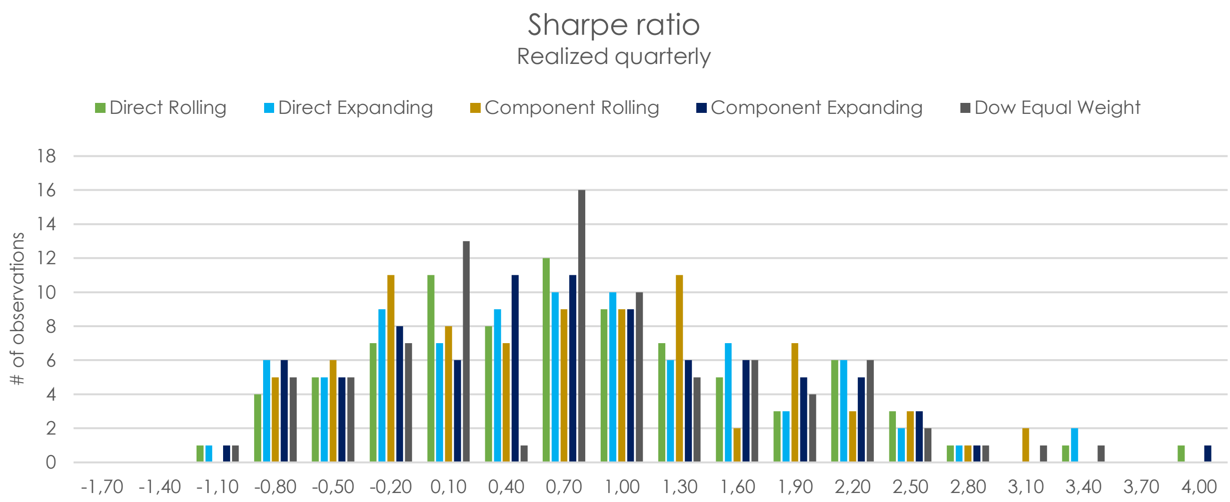

Figure 4.2.5: The distribution of realized quarterly Sharpe ratios for all five portfolios over the period starting 1997 Q2 and ending 2018 Q1

The Component Expanding Sharpe ratio distribution, found in the appendix figure 8.2.25, looks as if it is a mix between a normal distribution and a triangle distribution where the left tail is a bit short. The realized quarterly Sharpe ratio observations for the Component Expanding generally cluster as tightly as for the equally weighted Dow benchmark portfolio, which does have fewer observations in the 0.4 Sharpe ratio bin, which effectively creates a small hole in the distribution. The Component Expanding portfolio has the highest observation of 3.89.

The realized quarterly Sharpe ratios of the Component Rolling, located in the appendix figure 8.2.26, are uneven with two different peaks, one around -0.2 and the other around 1.3. Although it is uneven it does seem as a flat distribution in the sense, that it does not have any tails except a minor one towards the right side of the distribution. The Component Rolling observations are generally similar to the equally weighted Dow benchmark portfolio. The equally weighted Dow benchmark portfolio has the largest and smallest observations of the Sharpe ratio.

The Direct Expanding portfolio, seen in the appendix figure 8.2.27, resembles a somewhat flat normal distribution without a high peak, in other words it has short fat tails. The Direct Expanding portfolio only shares one observation in the minimum bin with the equally weighted Dow benchmark portfolio.

The Direct Rolling portfolio, portrayed in the appendix figure 8.2.28, has its realized quarterly Sharpe ratios distributed somewhat similar to a normal distribution with a longer fatter right tail. It is again very similar to the equally weighted Dow benchmark portfolio distribution. The Direct Rolling portfolio has the highest observation of 3.77 of the two portfolios.

Table 4.3.5 below display summary statistics for the five different portfolios related to the realized quarterly Sharpe ratios. The mean values for the portfolios are all quite close in the range from 0.61 for the equally weighted Dow benchmark portfolio to 0.66 for the Direct Rolling portfolio. The standard deviations for the Sharpe ratios of the portfolios are very similar with a range from 0.99 to 1.01. The maximum Sharpe ratios achieved vary quite a bit, with the portfolios in the range 3.10 for the Component Rolling portfolio to 3.89 for the Component Expanding portfolio, indicating significant difference between the portfolios maximum Sharpe ratio. For the 75th percentile values the range for the portfolios is from 1.12 for the Component Rolling portfolio to 1.32 for the Direct Expanding portfolio. The medians are somewhat close, the range for the portfolios are 0.50 for the Component Expanding to 0.56 for both the Component Rolling and the Direct Expanding portfolio. For the 25th percentile, the range for the portfolios are from -0.23 for the Component Rolling portfolio to -0.08 for the Direct Rolling portfolio. The minimum values achieved are quite close to each other and the range for the portfolios is from -1.19 for the Direct Rolling portfolio to -1.04 for the Component Rolling portfolio.

There is a small amount of variation in between the portfolios, with the biggest being present in the maximum achieved Sharpe ratios. Overall the portfolios can be said to be very similar.

Sharpe ratio	Direct Rolling	Direct Expanding	Component Rolling	Component Expanding	DOW Equal Weight
Mean	0.66	0.63	0.62	0.62	0.61
Standard deviation	1.01	1.01	0.99	1.01	0.99
Max	3.77	3.30	3.10	3.89	3.30
75th percentile	1.24	1.32	1.12	1.29	1.29
Median	0.53	0.56	0.56	0.50	0.53
25th percentile	-0.08	-0.17	-0.23	-0.13	-0.10
Min	-1.19	-1.11	-1.04	-1.14	-1.11

Table 4.3.5: Summary statistics for the realized quarterly Sharpe ratios for all five portfolios over the period starting 1997 Q2 and ending 2018 Q1

Risk statistics

To quantify the risk of the big losses by pursuing the portfolios the risk measures value at risk and expected shortfall are used.

For the value at risk measure at 10% significance level, that is the worst loss achieved 90% of the time, the equally weighted Dow benchmark portfolio has the biggest loss of 9.08%. The range for the four other portfolios is 7.61% for the Component Rolling to 8.66% for the Direct Expanding portfolio. The expected shortfalls at the 10% significance level, that is the expected loss when dealing with one of the 10% worst returns, are all similar between the Dow and the forecasted portfolios. The range spans from 11.48% for the Direct Rolling portfolio to 12.89% for both Direct Expanding and the Component Expanding. It is worth noting here that both rolling portfolios achieved an expected shortfall that were 1.2% points lower than the two expanding and the equally weighted Dow benchmark portfolio. There is more variance between the portfolios when it comes to the value at risk measure for a 5% significance level. The Direct Expanding portfolio is the highest outlier with an expected worst loss of 12.89%. The Direct Rolling portfolio has the lowest expected loss of 9.84 % given 95% certainty. This corresponds to a 30% difference between the highest and lowest value at risk. The rest of the observations are in the range from 11.07% for the Component Rolling portfolio to 11.63% for the Component Expanding portfolio. Interestingly for the expected shortfall there is a distinct grouping of the component, direct and Dow portfolios. The rolling portfolios achieve the lowest expected shortfall of 13.17% and 13.47% for the direct and component respectively. The Dow realizes 14.22%. The expanding portfolios realizes the worst expected shortfalls, 15.12% and 15.50% for the direct and component portfolios respectively

Overall it seems from these findings that the rolling portfolios have favorable risk characteristics as measured by value at risk and expected shortfall, achieving lower values for both.

Risk summary statistics	Direct Rolling	Direct Expanding	Component Rolling	Component Expanding	DOW Equal Weight
VAR 10%	-8.43%	-8.66%	-7.61%	-8.46%	-9.08%
ES 10%	-11.48%	-12.89%	-11.65%	-12.89%	-12.34%
VAR 5%	-9.84%	-12.91%	-11.07%	-11.63%	-11.11%
ES 5%	-13.17%	-15.12%	-13.47%	-15.50%	-14.22%

Table 4.3.6: Summary statistics for the risk measures for all five portfolios over the period starting 1997 Q2 and ending 2018 Q1

Model classification errors

To test the accuracy of the classifications made by the models, it was examined how often the rank based on predicted Sharpe ratio differed from the rank based on the ex post realized Sharpe ratios. This happens in two cases, either when the predicted rank excludes a stock that should be included, or when the rank includes a stock that should be excluded. Whenever a stock is falsely classified as either being in or out of the portfolio, this results in two wrong classifications, because in addition to the stock being wrongly classified, there is another stock that should be in its place. In a given period, the maximum number of misclassifications that each model can commit is $2 * 7 = 14$. This is because it can wrongly exclude 7 stocks that should be included, thereby at the same time including an additional 7 stocks that should be excluded.

In the table 4.3.7 below, are the misclassifications for all of the 84 forecasted quarters grouped per period for the four different forecasted portfolios. The table shows that the mean number of misclassified stocks per period are similar for the four different portfolios, all in the range from 10 for the Component Rolling portfolio to 10.83 for the Component Expanding portfolio. The spread in the standard deviation of misclassifications are quite similar for all the portfolios with a spread from 1.95 for the Direct Rolling portfolio to 2.17 for the Component Rolling portfolio. All portfolios have periods where they falsely classify all the 7 stocks that they exclude, that is they all have periods where they falsely classify 14 stocks.

Both the 75th percentile and median values are identical for all four portfolios, what is noteworthy is that they are both quite high values of 12 and 10 for the 75th percentile and median, when considering that the maximum is 14. There is some variation in the 25th percentile both the Component Rolling and Direct Expanding portfolios have 8 misclassifications while the Component Expanding and Direct Rolling portfolios have 10, once again quite high values. The minimum observed values are 6 except for the Component Rolling portfolio where the minimum is 4.

# of stocks wrongly classified per period	Direct Rolling	Direct Expanding	Component Rolling	Component Expanding
Mean	10.55	10.24	10.00	10.83
Standard deviation	1.95	2.02	2.17	2.11
Max	14	14	14	14
75th percentile	12	12	12	12
Median	10	10	10	10
25th percentile	10	8	8	10
Min	6	6	4	6

Table 4.3.7: Summary statistics of the classification errors for all five portfolios over the period starting 1997 Q2 and ending 2018 Q1

With the amount of misclassification being this high, it was investigated if there were any patterns of systematic misclassification, that is, if any stocks were more prone than others to be misclassified. To search for such patterns, it was computed how many times the different portfolios would misclassify each stock for all the 84 quarters. Table 4.3.8 found below shows the findings from which it seems that although there is variability, it is likely due to random sampling. There does not seem to be some consistently extreme outlier between the models. The only stock that does potentially come close to being an outlier is Walmart which has more than 40 misclassifications across all four portfolios.

To better visualize the data the misclassifications are graphed in the appendix in figures 8.2.29 to 8.2.32. The distribution of the misclassifications looks as if it is a uniform distribution with some random variability between stocks.

# of total misclassifications per stock	Direct Rolling	Direct Expanding	Component Rolling	Component Expanding
3M	32	36	27	41
AMERICAN EXPRESS	36	26	28	27
APPLE	42	28	29	30
BOEING	34	37	37	33
CATERPILLAR	31	36	32	32
CHEVRON	27	35	27	26
CISCO SYSTEMS	36	23	26	28
COCA COLA	26	32	30	42
WALT DISNEY	31	41	24	35
EXXON MOBIL	24	22	19	19
GENERAL ELECTRIC	32	26	41	27
HOME DEPOT	31	28	31	26
INTERNATIONAL BUS.MCHS.	37	25	35	39
INTEL	33	30	23	28
JOHNSON & JOHNSON	28	27	38	36
JP MORGAN CHASE & CO.	31	33	34	39
MCDONALDS	40	30	22	35
MERCK & COMPANY	37	42	43	39
MICROSOFT	41	31	25	32
NIKE 'B'	34	24	32	23
PFIZER	29	35	35	42
PROCTER & GAMBLE	35	30	35	38
TRAVELERS COS.	33	38	34	36
UNITED TECHNOLOGIES	18	19	35	29
UNITEDHEALTH GROUP	36	38	24	32
VERIZON COMMUNICATIONS	30	38	34	50
WALMART	42	50	40	46
Total	886	860	840	910

Table 4.3.8: Summary statistics of the classification errors for all 27 stocks over the period starting 1997 Q2 and ending 2018 Q1

Ranking consistency

Referencing the results from table 4.3.7 that shows the summary statistics for the classification errors for all the four portfolios over the forecasting period, there is no difference in the mean number of stocks that are wrongly classified each month for the different models. During the analysis of the four model's ability to predict the realized Sharpe ratio, they were found to not be valid. As such it is expected that the ranking will be inconsistent across time as well and not only across the stocks as explored in table 4.3.8.

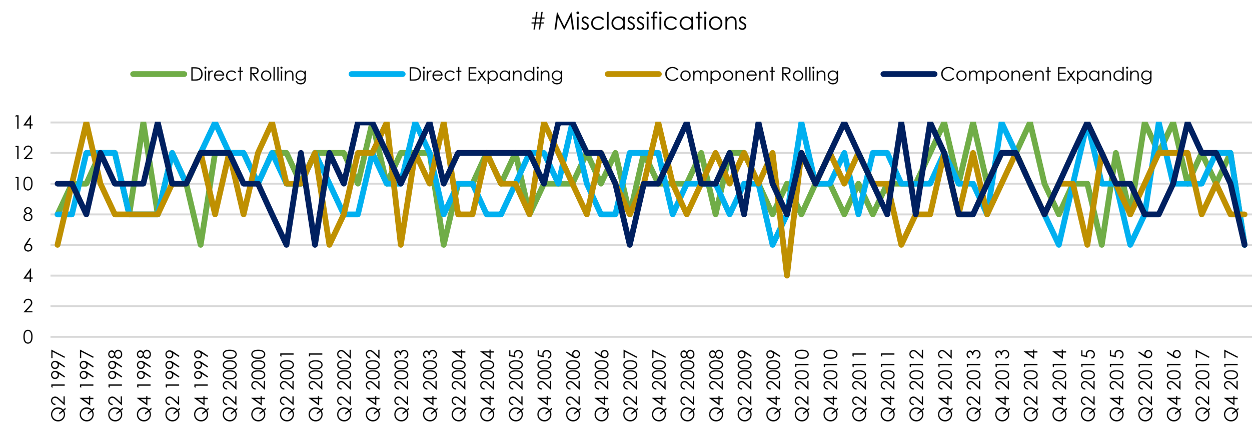

Figure 4.2.6: The number of misclassifications by each of the four models over the forecasting period

As seen in figure 4.2.6 there does not appear to be any systematic difference over time in the four models ability to correctly classify the 27 stocks with regards to whether the model correctly identifies the 20 stocks that will have the highest Sharpe ratio. The result is to be expected taking everything else that has been discovered about the models into consideration. Ability to classify perfectly would indicate that the models were able to perfectly estimate the Sharpe ratio of the individual stocks, which would be naive to expect the models to achieve.

The models poor ranking performance can be attributed to several factors, where the most integral factor is the missing linear relation between the chosen estimators; 12-month trailing earnings per share, 12-month forward consensus earnings per share, 3-month excess return, 6-month excess return, 9-month excess return, and 12-month excess return, and the dependent variables 3-month excess return, volatility, and Sharpe ratio, which was shown in table 3.3.1.

Volatility forecasting

Referring to the output from table 3.3.1 again, it was attempted to explain the variation in the actual Sharpe ratio, 3-month excess return, and volatility, by regressing the estimators it was found, that the estimators most successful in forecasting a linear relationship for volatility. The regression F-test were statistically significant at a 10% significance level for 13 of the 27 stocks.

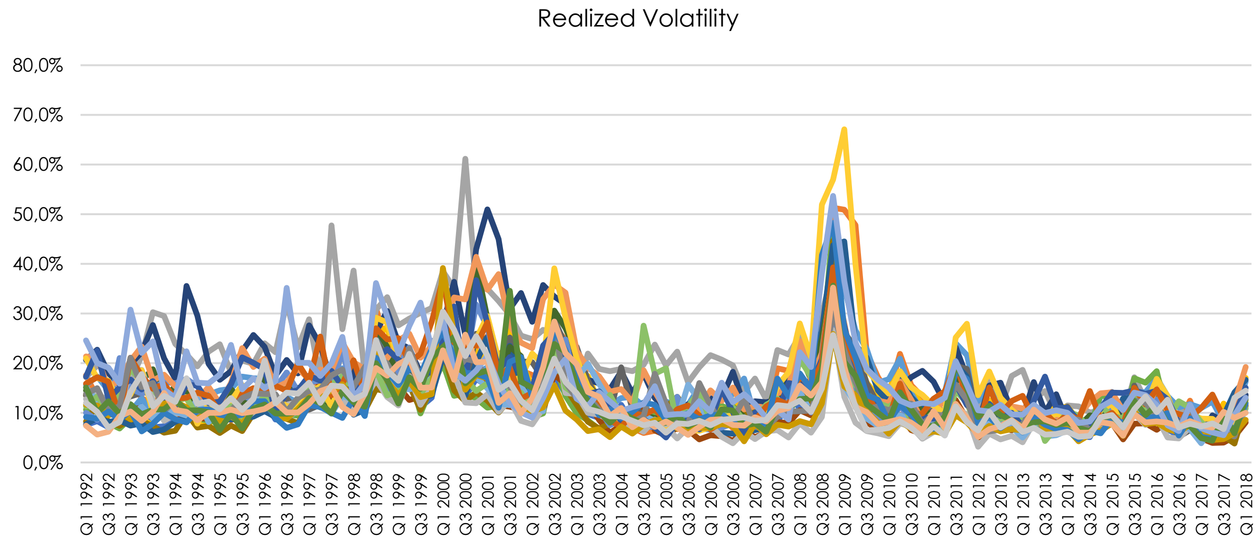

Figure 4.2.7: Realized volatility of the 27 stocks over the full sample period

By visualizing the realized volatility for all the 27 stocks over the period 1992 Q2 to 2018 Q1 in figure 4.2.7, it is seen that the data seems to move in cycles. In general, there seems to be four distinctive trends to be observed from the combined data with one possible new trend at the end of the sample period:

1. From 1992 Q2 to 2000 Q1: Overall volatility is climbing over the period with spikes for a few of the stocks.
2. From 2000 Q1 to 2017 Q4: After the dot-com bubble volatility is trending downwards, such that from 2003 Q3 and to 2007 Q3 there are four years with volatility centered around 10% across the 27 stocks. This period since it exhibits stable volatility could possibly be well described by a linear model.
3. From 2007 Q4 to 2009 Q4: During the financial crisis all the stocks volatility spike to high levels from 2007 Q4 to 2008 Q4, before it declines rapidly again until 2009 Q4. The relatively short period with extreme levels of change in the volatility is most likely not very well suited for a linear prediction model.

4. From 2010 Q1 to 2017 Q4: Following the financial crisis from 2007-2009 the 27 stocks' volatility is becoming increasingly stable and they track each other relatively well from 2013 and onwards. This relatively stable period, with a clear slow downward trend, most likely provides the best fit with a linear model of all the periods.
5. From 2017 Q4 to 2018 Q1: The volatility is trending upwards with the latest data available.

In general, all of the 27 stocks exhibit stable volatility in the period after the 2007 to 2009 financial crisis with very few stocks exhibiting swings in volatility not exhibited in the other stocks. Overall, it would be expected that the stocks are less volatile especially in the years close to the current time, as the companies are goliaths in their respective sectors and are thus established and mature companies.

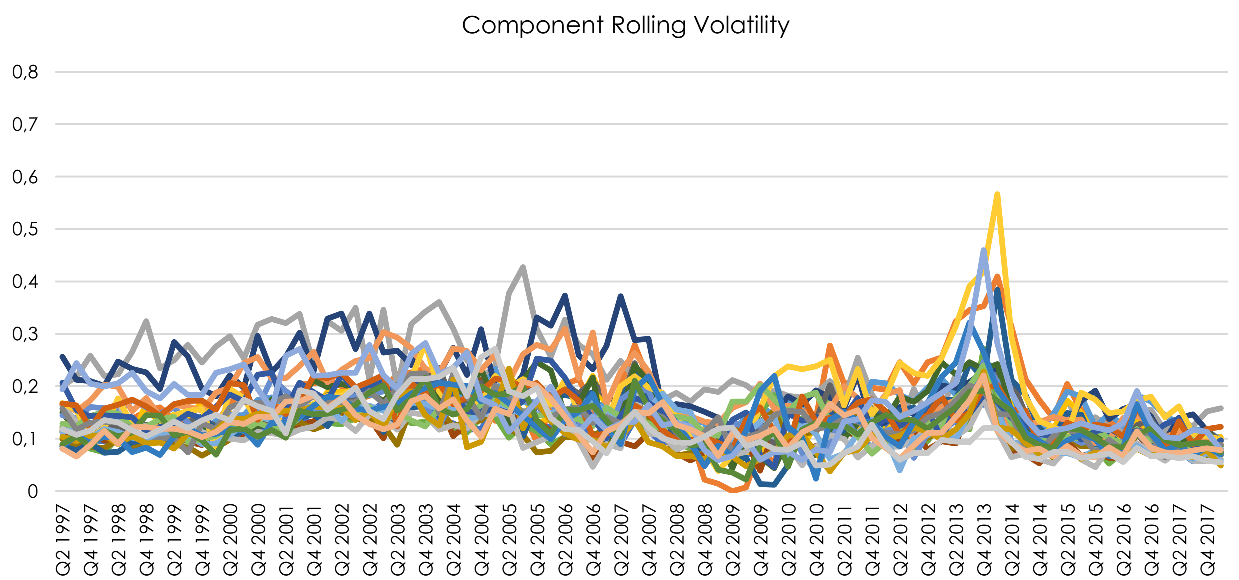

Figure 4.2.8: The forecasted volatility of the Component Rolling model over the full forecasting period for all 27 stocks

Examining the volatility that the two Component models forecast, it can be seen that one of them looks to be more accurate in relation to the realized observed volatility respectively figure 8.2.1 found in the appendix for the Component Expanding and figure 4.2.8 for the Component Rolling model

The Component Expanding model, that is the most accurate model according to the Diebold-Mariano test produces a very linear volatility pattern over the full sample period with very little variation. It is not able to forecast or identify many of the patterns observed in the realized data. The forecast model produces an upwards spike in volatility around 2013 Q3 and a downwards spike in 2014 Q4, which is most likely caused by the model incorporating the observations from the 2008 financial crisis with the effect seemingly being lagged by 20 quarters in the forecast.

The Component Rolling model, that was found to consistently have the largest squared errors and was ranked the least accurate by the Diebold-Mariano test, is able to some extent reproduce the pattern found in the realized volatility. However, the biggest difference is that the model reproduces the pattern observed in the realized data with a 15 to 20 quarter periods lag just like the Component Expanding model. As such the peak of volatility seen in the period 2003 to 2005, is most likely caused by the dot.com bubble around the year 2000. The swings in volatility happening in 20012 to 2014 in the model, is most likely attributed to the 2007 to 2009 financial crisis having a lagged effect on the model.

The Component Expanding model does not capture the variation in the volatility very well and simply produces an almost linear volatility across the full sample period. The Component Rolling model is able to capture the variation in the volatility to a much better degree, however it does so with such a long lag, that the forecasted data is very much out of synchronization with the actual observed data. This explains why the model's performance in the Diebold-Mariano test is extremely poor and forecasts very extreme outliers. If the time horizon is decreased the model may potentially be able to be less out of synchronization and produce forecasts that fit better. On the other hand, it will also be even more susceptible to predict extreme outliers when forecasting based on few observations.

Forecasting Sharpe ratio

Referring to table 3.3.9 that shows the summary statistics of the four regressions using forecasted Sharpe ratios as an estimator for the realized Sharpe ratio. It is found that the two component models had the highest number of significant estimators with 6 for the Component Expanding and 5 for the Component Rolling. When looking at the R^2 of the individual regressions with significant estimators in table 8.3.5 and 8.3.6 in the appendix, it is found that the Component Expanding can explain in the range of 4.0% to 9.3% of the variation in the actual Sharpe ratio and the Component Rolling can explain in the range of 3.5% to 8.3%, which is quite a bit above the average R^2 for each model across all 27 stocks.

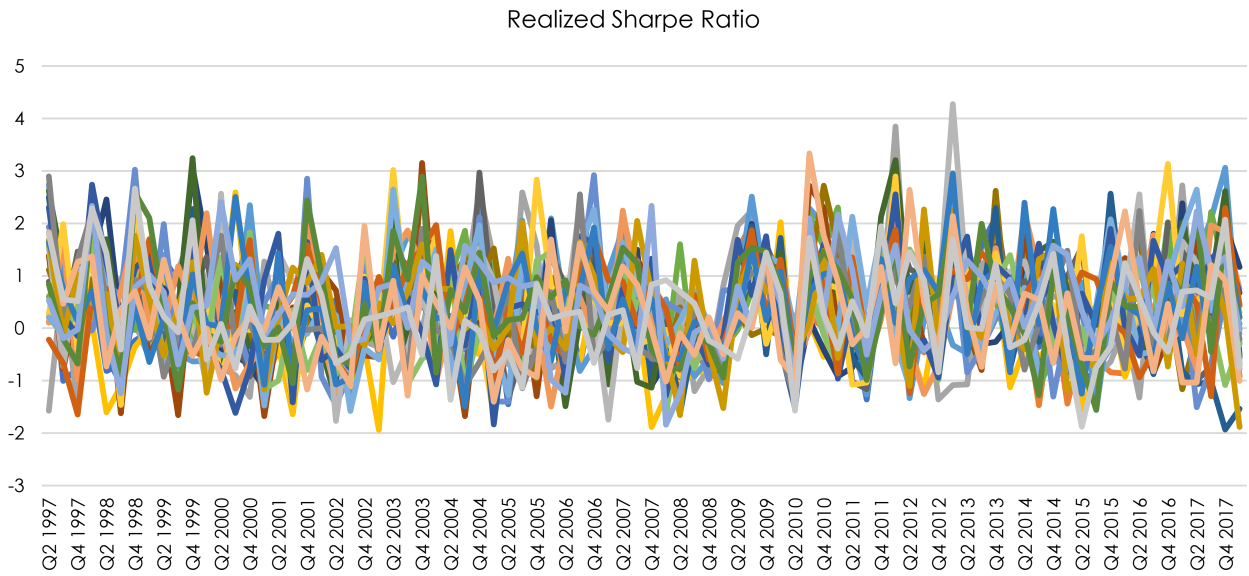

Figure 4.2.9: The realized Sharpe ratio over the full forecasting period for all 27 stocks

Examining the graph of the realized Sharpe ratio over the forecasting period in figure 4.2.9 the pattern is seemingly less linear, because there are no smooth trends in the data like the volatility observations in figure 4.2.7. The Component Expanding model is able to produce a Sharpe ratio with a similar pattern from 1997 Q2 to 2008 Q2 visualized in figure 8.2.3 in the appendix. However, the forecasted Sharpe ratio exhibits a much less volatile pattern with more observations clumped together. After 2008 Q2 the forecasted Sharpe ratio is exhibiting almost no volatility, except for around 2014 which is caused by the destabilization of the volatility due to the 2008 financial crisis. Consequently, the Component Expanding model may be better at correctly predicting Sharpe ratio from 1997 Q2 to 2008 Q2 or at least have the fewest misclassifications during that period. Referring to figure 4.2.6 however there is no visual indication of the Component Expanding model having fewer misclassifications during that specific time period.

The Component Rolling forecasted Sharpe ratio in figure 8.2.5 in the appendix is the second most significant model across all the 27 stocks, but the Sharpe ratio pattern produced is very volatile and has the largest outliers of all the forecasting models. The two direct models face the issue that their forecasted Sharpe ratios are only statistically significant as an estimator of the realized Sharpe ratio 2 times for the Direct Rolling and 1 time for the Direct Expanding as seen in table 3.3.9. Due to the expected imperfect multicollinearity of the model, as shown in the correlation matrix for the full sample explanatory variable in table 3.3.3, the examined partial effects of each estimator are most likely not valid.

The Direct Expanding model seems to suffer from the same problem that the Component Expanding model does when inspecting figure 8.2.7 in the appendix. It looks as if the model initially up until 2004 Q1 is able to produce a pattern that exhibits volatility like the realized Sharpe ratio. Going forward from 2004 Q1 the volatility in the Sharpe ratio drops very sharply. As it happens for both expanding models it seems that those models are not able to correctly incorporate the volatility of the Sharpe ratio when a sufficient amount of data is integrated into the regression. As opposed to the other models, the Direct Expanding model does not identify the major financial crises, as such there is only found minor lagged effects of these. There is a slight increase in volatility around the 2008 Q4 ending in 2010 Q2, which implies that the model's reaction is only lagged by around 2 quarters.

The Direct Rolling model as seen in figure 8.2.8 in the appendix produces a very volatile pattern. Unlike the volatility of the realized Sharpe ratio, the volatility of the Direct Rolling models' Sharpe ratio shows heteroscedastic tendencies. The pattern is not very consistent with the realized Sharpe ratio, especially when examining the data around the dot.com bubble and the 2008 financial crisis. In both historic scenarios the Direct Rolling model is out of synchronization by a lag of around 15 to 20 quarter periods just like both the Component models and to a lesser degree the Direct Expanding model.

Consequently, the models are able to produce some of the patterns exhibited by the realized Sharpe ratio, but are unable to reproduce it in a timely manner. It is likely that the models produce inverted or even nonsensical predictions, because they seemingly are predicting patterns that occurred 15 to 20 quarters ago.

Statistical significance test of realized model portfolios

To be able to make conclusions with regards to the results realized by the forecasting model portfolios, the performance measures need to be tested against each other and the equally weighted Dow benchmark portfolio. The test will specify whether there is a statistically significant difference in the performance, and afterwards if the performance is statistically significantly different from 0. The three different performance measures; excess return, volatility, and Sharpe ratio are the results that will be tested, because a superior performance in any of those parameters is desired ceteris paribus.

Test of difference in mean

The 1-way ANOVA tests all the possible combinations of portfolio pairs against each other and reveals if any of the pairs realize results that are different from one another with statistical significance. The test is computed for the three performance measures chosen above. If the test finds a statistically significant difference in means of either excess return, volatility, or Sharpe ratio between any of the portfolio pairs, a two-sided hypothesis test can be used to determine exactly which of the pairs have different mean values.

The ANOVA test the null hypothesis that all mean excess returns, volatilities, and Sharpe ratios, respectively, are equal across the four forecasting model portfolios and the equally weighted Dow portfolio:

$$H_0: \bar{r}^{excess}_{Rolling\ Direct} = \bar{r}^{excess}_{Expanding\ Direct} = \bar{r}^{excess}_{Rolling\ Component} = \bar{r}^{excess}_{Expanding\ Component} = \bar{r}^{excess}_{DOW\ Equal\ Weighted}$$

$$H_0: \bar{x}^{volatility}_{Rolling\ Direct} = \bar{x}^{volatility}_{Expanding\ Direct} = \bar{x}^{volatility}_{Rolling\ Component} = \bar{x}^{volatility}_{Expanding\ Component} = \bar{x}^{volatility}_{DOW\ Equal\ Weighted}$$

$$H_0: \bar{x}^{sharpe\ ratio}_{Rolling\ Direct} = \bar{x}^{sharpe\ ratio}_{Expanding\ Direct} = \bar{x}^{sharpe\ ratio}_{Rolling\ Component} = \bar{x}^{sharpe\ ratio}_{Expanding\ Component} = \bar{x}^{sharpe\ ratio}_{DOW\ Equal\ Weighted}$$

This is tested against the alternative hypothesis that any of the 10 possible pairs are not equal.

ANOVA summary	p-value of F-test
Excess return	1.00
Volatility	1.00
Sharpe ratio	1.00

Table 4.3.9: p-values of the ANOVA tests of the mean value for excess return, volatility, and Sharpe ratio

The ANOVA summary in table 4.3.9 reveals that the difference in the means of all three realized performance measures: excess return, volatility, and Sharpe ratio are not statistically different between any pairs of the model portfolios and the equally weighted Dow portfolio at a significance level of 10%. The p-value of 1.00 for all three tests indicate that the mean values are identical for all possible significance levels. As such the test will always be unable to reject the null hypothesis stating that the means are equal between all five portfolios.

Given the fact that a key assumption for the ANOVA test is normality, the test results are considered to be more validly interpretable for the output of excess return and Sharpe ratio compared to volatility. This is apparent when examining the histograms of the observed the distribution of excess returns in figure 4.2.3, the volatility in figure 4.2.4 and the Sharpe ratio in figure 4.2.5, because the volatility does not appear to be normally distributed.

Means different from zero

The output from the ANOVA tests produces the necessary statistics to calculate if the excess return, volatility, and Sharpe ratio means are statistically significantly different from 0. To compute if the realized means are different from 0 with statistical significance, a 90% confidence interval is calculated for each performance measure:

$$CI_{\alpha=10\%}^{excess\ return} = \bar{r}_{Portfolio\ x}^{excess\ return} \pm 1.645 * \sigma_{Portfolio\ x}^{excess\ return}$$

$$CI_{\alpha=10\%}^{volatility} = \bar{x}_{Portfolio\ x}^{volatility} \pm 1.645 * \sigma_{Portfolio\ x}^{volatility}$$

$$CI_{\alpha=10\%}^{sharpe\ ratio} = \bar{x}_{Portfolio\ x}^{sharpe\ ratio} \pm 1.645 * \sigma_{Portfolio\ x}^{sharpe\ ratio}$$

Where 1.645 is the critical Z-score value at the 10% significance level.

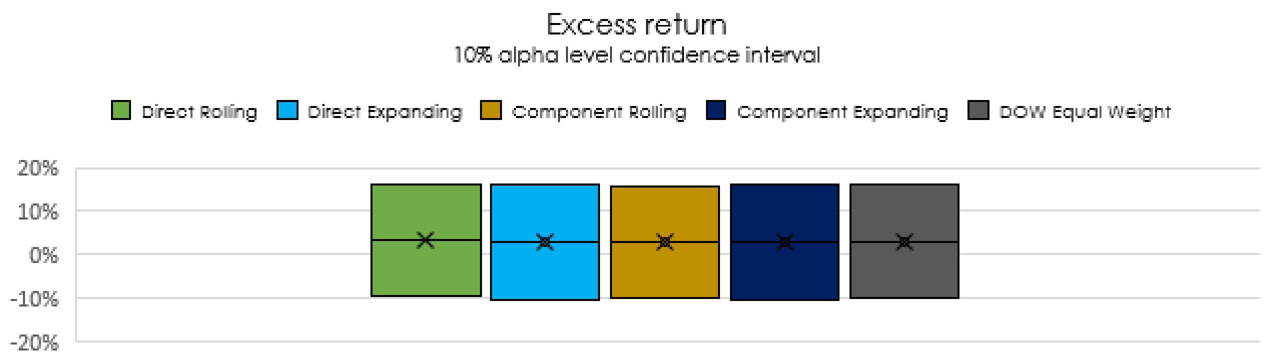

Figure 4.2.10: 90% confidence interval for the mean excess return of all five portfolios

When examining the mean excess returns in figure 4.2.10 across the 5 portfolios, it is apparent that the means are all almost identical. This confirms the results from the ANOVA test, which concluded none of the excess return means were different from the others with statistical significance. Another important observation is that the mean excess return across all 5 portfolios contain the value 0 at a significance level of 10%, which means that it cannot be rejected that the true value of the mean excess return is 0.

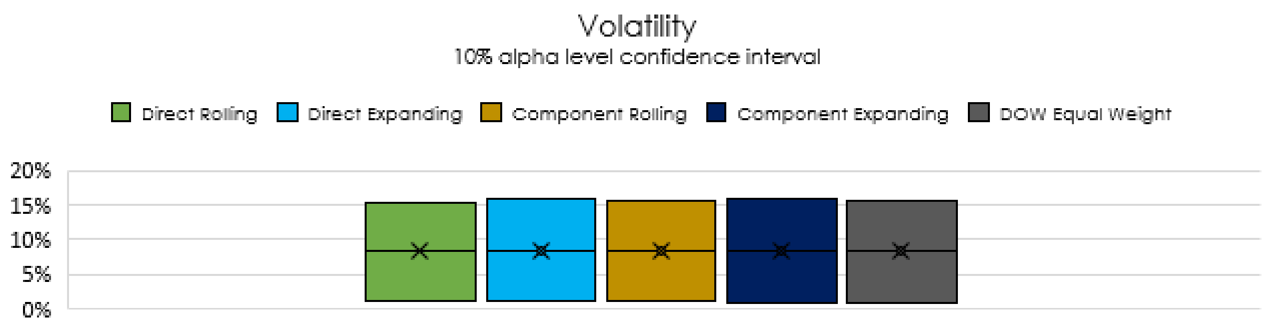

Figure 4.2.11: 90% confidence interval for the mean volatility of all five portfolios

The mean volatility visible in figure 4.2.11 is also depicting that the mean volatility is practically identical across the 5 portfolios, which again is in line with the findings from the ANOVA test. The confidence intervals assume only positive values, as such all the mean volatilities can be said to be different from 0. However, since the volatility data does not fulfill the normality assumptions, which the ANOVA test and confidence interval is dependent on, it is not possible to decisively conclude on the results.

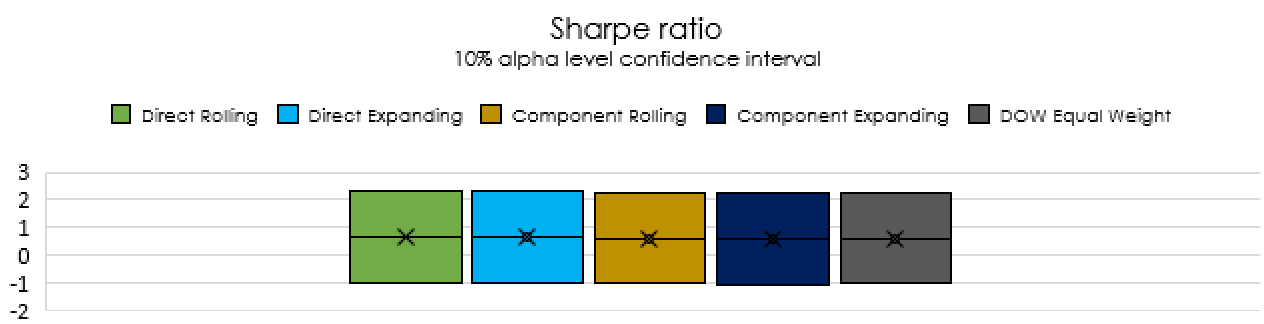

Figure 4.2.12: 90% confidence interval for the mean Sharpe ratio of all five portfolios

Regarding the Sharpe ratio visualized in figure 4.2.12, the interpretation is similar to that of the mean excess return, where the confidence intervals of the 5 portfolios overlap, meaning the mean Sharpe ratio is equal at a significance level of 10%. The confidence interval also overlaps with the value 0, such that the null hypothesis cannot be rejected implying the true mean Sharpe ratio is not different from 0 with statistical significance.

If either the ANOVA test or the confidence interval had shown the Sharpe ratio to be statistically different between any pairs of the 5 portfolios, that would likely have been an indication of a calculation error. Given the fact that the Sharpe ratio is constructed from the two components, excess return, and volatility, if none of the components are significantly different, then the combined product should not be either. As such it is noted that the results of the tests are consistent.

Ideal portfolio comparison

The ANOVA test of the excess return, volatility, and Sharpe ratio of the five different portfolios from table 4.3.9 showed, that there is no statistically significant difference in the mean value of all three different performance measures between any of the portfolios. From table 4.3.3 it was found that the quarterly mean excess return for the portfolios are in the interval of 2.99% for the equally weighted Dow portfolio to 3.27% for the Direct Rolling portfolio. However, when taking the volatility of the excess return into consideration, which is in the range of 7.77% for the Direct Rolling portfolio and 8.15% for the Component Expanding portfolio, it is not possible to prove that the portfolios produce mean excess returns that are different from at a significance level of 10%. Another important finding from table 4.3.3 and table 4.3.5, is that the volatility of the mean excess return and Sharpe ratio for all five portfolios has a numerical value that is higher than their mean value. This means that it is not possible to prove, that the mean values are different from 0, at a 10% significance level. Consequently, it cannot be rejected that the true value of the mean excess return of the Direct Rolling portfolio is 0.0% instead of the sample mean value of 3.27% The same holds true for all of the portfolio mean excess returns and Sharpe ratios.

After testing that all three different performance measures realized by the portfolios are not statistically different on average, the portfolio that executes the trading strategy perfectly is constructed to evaluate the theoretical upper bound of the trading strategy.

Excess return	Direct Rolling	Direct Expanding	Component Rolling	Component Expanding	DOW Equal Weight	Perfect SR Rank
Mean	3.27%	3.04%	3.07%	2.97%	2.99%	7.23%
Standard deviation	7.77%	8.13%	7.83%	8.15%	7.96%	8.11%
Max	25.23%	24.43%	20.96%	23.69%	22.09%	29.30%
75th percentile	7.12%	8.84%	7.57%	8.20%	7.94%	11.13%
Median	3.42%	2.97%	3.41%	3.17%	3.16%	7.52%
25th percentile	-0.67%	-1.23%	-1.63%	-0.84%	-1.06%	3.06%
Min	-18.30%	-17.62%	-18.11%	-21.09%	-19.20%	-15.78%

Table 4.3.10: Summary statistics of the excess return for the five portfolios and the portfolio that always ranks correctly

Located in table 4.3.10 are the summary statistics for the mean excess return for all the five portfolios and the portfolio that predicts Sharpe ratio perfectly. As a result, it correctly ranks and consistently invests in the 20 stocks with the highest Sharpe ratio. The Perfect Rank SR portfolio realizes a mean excess return of 7.23%, which is 1.2 times as high as the second largest at 3.27%. The volatility of the return of the Perfect SR Rank portfolio is in the range of the other five portfolios, and the numerical value of the excess return volatility is still higher than the numerical value of the mean excess return. Implying that at a significance level of 10%, it cannot be rejected that the true mean excess return of the Perfect SR Rank portfolio is 0.0%. However, the result is heavily skewed towards being positive, given that the 90% confidence interval would cover the values:

$$7.23\% \pm 1.645 * 8.11\% = [-6.11\% \,; 20.56\%]$$

As such at a significance level of 10% it cannot be rejected that the true mean excess return is -6.11%, but also cannot be rejected that it is 20.56%.

Table 8.3.3 in the appendix show that the Perfect Rank SR portfolio produces a mean volatility that is similar to the mean volatility that the 5 portfolios already analyzed produce. Table 8.3.4 in the appendix shows that the Perfect Rank SR portfolio produces a mean Sharpe ratio of 1.18, whereas the second best the Direct Rolling portfolio produces a Sharpe ratio of 0.66. The Perfect Rank SR portfolio thus has a Sharpe ratio that is 1.8 times higher than the second best. To be able to conclude if there is a statistically significant difference between the results, all of the mean values are tested using an ANOVA test.

ANOVA summary	p-value of F-test
Excess return	0.00
Volatility	1.00
Sharpe ratio	0.00

Table 4.3.11: p-values of the ANOVA tests of the mean value for excess return, volatility, and Sharpe ratio including the Perfect Rank SR portfolio

The ANOVA test summary in table 4.3.9 reveals three important results when comparing it to the results of the ANOVA in table 4.3.11:

1. There is a statistically significant difference in the mean excess return between at least 1 pair of portfolios with a p-value of 0.00.
2. There is still no statistically significant difference in the mean volatility between the 6 different portfolios with a p-value of 1.00.

3. There is a statistically significant difference in the mean Sharpe ratio between at least 1 pair of portfolios with a p-value of 0.00.

Touching on the first point; from the prior ANOVA tests in table 4.3.9 there was no statistically significant difference in the mean excess returns of any pairs of the 4 model portfolios and the equally weighted Dow portfolio. The same is true for the third point, that there was no statistically significant difference in mean Sharpe ratio. Consequently, it can then be deduced that since the test in table 4.3.11 show existence of pairs with a statistically significant difference in mean value, these pairs are all the previous five portfolios paired with the Perfect SR Rank portfolio. The p-value of 0.00 indicates that the null hypothesis is rejected for all possible significance levels.

The second point regarding mean volatility shows nothing new compared to the ANOVA test in table 4.3.9. There is no indication that any of the 6 portfolios realize a mean volatility that is different from any of the other portfolios with statistical significance. The p-value of 1.00 indicates the test is unable to reject the null hypothesis, that all mean volatilities are equal for all possible significance levels.

Summarizing the results from the ANOVA test yields that the strategy of ranking stocks by their predicted Sharpe ratio has the potential to produce mean excess returns and a Sharpe ratio, that is different from and better than the equal weighted portfolio with statistical significance. However, all the portfolios are expected to realize the same mean volatility.

Discussion

In this section we first discuss the potential shortcomings of the forecasting models and trading strategy. These are centered around overlooking a potential data seasonality, linear model shortcomings, forecast methodology, omitted variable bias and issues with forecasting the Sharpe ratio components. Secondly, we look to potential problems regarding the implementation of the models and trading strategy in a real-life scenario. These involve the problems related to sourcing earnings data in a timely manner and trading costs.

Model shortcomings

Since models are based on assumptions, they can only imperfectly reflect the real world, and often these assumptions are not congruent with the real world. The following section discusses some of the imperfections in the forecasting models used in the thesis.

Seasonality

By using a linear model fitted to past historical data across different quarters, potential seasonality in the data might be missed when using 5-year data in rolling models or a broader time window in the expanding models. Different times during the year might be of different significance to the dependent and independent variables of the multiple regression model. One hypothetical example regarding seasonality of the volatility could be that certain fiscal quarters carry more weight in terms of comparing investor expectations against reality. For example, the Christmas quarter might be very important for retailers, such as Walmart, and depending on the resulting performance there might be a spike in volatility, as investors assign more significance to results in this quarter and therefore are more prone to make big portfolio adjustments based on those results. Also, the fourth and final fiscal quarter might carry more importance than the others, since it marks the end of the fiscal year and it is possibly assigned more importance. Although seasonality effects might be present in the independent variables or dependent variable, it is not believed to have a significant effect. This is due to the nature of returns being based on investor expectations, that are adjusted randomly over the year as new data becomes available about the prospects of the companies.

In any case using a multiple linear regression model that does not take seasonality into account might lead to an oversimplification, that results in erroneous estimators and therefore false predictions.

Linear forecasting

While the estimators in the multiple regressions for the entire period did not show statistical significance with regards to forecasting the realized performance measures excess return, volatility, and Sharpe ratio, it might be possible that the estimators are significant in some of the individual regressions used to forecast the quarters. This is due to the fact that the only time the full data set is used in a regression with the actual forecasting models, as opposed to testing the entire period for significance, is with the two expanding models when forecasting the very last period in the sample 2018 Q1. As such the regressions based on different time periods may have significant estimators. The total quarterly regressions over the entire forecast period is 84 quarters times 27 stocks times 6 different regression models, equaling 13,608 – the 6 different regression models consist of excess return, volatility, and Sharpe ratio with both rolling and expanding time windows. Considering there is a total of 13,608 regressions between the four Sharpe ratio forecasting models it would be extremely difficult to interpret the statistical output in any aggregated form.

While the explanatory variables did not meet statistical significance levels, this only means that they were not able to reliably explain the dependent variable in a linear fashion. They might still be able to provide insights and be statistically significant in a non-linear model, in case they contain information that are useful for predicting future returns and volatility.

Lack of statistical significance

Referencing the literature review Tang and Whitelaw (2018) were able to find statistical significant estimators for excess returns, volatilities, and Sharpe ratio, however their chosen estimators were different than ours, as they used the variables; lagged volatility, the Baa-Aaa spread, the dividend yield, the one-year Treasury yield, and the commercial paper Treasury spread. They measured the effect on a large diversified stock index instead of individual stocks. Consequently, it is not possible to draw any parallels between their and our findings. This is due to the fact that a different data set and time period is used.

''On an out-of-sample basis, using l0-year rolling regressions, estimated conditional Sharpe ratios again show statistically and economically significant predictive power for realized Sharpe ratios." (Tang & Whitelaw, 2018)

They do however prove that a statistically significant relation can be found between their predicted Sharpe ratio and realized Sharpe ratio, which means that the realized Sharpe ratio must exhibits tendencies that makes it predictable with a linear model. However, while this is true for their model, it might only be significant for the specific data which they have used with the exact same time window. As such it is very hard to generalize these findings without being able to recreate the results on a wide range of data sets and across time.

Moskowitz, Ooi, and Pedersen (2011) predict the excess return of different financial instruments using only the financial instrument's lagged excess returns, where both variables are adjusted for their ex-ante volatility. Their regression is found as formula 1.1.1. They note that the regression results are similar without adjusting for the financial instrument's volatility. The regression on equity index futures shows that the lagged excess return is only significant when the lag specification is 4 and 9 months respectively. They are unable to find a consistently statistically significant relation between excess return and lagged excess return using a linear model. This is true even with a diversified data set, given that the equity data they use is based on 9 different developed equity indices. They specifically do not use the regressions coefficients to forecast, instead they use the t-statistics of the regression coefficients to show the relations of the momentum effect on the excess returns for different time windows.

Omitted variable bias

When deciding upon which explanatory variables to use, we are also making the decision to exclude all other variables from the forecasting models. The problem with this exclusion is that these variables might be determinants and have a significant explanatory power over the dependent variable in a linear model. This problem is known as omitted variable bias. It is defined as:

"Omitted variable bias is the bias in the OLS estimator, that arises when the regressor, X, is correlated with an omitted variable." (Watson & Stock, 2011).

As such we are fully aware that the forecasting models will suffer from omitted variable bias and will most likely be able to be improved by including other explanatory variables. Some of these potential variables can be found in the literature. Classic explanatory variables could be taken from the capital asset pricing model, the liquidity adjusted capital asset pricing model or Fama French 3 factor model.

Forecasting both components of the Sharpe ratio

When forecasting both the excess return and the volatility separately to calculate the expected Sharpe ratio there is the issue of the prediction being inconsistent with theoretical possibilities. This because the forecasted variables are not independent of each other due to the fact that volatility is derived from the return and as such there has to be consistency between the two. When these two variables move out of synchronization of each other, such that the variation in the forecasted excess return does not at all correspond to the forecasted volatility, the calculated Sharpe ratios can be extreme. The caveat here is that since the return is related to the first and last day in a period, while volatility is based on the inter-day returns, there can be seemingly big differences which are consistent moves between the two measures.

The differences become quite apparent when examining figure 8.2.4 showing the forecasted excess return, 8.2.5 showing the forecasted Sharpe ratio found in the appendix, and figure 4.2.8 showing the forecasted volatility for the Component Rolling model. It can be seen that there seemingly is no consistency between the swings in the excess return and the increase or decrease in volatility. It is very clear around the period 2012 Q4 to 2014 Q4, where the model seemingly makes a lagged prediction of the 2008 financial crisis when looking at the forecasted volatility. However, the excess returns in that period exhibit little volatility, except for one single outlier across all 27 stocks. As a result, the forecasted Sharpe ratios for that period are being pushed downwards as an effect of excess return being stable while the volatility spikes.

The excess return metric can of course be below -100% when the corresponding risk-free rate is subtracted from the return, but the gross return of a stock cannot. This leads to another inconsistency with financial theory, which is the fact that a linear model can forecast the excess return to be far below -100%, such that gross return is implied to be less than -100%. This is impossible because the value of a stock cannot carry a negative value - that is shareholder losses are limited to the capital invested. The linear model forecasts negative returns that are not at all realizable, however given that there is no interpretation of the realized values for any of the financial measures, the specific value that the models forecast is not of interest. The only interest is in the model ranking the stocks correctly, which means that the model can be wrong by any factor if the errors are systematic across the stocks, enabling correct relative ranking.

Taking the above into account would favor the approach of using a direct forecast model, as the values generated by this model always would have the total effect of excess returns and volatility integrated. This model cannot generate mathematically impossible values, since Sharpe ratios theoretically can assume the value of any real number, as opposed to volatility and gross returns, which both have mathematical limitations. Alternatively, a rule would have to be implemented in the case of negative volatilities and gross returns lower than -100%, but this could potentially affect the ranking procedure depending on the chosen rules.

Implementation

There are a many of problems that can occur when implementing the proposed trading strategy in a real-world scenario. In this section we explore some of those and discuss their potential implications.

Earnings per share release timing

In the analysis a very problematic discovery was made regarding both earnings per share measurements, namely that the availability of the data is not consistent across all stocks. This leads to an implementation problem, since the model cannot be provided new data related to all stocks with the same timing every quarter.

Consequently, the earnings-related coefficients are not correctly implemented. The model can be modified in either of four ways to accommodate reality:

1. The trading window of the model is changed such that all the stocks have released their 10-Q form before the regressions are computed and the rankings are made. The difficulty and complications of this method is that the companies do not all release the information at the same time and as such there is a big difference between JP Morgan Chase & Co using 12 days and Walt Disney using 37 days to disclose their most recent quarterly earnings found in table 8.3.1 in the appendix. Another element that makes it even more complicated is that companies are not consistent in the time lag between the end of the fiscal quarter and the public release of their earnings, which means that the frequency between announcement of recent earnings is not consistent over time. Chan, Jegadeesh, and Lakonishok (1996) find that there is a spread between those disclosing the worst and best quarterly earnings of 2.4% when the information is initially released. Waiting to trade on JP Morgan Chase & Co's earnings for 25 days, before all the companies have reported their quarterly earnings, might mean that the models cannot take advantage of the value from that information. In the worst case scenario, it is found that the information shows far superior earnings than anticipated and as such the stock price soars. When trades on this information are made 25 days later, it is possibly buying at a peak. Another complication is that the four stocks that do not follow the calendar quarters; Home Depot, Walmart, Nike and Cisco Systems would be more difficult to fit into the trading window and for any given combination the models would not be trading on the most recent earnings information for all of the companies.
2. Alternatively, the model can instead use earnings per share lagged by 2 quarters, such that it is that the information is available at the time of trading. The research by Chan, Jegadeesh and Lakonishok

(1996) found that the surprise factor in the earnings was carried forward throughout the following two earning announcements, but that the spread between the worst and the best was reduced to 0.8%, which means that 1.6% points of the 2.4% spread is gone. Given their research included all American stocks listed on NYSE, AMEX, and Nasdaq the portfolio they construct is much more diversified than ours and as such we cannot hope to replicate their results. However, there is no telling in advance what the result would be without implementing the actual alternative model. Given our dataset is only a subset of theirs and that it is based on a different sample period, we would most likely not arrive at the same conclusions.

3. As the third option the 12 months trailing and 12 months forward consensus earnings per share can be sourced with a daily instead of a quarterly frequency. The model can then be set up the model to take into consideration, that it should only execute trades when there is a change in the earnings between day *t* and day *t-1* as that would indicate that new earnings have been made available to the public for the given stock. This change however would require an alternative trading strategy from the one implemented in the thesis. The models would not be able to rank all the stocks simultaneously given that earnings are released with different frequencies and at different times.
4. The earnings per share as an estimator in the models could be completely disregarded. The regressions can be run with fewer estimators or other financial variables that have been found useful in predicting excess returns, volatility, or Sharpe ratio in the literature, such as lagged volatility used by Tang and Whitelaw (2018).

Trading costs

Aside from the technical difficulties with EPS, there are several important differences between the way the thesis has implemented the models and how they could be implemented in real life. One of these is that in the real world there are costs associated with making trades. These costs are the direct transaction fee paid to the broker as well as any potential order book spread paid.

The nature of transaction costs varies depending on the relevant trading strategy, the exchange, and the trade size. All trading strategies will eventually have increasing transaction costs when scaling them due to market impact. Pedersen (2015) identifies three types of transaction cost behavior: increasing transaction costs, constant transaction costs and decreasing transaction costs. Increasing transaction costs are usually due to market impact for large investors, since the buying or selling activity is enough to move prices. Constant transaction costs are when the average costs are constant for different trade size, usually in this scenario it is possible to trade the full amount within the order book, perhaps because a lot of market makers find the spread attractive. Decreasing order costs are usually seen in over-the-counter markets where the dealers charge higher percentage transaction costs for smaller orders than for larger orders.

Implementation of the thesis strategy is expected to experience constant trading costs when employing smaller amounts of capital, since the chosen stocks are very large and liquid. The strategy would most likely start to exhibit increasing trading costs once a larger amount of capital is committed to the strategy, since it would start to have a market impact. The way the portfolios are implemented assumes a constant transaction cost of zero.

The trading costs are an important consideration since the ideal amount of trading activity is about striking a balance between having the right portfolio allocations and saving on transaction costs.

The differences between the real-life portfolio and the theoretical paper portfolio comes down to two things: trading costs and opportunity costs. Where the opportunity costs are the costs incurred from having a different portfolio composition than the ideal paper portfolio that results in missing gains. The sum of these differences is called implementation shortfall, which is also the difference between the paper portfolio and the real-life portfolio:

$$Implementation\ shortfall\ (IS) = Trading\ costs + Opportunity\ costs$$

$$IS = Paper\ portfolio - \ Real\ portfolio$$

The goal for any given trading strategy is to scale it such that the implementation shortfall is as small as possible.

The models have the assumption that it is possible to trade without paying a transaction fee and that it is possible to transact at the last observed trade price. There is also the issue of the strategy being able to trade at use last quarters closing prices as the following quarters opening prices.

Conclusion

This thesis fails to find a consistently statistically significant linear relation between the 12-month forward and 12-month trailing earnings per share, the lagged 3-month excess return, the lagged 6-month excess return, the lagged 9-month excess return, and the lagged 12-month excess returns as estimators to forecast either the excess return, volatility, or Sharpe ratio. The empirical evidence of the thesis is based on 27 constituents of the Dow Industrial Average Index as of 28-03-2018 with the sample data from 28-03-1991 to 26-03-2018. The lack of a consistent statistically significant relation out of sample using either of four model approaches to forecast the Sharpe ratio is apparent in the classification errors of all the models. Furthermore, the portfolios based on ranking of the forecasted Sharpe ratios, realize the performance measures: excess return, volatility, and Sharpe ratio, which cannot be proven to be statistically different from each other or the equally weighted Dow portfolio at a 10% significance level.

Optimal implementation

In response to problem definition sub question 1 and 2, regarding which of the four suggested implementations is most optimal, the methods have been tested extensively with the tests yielding contradictory interpretations.

The Diebold-Mariano tests for a statistically significant difference in the models squared errors. The squared errors produced by the four models when predicting Sharpe ratio were compared in pairs against each other. The test found that for the majority of the 27 regressed stocks the four models produced significantly different squared errors. The Component Expanding and Direct Expanding models are the pair of models with the smallest squared errors, which are only statistically different for 3 of the 27 stocks. When examining the relative size of the squared errors of those 3 stocks the Component Expanding model is found to produce the smallest squared errors for all 3 and is as such marginally better than the Direct Expanding model in accuracy. The ranking of the models according to accuracy in squared error terms is thus the following according to the size of the squared errors in table 3.3.8:

1. Component Expanding
2. Direct Expanding
3. Direct Rolling
4. Component Rolling

Where the ranking between the two most accurate is marginally different with the third significantly behind and the fourth significantly far behind in terms of accuracy compared to the other three. The Diebold-Mariano test results are only relative and simply measures the size of the squared errors. Consequently, the Component Expanding model produces the smallest squared errors on average, although this does not mean that it is the best model to explain the variation observed in the realized Sharpe ratio.

The individual linear regressions with the forecasted Sharpe ratio for each of the four models regressed on the realized Sharpe ratio tests how much of the variation in the actual Sharpe ratio the models can predict. The regression results reveal that the models can produce statistically significant estimators when predicting the realized Sharpe ratio for 1 to 6 out of the 27 stocks. Ranking the models by their ability to significantly predict the actual Sharpe ratio by the output from table 3.3.9 yields the following:

1. Component Expanding: Statistically significant prediction of 6 stocks' Sharpe ratio with an average R^2 of 1.7%
2. Component Rolling: Statistically significant prediction of 5 stocks' Sharpe ratio with an average R^2 of 1.3%½
3. Direct Rolling: Statistically significant prediction of 2 stocks' Sharpe ratio with an average R^2 of 1.3%
4. Direct Expanding: Statistically significant prediction of 1 stock's Sharpe ratio with an average R^2 of 0.7%

The R^2 statistics that Tang and Whitelaw (2018) are able to produce in their study are in the range of 0.12% to 0.71% with their expanding models and 0.42% to 1.21% with their rolling models. Taking this into consideration the average R^2 statistics of the four models are in line with their findings, although the relative model type ranking is different. The model in their study that can explain most of the variation in the actual Sharpe ratio is a rolling model, whereas it is an expanding model in this thesis. The main takeaway from the test is the issue regarding lack of statistical significant predictions, which will lead to false predictions of the Sharpe ratio.

The false predictions are visualized in the table 4.3.7, where it can be seen that the mean number of classification errors is in the range of 10 to 10.83 out of a possible 14. Here classification errors refer to when a model either falsely includes or excludes a stock from the 20 stock portfolio for a given investment period, since it fails to correctly predict the relative Sharpe ratio ranking of all the 27 stocks. Plotting the number of misclassifications of the four models over time seen in figure 4.2.6 does also not provide any indication of there being certain time periods, where either of the models are consistently able to rank correctly.

The lack of statistically significant prediction power can be explained by looking at the regressions performed on the full sample data set. The results from these regressions in table 3.3.1 reveal that the chosen model estimators are only statistically significant for 1 stock when predicting Sharpe ratio and 4 stocks when predicting excess return at a 10% significance level. For volatility the estimators are significant for 13 of the stocks. Even in case of volatility, which has the most explanatory power, the adjusted R^2 is still only 7.9%, leaving most of the variation unexplained. The lack of explanatory power is consistent with the high number of misclassifications, because the Sharpe ratios will not be predicted with statistical significance.

Examining the pattern of the forecasted excess return, volatility, and Sharpe ratio reveals that the models poorly track the pattern of the realized variables. There are two big events that significantly affects the observable patterns during the data sample: The dot.com bubble and the 2008 financial crisis. Looking at the patterns of the models in the graphs, it seems as if the effect of the 2008 financial crisis is replicated by the Direct Rolling, Component Rolling and Component Expanding models with a significant lag of 15-20 periods. Consequently, these models predict that the volatility levels experienced around 2008 happens in 2012, which further confirms they are unable to significantly predict the realized excess return, volatility, and Sharpe ratio.

Furthermore, it was discovered that the release timing of the earnings data for the sample companies made the current implementation method using a time lag of 1 quarter unfeasible. The earnings are on average released to the public after the portfolios are formed, which means an alternative implementation of earnings is necessary or that the estimator should be replaced with an omitted variable.

Statistically different performance

An ANOVA F-test is used to evaluate the realized performance of the four forecasting model portfolios compared to each other and the equally weighted Dow portfolio. As seen in table 4.3.9 all three realized performance measures: mean excess return, mean volatility, and mean Sharpe ratio are not statistically different between any of the portfolios.

Since the model portfolios were not able to realize results that were statistically different from the equally weighted Dow portfolio, a Perfect Rank SR portfolio was created. The Perfect Rank SR portfolio is the portfolio that always predicts the realized Sharpe ratio correctly, and as such is able to include and exclude the correct stocks in all periods. The new portfolio is included in a new ANOVA F-test seen in table 4.3.11. The results reveal that the Perfect Rank SR portfolio is able to realize both a mean excess return and a mean Sharpe ratio, that is statistically significantly better than all the other portfolios. This implies that if it is possible to obtain more accurate forecasts, resulting in a better classification of relative stock ranks, then better performance can follow using the ranking method. Believing it possible to attain perfect forecasts is naive, however the performance of a model that produce predictions that are relatively better than the four thesis models and relatively worse than the Perfect Rank SR portfolio can possibly still be significantly better.

Summarizing the findings for each of the four models, Direct Rolling, Direct Expanding, Component Rolling, and Component Expanding, it is not possible to decisively conclude that any of them are the more optimal implementation of the thesis' strategy. This indecisiveness is attributable to the facts that while the Component Expanding model is found to significantly produce the least amount of squared errors and the highest number of significant estimators when regressing predicted Sharpe ratio on realized Sharpe ratio across all the 27 stocks, the ensuing realized performance is still not significantly different from that of the other three model portfolios nor the equally weighted Dow portfolio.

In conclusion to the problem statement regarding how a trading strategy that ranks and picks stocks based on their forecasted Sharpe ratio compare to an equally weighted portfolio, it was found that using the implementation methodologies discussed in this thesis, the trading strategy is not able to realize performance measures that are different with statistical significance. This is found for realized mean excess returns, mean volatility, and mean Sharpe ratios. The lack of significant results is attributed to the lack of linear relations between the dependent- and independent variables, resulting in forecasts that are statistically insignificant.

Outlook

It was not possible to demonstrate systematic and statistically significant linear relations across the different forecasted variables in this thesis. The volatility forecasts had the most significant coefficients but a very low explanatory power. It was shown that an ideal portfolio that performs the stock ranking based on the realized Sharpe ratio, would outperform the equally weighted Dow benchmark portfolio. The outperformance was statistically significant for excess return and Sharpe ratio. As such it would be attractive to be able to forecast more accurately and there are several ways to possibly do so.

We propose two general approaches to model changes, that could be used to potentially achieve better forecasting accuracy and performance results, while still using the overall methodology. These are, using a different forecasting model than a linear regression or changing the parameters of the model. Both can be performed in conjunction, giving rise to a lot of different potential implementations from the one in this thesis.

Instead of using a linear ordinary least squares regression, a non-linear model could be implemented as this might be better suited to predict the relations between the explanatory variables and dependent variables. Using an artificial neural network like Satinover and Sornette (2008) but using the methodology in this thesis with the same explanatory variables and ranking procedure, could be promising since it can model more complex relations between variables by tweaking the number of layers of neurons. Unfortunately, it does not come with the same interpretability of coefficients as a linear regression, since the predictions are based on adding a bias and multiple weight vectors being multiplied for each neuron layer. Another non-linear approach could be to transform the explanatory variables used in the linear regression as with a log transformation, this would allow for the ordinary least squares model to still be used, providing interpretable output for the estimators.

There are several different model parameters that could be changed which could potentially deliver different results. We have identified the following; data set, explanatory variables, model time window, rebalancing frequency, exclusion fraction, and weighting method.

Using a different data set might provide different findings. This could be due to using a bigger sample size than the 27 in our implementation, or because other stocks might have different characteristics that make it such that they are predictable using the model methodology outlined in this thesis.

Since omitted variable bias is a potentially large bias in the four forecasting models used in this thesis, using different or adding new explanatory variables should most likely result in better predictions. From table 3.3.2 it can be seen that the intercepts are much more likely to be statistically significant indicating omitted variable bias. Inspiration can be found in the literature, potential new implementations could include the Fama French high-minus-low and small-minus-big factors, the traditional systematic risk beta from the capital asset pricing model, the liquidity adjusted capital asset pricing model, the factors used by Tang and Whitelaw (2018), or an earnings surprise factor as Satinover and Sornette (2008).

When computing the forecasting model parameters, a larger or smaller sample size than the 20 that were used in this thesis could potentially provide different results. A smaller time window for a rolling time window forecasting model could maybe better reflect changing market patterns, while a longer time horizon could possibly reflect longer term trends.

The quarterly rebalancing chosen in this thesis were, as discussed earlier, a reflection of earnings data availability, but a different rebalancing frequency could potentially lead to different results. By rebalancing more often and thereby updating the model more frequently, the data used in the model would be more novel, which might lead to better results as a outcome of initial underreaction as described by (Pedersen, 2015).

In this thesis it was chosen to exclude 7 of the 27 stocks from the Dow Jones Industrial Average. But this could also be attempted with different proportions. Obviously, the discrepancies between the resulting model portfolio and the benchmark will become larger as the number of stocks excluded is increased, but if the model is accurate this also means that performance could potentially improve more.

Lastly the implementation in this thesis used an equally weighted approach for determining the portfolio weights once the ranks was assigned. There are many ways to chose portfolio weights such as a value weighted approach, larger weights to favorably ranked stocks, a significance weighted approach, etc. While these will not affect the accuracy of the forecasts and the ranking procedure they may improve performance by assigning more capital to the better performing stocks.

As such, going forward, there are many possible changes that can be made to the forecasting model parameters, alternatively different forecasting models than used in this thesis, such as non-linear transformations or artificial neural network can be implemented instead. Making these changes will result in new forecasted Sharpe ratio which are then ranked to construct portfolios, that can possibly realize statistically better results than an equally weighted portfolio.

Bibliography

Apple. (2018, 04 03). *FAQ*. Retrieved from Apple: http://investor.apple.com/faq.cfm

Bodie, Z., Kane, A., & Marcus, A. J. (2014). *Investments.* McGraw-Hill Education.

Chan, L. K., Jegadeesh, N., & Lakonishok, J. (1996). Momentum Strategies. *The Journal of Finance*, 1681-1713.

CRSP. (2018, 03 15). *CRSP Historical Indexes*. Retrieved from CRSP: http://www.crsp.com/products/research-products/crsp-historical-indexes

Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 253-263.

Hurst, B., Ooi, Y. H., & Pedersen, L. H. (2017). A Century of Evidence on Trend-Following Investing. *The Journal of Portfolio Management*, 2-16.

Jegadeesh, N., & Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market. *The Journal of Finance*, 65-91.

Jegadeesh, N., & Titman, S. (1999). *Profitability of Momentum Strategies: An Evaluation of Alternative Explanations.* National Bureau of Economic Research.

Moskowitz, T. J., & Grinblatt, M. (1999). Do Industries Explain Momentum? *The Journal of Finance*, 1249-1290.

Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2011). Time Series Momentum. *Journal of Financial Economics*, 228-250.

Newbold, P., Carlson, W. L., & Thorne, B. (2013). *Statistics for Business and Economics.* Pearson Education.

Pedersen, L. H. (2015). *Efficiently Inefficient.* Princeton University Press.

Ruppert, D. (2011). *Statistics and Data Analysis for Financial Engineering.* Springer Science+Business Media.

Satinover, J. B., & Sornette, D. (2008). *Anomalous Returns in a Neural Network Equity-Ranking Predictor.* Swiss Finance Institute.

SEC, U. (2018, 04 02). *Form 10-Q*. Retrieved from Securities and Exchange Commision: https://www.sec.gov/fast-answers/answersform10qhtm.html

Tang, Y., & Whitelaw, R. F. (2018, 03 05). *Time-Varying Sharpe Ratios and Market Timing.* Retrieved from SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1938613

Watson, M., & Stock, J. H. (2011). *Introduction to Econometrics.* Pearson Education.

Yahoo. (2018, 04 20). *Finance*. Retrieved from Yahoo Finance: https://finance.yahoo.com/

Appendix

Stock	Days to release Form 10-Q
3M	23
AMERICAN EXPRESS	22
APPLE	30
BOEING	29
CATERPILLAR	26
CHEVRON	26
CISCO SYSTEMS	15
COCA COLA	26
WALT DISNEY	37
EXXON MOBIL	26
GENERAL ELECTRIC	22
HOME DEPOT	14
INTERNATIONAL BUS.MCHS.	17
INTEL	25
JOHNSON & JOHNSON	16
JP MORGAN CHASE & CO.	12
MCDONALDS	29
MERCK & COMPANY	30
MICROSOFT	25
NIKE B	21
PFIZER	30
PROCTER & GAMBLE	29
TRAVELERS COS.	22
UNITED TECHNOLOGIES	29
UNITEDHEALTH GROUP	19
VERIZON COMMUNICATIONS	23
WALMART	15
Average	23.63

Table 8.3.1: Days used by the 27 companies before releasing their most recent earnings

Direct Expanding vs Component Expanding	Diebold-Mariano p-value	Direct Expanding Squared Errors	Component Expanding Squared Errors
3M	0.00	104.5	79.2
AMERICAN EXPRESS	0.35	91.9	196.7
APPLE	0.29	146.7	135.1
BOEING	0.07	116.7	102.1
CATERPILLAR	0.99	112.9	113.1
CHEVRON	0.65	82.5	79.9
CISCO SYSTEMS	0.42	95.6	89.4
COCA COLA	0.80	101.9	105.0
WALT DISNEY	0.04	86.9	71.8
EXXON MOBIL	0.35	66.1	62.3
GENERAL ELECTRIC	0.46	117.9	104.0
HOME DEPOT	0.35	122.3	106.6
INTERNATIONAL BUS.MCHS.	0.20	92.6	109.5
INTEL	0.52	92.2	86.5
JOHNSON & JOHNSON	0.66	99.4	96.1
JP MORGAN CHASE & CO.	0.97	112.7	112.2
MCDONALDS	0.75	91.6	89.4
MERCK & COMPANY	0.16	100.2	73.4
MICROSOFT	0.43	99.6	108.2
NIKE B	0.78	81.4	78.3
PFIZER	0.41	89.1	80.9
PROCTER & GAMBLE	0.22	88.9	76.3
TRAVELERS COS.	0.24	90.7	75.7
UNITED TECHNOLOGIES	0.47	100.5	93.6
UNITEDHEALTH GROUP	0.35	73.4	81.5
VERIZON COMMUNICATIONS	0.64	93.5	98.4
WALMART	0.32	94.6	79.3

Table 8.3.2: Testing if the volatility of the mean volatility is significantly different between the portfolio pairs

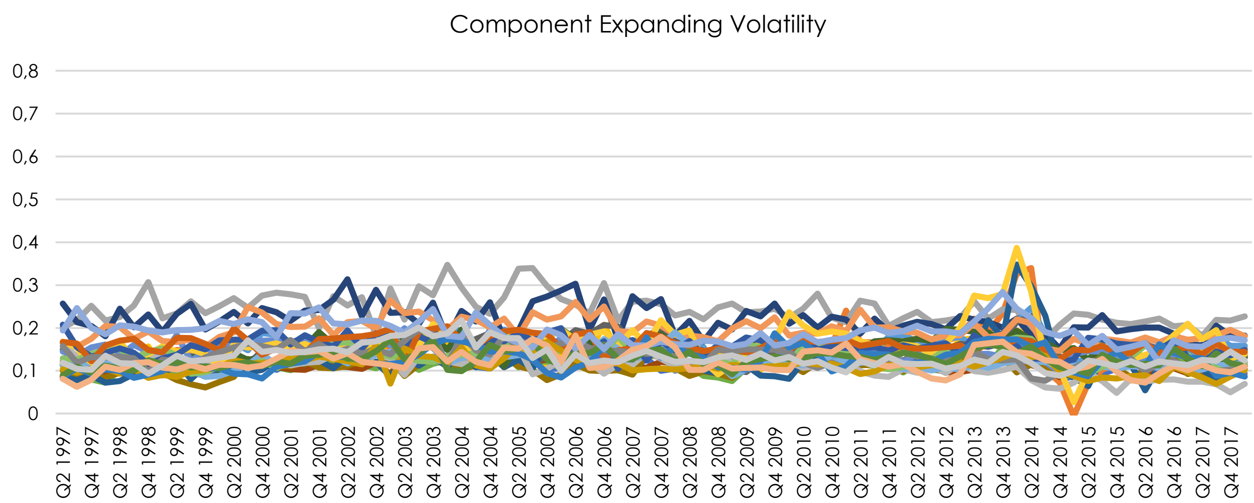

Figure 8.2.1: The forecasted volatility of the Component Expanding model over the full forecasting period

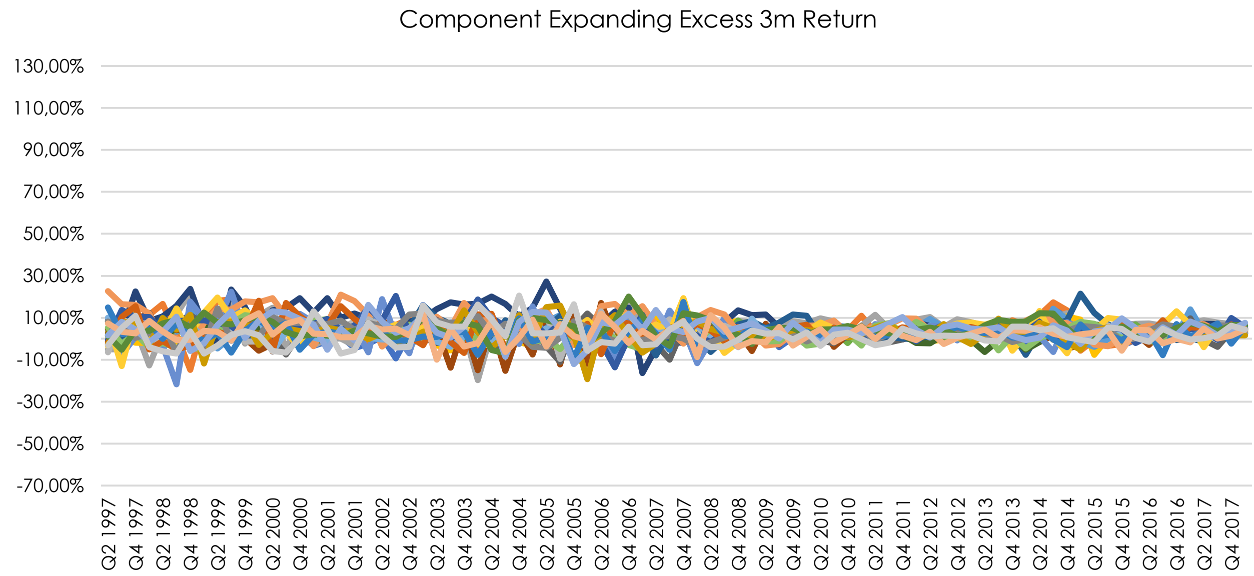

Figure 8.2.2: The forecasted excess 3m return of the Component Expanding model over the full forecasting period

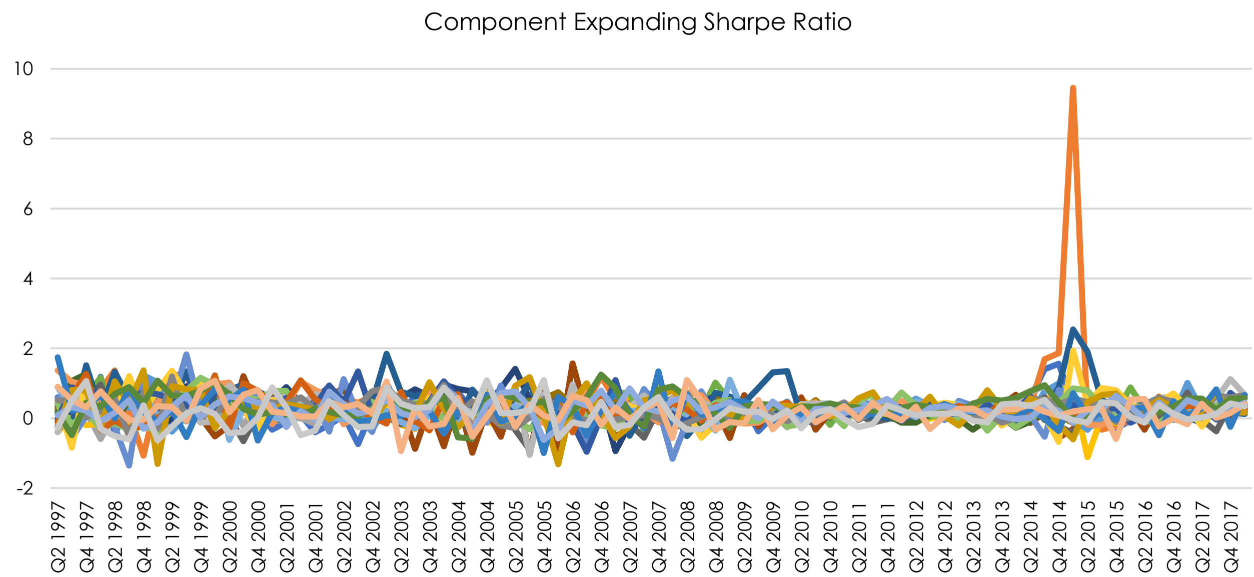

Figure 8.2.3: The forecasted Sharpe ratio of the Component Expanding model over the full forecasting period

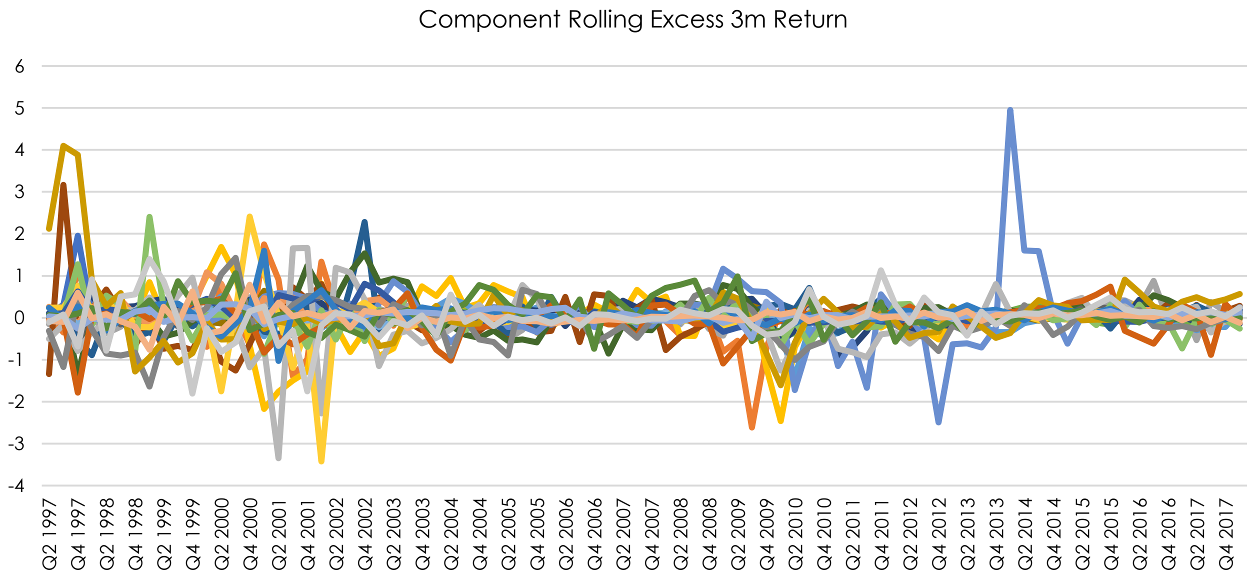

Figure 8.2.4: The forecasted excess 3m return of the Component Rolling model over the full forecasting period

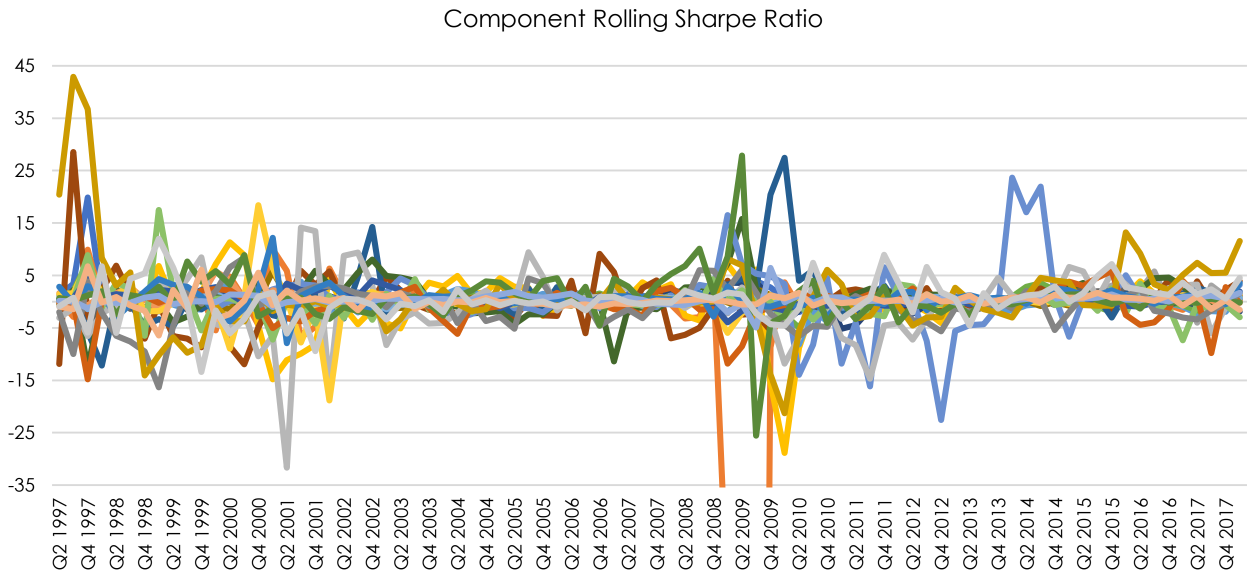

Figure 8.2.5: The forecasted Sharpe ratio of the Component Rolling model over the full forecasting period

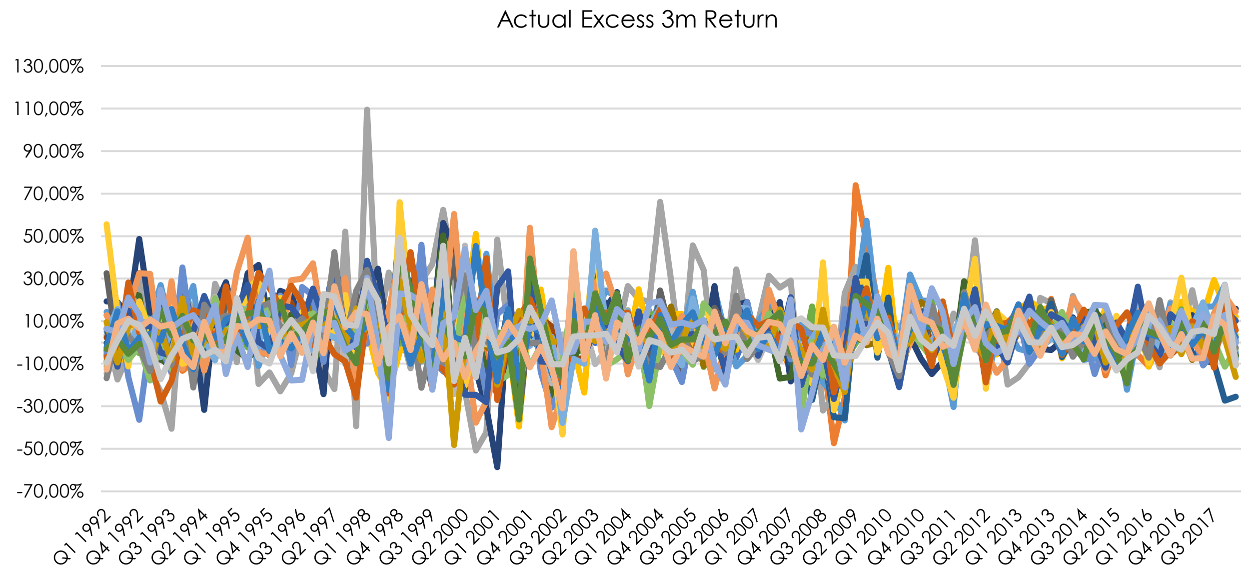

Figure 8.2.6: The actual excess 3m return over the full sample period

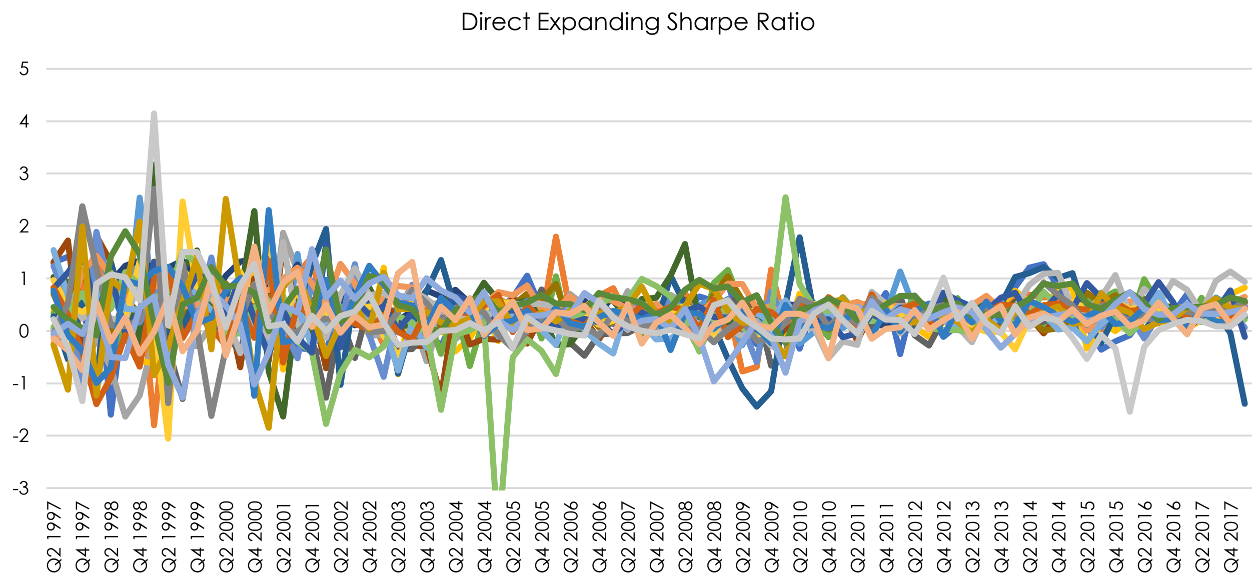

Figure 8.2.7: The forecasted Sharpe ratio of the Direct Expanding model over the full forecasting period

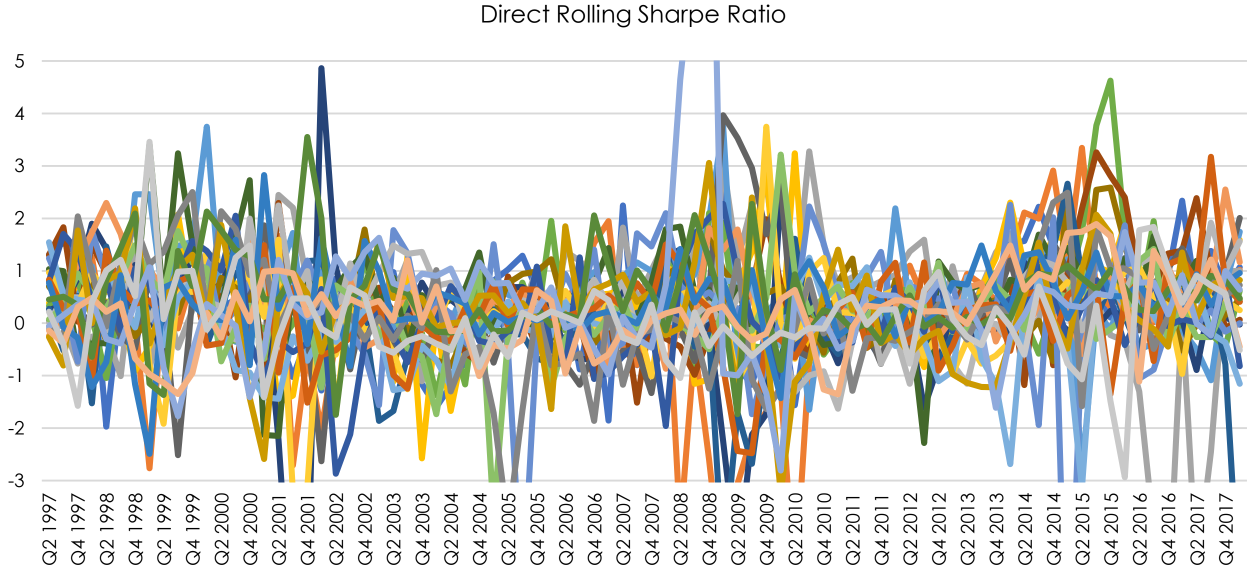

Figure 8.2.8: The forecasted Sharpe ratio of the Direct Rolling model over the full forecasting period

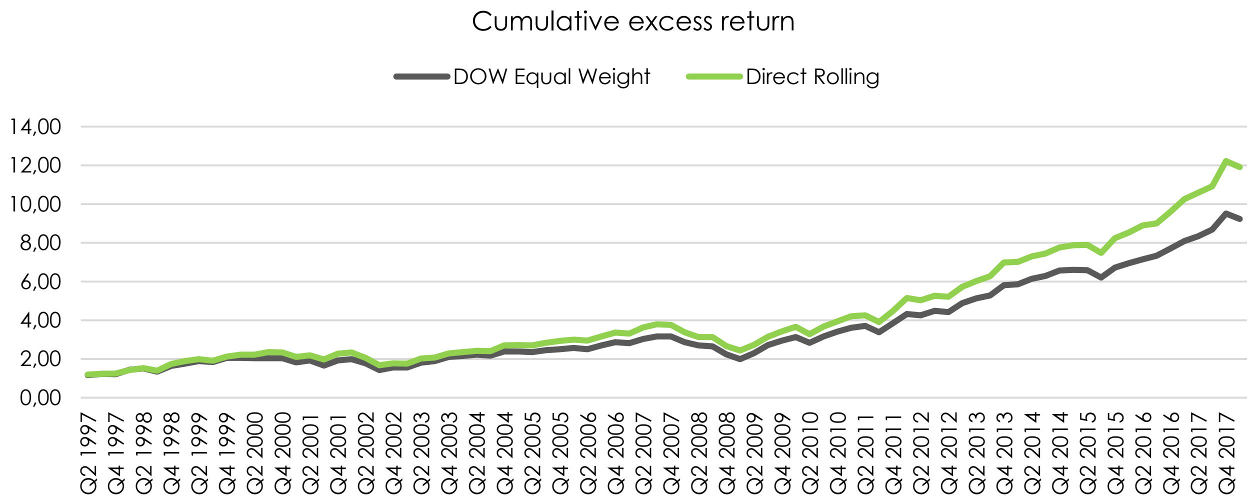

Figure 8.2.9: The cumulative excess return of the DOW Equal Weight and Direct Rolling portfolio

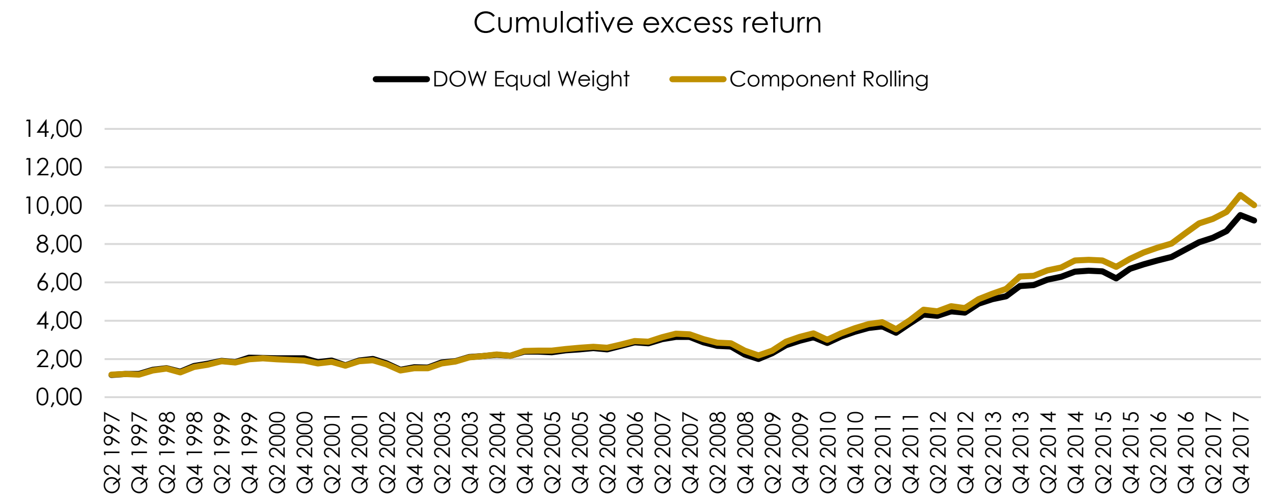

Figure 8.2.10: The cumulative excess return of the DOW Equal Weight and Component Rolling portfolio

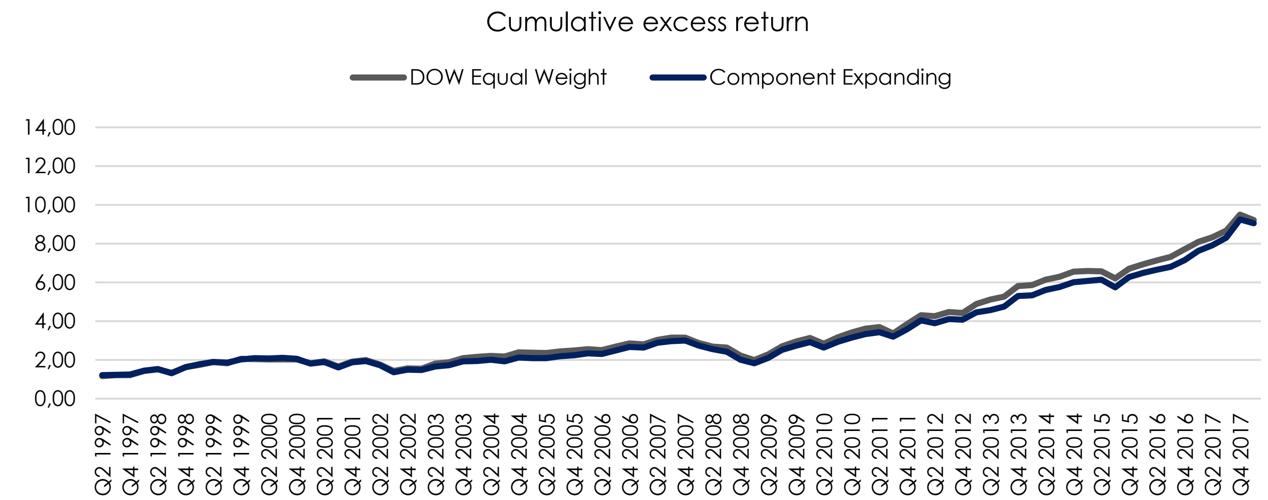

Figure 8.2.11: The cumulative excess return of the DOW Equal Weight and Component Expanding portfolio

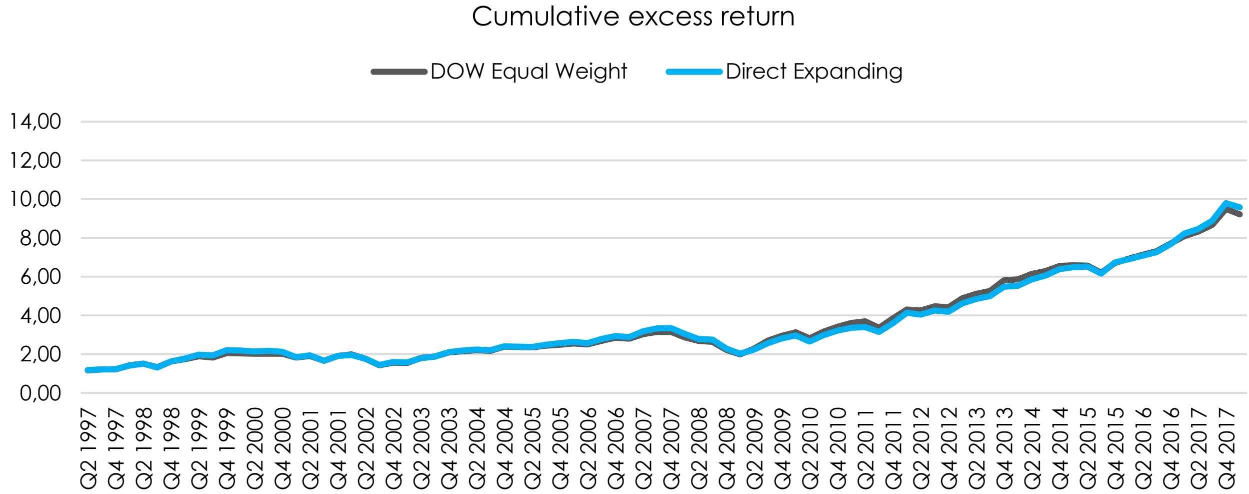

Figure 8.2.12: The cumulative excess return of the DOW Equal Weight and Direct Expanding portfolio

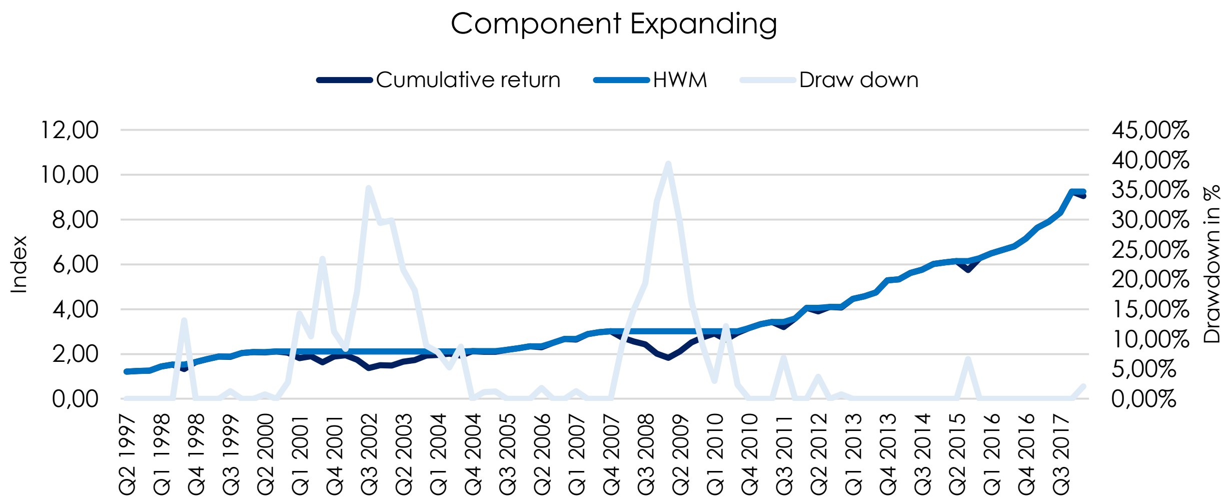

Figure 8.2.13: The highwater mark and drawdown for the Component Expanding portfolio

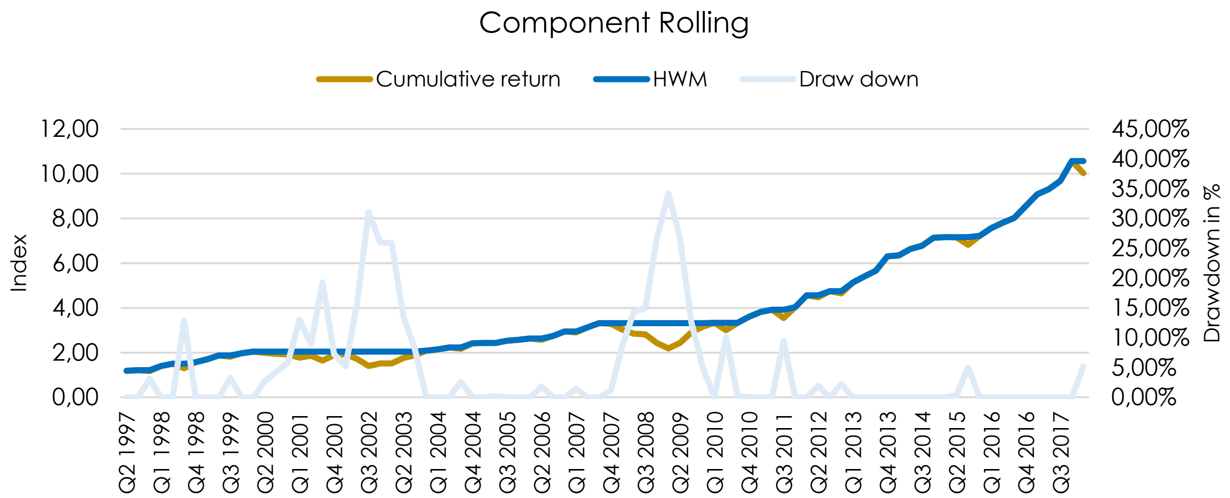

Figure 8.2.14: The highwater mark and drawdown for the Component Rolling portfolio

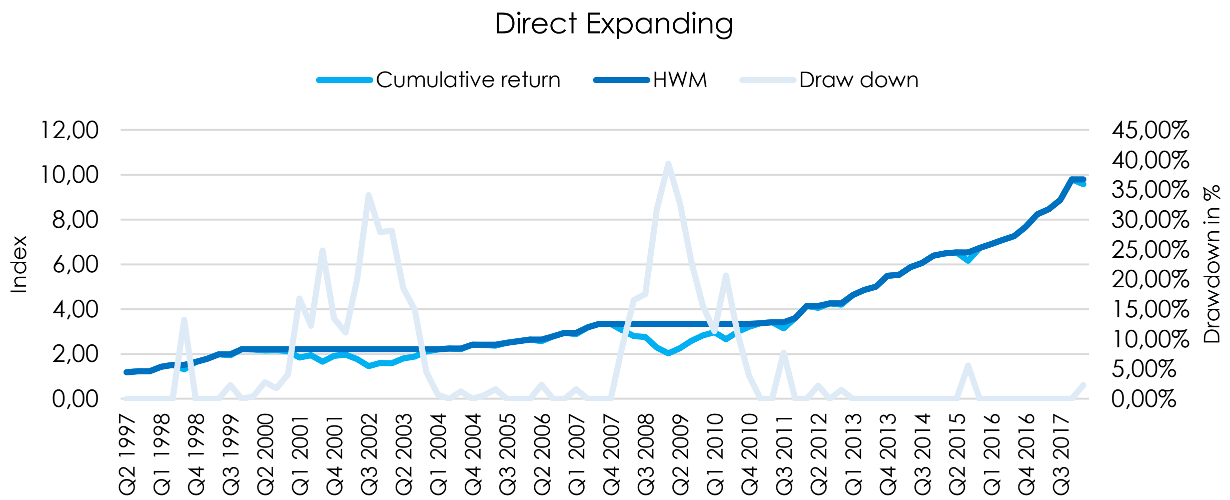

Figure 8.2.15: The highwater mark and drawdown for the Direct Expanding portfolio

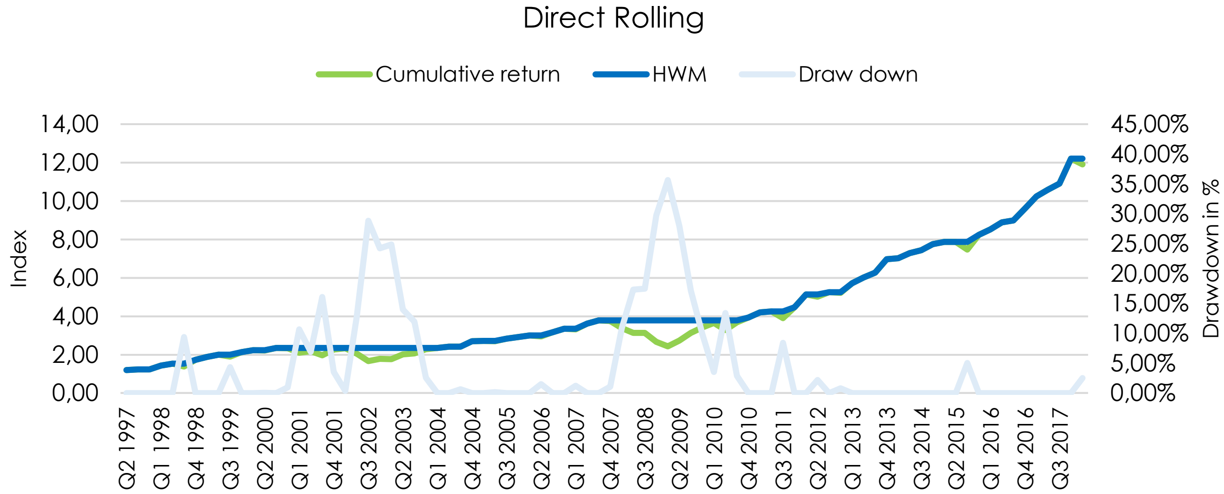

Figure 8.2.16: The highwater mark and drawdown for the Direct Rolling portfolio

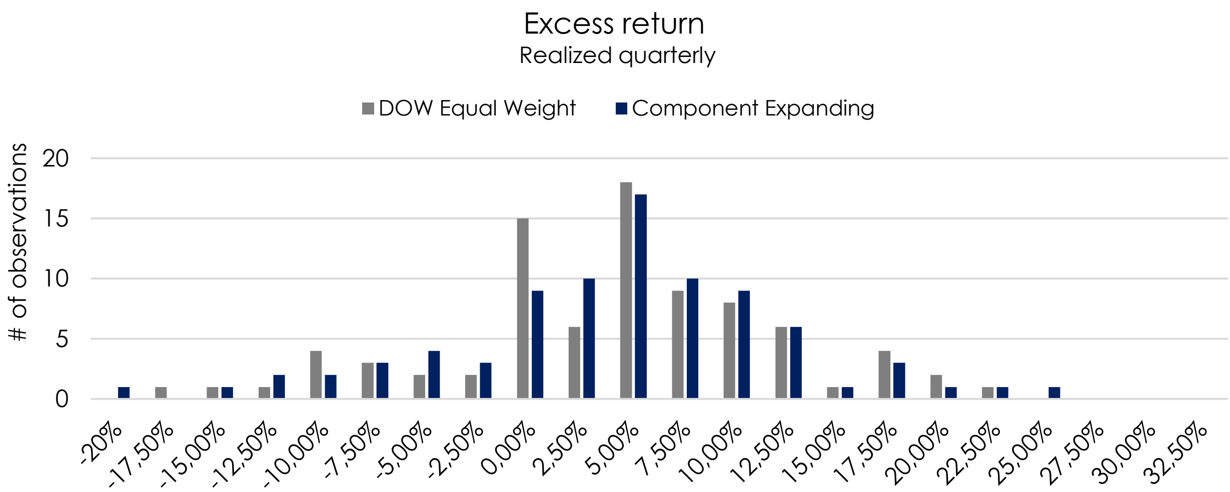

Figure 8.2.17: The distribution of realized quarterly excess returns for the DOW Equal Weight and Component Expanding portfolios

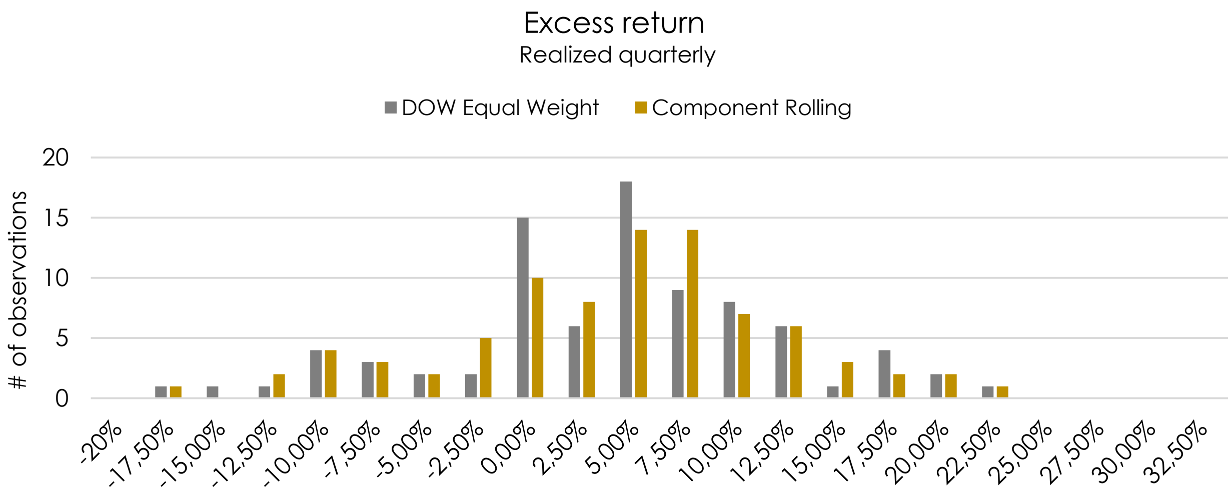

Figure 8.2.18: The distribution of realized quarterly excess returns for the DOW Equal Weight and Component Rolling portfolios

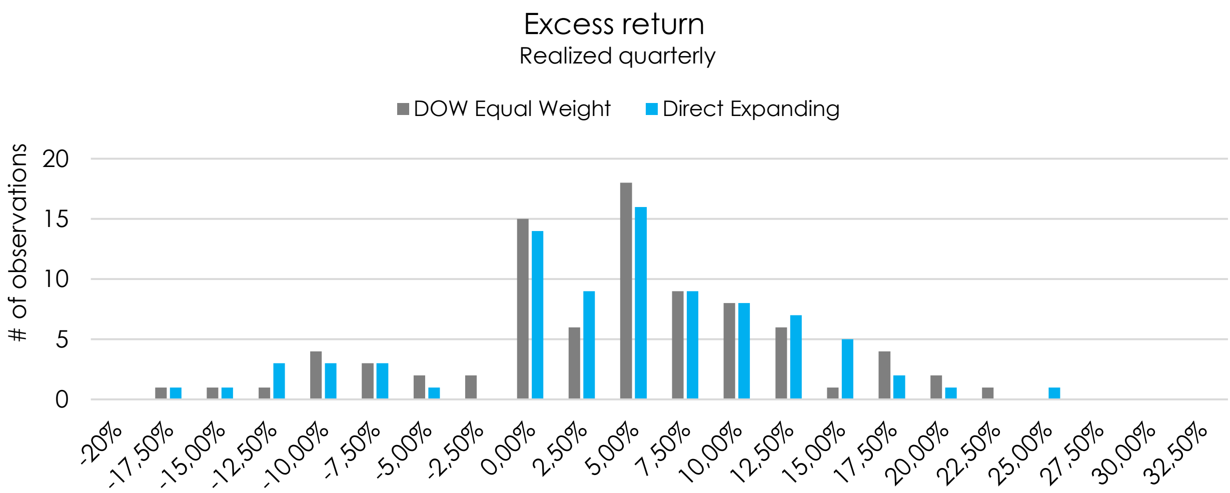

Figure 8.2.19: The distribution of realized quarterly excess returns for the DOW Equal Weight and Direct Expanding portfolios

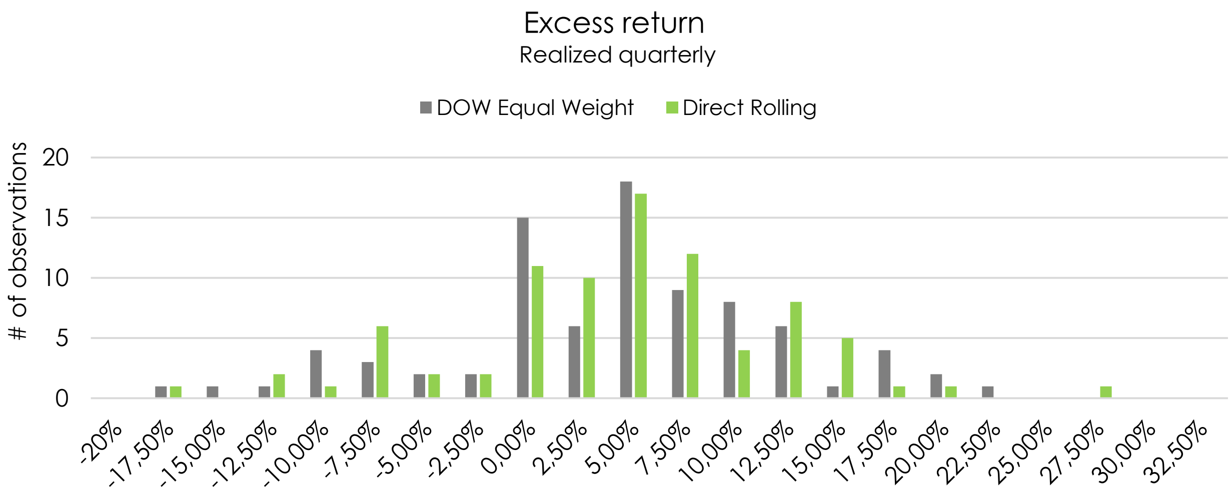

Figure 8.2.20: The distribution of realized quarterly excess returns for the DOW Equal Weight and Direct Rolling portfolios

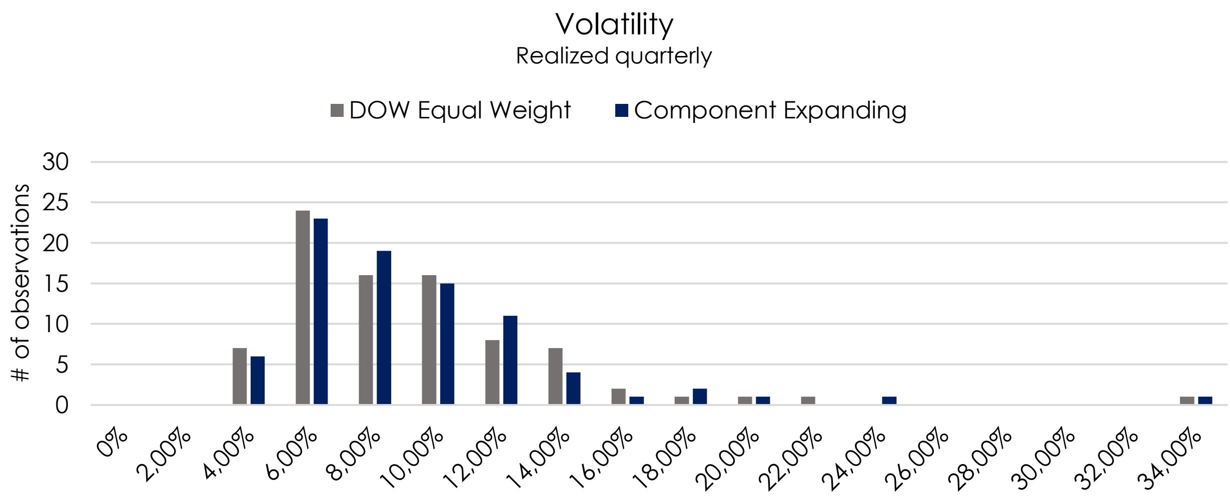

Figure 8.2.21: The distribution of realized quarterly volatility for the DOW Equal Weight and Component Expanding portfolio

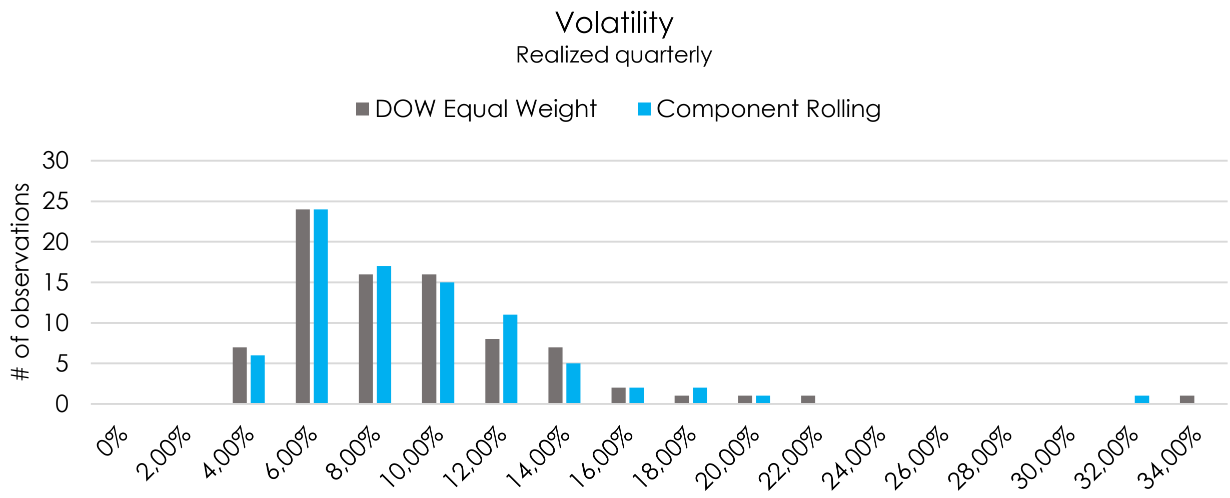

Figure 8.2.22: The distribution of realized quarterly volatility for the DOW Equal Weight and Component Rolling portfolio

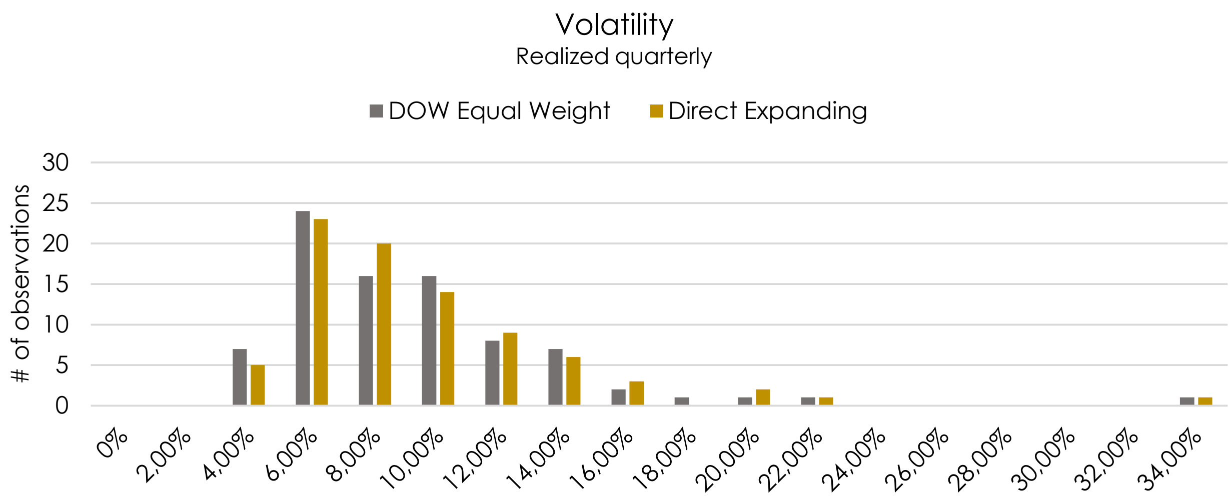

Figure 8.2.23: The distribution of realized quarterly volatility for the DOW Equal Weight and Direct Expanding portfolio

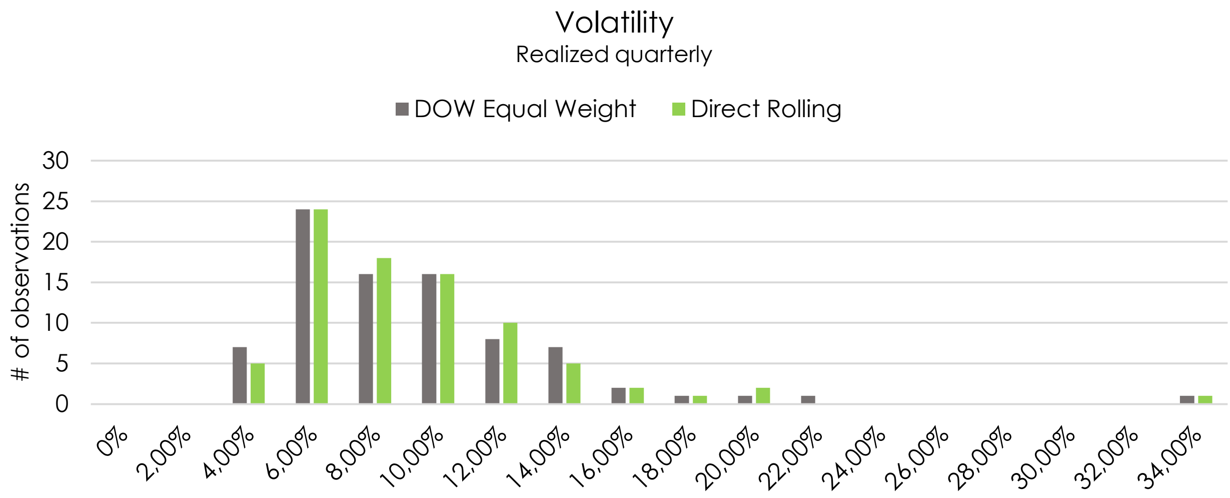

Figure 8.2.24: The distribution of realized quarterly volatility for the DOW Equal Weight and Direct Rolling portfolio

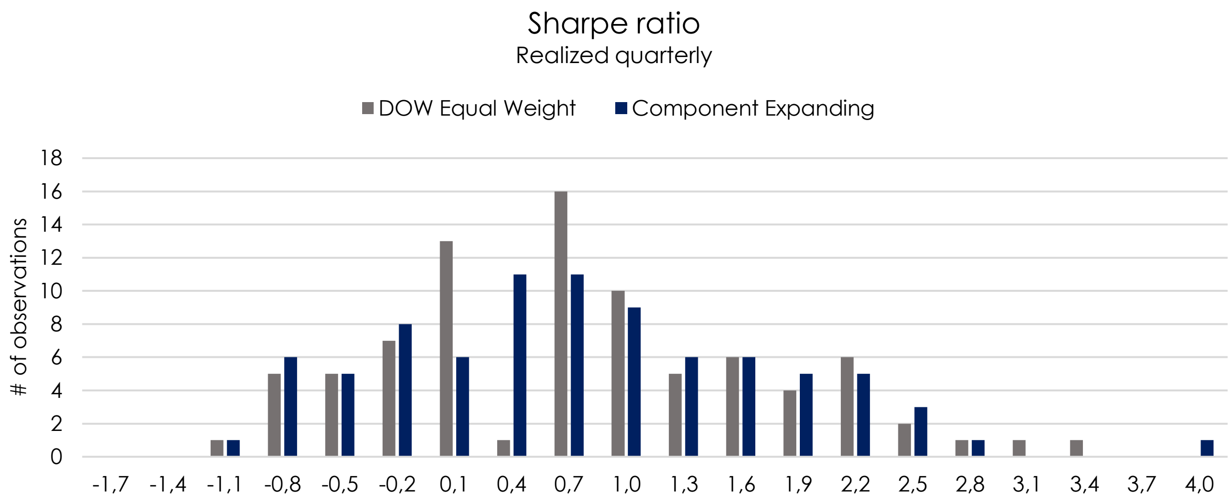

Figure 8.2.25: The distribution of realized quarterly Sharpe ratio for the DOW Equal Weight and the Component Expanding portfolio

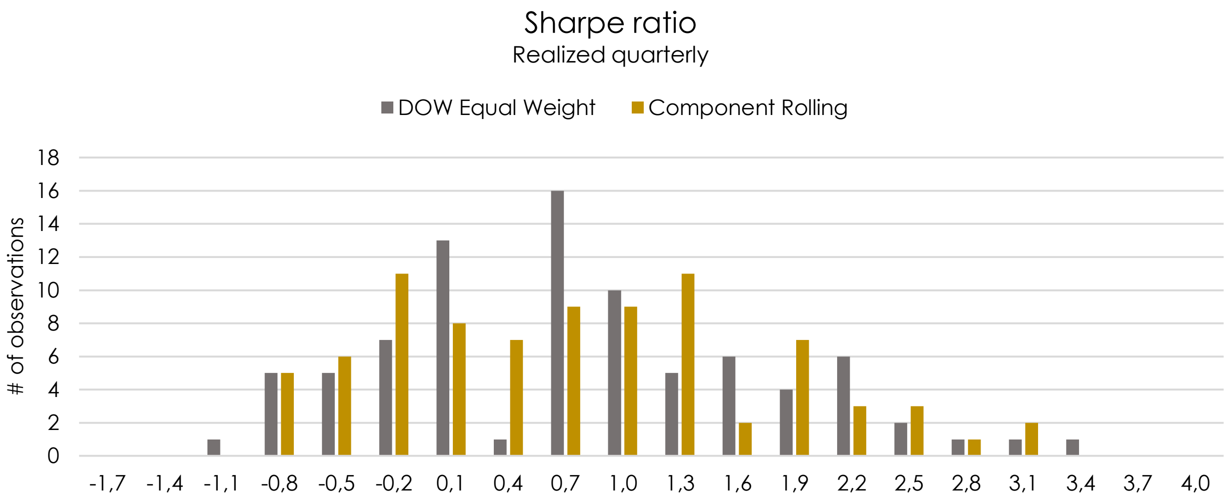

Figure 8.2.26: The distribution of realized quarterly Sharpe ratio for the DOW Equal Weight and the Component Rolling portfolio

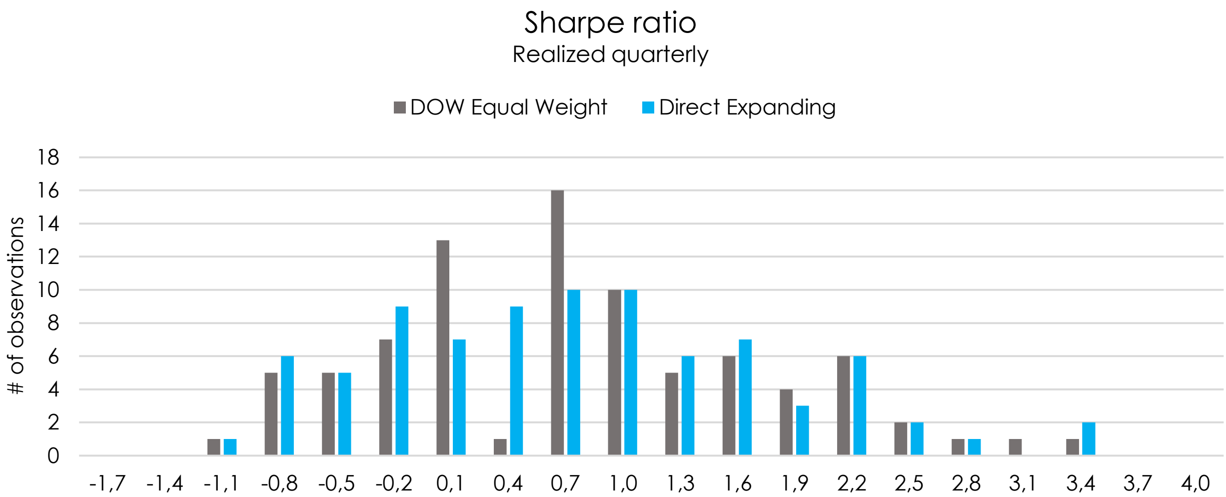

Figure 8.2.27: The distribution of realized quarterly Sharpe ratio for the DOW Equal Weight and the Direct Expanding portfolio

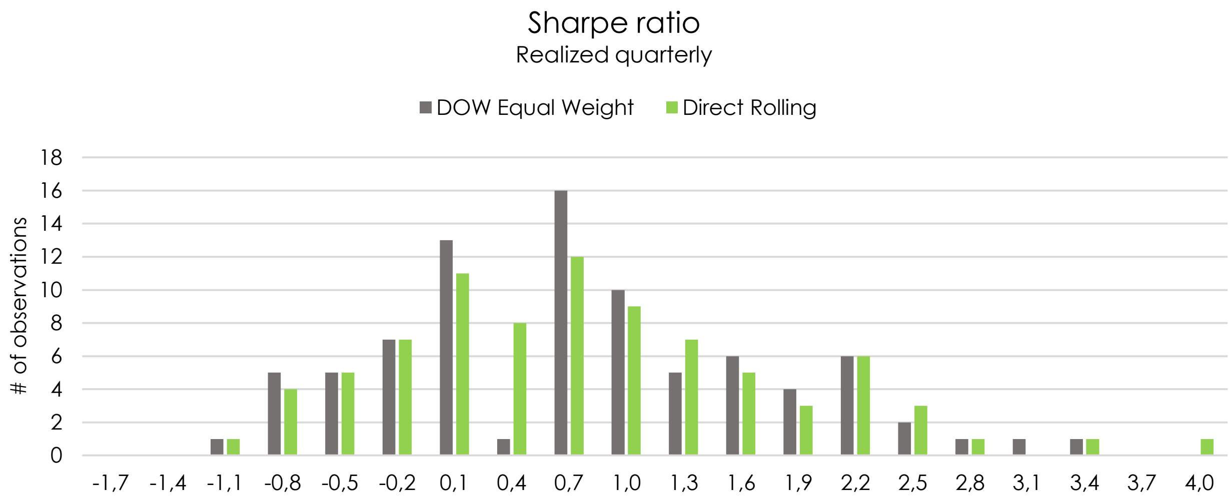

Figure 8.2.28: The distribution of realized quarterly Sharpe ratio for the DOW Equal Weight and the Component Direct portfolio

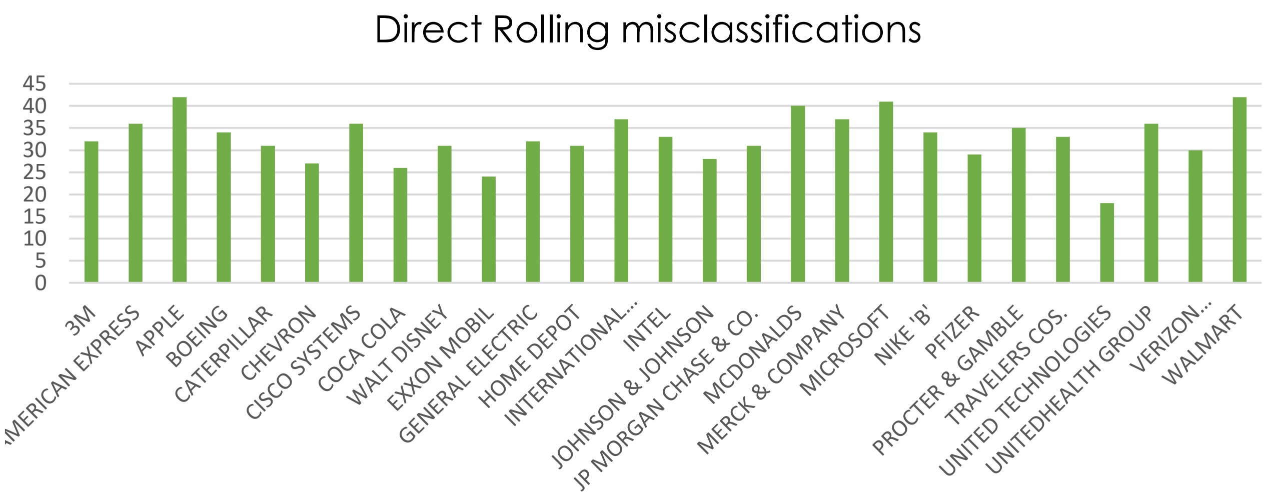

Figure 8.2.29: Misclassifications per stock for the Direct Rolling model

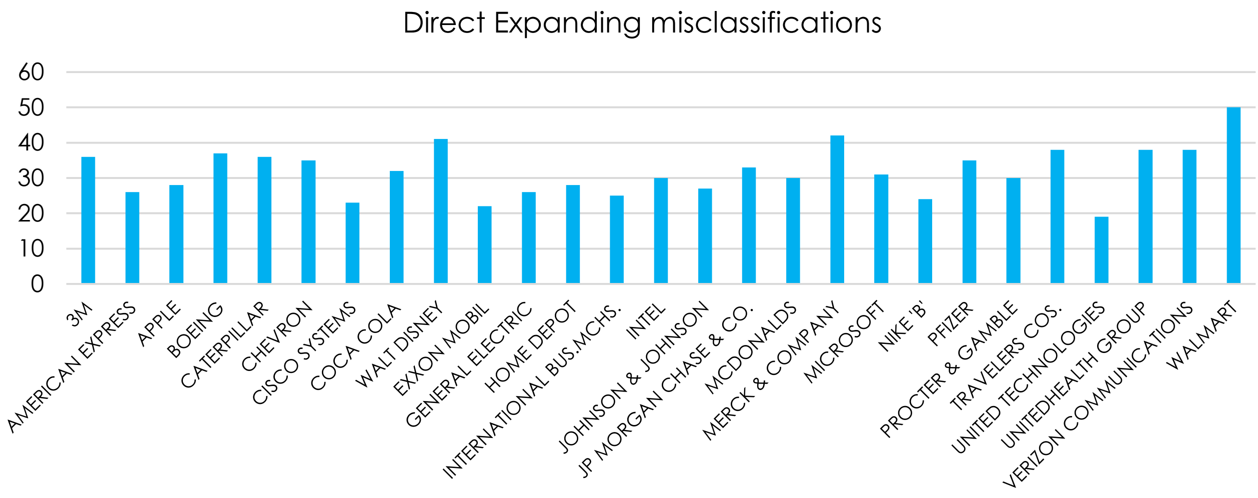

Figure 8.2.30: Misclassifications per stock for the Direct Expanding model

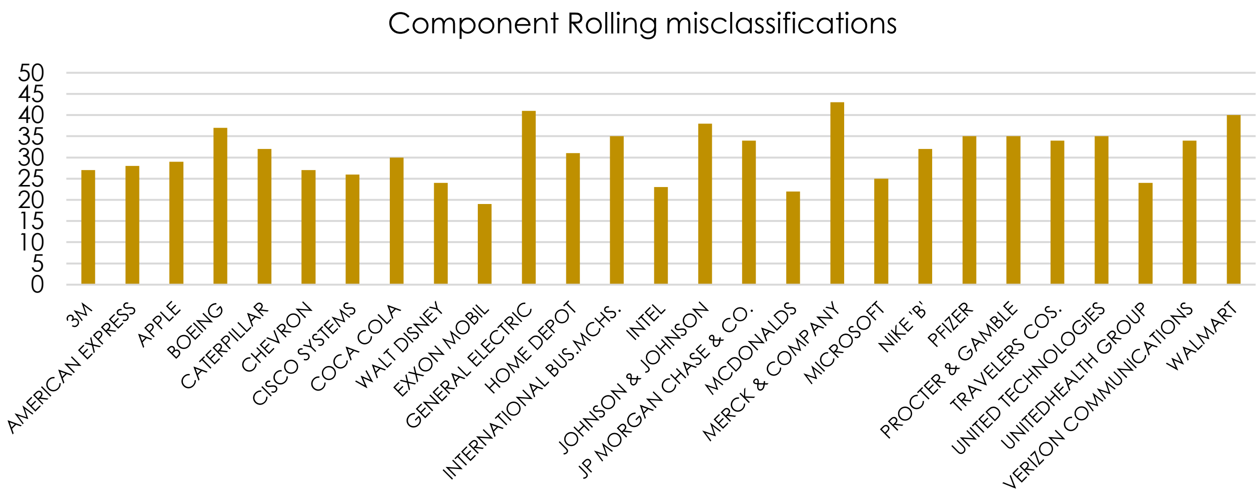

Figure 8.2.31: Misclassifications per stock for the Component Rolling model

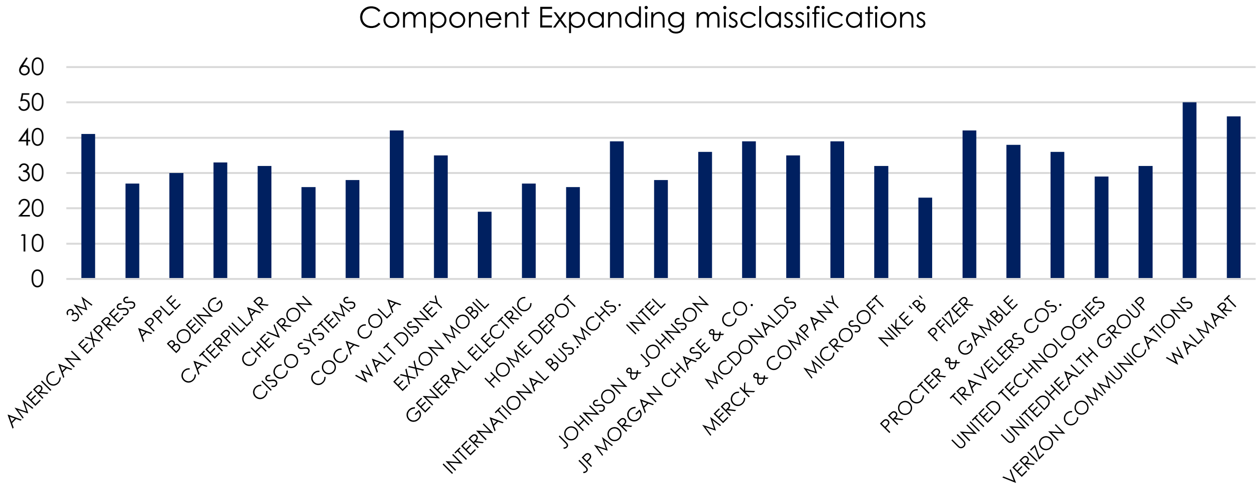

Figure 8.2.32: Misclassifications per stock for the Component Expanding model

Volatility	Direct Rolling	Direct Expanding	Component Rolling	Component Expanding	DOW Equal Weight	Perfect SR Rank
Mean	8.19%	8.43%	8.32%	8.29%	8.20%	8.29%
Standard deviation	4.39%	4.52%	4.37%	4.53%	4.46%	4.53%
Max	32.99%	32.39%	31.87%	33.09%	32.18%	33.09%
75th percentile	9.98%	10.21%	10.28%	10.00%	9.93%	10.00%
Median	7.14%	7.32%	7.39%	7.12%	7.13%	7.12%
25th percentile	5.37%	5.32%	5.48%	5.37%	5.32%	5.37%
Min	2.55%	2.61%	2.87%	2.89%	2.84%	2.89%

Table 8.3.3: Summary statistics of the volatility of the 5 portfolios and the portfolio that always ranks correctly

Sharpe ratio	Direct Rolling	Direct Expanding	Component Rolling	Component Expanding	DOW Equal Weight	Perfect SR Rank
Mean	0.66	0.63	0.62	0.62	0.61	1.18
Standard deviation	1.01	1.01	0.99	1.01	0.99	1.09
Max	3.77	3.30	3.10	3.89	3.30	4.71
75th percentile	1.24	1.32	1.12	1.29	1.29	1.87
Median	0.53	0.56	0.56	0.50	0.53	1.11
25th percentile	-0.08	-0.17	-0.23	-0.13	-0.10	0.47
Min	-1.19	-1.11	-1.04	-1.14	-1.11	-0.82

Table 8.3.4: Summary statistics of the Sharpe ratio of the 5 portfolios and the portfolio that always ranks correctly

Component rolling forecast on actual regression	Intercept p-value	Sharpe ratio p-value	R2
3M	0.00	0.56	0.42%
AMERICAN EXPRESS	0.00	0.09	3.49%
APPLE	0.00	0.52	0.51%
BOEING	0.00	0.54	0.47%
CATERPILLAR	0.00	0.19	2.08%
CHEVRON	0.00	0.45	0.68%
CISCO SYSTEMS	0.01	0.07	3.97%
COCA COLA	0.02	0.61	0.32%
WALT DISNEY	0.00	0.48	0.62%
EXXON MOBIL	0.01	0.67	0.23%
GENERAL ELECTRIC	0.07	0.61	0.32%
HOME DEPOT	0.00	0.96	0.00%
INTERNATIONAL BUS.MCHS.	0.02	0.56	0.43%
INTEL	0.04	0.27	1.46%
JOHNSON & JOHNSON	0.00	0.35	1.08%
JP MORGAN CHASE & CO.	0.00	0.67	0.22%
MCDONALDS	0.00	0.65	0.25%
MERCK & COMPANY	0.02	0.71	0.16%
MICROSOFT	0.00	0.86	0.04%
NIKE B	0.00	0.01	8.33%
PFIZER	0.05	0.65	0.25%
PROCTER & GAMBLE	0.00	0.65	0.25%
TRAVELERS COS.	0.00	0.09	3.45%
UNITED TECHNOLOGIES	0.00	0.19	2.05%
UNITEDHEALTH GROUP	0.00	0.08	3.78%
VERIZON COMMUNICATIONS	0.05	0.61	0.33%
WALMART	0.01	0.92	0.01%

Table 8.3.5: Output from regression of actual Sharpe ratio using forecasted Sharpe ratio as estimator for the Component Rolling model

Component expanding forecast on actual regression	Intercept p-value	Sharpe ratio p-value	R2
3M	0.05	0.02	6.00%
AMERICAN EXPRESS	0.00	0.06	4.30%
APPLE	0.00	0.59	0.35%
BOEING	0.00	0.96	0.00%
CATERPILLAR	0.00	0.26	1.55%
CHEVRON	0.01	0.38	0.95%
CISCO SYSTEMS	0.05	0.62	0.30%
COCA COLA	0.02	0.66	0.23%
WALT DISNEY	0.09	0.00	9.35%
EXXON MOBIL	0.03	0.43	0.77%
GENERAL ELECTRIC	0.25	0.64	0.26%
HOME DEPOT	0.00	0.35	1.09%
INTERNATIONAL BUS.MCHS.	0.01	0.42	0.79%
INTEL	0.14	0.65	0.25%
JOHNSON & JOHNSON	0.01	0.68	0.21%
JP MORGAN CHASE & CO.	0.00	0.06	4.10%
MCDONALDS	0.01	0.92	0.01%
MERCK & COMPANY	0.01	0.13	2.77%
MICROSOFT	0.00	0.07	4.00%
NIKE B	0.00	0.72	0.15%
PFIZER	0.17	0.69	0.20%
PROCTER & GAMBLE	0.10	0.09	3.54%
TRAVELERS COS.	0.02	0.21	1.93%
UNITED TECHNOLOGIES	0.01	0.62	0.30%
UNITEDHEALTH GROUP	0.00	0.21	1.93%
VERIZON COMMUNICATIONS	0.03	0.34	1.12%
WALMART	0.02	0.86	0.04%

Table 8.3.6: Output from regression of actual Sharpe ratio using forecasted Sharpe ratio as estimator for the Component Expanding model