UNIVERSITETET I AGDER

Timing the US Stock Market Using Moving Averages and Momentum Rules: An Extensive Study

HÅVARD LØSETH MODELL
LARS MAGNUS LYNNGÅRD

SUPERVISOR
Valeriy Zakamulin

University of Agder, 2017
School of Business and Law

Preface

This thesis concludes our Master of Science (M.Sc.) in Business Administration at the University of Agder. The goal of the thesis is to thoroughly investigate the performance of stock market timing using strategies based on moving averages and time-series momentum. We aim to do so by testing different strategies on data of the US stock market in the historical period from 1928 to the end of 2015. Inspired by the many academic papers encountered in the completion of our time as students, we have tried to structure this thesis as closely to a publishable academic article as possible. Finally we would like to give a special thanks to our supervisor Valeriy Zakamulin for giving guidance, constructive feedback and support whenever needed.

Contents

Timing the US Stock Market Using Moving Averages and Momentum Rules: An Extensive Study

Håvard Løseth Modell and Lars Magnus Lynngård *

May 29, 2017

Abstract

In this thesis we investigate the performance of moving average and momentum strategies by simulating returns, both in-sample and out-of-sample, while simultaneously taking into account important market frictions. We do so for two stock indices and four stock portfolios, at daily and monthly frequency, in the period from 1928 to 2015. This is carried out in order to examine if the active strategies outperform the passive benchmark on a risk-adjusted basis, and to see if the trading rules profitable when tested in-sample also are profitable out-of-sample. In addition, and for the first time, we examine the relevance of data frequencies in out-of-sample testing. A stationary block bootstrap methodology is adopted in order to evaluate the statistical significance of the risk-adjusted performance, measured by the Sharpe ratio. We find that in-sample profitable trading rules perform poorly when tested out-of-sample. However, we are able to find statistically significant outperformance when trading in small-cap stocks; yet, the outperformance disappeared in recent past. Moreover, we investigate how the performance depends on the split point between the in- and out-of-sample period and the length of the in-sample period. We find that the performance of an out-of-sample test highly depends on the choice of split point as well as in-sample period length. Consequently, the out-of-sample testing procedure is not a complete remedy for the "data mining bias". Finally, we are not able to find conclusive evidence suggesting any benefit of trading more frequently.

Key words: technical analysis, market timing, moving averages, time-series momentum, out-of-sample simulations, trading frequency

*School of Business and Law, University of Agder, Kristiansand, Norway

1 Introduction

Technical analysis represents a set of techniques for analyzing past prices (and volumes) in order to predict future price movements by discovering trends or reoccurring patterns. The methodology of technical analysis implicitly assumes that asset prices are predictable and investors can exploit and profit from the price information. In principle, there exist two branches of technical analysis. A quantitative form, where the *"technician"* typically trades using computerized technical trading systems, and a qualitative form, where the analyst use visual charts and inductive reasoning to form an opinion about future movements (Menkhoff and Taylor, 2007) .

Modern technical analysis and numerous technical trading rules can be traced back to the Dow Theory in the early 1900s, named after the long-time editor of *The Wall Street Journal* Charles Henry Dow. The theory was mainly developed after the death of Charles Dow by his successor William Peter Hamilton[1], and later made popular by Robert Rhea (1932) in his influential book "The Dow Theory". The Dow Theory, which is largely based of Dow's editorials, aims to identify long-term trends and predictable patterns in the stock market (Bodie, Kane, and Marcus, 2011, p. 394). Alfred Cowles (1933) was perhaps the first to quantify and test the Dow Theory, and his empirical study is considered significant in the early development of the Efficient Market Hypothesis (Brown, Goetzmann, and Kumar, 1998). Using the trading signals[2] obtained from Hamilton's editorials in the *Wall Street Journal* from 1904 to 1929, Cowles (1933) simulated returns to the active "Dow strategy" and compared it to a strategy based on investing 100% in the stock market. Cowles (1933) concluded that the Dow strategy performed worse then the passive strategy. Practitioners and academics have since debated the usefulness of the Dow Theory and technical analysis in general.

Technical analysis have historically mostly been used by practitioners, particularly in FOREX markets. For example, Menkhoff and Taylor (2007) find that almost all foreign exchange professionals use technical analysis to some degree. In fact, technical analysis is generally considered profitable in commodity and FOREX markets (See for example Zakamulin (2017)). In contrast to practitioners, academics tend to be skeptical about the

[1] See "The Stock Market Barometer" by William P. Hamilton (1922)

[2] The trading signals given by Hamilton was either "Bullish", "Bearish", or "Neutral/doubtful".

usefulness of technical analysis, and many consider it closer to *art* then science (see Park and Irwin (2007)). However, in recent decades, this sentiment has changed and the literature on technical analysis has experienced a renaissance. According to Park and Irwin (2007), the change in attitude can be linked in part to the availability of electronic databases and cheap computer power, and in part to the publication of numerous academic papers. In fact, a stream of recently published papers[3] have documented that trading based on simple moving average and momentum rules would have protected investors against the turbulent financial markets throughout the 2000s. Furthermore, the majority of these studies document that the active strategies outperforms the passive buy-and-hold strategy. Still, the results of these studies are varied and depend on a large number of variables including: trading rules, look-back period, sample period, statistical tests, asset class, simulation method etc. In practice, this leaves the researcher with an endless combination of parameters, which in turn raises a legitimate concern about the potential of "data mining bias" in the reported results.

The "data mining bias"[4] is a well known and important problem in financial studies. The fallacy can be explained as follows: When searching for outperformance, and at the same time being faced with a great number of possible combinations of parameters, one often end up reporting the optimal combination, yielding the best performance. But as the performance is given part by the true performance and part by a random component, chances are that the reported combination of parameters greatly benefits from the random component. In turn, this may lead to systematic overestimation of performance, and reported outperformance can exist entirely due to luck. The data mining bias is therefore known as an upward bias occurring when the same dataset is used over and over for the purpose of model selection. Even though the fallacy is regularly committed in the literature, the bias is by no means a new phenomenon. In fact, Jensen (1967) warned about the potential dangers of "mining" the data back in the late 1960s[5]. Despite being a widely recognized problem in financial studies, very few seem to adequately adjust for the bias. Peculiarly, only a minority of the studies on technical trading in the stock market use out-

[3]See for example Faber (2007), Metghalchi, Marcucci, and Chang (2012), Kilgallen (2012), Moskowitz, Ooi, and Pedersen (2012), Pätäri and Vilska (2014), and Glabadanidis (2016)

[4]"Data mining" is also known as "data snooping" and "selection bias".

[5]A further discussion on "data mining bias" will be given in Section 3. For an in-depth analysis of the problematic effects see Lo and MacKinlay (1990).

of-sample tests to obtain unbiased estimates of the performance (see for example Sullivan, Timmermann, and White (1999) and Zakamulin (2014, 2015)).

In this thesis, our goal is to remedy this issue and extend the existing literature by implementing the following seven extensions.

First, in addition to testing the profitability of the active strategies using an in-sample test, we also perform an out-of-sample test in order to obtain an unbiased and realistic estimate of the profitability. Furthermore, when carrying out an out-of-sample test, the "data window" is either rolled or expanded through time depending on assumptions regarding market dynamics. Many traders believe that there exist one optimal and time invariant trading rule. For example, Faber (2007) finds that the 10-month Simple Moving Average (SMA) outperforms the S&P 500 over a period of 100 years. If we assume that the best trading rule is time invariant, we would expect to see superior performance for the expanding estimation scheme. However, if we suspect parameter instability and changing market dynamics, the rolling estimation scheme is preferred. Both estimation methods will be used in this thesis.

Second, in our study we utilize two stock indicies and four stock portfolios at daily and monthly frequency that spans 88 years starting in January 1928 and ending in December 2015. First, we examine profitability for the S&P Composite xndex and the DJIA index. Second, in order to see if the results persist for portfolios with different characteristics, we examine profitability for value stocks, growth stocks, small stocks and large stocks.

Our third extension can be explained as follows. Numerous studies[6] in the literature on market timing do not adequately adjust for important market frictions. This can have unfortunate implications. Since the reported results overstates the true profitability, the reader is left with the impression that market timing is more profitable then what can be achieved in practice. Consequently, in order to obtain a realistic estimate of the real-life performance, we take into account the transaction costs incurring each time the active strategy switches.

Our fourth extension is motivated by the following question. Is there any benefit of trading more frequently? Intuitively, daily data provide more information then monthly

[6]E.g. Brock, Lakonishok, and LeBaron (1992), Glabadanidis (2016), Sullivan et al. (1999) and Gwilym, Clare, Seaton, and Thomas (2010)

data. Thus, a trader could potentially profit from the additional information. However, this advantage may be reversed if the profits from additional trading are offset by the added transaction costs. To our knowledge, there is only one paper to date that explicitly investigate the relevance of data frequencies in stock markets. Specifically, using in-sample testing, Clare, Seaton, Smith, and Thomas (2013) found that trading based on monthly data is superior compared to daily data when the underlying passive benchmark is the S&P 500 index.

Fifth, in previous studies researchers typically use a simple parametric test for statistical evaluation, despite the well known distributional characteristics of stock returns. With this in mind, we adopt a stationary block bootstrap methodology to evaluate the statistical significance of the risk-adjusted returns, measured by the Sharpe ratio, without making any distributional assumptions. In addition, the active strategies are further evaluated using the alpha of the Fama-French-Carhart (1997) 4-factor model. A bootstrap method is used to determine the statistical significance of the alpha.

Our sixth extension is motivated by the study by Zakamulin (2014) who documents that the performance of an out-of-sample test depends on the split point between the initial in-sample and out-of-sample period. This has an important implication. Since the performance is sensitive to the placement of the split point, the out-of-sample testing procedure is not free of "data mining" issues. We further investigate this characteristic and extends Zakamulin (2014)'s study by examining the dependence of split point for all datasets utilized in our study at both daily and monthly frequency. Additionally, we also examine how the performance depends on the length of the in-sample period.

Finally, our seventh extension is related to the choice of trading rule. In our study we use the time-series Momentum rule (MOM), the Moving Average Crossover (MAC) rule and the Moving Average Envelope (MAE) rule. In previous studies, the researchers usually only report the performance of each trading rule individually. However, in a real life setting, a trader can at each point in time analyze the performance of several rules and adjust accordingly. Thus, in addition to reporting the performance of each trading rule individually, we also report the performance of a combination (COMBI) rule where at each period the strategy chooses the best performing trading rule among all three rules. This allows us to examine if the combination rule outperforms the individual rules, and if a single trading rule consistently outperforms the other rules.

Our study makes several contributions to the existing literature and the results can be summarized as follows. First, our results generally reveal lower volatility and mean return for all trend following strategies. This result is by no means surprising since the active strategies holds the risk-free asset approximately 30% of the time. In fact, this characteristic of moving average and momentum strategies is heavily documented in the existing literature (see e.g. Faber (2007), Moskowitz et al. (2012) and Zakamulin (2015)). Furthermore, while most trading rules are able to avoid the largest negative returns, they are equivalently not able to capture the largest positive returns.

Second, in consensus with Sullivan et al. (1999), our empirical analysis reveal that statistically profitable trading rules discovered in-sample perform poorly when tested out-of-sample. Specifically, when the underlying benchmark is the S&P Composite and the DJIA index, we are not able to find any trading rule out-of-sample with statistical significant outperformance. However, we find that the technical trading rules perform slightly better on the S&P Composite compared to the DJIA index. Interestingly, when the underlying asset is small-cap stocks, we find evidence from the out-of-sample test of statistically significant outperformance for both daily and monthly data in the period 1960 to 2000. In fact, and contrary to the results on the other stock indices, we find that trading using daily data has performed significantly better then trading using monthly data. However, the performance has deteriorated significantly over the past two decades, and we no longer see statistically significant outperformance for neither daily, nor monthly data. We further find that daily trading no longer perform superior to monthly, even showing underperformance over the past 10 years.

Third, contrary to the findings of Clare et al. (2013), we are not able to find conclusive evidence suggesting that high frequency data (e.g. daily) deteriorates performance when the underlying asset is the S&P Composite index. However, by simulating each trading rule individually, we generally find that the Moving Average Envelope (MAE) rule performs best for daily data. In turn, this leads us to believe that the "percentage band" protects the trader against "whipsaw" trading, and thus preventing excessive accumulation of transaction cost for daily data.

Finally, and similar to Zakamulin (2014), we find that the performance of the active strategies is highly uneven over time for all datasets examined. Two implications follow. First, the performance of the active strategy is highly dependent on the choice of sample period and split point, and second, the out-of-sample testing procedure is not a complete remedy for the "data mining bias".

The rest of the thesis is structured as follows. Section 2 explores relevant literature on technical analysis and market timing from 1960 to 2014. In Section 3 we outline the methodology for in-sample and out-of-sample tests, trading rules, and statistical tests. Section 4 presents and describes the data used in our study. Section 5 reports the results of the empirical tests. In Section 6 we discuss the findings and relevant implications. Finally, Section 7 concludes and summarizes the thesis.

2 Literature Review

This section reviews previous literature on technical analysis and active trading strategies with particular emphasis on the profitability of momentum and moving average rules. Similar to Park and Irwin (2007), we divide the literature into *Early Studies*: 1960 - 1987, and *Modern Studies*: 1988 - 2014[7]. The motivation for the split point is threefold. First, many of the earlier studies do not conduct a statistical test, which in turn makes it difficult to assess the validity of the empirical findings. Second, many of the studies conducted in the late 1980s and forward significantly improve upon earlier studies in numerous ways. And third, a majority of the literature on technical analysis was published after the mid 1990s (Park and Irwin, 2007).

2.1 Early Studies: 1960 - 1987

In the early 1960s, Alexander (1961) began to develop filter rules in order to forecast market trends. A strategy based on filters generates a buy (sell) signal when the asset has increased (decreased), with a given percentage, from its recent low (high). Alexander tested his filter strategy on the S&P Composite index from 1929 to 1959 and DJIA index from 1897 to 1929, and concluded that a move in stock prices tends to persist once initiated. This result sparked an interest in the academia. Mandelbrot (1963) argued that the methodology applied by

[7]Extending the reviewed literature by 10 years

Alexander (1961) was flawed and could not be implemented in a realistic setting. Alexander (1964) corrected his calculation, and when adjusting for transaction cost, the trading profits disappeared.

Perhaps the most influential study categorized in the *early studies* is the seminal paper by Fama and Blume (1966), who applied Alexander's filter rules to individual stocks on DJIA from 1956 to 1962. They concluded that the filter rules would most likely be inferior to the buy-and-hold strategy when transaction costs and other fees were accounted for.

To forecast prices, Levy (1967) used a different approach termed "relative strength". In principle, a relative strength rule is based on buying assets that have performed *relatively* well previously. Using relative strength rules on stocks listed on NYSE from 1960 to 1965, Levy (1967) found evidence of superior performance, but concluded that the results were tentative due to the lack of statistical tests. Jensen and Benington (1970) extensively tested and replicated Levy's rules using stocks listed on NYSE in the period from 1926 to 1966. The authors divided the sample into seven 5-year sub-samples, where the final sub-sample was the same period as Levy (1967). After adjusting for risk and transaction cost, they argued that the rules based on relative strength did not yield significant profits greater then the buy-and-hold policy.

In the late 1960s, Van Horne and Parker (1967, 1968) published two papers investigating the profitability of moving average trading strategies and the validity of the random walk hypothesis. Using 30 stocks listed on the NYSE from January 1960 to June 1966 the authors calculated and tested the 100, 150 and 200 days moving average. By comparing the profits generated by the active strategies to the buy-and-hold strategy, the authors concluded that *technical trading rules are not as profitable as the buy-and-hold strategy.* However, the authors did not account for risk and only reported the amount of dollar each strategy generated.

Around the same time, James (1968) also investigated the random walk hypothesis using moving average strategies. The author generated returns using moving averages with different lengths and weights on stocks listed on NYSE in the period from 1926 to 1960. Using a simple parametric t-test for statistical evaluation, James (1968) concluded that the simple buy-and-hold strategy generally outperformed the moving average trading rules. However, similar to Van Horne and Parker (1967, 1968), the author only measured outperformance by dollar amount and did not account for risk. Thus, it is unclear how the trading rules performed adjusted for risk.

By the time the influential articles by Samuelson (1965) and Fama (1970) reached the mainstream, the Efficient Market Hypothesis became the leading paradigm in finance. As a consequence, very few studies on technical analysis were published in the 1970s and 1980s.

2.2 Modern Studies: 1988 - 2014

Similar to Park and Irwin (2007), we start the first modern study with the paper by Lukac, Brorsen, and Irwin (1988). This study is perhaps one of the first papers on technical analysis to implement an out-of-sample test, a statistical significance test, and adjusting for risk and market frictions. Specifically, the study use 12 different trading systems on a portfolio of 12 different futures markets in the period from 1978 to 1984. Using Jensen's alpha to estimate significant risk adjusted profits, Lukac et al. (1988) discovered that 4 out of 12 trading systems produce statistical significant positive net returns. Following up on their previous study, Lukac and Brorsen (1990) extended the study by using 23 trading systems on a longer sample for 30 futures markets. This time 7 out of 23 trading systems produced statistically significant net returns greater then zero.

Perhaps the most influential paper categorized in the *modern studies* is the paper by Brock et al. (1992). The authors applied 26 trading rules based on moving averages and trading range breaks to the DJIA from 1897 to 1986. By considering the well-known distributional characteristics of financial time series, the authors utilized a model-based bootstrap to validate the results. Brock et al. (1992) found that all trading rules considered had predictive power and generated significant profits. However, the deficiencies of this study are only simulating returns in-sample and not accounting for transaction costs. Using the same trading rules as Brock et al. (1992), Bessembinder and Chan (1998) tested for significant profits on DJIA from 1926 to 1991. A break-even transaction cost over the full sample was calculated to be 0.39%, suggesting that net returns from the study by Brock et al. (1992) were no longer significant (Park and Irwin, 2007).

Sullivan et al. (1999) used the same dataset and sample period for the in-sample test as Brock et al. (1992). In addition, the authors extended the sample by 10 years for an out-of-sample test. The full sample was divided into 5 sub-periods and the number of trading rules was extended to almost 8000. In addition, White's (2000) Reality Check bootstrap methodology was utilized to assess the degree of "data mining" from the back-test. The empirical findings were particularly compelling for two reasons. First, they were able to find similar statistical significant profits as Brock et al. (1992). And second, using White's Reality Check, the results were documented to be free of "data mining". However, Sullivan et al. (1999) were not able to find superior performance when the best trading rule was validated out-of-sample from 1987 to 1997. It should be noted that the out-of-sample period was a bull market and market timing rules generally underperform during bull markets.

More recently, Faber (2007) demonstrated that a simple moving average strategy applied to broad asset class indices would have produced superior performance compared to the buy-and-hold strategy. Specifically, Faber (2007) documented that a 10-month simple moving average strategy applied to the S&P 500 over a period of 100 years yielded higher annual returns and lower volatility, resulting in improved risk adjusted performance. Even though the author reported compelling and positive results, a number of weaknesses emerge. The author acknowledges that the 10-month SMA rule was chosen due to its known performance and the results were only simulated in-sample. In turn, this raises a legitimate "data mining" concern. In addition, the author did not account for transaction cost and no statistical test was conducted in order to assess the statistical significance of the results.

Gwilym et al. (2010) extended the study by Faber (2007) by simulating trading in-sample based on momentum and moving average rules on international equity markets. The authors reported statistically significant profits for the momentum rule, but did not account for transaction costs. However, the authors observed that the trading profit decreased towards the end of the sample. Additionally, Gwilym et al. (2010) confirmed the empirical results by Faber (2007), and reported superior risk adjusted performance for the moving average rule when compared to the buy-and-hold strategy.

Moskowitz et al. (2012) studied the effect of time-series momentum[8] across 58 futures contracts, in the period from 1985 to 2009, for major asset classes including equity markets, bond markets, currency markets and commodities markets. The authors were able to document a consistent and significant time series momentum effect across every asset

[8] It should be noted that "time-series momentum" is related but not the same as the cross-sectional "momentum" effect documented by Jegadeesh and Titman (1993) in the financial literature.

class examined. More precisely, Moskowitz et al. (2012) found that returns in the last 12 months was a positive predictor for future returns. The momentum effect was documented to persist for approximately one year before it partially reversed.

Kilgallen (2012) replicated the influential study by Faber (2007) with some modification. Specifically, instead of focusing on broad asset class indices, Kilgallen (2012) simulated the same strategies separately on individual assets. The author documented consistent lower volatility and higher annual returns for all individual assets examined. In fact, the volatility of the simple moving average strategy for individual currencies, equity indices and commodities was on average reported to be 27% lower then the passive benchmark. However, results are only simulated in-sample, and no statistical tests were conducted to validate the significance of the results.

Clare et al. (2013) and Hachemian, Tavernier, and Van Royen (2013) studied the effect of data frequencies on trend following strategies. Clare et al. (2013) documented that monthly trading on the S&P 500 index were superior to daily trading. Similarly, using daily and weekly data on 39 futures markets across 5 asset classes, Hachemian et al. (2013) were not able to find any significant advantage of using daily data.

In contrast to a majority of the *modern studies*, Zakamulin (2014) provided a more skeptical view of the superior performance of market timing strategies. The author utilized an out-of-sample test in order to assess the real-life performance of two popular market timing strategies. Two stock market indices and two bond market indices over a period of 87 years were used to evaluate if (1) the strategies performed over time, and (2) the performance was robust for different asset classes. Two results stand out. First, the empirical findings cast doubts on the results of previous research. Specifically, none of the active strategies statistically significantly outperformed the passive benchmark. And secondly, the empirical performance of market timing strategies was highly *non-uniform*. The significance of the latter is particularly important for investors with shorter investment horizons.

In summary, no general consensus can be found with respect to the profitability of market timing strategies based on moving averages and momentum rules. While the majority of the *modern studies* show positive and significant profits, particularly in currency and commodity markets, some deficiencies tend to persists. Specifically, many of the reviewed studies only test the trading rules in-sample without acknowledging or correcting for "data mining bias". Furthermore, several studies do not account for important market frictions like transaction costs. In turn, these persisting deficiencies makes it hard to assess the real-life profitability and applicability of the examined trading rules. Our main goal is to extend the existing literature on market timing by taking into account important market frictions and correcting for "data mining bias". Thus, we aim to provide an unbiased estimate of the risk-adjusted returns generated by momentum and moving average trading rules in a fashion that emulates real-life trading.

3 Methodology

3.1 Market Timing Rules and Moving Averages

A core principle among practitioners of technical analysis is that prices move in trends and that the trends can be uncovered in a timely manner. Thus, by identifying and "riding" a trend, traders can potentially profit from the subsequent price movement. In general, a simple trend following strategy can be implemented by buying (selling) assets that are trending upward (downward). However, due to price fluctuations, the simple concept of trend following can be difficult to implement in practice. In order to identify trends and filter out the noise from large price fluctuations, a trader can implement moving averages to "smooth" the price series.

3.1.1 Moving Averages

Moving averages are considered one of the simplest and most common ways to detect an underlying trend (Brock et al., 1992). Let P_t be the closing price of a given asset at time t. Note that the closing price is not adjusted for dividends. The general weighted moving average is computed as follows:

$$MA_t(k) = \frac{w_t P_t + w_{t-1} P_{t-1} \ldots + w_{t-k} P_{t-k}}{w_t + w_{t-1} + \ldots + w_{t-k}} = \frac{\sum_{j=0}^{k} w_{t-j} P_{t-j}}{\sum_{j=0}^{k} w_{t-j}}, \tag{1}$$

where w_t is the weight of price P_t, and $MA_t(k)$ is the general weighted moving average at time t of the last k observed prices.

There exist many types of moving average weighting schemes. The most common type of moving average is the Simple Moving Average (SMA) where each price observation is equally weighted. Many investors hold a belief that recent price observations contain more meaningful information about the future direction of the asset. In order to benefit from this idea the SMA can be substituted by the Linear Moving Average (LMA) where the weight of each price decreases in arithmetic sequence (Zakamulin, 2017). Following the same reasoning, investors can alternatively use the Exponential Moving Average (EMA) if they think the arithmetic weighting-scheme in the LMA is too rigid. As the name suggests, the price observations in EMA are weighted exponentially.

If the assumption that recent price information contains more relevant information is valid, one should expect the LMA and EMA to produce superior trading results compared to the simpler SMA. However, a recent empirical study by Zakamulin (2017) discovered that in many cases the SMA outperformed the LMA and EMA when tested on S&P 500. The author argues that in principle the choice of moving average is trivial in a practical sense. In fact, one could argue that the simplicity and understandability of the SMA makes it superior as a moving average compared to LMA and EMA. As a consequence, and in order to reduce the dimensionality of the thesis, only the SMA will be employed in the empirical analysis. The Simple Moving Average at time t of the last k observed prices are calculated as

$$SMA_t(k) = \frac{1}{k+1}\sum_{j=0}^{k} P_{t-j}. \tag{2}$$

3.1.2 Trading Rules

In this thesis, we employ several technical trading rules based on Momentum (MOM) and Moving Averages. The MOM rule constitutes one of the most basic market timing rules and is based on the assumption that prices that have increased in the last period, will continue to increase in the next period. Specifically, a Buy (Sell) signal is generated when the last closing price P_t are greater (less) then the closing price $t-k$ periods ago, P_{t-k}. The technical indicator for the k-month MOM rule at time t are computed according to:

$$\text{Momentum rule: Indicator}_t^{MOM(k)} = P_t - P_{t-k}. \tag{3}$$

The simplest and most popular trading rule based on moving averages is the simple moving average (P-MA) rule. In principle, the direction of a trend can be detected by comparing the value of the moving average MA_t to the last closing price P_t. Specifically, prices are trending upward when $MA_t < P_t$ and downward when $MA_t > P_t$. Thus, a Buy (Sell) signal is generated when the value of the moving average is less (greater) then the price. The technical indicator for the P-MA rule at time t is computed as

$$\text{Price minus Moving Average rule: Indicator}_t^{P-MA(k)} = P_t - MA_t(k), \tag{4}$$

where k denotes the averaging window and MA denotes the moving average.

However, as demonstrated by Zakamulin (2017), one of the major problems of the simple market timing rules are the false signals created by price fluctuations, known as "whipsaws". Specifically, when prices trend sideways, strategies based on simple market timing rules tend to produce many false signals which in turn lead to excessive accumulation of transaction costs. In order to avoid unnecessary trades caused by "whipsaws", a trader can alternatively use the Moving Average Crossover (MAC) rule and the Moving Average Envelope (MAE) rule.

The MAC rule consist of a shorter and a longer averaging window where a Buy (Sell) signal is generated when the shorter averaging window crosses above (below) the longer averaging window. The indicator for the MAC rule at time t is given by

$$\text{Moving Average Crossover rule: Indicator}_t^{MAC(s,l)} = MA_t(s) - MA_t(l), \tag{5}$$

where s and l in the MAC rule denotes the size of the shorter (fast) and longer (slow) averaging window respectively. It is worth emphasizing that when the slow averaging window is set to one, the MAC rule becomes the P-MA rule.

The MAE rule is somewhat different in nature and consists of an upper and lower boundary forming an envelope around the moving average. Specifically, no trading takes place as long as the price of the risky asset is within the upper and lower bound. In turn, a Buy (Sell) signal is produced when the price crosses the upper (lower) boundary. The upper and lower boundaries are computed according to:

$$\mathrm{L}_t = \mathrm{MA}_t(n) \times (1 - p), \qquad\qquad \mathrm{U}_t = \mathrm{MA}_t(n) \times (1 + p), \tag{6}$$

where the p in the MAE rule denotes, in terms of percentage, the distance from the moving average to the upper and lower boundary.

The MOM, MAC and MAE rule will all be used in the empirical part of this thesis[9]. The commonality for all trading rules employed in our study is a move to the risky asset (or stay invested in the risky asset) when a Buy signal is generated, and a move to the risk-free asset (or stay invested in the risk-free asset) if a Sell signal is generated. The trading signal for the MAC and MOM rule at time $t+1$ can formally be expressed as:

$$Signal_{t+1} = \begin{cases} Buy, & \text{if Indicator}_t > 0, \\ Sell, & \text{if Indicator}_t \leq 0. \end{cases} \tag{7}$$

Finally, the trading signal for the MAE rule at time $t+1$ can be computed according to following rule:

$$Signal_{t+1} = \begin{cases} Buy, & \text{if } \mathrm{P}_t > \mathrm{U}_t, \\ Sell, & \text{if } \mathrm{P}_t < \mathrm{L}_t, \\ Signal_t, & \text{if } \mathrm{L}_t \leq \mathrm{P}_t \leq \mathrm{U}_t. \end{cases} \tag{8}$$

3.1.3 Accounting for Transaction Costs

For a more realistic approach we need to account for the transaction costs incurring each time the active strategy switches. According to Zakamulin (2014), transaction costs consists of three main components: half-size of the quoted bid-ask spread, market impact cost, and commissions or brokerage fees[10]. However, the main components can vary significantly depending on numerous factors like: the size of the investor, volatility in the market, and liquidity of the asset. For example, large companies (large-cap) are considered more liquid then small companies (small-cap) and thus are facing a smaller bid-ask spread. Conse-

[9] Note that since the slow avering window for the MAC rule starts at 1, the P-MA rule is also included in the study.

[10] Other costs include: taxes and opportunity costs (Freyre-Sanders, Guobuzaite, and Byrne, 2004)

quently, some simplifications has to be made in order to model realistic transaction cost. In particular, we assume that transaction costs for the risky asset are proportional to the volume of trade. For the entirety of the thesis we assume an average one-way transaction cost of 0.25%. This is consistent with the findings by Chan and Lakonishok (1993) and Knez and Ready (1996), among others. The US Treasury Bill, particularly with shorter maturities (1-3 months), is considered highly liquid and with practically zero bid-ask spread (Zakamulin, 2017). We therefore assume that buying and selling the risk-free asset is costless.

Let $(R_1, R_2, ..., R_T)$ be the observed daily (monthly) total returns on the risky asset, and let $(r_{f1}, r_{f2}, ..., r_{fT})$ be the daily (monthly) returns on the risk-free asset proxied by the US Treasury Bill over the full sample period $[1, T]$. The average one-way transaction cost is given by τ. The post-transaction cost returns of the market timing strategy r_t can thus be expressed by:

$$r_t = \begin{cases} R_t, & \text{if } (Signal_t = Buy) \text{ and } (Signal_{t-1} = Buy) \\ R_t - \tau, & \text{if } (Signal_t = Buy) \text{ and } (Signal_{t-1} = Sell) \\ r_{ft}, & \text{if } (Signal_t = Sell) \text{ and } (Signal_{t-1} = Sell) \\ r_{ft} - \tau, & \text{if } (Signal_t = Sell) \text{ and } (Signal_{t-1} = Buy) \end{cases} \tag{9}$$

An alternative approach to this strategy is to sell short the risky asset when a Sell signal is generated. This will potentially allow the investor to earn profits from the short sale and simultaneously obtaining downside protection. While attractive and good in principle, the performance of this approach hinges on the strategy's ability to timely predict both upward and downward trends. This vulnerability makes the short selling strategy inherently more risky and results of empirical studies conducted by Zakamulin (2017) reveal that short selling substantially deteriorates performance. As a consequence, selling short the risky asset is restricted.

3.2 Testing the Profitability of Trading Rules

In our thesis, we employ two commonly used tests in the financial literature to evaluate the performance of market timing strategies. The back-test, commonly known as an in-sample test, and the forward-test, referred to as an out-of-sample test. Both will be presented in the following subsections.

3.2.1 Back-Testing (In-Sample)

A back-test constitutes a simulation technique in which a trader (or researcher) simulates how a strategy would have performed during a specific historical time period. A major benefit of a back-test, in contrast to the forward-test, is the possibility to utilize the full historical sample. However, the process of testing many different strategies and selecting the best performing trading rule is called “data mining”. Zakamulin (2017) defines “data mining” as “*the process of finding the best rule among a great number of alternative rules*”. By “mining” the data and finding the best trading rule, the performance of the trading rule will be biased upward. This upward bias increases the probability of Type 1 error (false positive) and is commonly referred to as the “data mining bias”. In order to understand why the bias occurs one has to recognize that the observed performance consist of two parts: the true performance and a random component. Thus, when a dataset is re-used several times with the purpose of selecting a trading rule, there is always a possibility that the best performing trading rule appeared due to chance alone and not superior performance. By “mining” the data, we tend to find the trading rule benefitting the most from the random component in the observed performance. It is possible to correct the “data-mining” problem from an in-sample test by adjusting the p-value of the test (see for example White (2000)). Alternatively, a common practice to mitigate “data mining bias” is to discount the reported Sharpe Ratios by 50% (Harvey and Liu, 2015).

3.2.2 Forward-Testing (Out-of-Sample)

Motivated by the discussion about the “data mining bias”, the rationale behind the forward-test is simple. Since the in-sample test systematically overstates the performance of the best observed trading rule, we need an additional segment of data to validate the trading rule. The out-of-sample performance method applied in this study is based on simulating returns in a way that resembles real-life trading. In contrast to the in-sample test, the out-of-sample performance is considered to be a more reliable estimate of the true performance (Zakamulin, 2015). We follow closely the exposition provided in Zakamulin (2014). For simplicity we shorten the exposition by outlining the methodology for the MOM(k) trading rule.

The procedure for the out-of-sample test starts with splitting the historical sample into two subsets. The first subset is denoted by in-sample $[1, s]$ and the second subset is denoted by out-of-sample $[s + 1, T]$. Where T denotes the final observation in the sample and s denotes the split point. The in-sample period (or the first subset) is used to find the best trading rule given some predetermined performance measure. The best trading rule is subsequently evaluated on the second subset (or the out-of-sample data).

In principle there are two alternative approaches to perform an out-of-sample test. The first approach is to use a rolling estimation scheme, where the size of the initial in-sample period is constant and rolled forward. If this estimation scheme is selected one usually assumes that the markets (and the optimal strategy) are dynamic and not constant through time. The optimal lookback period k^* for the market timing strategy when using the rolling estimation scheme is given by

$$k_t* = \arg \max_{k \in [k^{min},\ k^{max}]} PM(r_{s-n+1}, r_{s-n+2}, \ldots, r_s), \tag{10}$$

where k^{min} and k^{max} is the minimum and maximum value of k respectively, and PM is a performance measure representing the optimization criterion.

The second approach is an expanding estimation scheme where the size of the in-sample subset increases with each iteration. This approach is preferable if one assumes that market dynamics are constant through time. For example, it is widely believed that the 10-month SMA is a superior strategy[11]. The optimal lookback period k^* for the market timing strategy when using the expanding estimation scheme is given by

$$k_t* = \arg \max_{k \in [k^{min},\ k^{max}]} PM(r_1, r_2, \ldots, r_t). \tag{11}$$

In our study, both rolling (walk-forward test) and expanding (forward test) estimation schemes will be used to evaluate the out-of-sample performance.

The major problem in out-of-sample testing is deciding on a split point. One might think that the split point between the initial in-sample and out-of-sample subsets can be selected

[11] See for example Faber (2007)

arbitrarily, but this is not the case. In recent years, several researchers[12] have demonstrated that the choice of split point significantly alters the performance of the test. A similar result will also be demonstrated in Section 5 of the thesis. In general, Zakamulin (2014) argues that the market timing strategy usually outperforms (underperforms) the passive benchmark during bear (bull) markets. This characteristic makes the outperformance of the market timing strategy extremely non-uniform. In fact, we can easily obtain erroneous results that do not reflect the true performance of the market timing strategy by selecting split points before a major market crash (for example 1928, 1986, 1999, 2006). This means that the out-of-sample performance measurement is not a complete remedy for the "data mining bias". As a consequence, Zakamulin (2014) argues that the initial in-sample period should have a minimum length, and include both bull and bear markets.

In practice, there are three alternative ways to choose the split point. (1) We can choose a split point near the end of the sample in similar fashion to Sullivan et al. (1999), (2) we can set the split near the start of the sample like Zakamulin (2014), or (3) somewhere in between. Motivated by the discussion above, we report the out-of-sample performance with two different split points. First we report the performance when the split point is set to January 1953, leaving 25 years for the initial in-sample and the remaining 63 years for the out-of-sample period. Subsequently, we report the out-of-sample performance when the split point is set to January 1970, leaving 42 years for the initial in-sample and the remaining 46 years for the out-of-sample period. Finally, in order to examine how the performance depends on the choice of split point, plots reporting the performance for each possible split point will be presented in Section 5.

3.3 Performance Measures

There is a vast world of different performance measures in the finance literature (see e.g Cogneau and Hübner (2009)), all with different advantages and drawbacks. It may therefore be a difficult task to choose what measure to use, and not necessarily one single solution. The simplest measure of performance is the excess return, i.e. the return on the asset, portfolio or strategy less the risk-free rate of return. But in itself this is not a good measure, because it disregards risk. If risk is not a factor, the investor can borrow money to lever his portfolio

[12]See Rossi and Inoue (2012), Hansen and Timmermann (2013) and Zakamulin (2014)

and achieve greater expected mean return. However, according to conventional wisdom, investors require compensation for taking on risk. Accordingly, a risk-adjusted measure is more appropriate. In this thesis we apply the famous, and widely used Sharpe ratio. We also test for abnormal returns, using the alpha in the Fama-French-Carhart 4-factor asset pricing model as a measure.

3.3.1 The Sharpe Ratio

The Sharpe ratio was first introduced by William Sharpe in 1966 (Sharpe, 1966) as the reward-to-variability ratio. He later revisited and generalized the ratio (Sharpe, 1994) and referred to it with its then more common name, the Sharpe ratio. The ratio is calculated as the mean excess return of asset i over the standard deviation of excess return of asset i, and can formally be written as:

$$SR_i = \frac{\mu(r_i^e)}{\sigma(r_i^e)}, \tag{12}$$

where $\mu(r_i^e)$ is the mean of excess return on asset i and $\sigma(r_i^e)$ is the standard deviation of excess return on asset i.

Even though the Sharpe ratio is a widely used measure, it has some strong assumptions, which in real life rarely are fulfilled. E.g. it assumes the presence of a risk-free asset, and that one without frictions can lend and borrow unlimited amounts at its rate of return. It is also criticized for using the standard deviation as a measure of risk, because it penalizes upside potential in the same manner as downside risk. To cope with this issue Sortino and Price (1994) came up with an alternative measure which only accounts for the downside risk, namely the Sortino ratio. It should be noted that it builds on the same assumptions as the Sharpe ratio, and apart from not including upside potential as a risk, it retains all other weaknesses of the Sharpe ratio.

Even though there might be better, more suitable performance measures than the Sharpe ratio, several studies (see e.g. Eling and Schuhmacher (2007), Eling (2008) and Auer (2015)) have found that the choice of performance measure does not influence the evaluation of a variety of different risky portfolios. Zakamulin (2017) also tests how the choice of the best performing market-timing strategies varies when using excess return, Sharpe- and Sortino ratio, and finds rank correlations to lie between positive 0.97 and 1. We therefore choose to stick with the Sharpe ratio, because it is the most widely used performance measure in the finance literature, and presumably what potential readers are most familiar with.

3.3.2 The Fama-French-Carhart 4-Factor Model

As an estimate of abnormal returns we use the alpha of the Fama-French-Carhart 4-factor model. This is done in a similar fashion to what Jensen (1968) first intoduced with the one factor model. The 4-factor model was presented by Carhart (1997), and is an extension of the Fama-French 3-factor model (Fama and French, 1993). Carhart added a fourth factor to the 3-factor model motivated by the models inability to explain cross-sectional variations in momentum-sorted portfolio returns. He augmented the previous model by including the one-year momentum factor (PR1YR) of Jegadeesh and Titman (1993). We then get a model that explains the excess return of an asset, portfolio or trading strategy using four factors. The first being the excess return on the market portfolio (MKT), which simply gives the return of the market portfolio less the risk-free rate of return. The second, the size, or small minus big (SMB) risk factor, which gives the difference in return between small and large companies. The third, the value, or high minus low (HML) risk factor, which gives the difference in return between value and growth companies. And finally, the fourth, the one-year momentum factor (PR1YR), which considers the difference in expected return between stocks that performed well and poor in the prior year [13]. These factors are calculated at the end of either each trading day or each month, and will according to the model explain the excess return on our asset, portfolio or trading strategy. The estimation is performed by a multivariate ordinary least squares regression. Any linear asset pricing model states that if the market is in equilibrium, the model explains the returns, i.e. $\alpha = 0$. Thus, an alpha unequal zero represents an abnormal return. To estimate abnormal returns we therefore run a multivariate ordinary least square regression, with excess strategy returns as the dependent variable, and the mentioned four factors as independent variables. The interception, or constant, of the regression will represent the alpha, and thus the estimate of abnormal return. Formally the model can be expressed as:

$$r_{i,t}^e = \alpha_i + \beta_{1,i} \cdot MKT_t + \beta_{2,i} \cdot SMB_t + \beta_{3,i} \cdot HML_t + \beta_{4,i} \cdot PR1YR_t + \varepsilon_{i,t}, \tag{13}$$

[13]For more detailed explanations of the factors, see Carhart (1997)

where $r_{i,t}^e$ gives the excess return on strategy i at time t. α_i gives the intercept, and is the estimate of abnormal return for strategy i. MKT_t, SMB_t, HLM_t and $PR1YR_t$ gives the excess return on the market portfolio, small minus big risk factor, high minus low risk factor and the one-year momentum factor at time t respectively. $\beta_{1,i}$, $\beta_{2,i}$, $\beta_{3,i}$ and $\beta_{4,i}$ gives the corresponding exposure of strategy i to factors MKT, SMB, HML and PR1YR respectively. Finally $\varepsilon_{i,t}$ gives the regression residuals.

3.3.3 Drawdown

An alternative approach of measuring risk is to examine the downside risk, or historical drawdowns. Drawdown is a measure in percentage of the decline from a peak to a following trough. Formally, the drawdown at time t can be expressed as

$$D(t) = \frac{P_{peak} - P_{trough}}{P_{peak}}, \tag{14}$$

where P_{peak} is the price at the peak, and P_{trough} is the price at the following trough. In addition to reporting the average and median drawdown, we also present the maximum drawdown and the average maximum drawdown. While the maximum drawdown is the single largest drawdown, the average maximum drawdown refers to the mean of the 10 largest drawdowns during the sample period. The drawdown is a measure of financial risk and is particularly important for investors with shorter investment horizons.

3.4 Statistical Tests for Outperformance

There are several different methods for testing if our results are statistically significant. Broadly we can divide the tests into two sub-categories, namely parametric and non-parametric tests. Parametric tests builds on numerous assumptions, including that the data is normally distributed and without serial dependence. As the finance literature documents this is rarely the case for financial data, which often feature non-normal distributions, heteroscedasticity and serial dependence. This is also in line with what we have found in our data, with non-normal distributions of returns, difference in volatility over time and significant serial dependence, especially in the daily data. Consequently, assumptions underlying parametric tests are violated and these test will usually be invalid.

Non-parametric tests do not require making assumptions about the probability distributions because it uses randomization methods to estimate the distribution of the test statistic. They are therefore more suitable when analyzing financial data. The bootstrap methodology is the most popular non-parametric randomization method and is what we will employ in our study. The standard bootstrap method was introduced by Efron (1979) as a more primitive and robust method than the Quenouille-Tukey jackknife (as reviewed in Miller (1974)). It has later been proven unfit for data containing serial dependence, and other methods, such as overlapping and non-overlapping block bootstrap ((Carlstein, 1986) and (Kunsch, 1989)) and stationary block bootstrap (Politis and Romano, 1994) has been suggested.

3.4.1 Statistical Test for the Sharpe Ratio

As we here use the Sharpe ratio as the performance measure, let $\Delta = SR_{MT} - SR_{BH}$, where SR_{MT} gives the Sharpe ratio of the market timing strategy, SR_{BH} gives the Sharpe ratio of the passive benchmark (buy-and-hold) and consequently Δ gives the difference in performance. We then set the following null and alternative hypotheses:

$$H_0 : \Delta \leq 0 \qquad \text{versus} \qquad H_A : \Delta > 0. \tag{15}$$

A parametric test for the Sharpe ratio was first introduced by Jobson and Korkie (1981) with a later correction by Memmel (2003). The test assumes joint normality in the two return series, and is obtained via the test statistic

$$\tilde{z} = \frac{\Delta}{\sqrt{\frac{1}{T}\left[2(1-\hat{\rho}) + \frac{1}{2}(SR_{MT}^2 + SR_{BH}^2 - 2\hat{\rho}^2 SR_{MT} SR_{BH}\right]}}, \tag{16}$$

which is asymptotically distributed as a standard normal when the sample size is large. The advantage of using this test is the fast and fairly easy computation. Further, the test statistic follows a well known distribution, making it easy to find the p-value quickly. The problem is however the strong distributional assumptions of the return series. As previously mentioned, and also later documented, our return series are not normally distributed, hence such a test will be of little or no real value.

Instead we apply a bootstrap methodology, free of distributional assumptions. Because several of our datasets contain significant levels of serial dependence, we use one of the block bootstrap methods. When testing for outperformance the preferred method in the financial literature seems to be the stationary block bootstrap method (see e.g. Sullivan et al. (1999), Welch and Goyal (2008) and Kirby and Ostdiek (2012)), we will therefore follow the trend, and use the same method.

All block bootstrap methods resembles the standard bootstrap method of Efron (1979). The difference is that block bootstrap resamples blocks of observations instead of one-by-one. This is also why block bootstrap methods are more suitable for data containing serial dependence, because it does not break up the dependence in the data. While the overlapping and non-overlapping block bootstrap methods have fixed block sizes, the stationary method generates the block length from a geometric distribution. In this way stationary data keeps its stationary properties also after the resampling. Following the lines of Zakamulin (2017) the stationary block bootstrap method can be described as follows.

Let $\{r_t^e\}$ and $\{R_t^e\}$ be a paired sample of excess returns, where $t = \{1, 2, ..., T\}$. Then we draw N random samples with replacement, of blocks of the original sample. The block length ψ_i^b, where b denotes the index for the bootstrap number, is generated from a geometric distribution with probability φ. This gives an average block length of $\frac{1}{\varphi}$, thus φ is chosen according to $\varphi = \frac{1}{\psi}$, where ψ is the required average block length. The ith block begins from a random index i generated from the discrete uniform distribution on {1,2,...T}. Since the block length is not limited from above, $\psi_i^b \in [1, \infty)$, and the block can start at any time t, the data is wrapped around like a circle, so that 1 follows T and so on. The resampling is done so that the resampled psuedo-time series has the same number of observations as the original sample. Observe also that since we got a paired sample of excess returns, (r_t^e, R_t^e), the resampled data is also paired, i.e. the resampling process gives two new psuedo-time series. Consequently the historical correlation between the two return series remains.

A problem in any of the block bootstrap methods is choosing the block length, or here, the required average block length. In this study we apply the method proposed by Politis and White (2004) (with the correction made in Patton, Politis, and White (2009)) when selecting the required average block length.

After resampling the two excess return series, we want to test if the market timing strategy statistically outperforms the passive benchmark. To test the null hypothesis of equal or underperformance, we calculate the difference Δ^b between the Share ratio of $\{r^e_{t_b}\}$ and $\{R^e_{t_b}\}$. By repeating the resampling procedure N times and for each time calculating Δ^b, the bootstrap distribution of Δ is constructed. To estimate the significance level of the test, we count how many times Δ^b is less than zero. If the number of negative realizations of Δ^b is denoted as n, we compute the p-value as $\frac{n}{N}$.

3.4.2 Statistical Test for Alpha

When testing the performance using the Fama-French-Carhart 4-factor alpha as performance measure, we test if we have a statistically significant abnormal return. We therefore form the following null and alternative hypothesis:

$$H_0 : \alpha_i = 0 \qquad \text{versus} \qquad H_A : \alpha_i \neq 0 \tag{17}$$

We can easily test the hypothesis using a standard parametric t-test of the regression intercept coefficient. The test statistic for strategy i, $t_i^\diamond$ is then given by

$$t_i^\diamond = \frac{\hat{\alpha}_i}{SE_{\hat{\alpha}_i}}, \tag{18}$$

where $\hat{\alpha}_i$ is the estimated alpha of the Fama-French-Carhart 4-factor model for strategy i and $SE_{\hat{\alpha}_i}$ is the standard error of the distribution of the same alpha. As the test statistic follows a well known distribution, the p-value is easily found. However, the problem is that the test assumes normality in the data, an assumption that is often violated in financial data, leaving the results of the test invalid. We therefore once again see the need for a non-parametric test to cope with the non-normal data. Several studies (see e.g. Horowitz (2003), Kosowski, Timmermann, Wermers, and White (2006) and Fama and French (2010)) argue that bootstrap methods gives good evaluation of significance of alpha estimates in regression models. We therefore use a residual bootstrap method to test our hypothesis. We use notation similar to Kosowski et al. (2006), and describe the procedure as follows.

To prepare our bootstrap we run the Fama-French-Carhart 4-factor regression given in equation (13). We save the estimated coefficients $\left\{\hat{\alpha}_i, \hat{\beta}_{1,i}, \hat{\beta}_{2,i}, \hat{\beta}_{3,i}, \hat{\beta}_{4,i}\right\}$ and residuals $\{\hat{\varepsilon}_{i,t}, t = T_{i0}, ..., T_{i1}\}$, where T_{i0} and T_{i1} are the first and last observation of strategy i respectively. Using the standard bootstrap method, we draw a new sample with replacement of the stored residuals from the regression, creating a psuedo-time series of resampled residuals $\{\hat{\varepsilon}_{i,t_\varepsilon}, t_\varepsilon = z^b_{T_{i0}}, ..., z^b_{T_{i1}}\}$, for each strategy i. b is an index for the bootstrap number (i.e. $b = 1$ for bootstrap number 1, $b = 2$ for bootstrap number 2, and so on), and each of the time indices $z^b_{T_{i0}}, ..., z^b_{T_{i1}}$ is drawn randomly from $T_{i0}, ..., T_{i1}$. This is done in such a way that it reorders the original sample of $T_{i1} - T_{i0} + 1$ for strategy i. Next, we construct a pseudo-time series of excess returns, using the stored values of factor loadings and the bootstrapped residuals, while imposing our null hypothesis, $\alpha = 0$. The new series of excess return will then have a true alpha of zero by construction. We then run a new regression with the pseudo-time series of excess returns as the dependent variable, and the four factors in the Fama-French-Carhart 4-factor model as independent variables. We now have a new estimate for alpha. Repeating this for all bootstrap iterations $b = 1, ..., N$, we construct a distribution of the alphas, θ_i, for each strategy i. After the bootstrap procedure we test, and find that we cannot reject the null hypothesis of normality in the distribution of the bootstrapped alphas. We therefore use our estimated alphas and the standard error of the distribution of the bootstrapped alphas, and perform a standard t-test. The test statistic for strategy i, t^*_i, is then given by:

$$t^*_i = \frac{\hat{\alpha}_i}{SE_{\theta_i}}, \tag{19}$$

where $\hat{\alpha}_i$ is the estimated alpha of the Fama-French-Carhart 4-factor model for strategy i and SE_{θ_i} is the standard error of the distribution of bootstraped alphas for strategy i. We see that this test statistic is very similar to that of the parametric test, the only difference being the distribution of which the standard error is found. But as we now know that this distribution is normal, the test is valid. Once more the p-value is easily found in the Student's t-distribution.

4 Data

In our empirical study we use both daily and monthly data of two stock indices, two different stock portfolio composition schemes, as well as the risk-free rate of return and the risk factors in the Fama-French-Carhart 4-factor model. Motivated by the access to quality data, all data spans from around July 1st, 1926 to December 31st 2015. The first stock index is the Standard and Poor's Composite stock price index (S&P Composite). This is a value-weighted stock index, which from 1926 to 1956 consisted of only 90 stocks, whereas it form 1957 until today has consisted of 500 stocks. It has so been known as the S&P 500 index. The index is composed of large US companies, based on market size, liquidity and industry group representation, and is considered one of the best representations of the US stock market. The second stock index is the Dow Jones Industrial Average index (DJIA). It is a price-weighted stock index, consisting of 30 large US companies, selected to represent a cross-section of US industry. For both S&P Composite and DJIA we use both capital gain return, and the total return. Capital gain return is simply computed as the change in daily or monthly closing prices, while total return gives the sum of capital gain return and the dividend return. All data on the two stock indices are provided by Valeriy Zakamulin[14].

All stock portfolios consist of US companies, and the two different composition schemes are portfolios sorted on size, and portfolios sorted on book-to-market ratio. For both composition schemes we use the high and low quintile portfolios, which means we examine four different stock portfolios for both daily and monthly data. For portfolios sorted in size, the low quintile portfolio is denoted as *Small*, while the high quintile portfolio is denoted as *Large*. For portfolios sorted on book-to-market, the low quintile portfolio is denoted as *Growth*, whereas the high quintile portfolio is denoted as *Value*.

The four risk factors are the excess return on the market portfolio (MKT), the return on small minus big (SMB), the return on high minus low (HML) and the one-year momentum factor (PR1YR), all for the US market. The risk-free rate of returns are proxied by the Treasury Bill rate. Data for all stock portfolios, as well as the risk-free rate and the risk factors are retrieved from Kenneth French's online database[15]. For further explanation of the composition of stock portfolios or risk factors, see Kenneth French's online database. All

[14]To see where Zakamulin retrieves the data, see e.g. Zakamulin (2014)

[15]`http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

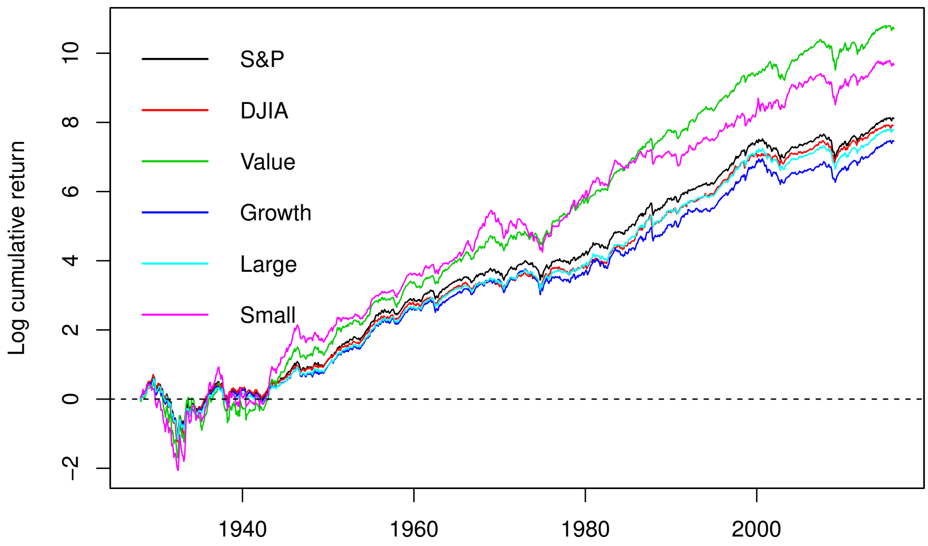


Figure 1: Time series plot of cumulative logarithmic returns of the six different risky assets used in the study. The plot is constructed by monthly total returns, i.e including and reinvesting dividends, and covers the period January 1926 to December 2015.

descriptive statistics given below includes data ranging from January 1st 1928 to December 31st 2015, a total of 88 years. The reason why we start reporting descriptive statistics from 1928 is that this is the starting point of our earliest in-sample period. All data before this date is only used in the initial moving average windows.

4.1 Standard and Poor's Composite Index

Table 1 summarizes the descriptive statistics for returns of the Standard and Poor's Composite index. Together with reporting statistics for the whole period we also divide the sample into two non-overlapping sub-periods of 44 years. The first ranging from January 1st 1928 to December 31st 1971, while second from January 1st 1972 to December 31st 2015. In doing this we may be able to see if the statistics of returns changes over the two periods.

From the Anderson-Darling normality test, we see that the null hypotheses of normal distribution are rejected for all return series. We also see that while the mean capital

S&P Composite

Statistics	1928-2015			1928-1971			1972-2015		
	CAP	**TOT**	**RF**	**CAP**	**TOT**	**RF**	**CAP**	**TOT**	**RF**
N	1056	1056	1056	528	528	528	528	528	528
Mean %	7.21	11.05	3.47	6.42	11.14	2.03	8.00	10.95	4.92
Std. dev. %	18.98	19.01	0.90	22.11	22.15	0.54	15.25	15.27	0.99
Min %	-29.94	-29.43	0.00	-29.94	-29.43	0.00	-21.76	-21.54	0.00
Max %	42.22	42.91	1.36	42.22	42.91	0.66	16.30	16.78	1.36
Skewness	0.30	0.38	1.04	0.55	0.62	0.83	-0.45	-0.44	0.53
Kurtosis	12.32	12.58	4.18	12.33	12.56	2.81	4.96	4.99	3.31
Anderson-Darling	**13.55**	**13.52**	**23.83**	**10.41**	**10.37**	**17.68**	**2.05**	**2.07**	**5.99**
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 1: Descriptive statistics for monthly S&P Composite index returns for three sample periods. **TOT** denotes the total market return, **CAP** denotes the the capital gain return and **RF** denotes the risk-free return. The mean and standard deviation of returns are annualized and reported in percentages. **Anderson-Darling** represent the test statistic in the normality test and the p-value of the test are reported in brackets below. Values in bold text are statically significant at the 5% level.

returns seem to increase from the first to the second sub-period, the mean total returns remain practically the same. The standard deviation how ever, seems to lower for the second sub-period. This implies that the risk of the index has changed over time. Also interesting is the fact that the first sub-period exhibits positive skewness, while for the second it is negative. We see that the kurtosis is far less in the second sub-period compared with the first, indication that the tails of the distribution are slimmer for the second sub-period, and hence returns are more centralized around the mean. It is also worth noting that the risk-free rate is substantially higher in the second sub-period

4.2 Dow Jones Industrial Average Index

Table 2 summarizes the descriptive statistics for returns of the Dow Jones Industrial Average index. Also here we report descriptive statistics for both the complete period, and two non-overlapping sub-periods of 44 years.

DJIA

Statistics	1928-2015			1928-1971			1972-2015		
	CAP	**TOT**	**RF**	**CAP**	**TOT**	**RF**	**CAP**	**TOT**	**RF**
N	1056	1056	1056	528	528	528	528	528	528
Mean %	6.78	10.69	3.37	5.62	10.14	1.93	7.94	11.23	4.81
Std. dev. %	18.40	18.35	0.89	21.17	21.09	0.53	15.15	15.13	0.98
Min %	-30.70	-29.88	-0.06	-30.70	-29.88	-0.06	-23.22	-22.90	0.00
Max %	40.18	40.46	1.35	40.18	40.46	0.64	14.41	14.75	1.35
Skewness	0.02	0.07	1.04	0.23	0.27	0.86	-0.48	-0.47	0.53
Kurtosis	10.90	10.95	4.17	11.10	11.16	2.87	5.30	5.34	3.32
Anderson-Darling	**13.88**	**13.74**	**24.12**	**10.91**	**10.74**	**18.49**	**2.22**	**2.30**	**5.44**
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 2: Descriptive statistics for monthly DJIA returns for three sample periods. **TOT** denotes the total market return, **CAP** denotes the the capital gain return and **RF** denotes the risk-free return. The mean and standard deviation of returns are annualized and reported in percentages. **Anderson-Darling** represent the test statistic in the normality test and the p-value of the test are reported in brackets below. Values in bold text are statically significant at the 5% level.

In the descriptive statistics we see much the same as for the S&P Composite index. Once again the null hypotheses of normality are rejected across all return series. We also see that mean returns seem to be higher in the second sub-period, and here also for the total return. Further we see that standard deviation is lower in the second sub-period and that it also here seems to be a difference in index volatility between the two sub-periods. When it comes to skewness we have the same as for the S&P Composite index, positive in the first sub-period, and negative in the second. For kurtosis we observe larger values in the first sub-period.

4.3 Portfolios Sorted on Size and Value

Table 3 summarizes the descriptive statistics for a total of four stock portfolios sorted on size and value. Additionally, the descriptive statistics for the risk-free rate is provided in the far-right column.

From Table 3 we see the Value portfolio display a mean and standard deviation of 15.94% and 28.25%, and the Small portfolio display a mean of 15.47% and the largest standard deviation of 31.41%. Particularly noteworthy is the max returns for the Small and Value portfolio of 95.92% and 82.48% respectively. The return distributions for all portfolios are leptokurtic with values of kurtosis ranging from 7.82 to 29.39. By examining the value of the Anderson-Darling test and its associated p-value, it becomes evident that we reject the null hypothesis of normality for all portfolios.

Portfolios sorted on Size and Book-to-Market

Statistics	Small	Large	Growth	Value	RF
N	1056	1056	1056	1056	1056
Mean %	15.47	10.48	10.32	15.94	3.37
Std. dev. %	31.41	17.97	19.01	28.25	0.89
Min %	-33.07	-28.69	-29.27	-38.00	-0.06
Max %	95.92	36.95	32.76	82.48	1.35
Skewness	2.67	0.14	-0.17	2.11	1.04
Kurtosis	29.39	10.41	7.87	25.24	4.17
Anderson-Darling	**27.93**	**11.77**	**8.33**	**30.81**	**24.12**
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 3: Descriptive statistics for monthly returns of portfolios sorted on size and book-to-market from January 1928 to December 2015. **Small** denotes the total market return of a portfolio consisting of the 20% smallest US stocks sorted on size, **Large** denotes the total market return of a portfolio consisting of the 20% largest US stocks sorted on size. **Growth** denotes he total market return of a portfolio consisting of the lowest 20% of US stocks sorted on book-to-market, **Value** denotes he total market return of a portfolio consisting of the highest 20% of US stocks sorted on book-to-market and **RF** denotes the risk-free rate of return. The mean and standard deviation of returns are annualized and reported in percentages. **Anderson-Darling** gives the test statistic in the normality test and the p-value of the test are reported in brackets below. Values in bold text are statically significant at the 5% level.

4.4 Risk Factors

Table 4 presents the descriptive statistics for the four Fama-French-Carhart factors over the full sample period and two sub-periods. By examining the full sample, we observe that the factors MKT and PR1YR displays the greatest mean return and standard deviation for monthly data. This characteristic can also be seen for the two sub-periods. All factors except PR1YR displays greater mean return in the first sub-period. In the second sub-period we see the opposite. From Panel B we see that the market premium in the full sample displays the highest standard deviation with 18.77%. Further, we note that the standard deviation for all factors is higher in the first sub-period and lower in the second sub-period when compared to the full sample. This indicates that the period from 1928 to 1971 was considerably more volatile then the period from 1972 to 2015.

5 Empirical Results

In this section we present the results of our empirical analysis. We employ the methods described in Section 3, and present results both for in-sample and out-of-sample simulations. All results are found and presented using the R software. By combining the SMA with all

Risk Factors

	MKT	**SMB**	**HML**	**PR1YR**
Panel A: Mean Return %				
1928 - 2015	7.52	2.63	4.74	8.11
1928 - 1971	8.72	3.31	5.12	7.62
1972 - 2015	6.31	1.94	4.35	8.61
Panel B: Std. Dev. %				
1928 - 2015	18.77	11.22	12.19	16.50
1928 - 1971	21.35	11.65	13.95	17.60
1972 - 2015	15.78	10.78	10.13	15.34

Table 4: Descriptive statistics for monthly factor returns used in the Fama-French-Carhart 4-factor model. **MKT** denotes the excess return of the market portfolio. **SMB** and **HML** denotes the premium on the size and book-to-market factor, respectively. **PR1YR** denotes the one-year momentum factor of Jegadeesh and Titman (1993). The mean and standard deviation of returns are annualized and reported in percentages.

technical trading rules, we obtain the following set of rules for the monthly data:

MOM(k) for $k \in [2, 18]$, a total of 17 strategies;

MAC(s, l) for $s \in [1, 7]$ and $l \in [2, 18]$, a total of 98 strategies;

MAE(k, p) for $k \in [2, 18]$ and $p \in [0{,}25, 0{,}5, \ldots, 4]$, a total of 272 strategies;

This amount to a combination of 387 tested strategies for monthly data. For daily data we obtain the following set of rules:

MOM(k) for $k \in [50, 100, 150, 170, 180, \ldots, 260, 280, \ldots, 360]$, a total of 18 strategies;

MAC(s, l) for $s \in [1, 2, \ldots, 6, 8, \ldots, 16, 20, 40, \ldots, 100]$ and $l \in [50, 100, 150, 170, 180, \ldots, 260, 280, \ldots, 360]$, a total of 284 strategies;

MAE(k, p) for $k \in [50, 100, 150, 170, 180, \ldots, 260, 280, \ldots, 360]$ and $p \in [0{,}25, 0{,}5, \ldots, 4]$, a total of 288 strategies;

This amount to a combination of 590 unique strategies for daily data.

5.1 In-Sample Tests

First we take a closer look at the results of the in-sample simulations. For each dataset we report summary statistics for the best performing trading rule and its equivalent buy-and-hold benchmark. In addition to presenting the results for the full sample, we also provide results for three non-overlapping sub-periods.

5.1.1 S&P Composite and DJIA

By studying the results for both panels in Table 5, it becomes evident that when the underlying index is the S&P Composite, the mean returns and standard deviations are generally lower for the market timing strategies. However, judging by the Sharpe ratios, the reduction in mean return is *less* than the reduction in standard deviation, resulting in improved risk-adjusted performance. The only exception can be found in the sub-period from 1958 to 1987, which report greater mean return than the corresponding buy-and-hold strategy. Looking closer at Δ Sharpe, all reported market timing strategies outperformed its corresponding benchmark. In fact, with the exception of the period from 1988 to 2015, all Sharpe ratios in Panel A are statistically significantly greater then the Sharpe ratio of its benchmark. It is also noteworthy that the moving average envelope produce the best trading rule for all periods tested in Panel A. Looking at the results in Panel B, we observe that the trading rule MAC(2, 10) produce statistically significant outperformance measured by Δ Sharpe for the full sample. For the three sub-periods in Panel B, we find Sharpe ratios ranging from 36% to 63% higher then the Sharpe ratio of the passive benchmark. However, using conventional statistical levels, we cannot reject the null hypothesis. Its important to note that the test results depends on the number of observations. In that sense, we expect daily data to produce more statistical significant results.

S&P Composite: In-Sample

		Mean %	Std.dev. %	Sharpe	Δ Sharpe	p-value	Best Strategy
Panel A: Daily							
1928-2015	MT	10.88	12.06	0.62	**0.21**	(0.01)	MAE(230,2.5)
	BH	11.02	18.76	0.41			
1928-1957	MT	10.82	13.73	0.71	**0.29**	(0.04)	MAE(320,2.25)
	BH	10.95	23.47	0.42			
1958-1987	MT	11.91	8.82	0.69	**0.32**	(0.03)	MAE(150,0.25)
	BH	10.82	13.63	0.37			
1988-2015	MT	11.01	11.68	0.67	0.21	(0.06)	MAE(240,3.5)
	BH	11.31	17.84	0.45			
Panel B: Monthly							
1928-2015	MT	10.52	12.05	0.58	**0.19**	(0.02)	MAC(2,10)
	BH	11.05	19.01	0.40			
1928-1957	MT	10.33	14.08	0.65	0.25	(0.08)	MAE(14,2)
	BH	11.32	25.44	0.40			
1958-1987	MT	11.35	10.78	0.49	0.16	(0.11)	MAC(2,10)
	BH	11.00	14.87	0.33			
1988-2015	MT	10.52	10.19	0.72	0.19	(0.13)	MAE(8,1.5)
	BH	10.80	14.42	0.53			

Table 5: Summary statistics and performance of the best active market timing strategies from in-sample simulations. **Best strategy** gives the best performing strategy for the given period, and for which we present summary statistics. Means and standard deviations are annualized and reported in percentages. **Sharpe** denotes the annualized Sharpe ratios, while Δ **Sharpe** denotes the difference in annualized Sharpe ratios between the active strategy and the passive benchmark. It is calculated as $\Delta = SR_{MT} - SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. For each Sharpe ratio we test the null hypothesis $H_0 : SR_{MT} \leq SR_{BH}$. P-values are reported in brackets and values in bold text are statically significant at the 5% level.

The results presented in Table 6 reports the best trading rules when the passive benchmark is the DJIA index. Some remarkable differences emerge. In particular, in Panel B we no longer find statistical evidence of outperformance over the full sample. Instead, the MOM(18) rule produce superior performance in the period from 1928 to 1957, with a statistically significant Δ Sharpe of 0.33. In general, we observe that the best trading rules for S&P Composite in Table 5 produce higher Sharpe ratios then the best trading rules for DJIA in Table 6. We also note that, similar to the S&P Composite, the moving average envelope generally produce the best trading rule, in particular for daily data.

DJIA: In-Sample

		Mean %	Std.dev. %	Sharpe	Δ Sharpe	p-value	Best Strategy
Panel A: Daily							
1928-2015	MT	9.82	11.49	0.56	**0.16**	(0.05)	MAE(260,3)
	BH	10.66	18.20	0.40			
1928-1957	MT	11.71	12.56	0.85	**0.44**	(0.01)	MAE(260,2.75)
	BH	10.15	22.24	0.41			
1958-1987	MT	11.54	9.43	0.60	**0.30**	(0.02)	MAE(200,0.5)
	BH	10.14	14.29	0.30			
1988-2015	MT	10.96	12.53	0.62	0.12	(0.19)	MAC(100,320)
	BH	11.77	17.11	0.50			
Panel B: Monthly							
1928-2015	MT	9.53	12.21	0.50	0.11	(0.13)	MAE(8,3)
	BH	10.69	18.35	0.40			
1928-1957	MT	11.11	13.94	0.72	**0.33**	(0.03)	MOM(18)
	BH	10.54	24.09	0.39			
1958-1987	MT	10.82	11.07	0.45	0.15	(0.13)	MAC(3,9)
	BH	10.22	14.87	0.29			
1988-2015	MT	11.34	11.32	0.72	0.14	(0.14)	MAC(6,17)
	BH	11.35	14.18	0.58			

Table 6: Summary statistics and performance of the best active market timing strategies from in-sample simulations. **Best strategy** gives the best performing strategy for the given period, and for which we present summary statistics. Means and standard deviations are annualized and reported in percentages. **Sharpe** denotes the annualized Sharpe ratios, while Δ **Sharpe** denotes the difference in annualized Sharpe ratios between the active strategy and the passive benchmark. It is calculated as $\Delta = SR_{MT} - SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. For each Sharpe ratio we test the null hypothesis $H_0 : SR_{MT} \leq SR_{BH}$. P-values are reported in brackets and values in bold text are statically significant at the 5% level.

5.1.2 Portfolios Sorted on Size and Value

Table 7 below presents the in-sample results when the underlying risky asset is a portfolio of small and large stocks for daily and monthly data. The results in Panel A are striking. Mean returns are significantly higher and standard deviations are considerably lower, when compared to its corresponding buy-and-hold strategy. In particular, the best trading rule for the sub-period between 1958 to 1987 are able to produce extraordinary risk-adjusted performance with a Sharpe ratio of 2.5. Moreover, it is noteworthy to observe that all the best trading rules are practically identical. Looking at Panel B, C and D, we observe that all measures of Δ Sharpe except one are statistically significant at the 5% level. The only Δ Sharpe not statistically significant can be located in Panel D for the sub-period 1958 to 1987. However, looking at the p-value, the trading rule would be statistically significant if the statistical threshold level was 7%. Once more, the moving average envelope rule generally produces the best trading rule.

Further we study the results from the portfolios sorted on value. Table 8 presents the best trading rule for portfolios of growth and value stocks. Similar to the results presented in Table 5 and 6, every market timing strategies reports lower standard deviation then its corresponding benchmark. In addition, 75% of the best trading rules produce mean return greater then its passive benchmark. In turn, this results in statistically significant measures of Δ Sharpe for a majority of the best trading rules. However, there are some noteworthy exceptions. Specifically for daily data in Panel A and C, the best trading rule is not able to produce a statistically significant Δ Sharpe in the period from 1988 to 2015. Similarly, for monthly data in Panel B and D, the best trading rule are not able to produce statistically significant Δ Sharpe in the period from 1958 to 1987. Moreover, by examining the best trading rule across all value-sorted portfolios, 69% were generated by the moving average envelope rule.

Size-sorted: In-Sample

		Mean %	Std.dev. %	Sharpe	Δ Sharpe	p-value	Best Strategy
Panel A: Small Daily							
1928-2015	MT	20.20	12.44	1.37	**0.91**	(0.00)	MAE(50,0.25)
	BH	12.41	19.99	0.46			
1928-1957	MT	21.55	16.06	1.28	**0.86**	(0.00)	MAE(50,0.25)
	BH	12.00	25.92	0.43			
1958-1987	MT	24.64	7.51	2.50	**1.88**	(0.00)	MAE(50,0.25)
	BH	13.54	12.29	0.63			
1988-2015	MT	14.52	11.56	0.98	**0.51**	(0.00)	MAE(50,0.75)
	BH	11.71	18.29	0.46			
Panel B: Small Monthly							
1928-2015	MT	16.90	23.22	0.58	**0.20**	(0.00)	P-MA(2)
	BH	15.47	31.41	0.38			
1928-1957	MT	20.85	33.59	0.59	**0.20**	(0.03)	MAE(2,3.5)
	BH	18.48	44.96	0.39			
1958-1987	MT	17.75	13.34	0.89	**0.45**	(0.00)	P-MA(2)
	BH	15.43	21.80	0.44			
1988-2015	MT	14.21	14.72	0.75	**0.31**	(0.03)	P-MA(2)
	BH	12.29	20.70	0.44			
Panel C: Large Daily							
1928-2015	MT	10.25	11.46	0.61	**0.22**	(0.01)	MAE(210,2.25)
	BH	9.91	17.20	0.39			
1928-1957	MT	10.00	12.18	0.74	**0.36**	(0.01)	MOM(280)
	BH	8.36	19.55	0.38			
1958-1987	MT	11.46	9.48	0.59	**0.26**	(0.04)	MAE(210,2)
	BH	10.35	13.51	0.33			
1988-2015	MT	11.21	11.61	0.69	**0.23**	(0.05)	MAE(170,3)
	BH	11.33	17.67	0.46			
Panel D: Large Monthly							
1928-2015	MT	10.56	12.10	0.59	**0.20**	(0.01)	MAE(10,3)
	BH	10.48	17.97	0.39			
1928-1957	MT	11.14	15.51	0.65	**0.26**	(0.05)	MOM(11)
	BH	10.10	23.27	0.39			
1958-1987	MT	11.40	11.05	0.50	0.18	(0.07)	MAE(10,3)
	BH	10.53	14.70	0.32			
1988-2015	MT	11.80	11.05	0.78	**0.25**	(0.05)	MOM(11)
	BH	10.85	14.31	0.53			

Table 7: Summary statistics and performance of the best active market timing strategies from in-sample simulations. **Small** denotes a portfolio consisting of the smallest 20% of US stocks sorted on size, while **Large** denotes a portfolio consisting of the largest 20%. **Best strategy** gives the best performing strategy for the given period, and for which we present summary statistics. Means and standard deviations are annualized and reported in percentages. **Sharpe** denotes the annualized Sharpe ratios, while **Δ Sharpe** denotes the difference in annualized Sharpe ratios between the active strategy and the passive benchmark. It is calculated as $\Delta = SR_{MT} - SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. For each Sharpe ratio we test the null hypothesis $H_0 : SR_{MT} \leq SR_{BH}$. P-values are reported in brackets and values in bold text are statically significant at the 5% level.

Value-sorted: In-Sample

		Mean %	Std.dev. %	Sharpe	Δ Sharpe	p-value	Best Strategy
Panel A: Growth Daily							
1928-2015	MT	9.78	11.85	0.55	**0.19**	(0.02)	MAE(250,1.75)
	BH	9.69	17.91	0.36			
1928-1957	MT	9.80	12.34	0.72	**0.35**	(0.01)	MOM(280)
	BH	8.31	19.94	0.37			
1958-1987	MT	12.04	9.44	0.66	**0.41**	(0.01)	MAE(50,0.25)
	BH	9.46	14.95	0.24			
1988-2015	MT	11.75	13.48	0.63	0.17	(0.09)	MAC(60,240)
	BH	11.62	18.21	0.46			
Panel B: Growth Monthly							
1928-2015	MT	10.25	13.39	0.51	**0.15**	(0.05)	MAE(14,4)
	BH	10.32	19.01	0.36			
1928-1957	MT	11.18	15.14	0.67	**0.29**	(0.03)	MAE(18,3)
	BH	9.95	23.22	0.38			
1958-1987	MT	10.94	11.34	0.45	0.22	(0.09)	MAE(2,0.75)
	BH	9.86	17.30	0.23			
1988-2015	MT	12.63	12.91	0.73	**0.21**	(0.02)	MAC(7,16)
	BH	11.21	15.46	0.52			
Panel C: Value Daily							
1928-2015	MT	14.47	15.07	0.75	**0.27**	(0.00)	MAE(100,0.75)
	BH	14.29	23.03	0.48			
1928-1957	MT	14.54	19.48	0.70	**0.32**	(0.02)	MAE(100,1.25)
	BH	12.38	30.26	0.38			
1958-1987	MT	17.96	9.04	1.34	**0.55**	(0.00)	MAE(50,0.25)
	BH	16.27	13.16	0.79			
1988-2015	MT	13.42	13.23	0.77	0.24	(0.08)	MAC(60,220)
	BH	14.49	21.07	0.54			
Panel D: Value Monthly							
1928-2015	MT	14.20	19.17	0.56	**0.12**	(0.05)	MAE(4,3.5)
	BH	15.94	28.25	0.44			
1928-1957	MT	17.55	27.34	0.60	**0.23**	(0.05)	MAE(4,4)
	BH	16.74	41.69	0.38			
1958-1987	MT	15.25	13.03	0.72	0.08	(0.27)	MAC(5,6)
	BH	17.00	17.53	0.63			
1988-2015	MT	13.85	12.11	0.88	**0.28**	(0.04)	MAE(5,0.25)
	BH	13.94	17.94	0.60			

Table 8: Summary statistics and performance of the best active market timing strategies from in-sample simulations. **Growth** denotes a portfolio consisting of the lowest 20% of US stocks sorted on book-to-market, while **Value** denotes a portfolio consisting of the highest 20%. **Best strategy** gives the best performing strategy for the given period, and for which we present summary statistics. Means and standard deviations are annualized and reported in percentages. **Sharpe** denotes the annualized Sharpe ratios, while Δ **Sharpe** denotes the difference in annualized Sharpe ratios between the active strategy and the passive benchmark. It is calculated as $\Delta = SR_{MT} - SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. For each Sharpe ratio we test the null hypothesis $H_0 : SR_{MT} \leq SR_{BH}$. P-values are reported in brackets and values in bold text are statically significant at the 5% level.

5.2 Out-of-Sample Tests

5.2.1 Results of the Tests

We now take a look at the empirical results of the out-of-sample tests. First we present summary statistics of the strategy returns, as well as two different performance measures with corresponding statistical tests. We also include descriptive statistics for the passive benchmark, to compare. Our three different trading rules will be simulated separately as well as combined. In doing so, we get a general impression of the rules' performance on their own, compared to when we combine them. We may also compare them to one another, as they are all tested against the same benchmark. To indicate what one may expect with more realistic investment horizons, we also report 5- and 10-year outperoformance probabilities. Further, tests are run using both a rolling and an expanding in-sample window. We also present results for different lengths of the rolling in-sample window. Finally, to investigate if the performance is stable over time, we present plots of rolling 10-year outperformance.

Daily Data on the S&P Composite Index

We start by looking at daily data of the S&P Composite index. In Table 9, we find summary statistics, and results of statistical tests, when simulating the active strategies, with two different split points between the initial in-sample and out-of-sample periods. Panel A, gives results for when the initial in-sample period spans from January 1928 to December 1952, and the initial out-of-sample period spans from January 1953 to December 2015. Similarly Panel B gives results for when the initial in-sample period spans from January 1928 to December 1969, and out-of-sample from January 1970 to December 2015. Simulations are run for both choosing from a pool of all trading rules combined, and each trading rule by itself. I.e. the results denoted "COMBI" is found when choosing from all 590 unique trading rules, whereas the results under for instance "MAC" is found when choosing from the 284 unique MAC trading rules. We report results for both the rolling and expanding estimation scheme. Observe that as the split points for Panel A and Panel B are 1953 and 1970 respectively, the rolling in-sample window in the rolling estimation scheme will be 25 and 42 years respectively.

Looking at the results, we see that a clear pattern emerges in the mean returns and standard deviations across both panels. Mean returns are slightly lower for all strategies compared to the benchmark, while standard deviations are quite substantially lower for all strategies. This in turn leads to greater Sharpe ratios when one chooses from all the trading rules (COMBI, both rolling and expanding), and for three out of the six when we choose among each trading rule by itself (MAC rolling, and MAE rolling and expanding). This is easily seen when looking at the values of Δ Sharpe. Even though several of the trading rules seems to outperform the passive benchmark, none of them are statistically significant at the 5% level. The same is seen when looking at the Fama-French-Carhart 4-factor alphas. Several strategies yield abnormal positive returns, but none are even close to being statistically significant. In fact, the most significant alpha found is that for MOM expanding in panel A, which exhibits a quite large negative abnormal return.

Also interesting are the minimum and maximum returns. For Panel A, we see that all strategies but the MAC rule manage to avoid the largest negative return in the period. This in turn leads to a much fatter tail in the distribution, illustrated by a very high kurtosis, and a larger negative skewness. The same is seen for the MAC rule in the expanding estimation scheme in Panel B. Further we see that none of the strategies manage to capture the largest return in the period. Looking at the skewness for all other strategies, we see that it resembles that for the passive benchmark, all being slightly negative. The kurtosis on the other hand are lower for the active strategies, implying that returns are less turbulent, with fewer observations far away from the mean.

The rolling 5- and 10-year outperformance probability gives the probability that the active strategy outperforms the passive benchmark over a 5- and 10-year investment horizon respectively. We see that most active strategies outperform the passive between 40% and 60% of the time, while the best strategy (MAE rolling) outperformes the passive benchmark 78% of the time. The outperformance probability looks to increase when the investment horizon increases from 5 to 10 years.

Further investigating the 10-year outperformance, we can look at Figure 2, showing the rolling 10-year outperformance of the active strategies. The blue line gives the outperformance of using the active strategies over the past 10 years measured by Δ Sharpe. The date index on the x-axis gives the end of each 10-year period. What we see is that the outperformance of the strategies varies significantly over time. For some periods we see

S&P Composite Daily: Out-of-Sample

	BH	COMBI		MAC		MAE		MOM	
		roll	exp	roll	exp	roll	exp	roll	exp
Panel A: Period 1928-1953-2016									
Mean %	11.24	10.15	9.67	9.74	8.89	10.38	9.67	8.47	8.84
Standard Deviation %	15.55	10.26	10.30	10.83	10.53	10.26	10.30	10.45	10.71
Skewness	-0.63	-0.38	-0.44	-2.07	-2.20	-0.38	-0.44	-0.55	-0.49
Kurtosis	20.85	7.41	8.07	58.20	63.56	7.41	8.07	8.40	7.83
Minimum %	-20.45	-6.86	-6.86	-20.45	-20.45	-6.86	-6.86	-6.86	-6.86
Maximum %	11.59	5.12	5.12	5.12	5.12	5.12	5.12	5.12	5.12
Sharpe	0.44	0.56	0.52	0.50	0.43	0.59	0.52	0.39	0.42
Δ Sharpe		0.12	0.08	0.06	-0.01	0.15	0.08	-0.05	-0.02
p-value		(0.10)	(0.25)	(0.32)	(0.55)	(0.10)	(0.25)	(0.71)	(0.60)
Alpha		0.72	0.01	-0.52	-1.02	0.96	0.01	-1.49	-1.60
p-value		(0.46)	(1.00)	(0.58)	(0.27)	(0.30)	(1.00)	(0.12)	(0.07)
Rolling 5-year Win		0.56	0.47	0.51	0.41	0.59	0.47	0.42	0.46
Rolling 10-year Win		0.70	0.55	0.62	0.47	0.78	0.55	0.38	0.45
Panel B: Period 1928-1970-2016									
Mean %	11.13	10.28	9.91	9.83	8.80	10.26	9.91	8.20	8.73
Standard Deviation %	16.96	10.87	10.97	11.05	11.25	10.87	10.97	10.96	11.48
Skewness	-0.63	-0.32	-0.42	-0.36	-2.39	-0.32	-0.42	-0.37	-0.49
Kurtosis	19.37	5.97	6.94	6.94	65.01	5.97	6.94	6.02	6.71
Minimum %	-20.45	-6.86	-6.86	-6.86	-20.45	-6.86	-6.86	-6.86	-6.86
Maximum %	11.59	5.12	5.12	5.12	5.12	5.12	5.12	5.12	5.12
Sharpe	0.37	0.50	0.46	0.45	0.35	0.50	0.46	0.31	0.34
Δ Sharpe		0.13	0.09	0.08	-0.02	0.13	0.09	-0.06	-0.03
p-value		(0.15)	(0.21)	(0.24)	(0.56)	(0.14)	(0.21)	(0.71)	(0.58)
Alpha		0.98	0.45	0.07	-0.99	0.95	0.45	-1.27	-1.61
p-value		(0.42)	(0.69)	(0.95)	(0.39)	(0.41)	(0.70)	(0.29)	(0.16)
Rolling 5-year Win		0.55	0.45	0.51	0.37	0.54	0.45	0.29	0.42
Rolling 10-year Win		0.65	0.50	0.62	0.40	0.64	0.50	0.16	0.35

Table 9: Descriptive statistics and performance of the active market timing strategies and the corresponding passive benchmark. For each active strategy we simulate daily returns out-of-sample for both rolling and expanding estimation schemes. The descriptive statistics for means and standard deviation are annualized and reported in percentages. The rolling and expanding estimation schemes are denoted by **Roll** and **Exp**. Sharpe denotes the annualized Sharpe ratios and **Alpha** denotes the annualized alpha in the Fama-French-Carhart 4-factor model. The values of alpha, minimum and maximum are reported in percentages. For each Sharpe ratio we test the null hypothesis $H_0 : SR_{MT} \leq SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. For each alpha, the following null hypothesis is tested $H_0 : \alpha = 0$. P-values are reported in brackets and values in bold text are statically significant at the 5% level. **BH** (Buy and Hold) denotes the passive benchmark. **COMBI** denotes a combination of all possible strategies. **MAC** denotes the Moving Average Crossover rule. **MAE** denotes the Moving Average Envelope rule. **MOM** denotes the time-series Momentum rule. The rolling 5- and 10-year window denotes the probability of outperformance over a 5- and 10-year horizon.

quite large outperformance, whereas other periods shows underperformance. We also see that only slight differences in when you start using the active strategies can result in large differences in performance. A similar plot for monthly data, as well as plots for daily and monthly data on the DJIA index and portfolios sorted on value can be found in Appendix I.

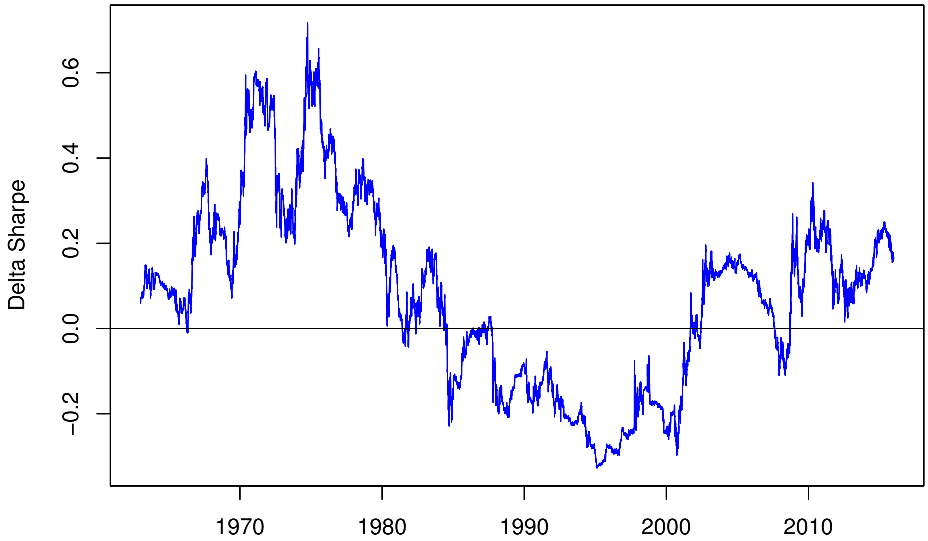


Figure 2: Rolling 10-year outperformance of daily trading on the S&P Compostie index over the period January 1953 to December 2015. The results are simulated out-of-sample using the **COMBI** rule for the rolling estimation scheme. Delta Sharpe denotes the difference between the Sharpe ratio of the active strategy and the passive benchmark and is calculated $\Delta = SR_{MT} - SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. The solid black line gives where $SR_{MT} = SR_{BH}$.

To conclude, we generally see that the active strategies produce less volatile, though also slightly lower returns than the passive counterpart. Looking at our two measures of outperformance we see that some trading rules perform quite well, especially the MAE rule, when we use the rolling estimation scheme. However, none of the trading rules has statistically significantly outperformed the passive buy-and-hold. Generally, it looks like market timing strategies have performed slightly better in the period presented in Panel A in Table 9. From Figure 2 we also saw that the performance of the strategies is highly non-uniform.

Monthly Data on the S&P Composite Index

Table 10 provides the same statistics and information as Table 9, only for monthly data. Also here we see a similar pattern when it comes to annual mean returns and standard deviations. Mean returns are slightly lower for the active strategies compared to the pas-

S&P Composite Monthly: Out-of-Sample

	BH	COMBI		MAC		MAE		MOM	
		roll	exp	roll	exp	roll	exp	roll	exp
Panel A: Period 1928-1953-2016									
Mean %	11.12	9.57	10.25	9.71	10.56	10.01	10.22	10.05	9.45
Standard Deviation %	14.55	10.83	10.94	10.95	10.91	10.72	10.98	10.96	10.88
Skewness	-0.40	-0.48	-0.44	-0.59	-0.50	-0.43	-0.45	-0.47	-0.54
Kurtosis	4.73	7.93	7.83	7.80	8.10	8.36	7.76	8.11	8.24
Minimum %	-21.54	-21.54	-21.54	-21.54	-21.54	-21.54	-21.54	-21.54	-21.54
Maximum %	16.78	12.17	13.46	11.92	13.46	13.46	13.46	13.21	13.21
Sharpe	0.45	0.47	0.52	0.48	0.55	0.51	0.52	0.51	0.46
Δ Sharpe		0.01	0.07	0.02	0.10	0.06	0.07	0.05	0.00
p-value		(0.48)	(0.23)	(0.43)	(0.13)	(0.28)	(0.24)	(0.28)	(0.50)
Alpha		-0.54	-0.35	-0.60	0.33	-0.06	-0.45	-0.15	-0.89
p-value		(0.56)	(0.70)	(0.51)	(0.72)	(0.95)	(0.62)	(0.88)	(0.35)
Rolling 5-year Win		0.46	0.46	0.51	0.53	0.49	0.45	0.52	0.38
Rolling 10-year Win		0.63	0.58	0.59	0.64	0.60	0.62	0.62	0.50
Panel B: Period 1928-1970-2016									
Mean %	10.92	10.02	10.94	10.40	10.80	9.55	10.73	9.65	9.68
Standard Deviation %	15.33	11.18	11.36	11.15	11.29	11.25	11.44	11.45	11.23
Skewness	-0.43	-0.65	-0.57	-0.64	-0.65	-0.72	-0.57	-0.63	-0.68
Kurtosis	1.80	6.15	5.51	6.17	5.87	5.82	5.33	5.54	6.06
Minimum %	-21.54	-21.54	-21.54	-21.54	-21.54	-21.54	-21.54	-21.54	-21.54
Maximum %	16.78	13.46	13.46	13.46	13.46	13.46	13.46	13.21	13.21
Sharpe	0.39	0.45	0.53	0.49	0.52	0.41	0.51	0.41	0.42
Δ Sharpe		0.06	0.14	0.10	0.13	0.02	0.12	0.02	0.03
p-value		(0.29)	(0.12)	(0.20)	(0.12)	(0.43)	(0.15)	(0.42)	(0.37)
Alpha		0.25	0.65	0.42	0.81	-0.33	0.31	-0.41	-0.43
p-value		(0.83)	(0.55)	(0.70)	(0.47)	(0.77)	(0.78)	(0.72)	(0.72)
Rolling 5-year Win		0.43	0.49	0.49	0.52	0.40	0.46	0.42	0.43
Rolling 10-year Win		0.42	0.62	0.48	0.60	0.42	0.62	0.41	0.59

Table 10: Descriptive statistics and performance of the active market timing strategies and the corresponding passive benchmark. For each active strategy we simulate monthly returns out-of-sample for both rolling and expanding estimation schemes. The descriptive statistics for means and standard deviation are annualized and reported in percentages. The rolling and expanding estimation schemes are denoted by **Roll** and **Exp**. Sharpe denotes the annualized Sharpe ratios and **Alpha** denotes the annualized alpha in the Fama-French-Carhart 4-factor model. The values of alpha, minimum and maximum are reported in percentages. For each Sharpe ratio we test the null hypothesis $H_0 : SR_{MT} \leq SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. For each alpha, the following null hypothesis is tested $H_0 : \alpha = 0$. P-values are reported in brackets and values in bold text are statically significant at the 5% level. **BH** (Buy and Hold) denotes the passive benchmark. **COMBI** denotes a combination of all possible strategies. **MAC** denotes the Moving Average Crossover rule. **MAE** denotes the Moving Average Envelope rule. **MOM** denotes the time-series Momentum rule. The rolling 5- and 10-year window denotes the probability of outperformance over a 5- and 10-year horizon.

sive benchmark, while the standard deviations are considerably lower. Again this leads to positive realizations of Δ Sharpe, which indicates that the active strategies outperforms the passive benchmark. Once more the outperformances are very small, and none are statistically significant at the 5% level. Looking at the Fama-French-Carhart 4-factor alphas we also see that none of the strategies in either period produce statistically significant abnormal returns. The skewness of all active strategies resemble, and seems slightly greater negative than that of the passive buy-and-hold. We also see slightly greater negative skewness in Panel B, compared to Panel A. Comparing kurtosis we see the opposite for monthly data as we did for daily. The market timing strategies exhibits greater kurtosis than the passive benchmark for both periods, indicating that the returns are more spread out from the mean. Further, looking at the minimum and maximum returns, we see that none of the active strategies in either of the periods manage to avoid the largest negative return. Neither do they capture the maximum. Looking at the 5- and 10-year outperformance probability we once again see that most lie in the range 40% to 60%, with the highest across the two panels being 64%. Again the outperformance probability seems to be greater when we have a 10-year investment horizon.

Daily Data on the DJIA Index

Table 11 is similar to Table 9 and 10, only for daily data of the DJIA index. Once again both annual mean returns and standard deviations are lower for the active strategies than for the passive benchmark. This time we also see a quite severe drop in the mean returns as well as in the standard deviations. Futher, this leads to poorer performance than what we saw for data on the S&P Composite index, and none of the reported values of Δ Sharpe exceeds 0.01. Looking at the Fama-French-Carhart 4-factor alphas we see that all but one of the strategies yield negative abnormal returns, where three of these are statistically significant at the 5% level. For both panels we see that all but one strategy managed to avoid the largest negative return, while once again none manage to capture the maximum. The distributions, described by the skewness and kurtosis, looks similar to what we saw for daily data of the S&P Composite index. Looking at the 5- and 10-year outperformance probabilities we see that the active strategies seldom perform well. Several strategies outperform the passive benchmark less than 25% of the time, and the best strategy outperforms the buy-and-hold only 54% of the time.

What we generally see is that the active strategies performs worse when trading on the DJIA index than when trading on the S&P Composite index. Though we also here see less volatility, the mean returns decreases more, resulting in poorer risk-adjusted performance and underperformance for almost all trading rules.

DJIA Daily: Out-of-Sample

	BH	COMBI		MAC		MAE		MOM	
		roll	exp	roll	exp	roll	exp	roll	exp
Panel A: Period 1928-1953-2016									
Mean %	11.12	8.30	8.21	9.02	8.14	8.86	8.21	8.11	6.69
Standard Deviation %	15.40	10.50	10.35	10.58	10.32	10.10	10.35	10.68	10.81
Skewness	-0.75	-0.41	-0.37	-0.41	-0.37	-0.39	-0.37	-0.35	-2.68
Kurtosis	26.50	7.13	6.65	7.11	6.69	7.05	6.65	7.24	82.99
Minimum %	-22.60	-7.18	-7.18	-7.18	-7.18	-7.18	-7.18	-7.18	-22.60
Maximum %	11.09	4.51	4.51	4.51	4.51	4.51	4.51	4.71	4.71
Sharpe	0.44	0.38	0.37	0.44	0.37	0.45	0.37	0.35	0.22
Δ Sharpe		-0.06	-0.07	0.00	-0.07	0.01	-0.07	-0.09	-0.22
p-value		(0.74)	(0.75)	(0.48)	(0.75)	(0.46)	(0.74)	(0.81)	(0.99)
Alpha		-1.15	-1.05	-0.62	-1.05	0.03	-1.05	-1.06	**-3.11**
p-value		(0.22)	(0.29)	(0.54)	(0.29)	(0.97)	(0.27)	(0.29)	(0.00)
Rolling 5-year Win		0.34	0.41	0.51	0.45	0.44	0.41	0.35	0.30
Rolling 10-year Win		0.35	0.37	0.54	0.40	0.51	0.37	0.36	0.18
Panel B: Period 1928-1970-2016									
Mean %	11.41	8.90	7.92	8.38	7.88	8.79	7.92	6.75	6.07
Standard Deviation %	16.85	11.04	11.19	11.55	11.13	10.91	11.19	11.36	11.76
Skewness	-0.75	-0.30	-0.35	-0.40	-0.31	-0.30	-0.35	-0.34	-2.76
Kurtosis	24.63	5.62	5.75	6.38	5.52	5.88	5.75	6.09	79.82
Minimum %	-22.60	-7.18	-7.18	-7.18	-7.18	-7.18	-7.18	-7.18	-22.60
Maximum %	11.09	4.27	4.27	4.27	4.27	4.27	4.27	4.71	4.71
Sharpe	0.39	0.37	0.28	0.31	0.27	0.36	0.28	0.17	0.11
Δ Sharpe		-0.02	-0.11	-0.08	-0.12	-0.03	-0.11	-0.22	-0.28
p-value		(0.57)	(0.85)	(0.77)	(0.84)	(0.60)	(0.84)	(0.98)	(1.00)
Alpha		-0.07	-1.27	-1.24	-1.25	-0.07	-1.27	**-2.60**	**-3.76**
p-value		(0.96)	(0.31)	(0.31)	(0.32)	(0.96)	(0.33)	(0.04)	(0.00)
Rolling 5-year Win		0.49	0.33	0.45	0.36	0.48	0.33	0.28	0.18
Rolling 10-year Win		0.51	0.20	0.50	0.24	0.49	0.20	0.15	0.07

Table 11: Descriptive statistics and performance of the active market timing strategies and the corresponding passive benchmark. For each active strategy we simulate daily returns out-of-sample for both rolling and expanding estimation schemes. The descriptive statistics for means and standard deviation are annualized and reported in percentages. The rolling and expanding estimation schemes are denoted by **Roll** and **Exp**. Sharpe denotes the annualized Sharpe ratios and **Alpha** denotes the annualized alpha in the Fama-French-Carhart 4-factor model. The values of alpha, minimum and maximum are reported in percentages. For each Sharpe ratio we test the null hypothesis $H_0 : SR_{MT} \leq SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. For each alpha, the following null hypothesis is tested $H_0 : \alpha = 0$. P-values are reported in brackets and values in bold text are statically significant at the 5% level. **BH** (Buy and Hold) denotes the passive benchmark. **COMBI** denotes a combination of all possible strategies. **MAC** denotes the Moving Average Crossover rule. **MAE** denotes the Moving Average Envelope rule. **MOM** denotes the time-series Momentum rule. The rolling 5- and 10-year window denotes the probability of outperformance over a 5- and 10-year horizon.

Monthly Data on the DJIA Index

Table 12 gives the same descriptive statistics and tests as Table 9, 10 and 11, this time for monthly data in the DJIA index. Like we saw for the daily data, the drop in annual mean returns for the active strategies, compared to the passive are more severe for this index, than for the S&P Composite. Here we also see less drop in annual standard deviations.

DJIA Monthly: Out-of-Sample

	BH	COMBI		MAC		MAE		MOM	
		roll	exp	roll	exp	roll	exp	roll	exp
Panel A: Period 1928-1953-2016									
Mean %	10.98	8.51	8.32	8.97	8.55	8.23	8.62	9.06	8.13
Standard Deviation %	14.37	11.11	10.94	11.19	10.84	11.03	11.18	11.09	11.02
Skewness	-0.43	-0.69	-0.71	-0.55	-0.79	-0.65	-0.57	-0.62	-0.66
Kurtosis	5.04	8.68	8.99	8.98	9.48	8.78	8.86	8.80	8.86
Minimum %	-22.90	-22.90	-22.90	-22.90	-22.90	-22.90	-22.90	-22.90	-22.90
Maximum %	14.75	14.12	13.87	14.75	13.87	14.12	14.75	14.12	14.12
Sharpe	0.46	0.37	0.36	0.41	0.39	0.35	0.38	0.42	0.34
Δ Sharpe		-0.09	-0.10	-0.05	-0.07	-0.11	-0.08	-0.04	-0.12
p-value		(0.85)	(0.86)	(0.71)	(0.80)	(0.90)	(0.80)	(0.65)	(0.91)
Alpha		-0.66	-0.77	-0.70	-0.31	-0.77	-1.48	-0.02	-0.56
p-value		(0.53)	(0.44)	(0.47)	(0.77)	(0.46)	(0.14)	(0.99)	(0.58)
Rolling 5-year Win		0.29	0.33	0.32	0.43	0.29	0.26	0.34	0.16
Rolling 10-year Win		0.23	0.33	0.37	0.48	0.28	0.23	0.44	0.06
Panel B: Period 1928-1970-2016									
Mean %	11.17	8.53	8.55	9.04	8.50	8.84	9.08	8.85	8.45
Standard Deviation %	15.11	11.66	11.35	11.84	11.31	11.69	11.65	11.37	11.38
Skewness	-0.46	-0.72	-0.82	-0.64	-0.91	-0.65	-0.67	-0.81	-0.78
Kurtosis	5.20	9.49	9.84	9.14	10.20	9.45	9.59	9.90	9.77
Minimum %	-22.90	-22.90	-22.90	-22.90	-22.90	-22.90	-22.90	-22.90	-22.90
Maximum %	14.75	14.50	13.87	14.50	13.87	14.75	14.75	14.12	14.12
Sharpe	0.42	0.32	0.33	0.35	0.32	0.34	0.36	0.35	0.32
Δ Sharpe		-0.10	-0.09	-0.06	-0.09	-0.08	-0.06	-0.07	-0.10
p-value		(0.83)	(0.80)	(0.73)	(0.80)	(0.76)	(0.71)	(0.70)	(0.81)
Alpha		-1.21	-0.27	-0.76	-0.13	-0.68	-0.81	-0.30	0.25
p-value		(0.33)	(0.83)	(0.54)	(0.92)	(0.61)	(0.52)	(0.81)	(0.84)
Rolling 5-year Win		0.25	0.34	0.27	0.39	0.28	0.29	0.23	0.14
Rolling 10-year Win		0.22	0.35	0.23	0.37	0.20	0.20	0.17	0.06

Table 12: Descriptive statistics and performance of the active market timing strategies and the corresponding passive benchmark. For each active strategy we simulate monthly returns out-of-sample for both rolling and expanding estimation schemes. The descriptive statistics for means and standard deviation are annualized and reported in percentages. The rolling and expanding estimation schemes are denoted by **Roll** and **Exp**. Sharpe denotes the annualized Sharpe ratios and **Alpha** denotes the annualized alpha in the Fama-French-Carhart 4-factor model. The values of alpha, minimum and maximum are reported in percentages. For each Sharpe ratio we test the null hypothesis $H_0 : SR_{MT} \leq SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. For each alpha, the following null hypothesis is tested $H_0 : \alpha = 0$. P-values are reported in brackets and values in bold text are statically significant at the 5% level. **BH** (Buy and Hold) denotes the passive benchmark. **COMBI** denotes a combination of all possible strategies. **MAC** denotes the Moving Average Crossover rule. **MAE** denotes the Moving Average Envelope rule. **MOM** denotes the time-series Momentum rule. The rolling 5- and 10-year window denotes the probability of outperformance over a 5- and 10-year horizon.

This leads to poorer performance, and all reported values of Δ Sharpe are negative. Similar observations are made for the Fama-French-Carhart 4-factor alpha, with all but one of the active strategies yielding negative abnormal returns. Like we saw for monthly data on the S&P Composite index, none of the strategies manage to avoid the largest negative return. On the other hand, we actually see that two of the strategies in both periods manage to capture the maximum return. Like for monthly data on the S&P Composite, all active strategies has a higher kurtosis than the buy-and-hold. We also see generally greater kurtosis than we have seen for the other datasets, and also slightly greater negative skewness. The 5- and 10-year outperformance probabilities substantiates the impression of poor performance. 14 out of 16, and 13 out of 16 reported 5- and 10-year outperformance probabilities respectively, are less than 35%.

Portfolios Sorted on Size

Table 13 gives summary statistics, and results of statistical tests, when simulating our three different trading rules on two different stock portfolios sorted on size. Small and Large denotes the low and high quintile portfolios sorted on size respectively. The initial in-sample period spans from January 1928 to December 1952, and the initial out-of-sample period spans from January 1953 to December 2015. Simulations are run for both choosing from all possible trading rules, denoted by COMBI, and each trading rule type by itself. In simulating the results, we use the rolling estimation scheme only. Also included are the descriptive statistics for the passive benchmark, to compare. Panel A and B gives statistics and results for daily and monthly data respectively. Observe that as the split point is 1953, the rolling in-sample window is 25 years.

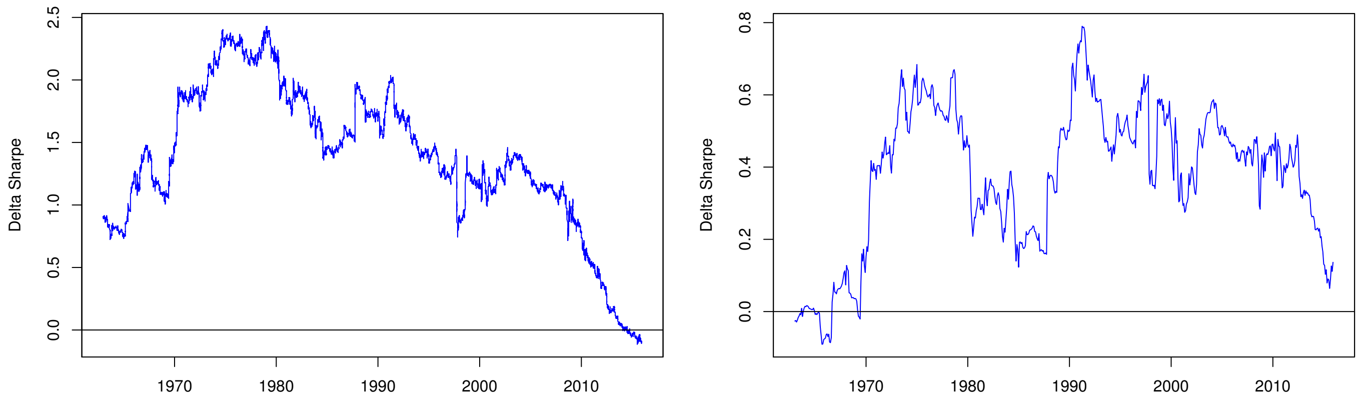
We start by looking at the Small-portfolio, which obviously has performed very well. As we for previous datasets have seen that annual mean returns has been lower for the active strategies, compared to the passive benchmark, we here see the complete opposite. Together with all mean returns being considerably greater, we also see that annual standard deviations are considerably lower for the active strategies. This in turn gives high values of outperformance measured by Δ Sharpe, and all active strategies significantly outperform the passive on a 1% level. This is the case for both daily and monthly data. We also see very high levels of positive abnormal return, measured by the Fama-French-Carhart 4-factor alpha. Also these being statistically significant at the 1% level. Even though we see significant outperformance for both daily and monthly data, we clearly see that daily performs best, having both higher annual mean return, and lower annual standard deviation. Looking at the shape of the distributions, we see much the same as for daily data on the two stock indices, while the monthly data actually exhibits positive skewness for three of the

Size Sorted: Out-of-Sample

	Small					Large				
	BH	COMBI	MAC	MAE	MOM	BH	COMBI	MAC	MAE	MOM
Panel A: Daily										
Mean %	12.55	18.98	18.82	19.03	14.79	11.02	10.30	9.52	10.18	9.83
Standard Deviation %	15.08	9.52	9.52	9.51	9.91	15.34	10.41	10.66	10.33	10.95
Skewness	-0.67	-0.62	-0.60	-0.63	-0.83	-0.54	-0.42	-0.38	-0.44	-0.41
Kurtosis	11.69	10.46	10.34	10.48	11.76	18.82	7.01	7.28	7.20	7.27
Minimum %	-10.81	-6.27	-6.27	-6.27	-6.27	-19.27	-6.91	-6.69	-6.91	-6.91
Maximum %	8.42	5.28	5.28	5.28	6.80	11.82	4.89	4.89	4.89	4.89
Sharpe	0.54	1.54	1.52	1.55	1.05	0.44	0.57	0.48	0.56	0.50
Δ Sharpe		**0.99**	**0.98**	**1.00**	**0.51**		0.14	0.05	0.13	0.07
p-value		(0.00)	(0.00)	(0.00)	(0.00)		(0.07)	(0.31)	(0.07)	(0.24)
Alpha		**10.76**	**10.60**	**10.83**	**5.13**		0.90	-0.35	0.92	-0.33
p-value		(0.00)	(0.00)	(0.00)	(0.00)		(0.33)	(0.70)	(0.31)	(0.71)
Rolling 5-year Win		0.94	0.92	0.94	0.87		0.56	0.52	0.56	0.50
Rolling 10-year Win		0.97	0.96	0.96	0.93		0.70	0.53	0.68	0.71
Panel B: Monthly										
Mean %	13.75	16.33	15.57	15.15	14.76	10.91	10.10	9.88	10.35	10.20
Standard Deviation %	20.71	15.05	14.67	15.58	14.50	14.35	11.17	11.03	11.07	10.98
Skewness	-0.20	0.33	0.27	-0.12	0.33	-0.36	-0.46	-0.46	-0.43	-0.53
Kurtosis	2.52	4.85	5.10	6.80	5.00	1.66	3.92	4.19	3.84	3.97
Minimum %	-29.63	-20.82	-20.82	-29.63	-20.82	-20.31	-20.31	-20.31	-20.31	-20.31
Maximum %	27.70	27.70	27.70	27.70	27.70	18.12	13.16	13.16	13.16	12.07
Sharpe	0.45	0.79	0.76	0.69	0.72	0.45	0.51	0.50	0.54	0.53
Δ Sharpe		**0.34**	**0.31**	**0.24**	**0.26**		0.06	0.04	0.08	0.07
p-value		(0.00)	(0.00)	(0.00)	(0.01)		(0.25)	(0.32)	(0.18)	(0.22)
Alpha		**5.69**	**4.90**	**3.48**	**5.04**		0.17	-0.11	0.11	0.32
p-value		(0.00)	(0.00)	(0.01)	(0.00)		(0.85)	(0.90)	(0.90)	(0.72)
Rolling 5-year Win		0.83	0.83	0.74	0.80		0.45	0.50	0.50	0.48
Rolling 10-year Win		0.95	0.95	0.88	0.88		0.55	0.65	0.58	0.59

Table 13: Descriptive statistics and performance of the active market timing strategies and the corresponding passive benchmark. **Small** denotes a portfolio consisting of the smallest 20% of US stocks sorted on size, while **Large** denotes a portfolio consisting of the largest 20%. For each active strategy we simulate daily and monthly returns out-of-sample for the rolling estimation scheme from January 1953 to Desember 2015. The descriptive statistics for means and standard deviation are annualized and reported in percentages. Sharpe denotes the annualized Sharpe ratios and **Δ Sharpe** denotes the difference in Sharpe ratios between the active strategy end the passive benchmark. **Alpha** denotes the annualized alpha in the Fama-French-Carhart 4-factor model. The values of alpha, minimum and maximum are reported in percentages. For each Sharpe ratio we test the null hypothesis $H_0 : SR_{MT} \leq SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. For each alpha, the following null hypothesis is tested $H_0 : \alpha = 0$. P-values are reported in brackets and values in bold text are statically significant at the 5% level. **BH** (Buy and Hold) denotes the passive benchmark. **COMBI** denotes a combination of all possible strategies. **MAC** denotes the Moving Average Crossover rule. **MAE** denotes the Moving Average Envelope rule. **MOM** denotes the time-series Momentum rule. The rolling 5- and 10-year window denotes the probability of outperformance over a 5- and 10-year horizon.

four reported return series. For daily data, we see that the active strategies manage to avoid the largest negative return, but also fail to capture the largest positive return. For monthly data, we see that three of the four trading rules avoid the largest negative return, while all actually manage to capture the largest positive return. This is probably also the reason why we see positive skewness for these strategies. Looking at the 5- and 10-year outperformance probabilities, it once more becomes clear that the active strategies has outperformed the passive. Especially for daily data, where we see 5- and 10-year outperformance probabilities as high as 94% and 97% respectively.



(a) Daily

(b) Monthly

Figure 3: Rolling 10-year outperformance in trading small stocks over the period January 1953 to December 2015. The results are simulated out-of-sample using the **COMBI** rule for the rolling estimation scheme. Delta Sharpe denotes the difference between the Sharpe ratio of the active strategy and the passive benchmark and is calculated $\Delta = SR_{MT} - SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. The solid black line gives where $SR_{MT} = SR_{BH}$.

Figure 3 plots the 10-year rolling outperformance when trading small stocks using both daily and monthly data. We see that the active strategy has outperformed its passive counterpart over the period 1960 to 2000, especially good performance is seen when using daily data. However, the performance deteriorates dramatically over the last 10 years or so. In the end of the sample the outperformance has disappeared completely both for daily and monthly data. For daily we even see underperformance in the very end. Thus in this period we actually see monthly trading performing better than daily.

Looking at the Large-portfolio, we see that the results very much resembles what we saw for the S&P Composite in Panel A of both Table 9 and 10. This is though to be expected as many of the same stocks will be included in a portfolio consisting of the 20% largest companies in the US, as in the S&P Composite index. We therefore do not comment the results further.

To conclude, we see that the active strategies historically have performed very well when trading on small stocks. We find statistically significant outperformance when using both daily and monthly data, though daily seems to have performed significantly better. Interestingly, the performance has deteriorated severely over the past decade, and the statistically significant outperformance has disappeared. Towards the end of the sample we also see that daily trading no longer perform better than monthly. We even see underperformance for daily data over the past 10 years or so. For large stocks we see much the same as we did for the S&P Composite index, with no statistically significant outperformance.

Portfolios Sorted on Value

Similar to Table 13, Table 14 gives summary statistics, and results of statistical tests, when simulating our three different trading rules on two different portfolios, only this time sorted on book-to-market ratio. Growth and Value denotes the low and high quintile portfolios sorted on value respectively. Except different data, Table 14 is equal to Table 13.

Looking at the results of the Growth-portfolio for daily data, we once again see the pattern that annual mean returns of the active strategies are slightly lower than that for buy-and-hold. We also see that annual standard deviation are considerably lower, and that this leads to an outperformance measured by Δ Sharpe. These levels of Δ Sharpe are quite high, and though none of the four are statistically significant at the 5% level, three of them are at the 10% level. We also see quite large positive abnormal returns, one being statistically significant at the 5% level. Both the shape of the distribution and minimum and maximum returns have similar characteristics as that we have seen for the other daily datasets. Further, we see that the 10-year outperformance probabilities are quite high, two of them being 82%. The best strategy also outperforms the passive buy-and-hold 73% of the time when we have a 5-year investment horizon.

For the Value-portfolio, we see that the annual mean returns are relatively less reduced when we apply the market timing strategies, while the standard deviations still decline quite a lot. Consequently values for Δ Sharpe are greater, two of them statistically significant at the 5% level, while another at the 10% level. We also see high positive abnormal returns for three of the strategies, where one is significant at the 5% level, and the other two at the 10% level. The kurtosis of these strategies are slightly higher than what we have seen for

Value Sorted: Out-of-Sample

	Growth					Value				
	BH	COMBI	MAC	MAE	MOM	BH	COMBI	MAC	MAE	MOM
Panel A: Daily										
Mean %	10.75	10.28	10.27	10.56	9.58	14.87	13.24	13.38	13.21	11.88
Standard Deviation %	16.26	10.98	11.17	10.94	11.36	17.17	11.35	11.38	11.17	11.50
Skewness	-0.30	-0.43	-0.38	-0.45	-0.38	-0.48	-0.81	-0.64	-1.10	-0.55
Kurtosis	12.75	8.24	8.40	8.39	6.63	18.07	12.99	11.08	17.18	7.45
Minimum %	-17.27	-8.14	-8.14	-8.14	-6.91	-16.90	-10.24	-10.24	-13.32	-7.61
Maximum %	12.67	5.42	5.42	5.42	5.42	11.96	7.26	7.26	6.30	6.30
Sharpe	0.39	0.54	0.53	0.57	0.46	0.61	0.78	0.79	0.79	0.65
Δ Sharpe		0.15	0.14	0.17	0.07		**0.17**	**0.18**	0.18	0.04
p-value		(0.08)	(0.10)	(0.07)	(0.24)		(0.05)	(0.04)	(0.06)	(0.36)
Alpha		1.65	1.05	**2.00**	-0.17		1.90	1.81	**2.26**	-0.40
p-value		(0.11)	(0.29)	(0.05)	(0.87)		(0.08)	(0.09)	(0.04)	(0.70)
Rolling 5-year Win		0.66	0.59	0.73	0.49		0.63	0.67	0.62	0.47
Rolling 10-year Win		0.82	0.67	0.82	0.60		0.72	0.73	0.71	0.53
Panel B: Monthly										
Mean %	10.77	9.78	9.94	9.93	7.43	14.99	12.16	11.91	12.58	12.93
Standard Deviation %	16.20	12.33	12.92	11.53	12.22	17.55	12.99	12.52	13.35	13.50
Skewness	-0.33	-0.45	-0.43	-0.05	-0.60	-0.22	-0.23	-0.62	-0.20	-0.22
Kurtosis	1.66	4.19	3.73	1.76	4.46	2.22	4.39	3.86	4.04	4.27
Minimum %	-23.76	-23.76	-23.76	-14.33	-23.76	-20.99	-20.99	-20.99	-20.99	-20.99
Maximum %	22.03	13.14	14.82	13.14	13.14	25.17	22.76	14.10	22.76	22.51
Sharpe	0.39	0.44	0.43	0.48	0.25	0.60	0.60	0.60	0.61	0.63
Δ Sharpe		0.04	0.03	0.09	-0.14		-0.01	-0.00	0.01	0.03
p-value		(0.37)	(0.36)	(0.22)	(0.94)		(0.58)	(0.49)	(0.51)	(0.39)
Alpha		0.28	-0.37	0.95	**-2.24**		-0.33	-0.08	0.22	0.10
p-value		(0.80)	(0.71)	(0.37)	(0.03)		(0.78)	(0.94)	(0.85)	(0.93)
Rolling 5-year Win		0.51	0.41	0.56	0.24		0.46	0.38	0.50	0.38
Rolling 10-year Win		0.50	0.48	0.71	0.24		0.46	0.44	0.47	0.48

Table 14: Descriptive statistics and performance of the active market timing strategies and the corresponding passive benchmark. **Growth** denotes a portfolio consisting of the lowest 20% of US stocks sorted on book-to-market, while **Value** denotes a portfolio consisting of the highest 20%. For each active strategy we simulate daily returns out-of-sample for the rolling estimation schemes from January 1953 to Desember 2015. The descriptive statistics for means and standard deviation are annualized and reported in percentages. Sharpe denotes the annualized Sharpe ratios and **Δ Sharpe** denotes the difference in Sharpe ratios between the active strategy end the passive benchmark. **Alpha** denotes the annualized alpha in the Fama-French-Carhart 4-factor model. The values of alpha, minimum and maximum are reported in percentages. For each Sharpe ratio we test the null hypothesis $H_0 : SR_{MT} \leq SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. For each alpha, the following null hypothesis is tested $H_0 : \alpha = 0$. P-values are reported in brackets and values in bold text are statically significant at the 5% level. **BH** (Buy and Hold) denotes the passive benchmark. **COMBI** denotes a combination of all possible strategies. **MAC** denotes the Moving Average Crossover rule. **MAE** denotes the Moving Average Envelope rule. **MOM** denotes the time-series Momentum rule. The rolling 5- and 10-year window denotes the probability of outperformance over a 5- and 10-year horizon.

other datasets, and the skewness slightly more negative. The strategies avoid the largest negative returns, but also miss the greatest positive returns. The three best strategies have outperformance probabilities in the range of 62% to 67% and 71% and 73% for 5- and 10-year horizons respectively.

For monthly data we see that both portfolios perform much worse. Both with relatively larger decreases in annual mean returns, and less decrease in standard deviation when applying the strategies, yielding lower values of Δ Sharpe. We see little or no outperformance, and far from any statistically significant outperformance. No statistically significant positive abnormal returns are observed, and several strategies yield negative abnormal returns. One of which also statistically significant at the 5% level. Outperformance probabilities are also relatively low, ranging in the area between 40% and 50% for most strategies, both with 5- and 10-year horizons.

To conclude, we see that the active strategies has performed quite well for value stocks when trading daily, with several performance measures being statistically significant. For growth stocks the performance is worse, though we do find a statistically significant abnormal return when using the MAE rule on daily data. We also see a clear tendency that the strategies have performed superior when using daily data compared to monthly.

Summarizing our findings so far, a clear pattern emerges when looking at mean returns and standard deviations of the active strategies. Standard deviation decreases substantially across all datasets, but also the mean returns decrease for almost all datasets. The only exception is daily data on small stocks. This leads to varying risk-adjusted performance, where some outperform, whereas other underperform the passive benchmark. However, almost none of the observed outperformances are statistically significant. Statistically significant outperformance is seen for small stocks, though the outperformance deteriorates and disappears over the last 10 years of the sample. Further looking at small stocks, we see that daily trading performed much better than monthly in earlier years. This is no longer the case, and we now actually see the opposite. This is also the case for the other stock portfolios. In earlier years we see superior performance when trading daily, whereas more recently, we see that monthly trading has performed just as good, or even better. We further see that the performance is highly non-uniform, varying significantly over time.

5.2.2 Drawdown Analysis

To further analyse the risk, we can look at the drawdowns in the strategy returns. It is interesting to see how strategies may reduce risk by reducing the largest drawdowns, and it

Drawdown Analysis

	BH	COMBI		MAC		MAE		MOM	
		roll	exp	roll	exp	roll	exp	roll	exp
Panel A: S&P Composite Daily									
Average %	2.23	2.10	2.10	2.12	2.13	2.08	2.06	2.20	2.26
Median %	0.73	0.76	0.84	0.74	0.77	0.79	0.83	0.74	0.76
Average max %	34.68	21.98	20.16	23.55	23.20	20.84	18.55	25.38	23.10
Max %	55.23	53.87	49.03	56.51	53.37	51.01	32.97	44.12	44.67
Panel B: S&P Composite Monthly									
Average %	6.41	5.26	4.85	5.68	4.93	5.14	4.71	4.97	5.11
Median %	3.12	3.00	3.16	3.61	2.97	3.09	3.01	3.12	3.09
Average max %	31.27	20.45	17.31	19.86	19.10	19.17	17.63	17.75	18.42
Max %	50.96	43.95	40.29	37.68	37.12	41.96	41.35	43.15	40.29
Panel C: DJIA Daily									
Average %	2.14	2.04	1.98	1.97	2.03	2.08	1.97	2.16	2.10
Median %	0.75	0.77	0.79	0.77	0.82	0.84	0.78	0.86	0.83
Average max %	32.17	22.88	19.92	20.96	20.35	20.28	19.45	21.41	23.54
Max %	51.64	44.26	45.17	43.01	44.59	41.47	45.17	38.55	54.60
Panel D: DJIA Monthly									
Average %	6.14	5.46	5.66	4.87	5.45	5.62	5.43	5.68	5.81
Median %	3.17	3.50	3.26	2.69	3.11	3.88	3.24	2.84	3.18
Average max %	28.20	18.84	20.49	19.58	21.73	18.71	19.80	21.27	20.19
Max %	46.96	43.05	39.35	41.52	42.75	27.89	31.94	39.35	39.35

Table 15: Summary statistics for the drawdown in the out-of-sample simulation from Desember 1937 to Desember 2015. All statistics are reported in percentages. Average max drawdown denotes the average of the 10 largest drawdowns.

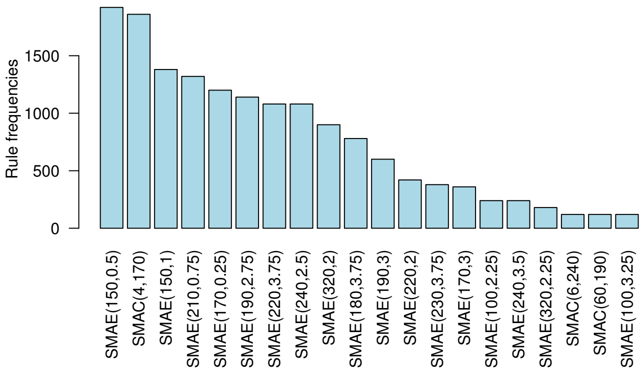
is a much used risk measure by practitioners. We will therefore look at the average, median and maximum drawdown, as well as the average of the ten largest drawdowns.

Table 15 shows descriptive statistics of drawdowns of the passive benchmark and the active strategies. We see that the average drawdowns of almost all strategies, for all datasets are less than that for the passive, with only two exceptions. Medians tends to be slightly greater for the active strategies, but not severely. More importantly, we see that the average of the ten greatest drawdowns are significantly lower for the active strategies. This means that the active strategies often manage to avoid the largest drops in the market. Also looking at the maximum drawdowns, we see that most strategies manage to avoid the largest drop. Interestingly we see that two of the reported strategies actually have a greater maximum drawdown than its passive benchmark.

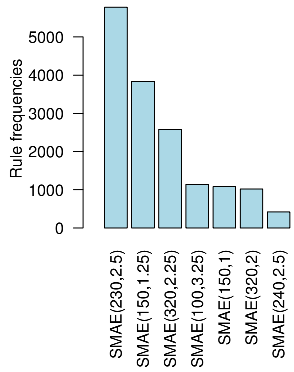
What we generally see is that the strategies reduces risk. Especially this becomes evident looking at the average of the ten largest drawdowns, with all strategies reducing this measure significantly. Once again it looks like the MAE rule produces the lowest risk, especially for daily data.

5.2.3 Analysis of Trading Rules

We have now seen how the strategies perform, but which trading rules are used to get these results? Here we take a closer look at which rules that are being used. Both to see which rules that are used most frequently, but also to compare how the two test procedures choose trading rules. To illustrate this, we provide barplots showing the most used trading rules, and their frequency of use. Further, to see the relationship between the different trading rules, we also produce a cluster dendrogram. We do so using both a rolling and an expanding in-sample window.



(a) Barplot: Rolling



(b) Barplot: Expanding

(c) Dendrogram: Rolling

(d) Dendrogram: Expanding

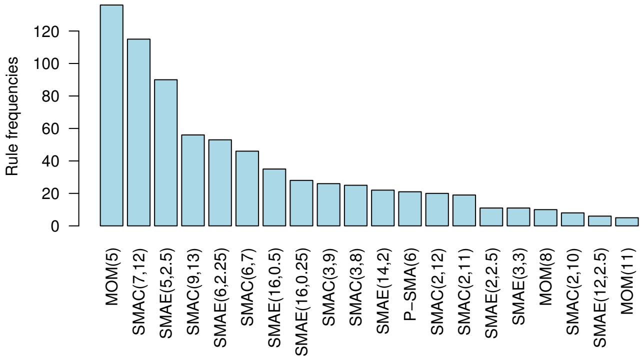
Figure 4: Barplots and cluster dendrograms of the most frequent strategies in the out-of-sample simulation on daily data on the S&P Composite index from January 1953 to Desember 2015. The barplots in (a) and (b) illustrates the most frequent strategies for both the rolling and expanding estimation schemes. The cluster dendrograms in (c) and (d) is a graphic representation of the correlation matrix and illustrate the relationship between the most frequent strategies.

Figure 4 shows barplots and dendrograms of which strategies that has been used when trading daily on the S&P Composite index. We include graphics for both the rolling and expanding estimation scheme. For rolling, we only present the 20 most frequently used trading rules. The initial in-sample period spans from January 1928 to December 1952, and the initial out-of-sample period spans from January 1953 to December 2015. The barplots shows the frequency of use, while the dendrogram shows the relationship between the different trading rules. The distance given on the y-axis of the dendrogram is a measure of relationship between two trading rules. It is calculated as $(1 - \rho_{i,j}) \cdot 100$, where $\rho_{i,j}$ is the correlation between trading rule i and j. Consequently, a distance of 5 means that the trading rules have a positive correlation of 0.95.

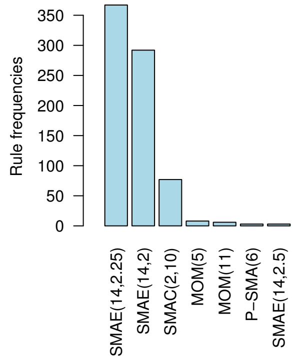
For daily data, we see that the rolling estimation scheme changes strategy more often, while the expanding stays more with the same trading rules throughout the period. For both schemes the different trading rules are highly correlated, and some rules produce almost identical returns. However, the similarities in trading rules comes as no surprise as many of them are practically the same. The only difference might be 10 or 20 days in the averaging window, or 0.25 percent in the percentage band (MAE rule). We observe that different versions of the MAE rule is highly represented, both for rolling and expanding, and it once again becomes clear that this is the best performing trading rule when using daily data.

Figure 5 is equal to Figure 4 only for monthly data. Once again we see that when using a rolling in-sample window, many different trading rules are used, while when using an expanding window, we only use a few. Different from for daily data, we here see quite large distances between the trading rules. We also see that different types of trading rules are used, especially when using a rolling in-sample window. Whereas we for daily data saw mainly the MAE rule, we here see that both the MAC and the MOM rule is well represented. It is interesting to see that it is far from one single rule that is the best performing trading rule over different in-sample periods. Similar results and implications are also found for the other datasets used in the thesis.

To conclude, we see that the rolling estimation scheme changes trading rules often, whereas the expanding scheme stays more with the same rules. Even though many of the trading rules produce similar returns, there are still differences in the best trading rules over different in-sample periods. This advocates for a dynamic strategy, which can switch



(a) Barplot: Rolling



(b) Barplot: Expanding

(c) Dendrogram: Rolling

(d) Dendrogram: Expanding

Figure 5: Barplots and cluster dendrograms of the most frequent strategies in the out-of-sample simulation on monthly data on the S&P Composite index from January 1953 to Desember 2015. The barplots in (a) and (b) illustrates the most frequent strategies for both the rolling and expanding estimation schemes. The cluster dendrograms in (c) and (d) is a graphic representation of the correlation matrix and illustrate the relationship between the most frequent strategies.

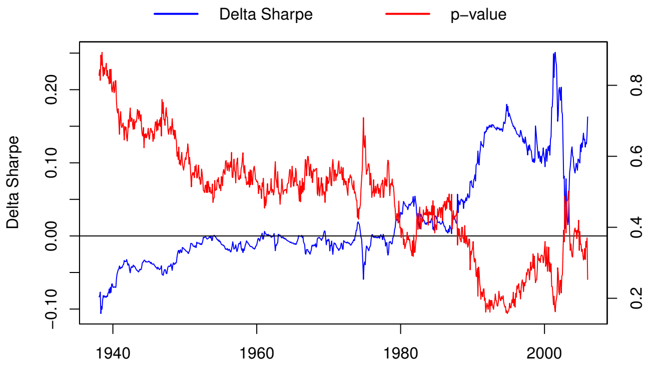
between different trading rules, in contrast to applying just one rule over a long period. Further, we see that the MAE rule is preferred for all our daily datasets. For monthly data the story is different, here we see a wider mix of rules being used, including MOM, MAC, as well as MAE.

5.2.4 Dependence on Split Point

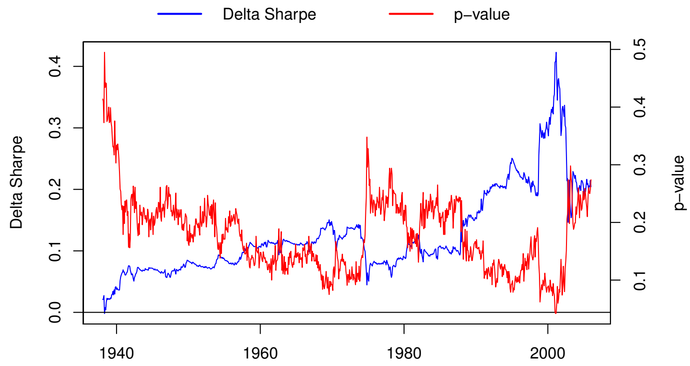
We have previously seen how the active strategies perform versus the passive buy-and-hold. Though we have only seen the performance for two different, arbitrarily chosen split points between the in-sample and out-of-sample periods. We now want to see if the strategies performance is dependent on the choice of the split point, and if so, how significant the dependence is. This is interesting to see, because if the performance is severely dependent on the split point, it will question the robustness of the strategies' outperforming capabilities. To study this, we provide plots showing the outperformance, measured by Δ Sharpe, and its corresponding p-value for different split points through time.

Figure 6 shows the outperformance measured by Δ Sharpe and its corresponding p-value of monthly trading on the S&P Composite index, when the split point is rolled through time. Looking at this we can see if, and to what degree the choice of split point affects the performance of the active strategies. Estimations are done when using the COMBI rule. The first split point is January 1938, while the last is December 2005. The initial in-sample period starts in January 1928, and is fixed for the expanding estimation scheme. The start of the in-sample period for the rolling estimation scheme rolls forward as the split point rolls forward, resulting in a constant in-sample period of 10 years for the rolling estimation scheme. For both estimation schemes, the end of the out-of-sample period is fixed to December 2015. This means that each point on the blue line and each point on the red line gives the Δ Sharpe and its corresponding p-value respectively, where the split point is given by the x-axis and the end of the out-of-sample period is December 31st, 2015. E.g. 1953 and 1970 in Figure 6 (b) gives exactly the same Δ Sharpe and p-value as we see for COMBI expanding in Table 9, Panel A and Panel B respectively. This is however not the case for Figure 6 (a) and Table 9 COMBI rolling, as the rolling in-sample periods in Table 9 are 25 and 42 years, whereas it in Figure 6 (a) is only 10 years.

From Figure 6, we see that the outperformance of the tested market timing strategies varies significantly with different choices of split points. Especially the rolling estimation scheme is very dependent on the choice of split point, whereas the expanding scheme seems to be somewhat more stable, at least over certain periods. We see that when using an expanding in-sample period, the performance is better than when we use a 10-year rolling in-sample period, and it has a positive Δ Sharpe for the whole period. Even though it seem to outperform the passive counterpart, the outperformance is almost never statistically significant. The exceptions are only a few split points in a tiny period around year 2000. For the rolling estimation scheme the performance is generally worse than that for the



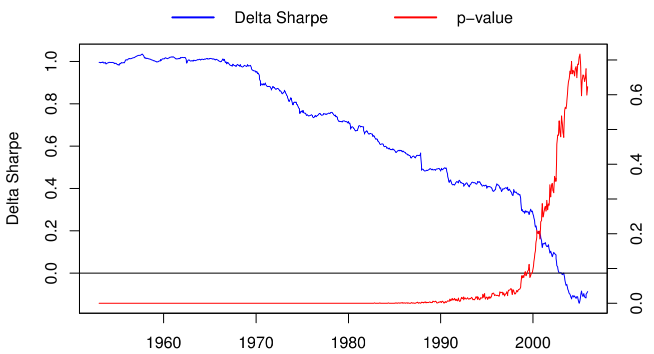
(a) Rolling



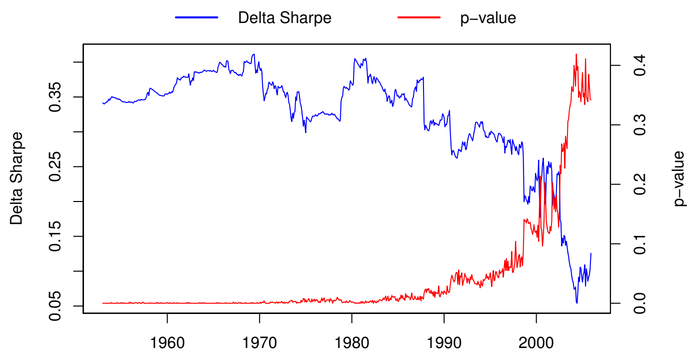
(b) Expanding

Figure 6: Delta Sharpe and p-value for different choices of split points from January 1938 to Desember 2005. The results are simulated out-of-sample using the **COMBI** rule for both rolling and expanding estimation schemes. Delta Sharpe denotes the difference between the Sharpe ratio of the active strategy and the passive benchmark and is calculated $\Delta = SR_{MT} - SR_{BH}$. For each Sharpe ratio we test the null hypothesis $H_0 : SR_{MT} \leq SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. The solid black line gives where $SR_{MT} = SR_{BH}$.

expanding estimation scheme, and we see underperformance for almost half the sample. Nor is it ever close to statistically significantly outperform the passive buy-and-hold.



(a) Daily



(b) Monthly

Figure 7: Delta Sharpe and p-value for different choices of split points from January 1953 to Desember 2005. The results are simulated out-of-sample using the **COMBI** rule for both rolling and expanding estimation schemes. Delta Sharpe denotes the difference between the Sharpe ratio of the active strategy and the passive benchmark and is calculated $\Delta = SR_{MT} - SR_{BH}$. For each Sharpe ratio we test the null hypothesis $H_0 : SR_{MT} \leq SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. The solid black line gives where $SR_{MT} = SR_{BH}$.

Figure 7 shows the outperformance of the market timing strategies for different split points when the underlying passive benchmark is small-cap stocks. Note that we here only see plots when we use the rolling estimation scheme, and that the initial in-sample period spans from January 1928 to Desember 1952, leaving a 25-year rolling in-sample window. We see statistically significant outperformance for most of the sample, but for split points towards the end, the performance deteriorates significantly. This is also in line with what we saw for the 10-year rolling outperformance. Even though the outperformance stays statistically significant over a long period, the performance varies a lot. Thus the performance is also here dependent on the choice of split point.

Similar plots for all our datasets leave us with the same impression relating to the dependence on split point. No matter if the performance is relatively good or bad, it varies significantly with different choices of split point. Plots for all other datasets can be found in Appendix II

To conclude, we generally see that the performance of the trading strategies are dependent on the choice of split point. This is the case across all our datasets, both when the general performance has been relatively good and relatively poor. This implies that *when* we start investing using these strategies may decide whether we will outperform or underperform the passive benchmark. Further, it questions whether the tested strategies has outperformance capabilities, or if outperformance merely is a result of luck. It also shows that the out-of-sample testing procedure is not a complete remedy for the data mining bias, as the results can be altered by the choice of split point.

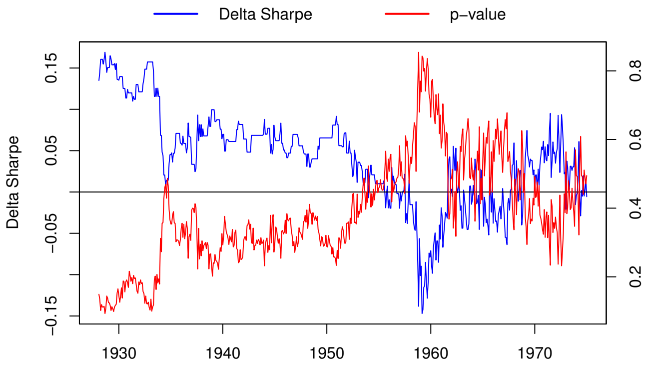
5.2.5 Dependence on In-Sample Period Length

We have seen that the performance of the active strategies is dependent on the split point, now we also want to investigate if it is dependent on the choice of in-sample period length. It is interesting to see if also this parameter alters the performance significantly, making the performance even more unpredictable. Also this may give an indication on what the minimum length of the in-sample window should be. In turn, it also indicates how much data that is required to apply the active strategies.

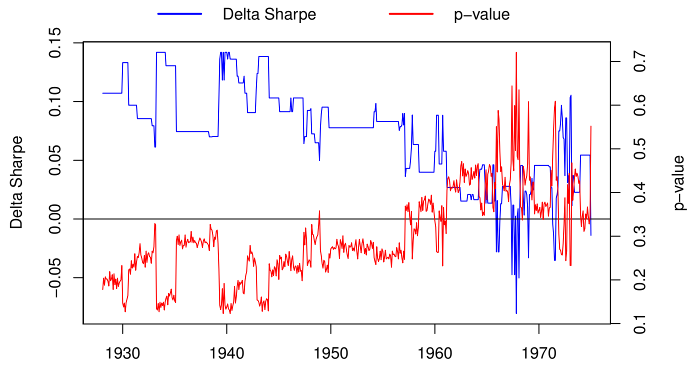
To illustrate the dependence between market timing strategy performance, and the choice of in-sample period, we plot the outperformance of a fixed out-of-sample period, with different start points for the initial in-sample period. That is, the start and end of the out-of-sample period is fixed, while the start of the initial in-sample period is rolled through time. Figure 8 shows such plots for both the rolling and expanding in-sample windows, for two different choices of out-of-sample periods. (a) and (b) gives plots when the out-of-sample period spans from January 1985 to December 2015, while (c) and (d) when the out-of-sample period spans from January 2000 to December 2015. (a) and (c) gives for the rolling estimation scheme, whereas (b) and (d) gives for the expanding. The blue and red lines gives the outperformance measured by Δ Sharpe, and its corresponding p-value respectively. The plotted lines gives the measures for the out-of-sample period, when the start of the initial in-sample period is given by the x-axis. The start date of the initial in-sample periods spans from January 1928 to ten years before the fixed split date, i.e. January 1975 for (a) and (b), and January 1990 for (c) and (d). All results are found using monthly data of the S&P Composite index.

Looking at the first fixed out-of-sample period, in (a) and (b), we see that the performance of both the rolling and expanding in-sample window varies quite a lot for the different choices of initial in-sample periods. With relatively long in-sample periods, we manage to stably outperform the passive benchmark, but still the size of the outperformance varies quite a bit. As the in-sample period becomes shorter, the performance becomes poorer, and also more unstable. For in-sample periods less than 20 years, we see that the performance is very dependent on the choice of the in-sample period. Only a few months difference in the in-sample period start date, can determine if one either performs better or worse than the passive benchmark. Poor performance seems to come "earlier" with the rolling estimation scheme, and in general performance looks to be better for the expanding estimation scheme. We note that even though the performance varies a lot, no choice of in-sample period creates statistically significant outperformance at the 5% level.

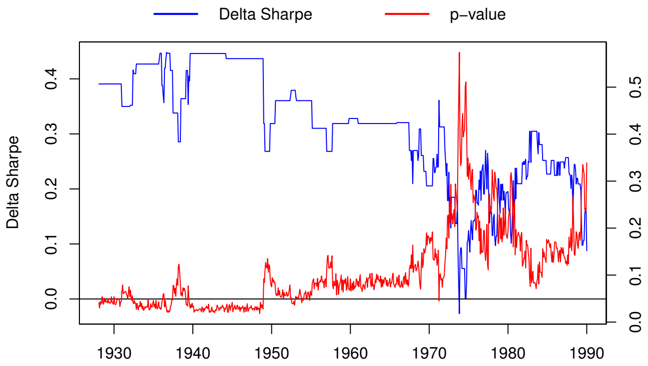
For the second fixed out-of-sample period, we see much the same pattern as in the first, only with generally better performance. This is in line with earlier presented results, where we have seen the time around year 2000 to be the best split point between the in-sample and out-of-sample periods. Once again we see that market timing strategies seems to perform better when chosen with longer in-sample periods. We also see that the rolling in-sample window needs to be longer than the expanding. Though when the initial in-sample period



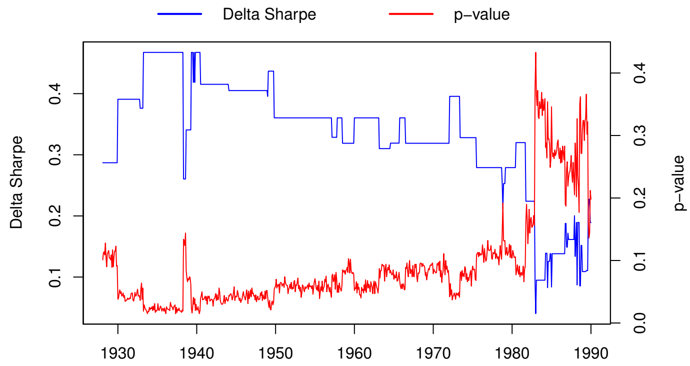
(a) Rolling, 1985-2015



(b) Expanding, 1985-2015



(c) Rolling, 2000-2015



(d) Expanding, 2000-2015

Figure 8: Delta Sharpe and p-value for out-of-sample simulations using the **COMBI** rule, when the out-of-sample window is fixed. The plots shows the performance in the fixed window for different choices of in-sample periods. For plots (a) and (b) the fixed window is set from January 1985 to December 2015, and for plots (c) and (d) the fixed window is set from January 2000 to December 2015. The dates on the x-axis represents the choice of start date of the initial in-sample period.

is less than 20 years, it looks as though the rolling estimation scheme performs superior to the expanding scheme. For this period we see several choices of in-sample periods where market timing strategies significantly outperforms the passive buy-and-hold, on both the 10% and 5% level. This is though only the case when the initial in-sample period is longer than 25 and 15 years for the rolling and expanding estimation schemes respectively.

Even though we see improved performance for longer in-sample periods, we see that the performance can vary significantly with only small differences in in-sample period length. That is even when the investment period is exactly the same, only months difference in when you start the initial in-sample period can affect if you will outperform or underperform the passive benchmark. Similar results and implications are found for monthly data on the DJIA index, and the plots can be found in Appendix III.

To conclude, we see that the performance of the active strategies is dependent on the length of the in-sample period. Using both a rolling and an expanding in-sample window, we see that the performance increases with increased in-sample period length. We see that the rolling window requires a longer initial in-sample window than the expanding, and from the plots, it looks like the minimum length should be about 25 to 30 years. As the expanding window expands, and the performance increases with increased in-sample period lengths, it makes sense that the expanding scheme requires a shorter initial in-sample period. However, we see that also this scheme requires a minimum of about 15 to 20 years. We further see that even though the performance is better for longer in-sample periods, only small differences in in-sample period starting point may lead to significant changes in performance.

6 Discussion

In this thesis we have tested a vast number of market timing strategies, both in- and out-of-sample. We have done so for two stock indices and four different stock portfolios, with both daily and monthly data frequency. In all of our results, we see that employing the active strategies, decreases the risk substantially. It looks like the active strategies manages to avoid the largest drops in the market, which also becomes evident in our drawdown analysis. Similarly, it seems that the strategies fail to capture the largest positive returns. A possible explanation is that large positive returns often follow sever market drops, and that our strategies do not manage to get invested in the risky asset in time. Consequently we generally see drops in mean annual return, though for the most part not so severe. The results in many ways follows what Faber (2007) calls *equity-like returns with bond-like volatility*, though our results do not advocate the same optimism towards the strategies' outperformance capabilities as Faber.

Generally, the empirical results leaves us the impression that market timing strategies statistically significantly outperform their passive counterpart when testing in-sample, whereas they do not when tests are run out-of-sample. These results are in line with those of Zakamulin (2014), and substantiates his claim that data mining is a problem in studies finding market timing strategies to hold outperforming capabilities over their passive counterpart. That is, testing an extensive amount of different strategies over a given period, and reporting the best (in-sample testing), is not a good way to measure if the strategies, in a real-life setting, outperform the passive benchmark. The probability that at least one of the tested strategies outperforms the passive is relatively large, but there is no guarantee that it also will perform well the following period. Many studies also back-test the *most popular trading rules*, leaving the same data-mining bias. This is because practitioners most likely already have used an in-sample test to choose these strategies. We therefore focus on the out-of-sample results to assess the outperformance of the strategy.

Out-of-sample results for the two tested stock indices shows few signs that the active strategies outperform the passive. Especially results for the DJIA index shows very poor performance, and leaves no doubt that market timing strategies do not statistically outperform the passive buy-and-hold. For the S&P Composite index we see better performance, though almost none of the reported outperformances are statistically significant at the 5% level. Like Zakamulin (2014), we see that the performance is highly non-uniform. We also find that it depends heavily on both the choice of split point and the choice of initial in-sample period. In fact, looking at Figure 6, we see that if we had chosen to report results with split point around year 2000, we could have found statistically significant outperformance when trading monthly on the S&P Composite index. Though looking at Figure 8, we see that this would have required starting the initial in-sample period before 1950 if using the expanding estimation scheme, and even earlier if using the rolling. Even then we could have been "unlucky", choosing to start the in-sample period in the late 1940s, resulting in non-significant results. All else equal, using the rolling scheme and starting the initial in-sample period around 1973, we actually see that the active strategy underperform the buy-and-hold[16]. Also in other results, we have seen that only slight changes in period choices can effect the performance dramatically. These are good examples of the non-uniformity of the out-of-sample performance of market timing strategies, and shows that outperformance may be due to luck rather than science.

Looking at for which data frequency the active strategies has performed best for the

[16]Commented results are given that the end of the out-of-sample period is set to December 2015

stock indices, our results are inconclusive. For some results, the daily data looks to perform the best, while for other, monthly data seem to perform the best. Though none of the differences result in a change of conclusion on whether the active strategies statistically outperform the passive or not. We also find somewhat conflicting results for the different stock portfolios. Results in our tables show a clear tendency across all four portfolios that the active strategies perform superior when using daily data instead of monthly. For the value sorted portfolios, we even see that the difference lead to a change in the conclusion on whether some of the active strategies statistically outperform the passive. Though when looking at the performance plots with different split points, we see that the best performing data frequency varies over different periods. Specifically we see that early in our sample, daily trading looks to have performed superior for all portfolios, whereas it more recently seems that monthly trading has performed the best. Still our results are in contrast to the conclusions of Clare et al. (2013), who finds that market timing strategies performs superior, or at least equal, when using monthly, compared to higher frequency data.

Interestingly, the active strategies looks to have performed very well for some of the stock portfolios. Especially good performances are seen for the low quintile portfolio sorted on size in the period from 1960 to 2000. Statistically significant outperformance is found when using both daily and monthly data, where daily seems to perform superior for most of the sample. However, in the plots in Figure 3, we see that the outperformance deteriorates severely towards the end of the sample. For daily trading we even see underperformance in most recent years. This also leads to monthly trading performing better than daily in the very end of the sample. Similar tendencies is seen when looking at the plots of different split points in Figure 7. For split points later than 1995 the outperformance is no longer statistically significant for neither daily nor monthly data, and in the 2000s, we even see underperformance for daily trading. Also our in-sample results shows similar signs, with poorer performance in the last sub-period from 1988 to 2015. In Table 14, statistically significant outperformances are also found for the high quintile portfolio sorted on value. However, looking at the performance plot for different split points found in Appendix II, we see that this statistically significant outperformance only occur for split points between 1953 and 1958. For split points later than that, outperformance deteriorates and for splits after 1980, all realizations of Δ Sharpe are negative. Again, this shows the data mining problem, and how misleading reporting results from only one split point can be. Looking only at Table 14, we conclude that market timing strategies using daily data on value stocks statistically outperform the passive benchmark. Though, when looking at the plot, we see that this is far from the case.

Results of portfolios consisting of small-cap stocks are the most compelling evidence we have seen for real-life outperformance of market timing strategies. Still, the deterioration of performance over the past two decades indicates that outperforming the passive benchmark is no longer probable for these stocks either. There is also a problem of how to efficiently trade such portfolios. Composing and trading such portfolios stock by stock is likely to be very costly, and may in that way deteriorate all possible outperformance. Small-cap stocks may also be quite illiquid, and in that way incur larger transaction costs. A solution might be trading in exchange traded funds (ETFs) who follows e.g. a small cap index. However, studies show ETFs to contain significant tracking errors (see Shin and Soydemir (2010)), which might also deteriorate outperformance.

Further we have to decide which trading rules to include when finding which rule to use in the next period. Our results shows that the optimal trading rules depends on what data frequency we use. For daily data, we generally see that the MAE rule is the best performing. This is seen both in the in-sample results, the out-of-sample tables, as well as the barplots in Figure 4. In turn, the good performance of the MAE rule leads us to believe that the "percentage band" protects the trader against "whipsaws", and thus prevents excessive accumulation of transaction cost for daily data. To support this argument, experiments have shown that restricting the MAE rule significantly deteriorates performance for daily data. The same performance deterioration is not found for monthly data. The MAC rule has also performed well, and at times even better than the MAE. It might therefore be more suitable at times and should be included when choosing trading rules. The question then becomes if to include the MOM rule. The performance on its own has been variable, though it actually appear twice as the best performing strategy in the in-sample simulations. It seems to be the best performing under certain market states, and should arguably also be included. The possible damage of including it also seems to be small, as our results shows that the performance of the combination rule seldom is worse than that for the best trading rule on its own. For monthly data the MAC rule looks to be the best performing. Again this is seen both in the in-sample and out-of-sample results. Further we see that the MAE rule performs only slightly worse, and that the MOM rule seem to work better for monthly than for daily data. From the barplot in Figure 5 we also see that MOM(5) is the most frequently used rule in the rolling out-of-sample test, indicating that it has been the best performing rule over several in-sample periods. Reading our results, we once again argue that the combination rule gives the best chance of good performance. It is also most realistic to expect an investor to test several different trading rules instead of just one single rule by itself.

We also have to decide whether to use a rolling or an expanding in-sample window. Our result on the matter seems to be varying and difficult to read. Over some periods it looks as though a rolling window perform superior, whereas for others, an expanding window seems to perform the best. Even though we see that the performances are similar, we also see that the two procedures chose strategies quite differently. The fact that the rolling window changes trading rule more often indicates that different trading rules are preferred over different periods. If market dynamics changes substantially over time, such a strategy are likely to perform superior, by not including too old, irrelevant data. Further we have the problem of choosing the length of the rolling in-sample window. The plots in Figure 8 shows that the rolling window should be quite long, preferably at least 25 years. We therefore use a 25-year rolling window in most of our out-of-sample tests. This means that the two test procedures have only slight differences in in-sample periods for a long period in the beginning of our sample. This, together with the fact that the underlying asset is the same, may also explain why the performances are similar.

7 Conclusion

Motivated by reoccurring deficiencies in the literature on technical analysis regarding "data mining" and negligence of transaction cost, this thesis critically investigates the profitability of moving average and momentum strategies in several stock portfolios. We simulate daily and monthly returns both in-sample and out-of-sample on two stock indices and four stock portfolios from 1928 to 2015, while simultaneously taking into account important market frictions. This is carried out in order to examine if the active strategies outperform the passive benchmark on a risk-adjusted basis, and to see if the profitable trading rules discovered in-sample also are profitable when tested out-of-sample. In addition, and for the first time, we examine the relevance of data frequency in an out-of-sample test. We adopt a stationary block bootstrap methodology in order to evaluate the statistical significance of the risk-adjusted performance, measured by the Sharpe ratio, without making any distributional assumptions. In addition, we further evaluate the active strategies using the alpha of the Fama-French-Carhart (1997) 4-factor model. A regression bootstrap is used to determine the statistical significance of the alpha.

The empirical analysis from this thesis reveal that trading rules profitable when tested in-sample, perform poorly when tested out-of-sample. Specifically, when the underlying benchmark is the S&P Composite and the DJIA index, we are not able to find any trading rule out-of-sample with statistical significant risk-adjusted outperformance. However, it is worth emphasizing that the performance of the active strategies are dataset specific. For instance, the benefit from timing the market are more distinct on the S&P Composite compared to the DJIA index. Overall, these results stands in contrast to numerous studies published in recent decades reporting that "market timing works" in stock markets (see e.g. Brock et al. (1992), Faber (2007) and Kilgallen (2012)). On the contrary, our results coincide with the empirical findings by Zakamulin (2014), suggesting that previous reported performance is highly overstated. We find generally lower volatility and mean return for all trend following trading rules. Additionally, from the drawdown analysis of the S&P Composite and DJIA, we see that the average drawdown for the active strategies generally are lower compared to the corresponding passive benchmark. However, these results are by no means surprising since the active strategy holds the risk-free asset approximately 30% of the time.

Whether to use a rolling or an expanding in-sample window remains somewhat unclear. We see varying performance over different periods, and in general, neither of the two test procedures stands out as superior to the other. What we find is that when using a rolling window, the window size must be quite large, and preferably larger than 25 years. We further see that a rolling window better adopts to changes in the market, and thus is likely to outperform an expanding window if we have changing market dynamics.

Interestingly, when the underlying asset is small-cap stocks, we are able to find statistically significant outperformance for both daily and monthly data from the out-of-sample test. In fact, and contrary to the results on the S&P Composite and DJIA index, we find that trading small-stocks using daily data performs significantly better than trading using monthly data. However, towards the end of the sample we see a reversion where trading using monthly data perform better than trading using daily data. Further, the good performance of trading small stocks appears to have decayed and disappeared in recent years. In turn, we consider future profitability from trading small stocks highly unlikely. For large stocks we find similar results as for the S&P Composite index. Statistically significant outperformances are also found for value stocks at daily frequency. However, looking at the performance plot for different split points, we see that this statistically significant outperformance only occur for split points between 1953 and 1958. This shows the "data mining" problem and illustrates how misleading reporting results from only one split point can be. When the underlying portfolio is growth stocks, we are not able to find any trading rule out-of-sample with statistically significant outperformance measured by the Sharpe ratio. However, we report a statistical significant positive alpha for the MAE rule at daily frequency.

Moreover, we find that the performance of the active strategies is highly uneven over time. This have two important implications. First, the performance of the active strategy is highly dependent on the choice of sample period and split point, and second, the out-of-sample testing procedure is not a complete remedy for the "data mining bias".

Finally, contrary to the findings of Clare et al. (2013), we are not able to find conclusive evidence suggesting that high frequency data (e.g. daily) deteriorates performance when the underlying asset is the S&P Composite index. However, by simulating each trading rule individually, we generally find that the Moving Average Envelope (MAE) rule performs best for daily data. In turn, this leads us to believe that the "percentage band" protects the trader against "whipsaw" trading, and thus prevents excessive accumulation of transaction cost for daily data. Overall, our results shows that the performance of the COMBI rule, at both daily and monthly frequency, seldom is worse than the best trading rule individually. Therefore, we argue that choosing the COMBI rule will give the best chance of good performance.

While the results of this thesis sheds light on some interesting aspects of technical trading rules, some unanswered questions remain. In particular, a similar in-depth study can be conducted for other asset classes like e.g. commodities, bonds, currencies etc. Testing the profitability of trading currencies is an intriguing topic for two reasons. First, as documented by Menkhoff and Taylor (2007), technical analysis in FOREX markets is widely used in practice. And Second, Park and Irwin (2007) finds that a majority of studies in FOREX markets from 1976 to 1991 report positive trading profits. While the authors note that the profits disappeared in the 1990s, Zakamulin (2017) argues that this conclusion was premature, and that the poor performance can be explained by the strengthening of the US dollar in the period from 1995 to 2005.

References

Alexander, S. S. (1961). "Price Movements in Speculative Markets: Trends or Random Walks", *Industrial Management Review*, *2*, 7–26.

Alexander, S. S. (1964). "Price Movements in Speculative Markets: Trends or Random Walks No. 2", *Industrial Management Review*, *5*, 25–46.

Auer, B. R. (2015). "Does the Choice of Performance Measure Influence the Evaluation of Commodity Investments?", *International Review of Financial Analysis*, *38*, 142–150.

Bessembinder, H. and Chan, K. (1998). "Market Efficiency and the Returns to Technical Analysis", *Financial Management*, 5–17.

Bodie, Z., Kane, A., and Marcus, A. J. (2011). *Investment and Portfolio Management*. McGraw-Hill Irwin.

Brock, W., Lakonishok, J., and LeBaron, B. (1992). "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns", *Journal of Finance*, *47*(5), 1731–1764.

Brown, S. J., Goetzmann, W. N., and Kumar, A. (1998). "The Dow Theory: William Peter Hamilton's Track Record Reconsidered", *Journal of Finance*, *53*(4), 1311–1333.

Carhart, M. M. (1997). "On Persistence in Mutual Fund Performance", *Journal of Finance*, *52*(1), 57–82.

Carlstein, E. (1986). "The Use of Subseries Values for Estimating the Variance of a General Statistic from a Stationary Sequence", *The Annals of Statistics*, *14*(3), 1171–1179.

Chan, L. K. and Lakonishok, J. (1993). "Institutional Trades and Intraday Stock Price Behavior", *Journal of Financial Economics*, *33*(2), 173–199.

Clare, A., Seaton, J., Smith, P. N., and Thomas, S. (2013). "Breaking Into the Blackbox: Trend Following, Stop Losses and the Frequency of Trading–The Case of the S&P500", *Journal of Asset Management*, *14*(3), 182–194.

Cogneau, P. and Hübner, G. (2009). "The (More Than) 100 Ways to Measure Portfolio Performance. Part 1: Standardized Risk-Adjusted Measures", *Journal of Performace Measurment*, *48*, 56–71.

Cowles, A. (1933). "Can Stock Market Forecasters Forecast?", *Econometrica: Journal of the Econometric Society*, 309–324.

Efron, B. (1979). "Bootstrap Methods: Another Look at the Jackknife", *The Annals of Statistics*, *7*(1), 1–26.

Eling, M. (2008). "Does the Measure Matter in the Mutual Fund Industry?", *Financial Analysts Journal*, *64*(3), 54–66.

Eling, M. and Schuhmacher, F. (2007). "Does the Choice of Performance Measure Influence the Evaluation of Hedge Funds?", *Journal of Banking & Finance*, *31*(9), 2632–2647.

Faber, M. T. (2007). "A Quantitative Approach to Tactical Asset Allocation", *Journal of Wealth Management*, *9*(4), 69–79.

Fama, E. F. (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work", *Journal of Finance*, *25*(2), 383–417.

Fama, E. F. and Blume, M. E. (1966). "Filter Rules and Stock-Market Trading", *Journal of Business*, *39*(1), 226–241.

Fama, E. F. and French, K. R. (1993). "Common Risk Factors in the Returns on Stocks and Bonds", *Journal of Financial Economics*, *33*(1), 3–56.

Fama, E. F. and French, K. R. (2010). "Luck Versus Skill in the Cross-Section of Mutual Fund Returns", *Journal of Finance*, *65*(5), 1915–1947.

Freyre-Sanders, A., Guobuzaite, R., and Byrne, K. (2004). "A Review of Trading Cost Models: Reducing Transaction Costs", *Journal of Investing*, *13*(3), 93–115.

Glabadanidis, P. (2016). "Timing the Market with a Combination of Moving Averages", *International Review of Finance*. `http://dx.doi.org/10.1111/irfi.12107`.

Gwilym, O., Clare, A., Seaton, J., and Thomas, S. (2010). "Price and Momentum as Robust Tactical Approaches to Global Equity Investing", *Journal of Investing*, *19*(3), 80–91.

Hachemian, F., Tavernier, S., and Van Royen, A.-S. (2013). "The Significance of Trading Frequency and Stop Loss in Trend Following Strategies", *SSRN Electronic Journal*.

Hamilton, W. P. (1922). *The Stock Market Barometer; a Study of its Forecast Value Based on Charles H. Dow's Theory of the Price Movement*. Harper & Bros.

Harvey, C. R. and Liu, Y. (2015). "Backtesting", *Journal of Portfolio Management*, *42*(1), 13–28.

Horowitz, J. L. (2003). "Bootstrap Methods for Markov Processes", *Econometrica*, *71*(4), 1049–1082.

James, F. (1968). "Monthly Moving Averages an Effective Investment Tool", *Journal of Financial and Quantitative Analysis*, *3*(03), 315–326.

Jegadeesh, N. and Titman, S. (1993). "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency", *Journal of finance*, *48*(1), 65–91.

Jensen, M. C. (1967). "Random Walks: Reality or Myth: Comment", *Financial Analysts Journal*, 77–85.

Jensen, M. C. (1968). "The Performance of Mutual Funds in the Period 1945–1964", *Journal of finance*, *23*(2), 389–416.

Jensen, M. C. and Benington, G. A. (1970). "Random Walks and Technical Theories: Some Additional Evidence", *Journal of Finance*, *25*(2), 469–482.

Jobson, J. D. and Korkie, B. M. (1981). "Performance Hypothesis Testing with the Sharpe and Treynor Measures", *Journal of Finance*, *36*(4), 889–908.

Kilgallen, T. (2012). "Testing the Simple Moving Average Across Commodities, Global Stock Indices, and Currencies", *Journal of Wealth Management*, *15*(1), 82–100.

Kirby, C. and Ostdiek, B. (2012). "It's All in the Timing: Simple Active Portfolio Strategies that Outperform Naive Diversification", *Journal of Financial and Quantitative Analysis*, *47*(02), 437–467.

Knez, P. J. and Ready, M. J. (1996). "Estimating the Profits from Trading Strategies", *Review of Financial Studies*, *9*(4), 1121–1163.

Kosowski, R., Timmermann, A., Wermers, R., and White, H. (2006). "Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis", *Journal of finance*, *61*(6), 2551–2595.

Kunsch, H. R. (1989). "The Jackknife and the Bootstrap for General Stationary Observations", *The Annals of Statistics*, *17*(3), 1217–1241.

Levy, R. A. (1967). "Relative Strength as a Criterion for Investment Selection", *Journal of Finance*, *22*(4), 595–610.

Lo, A. W. and MacKinlay, A. C. (1990). "Data-Snooping Biases in Tests of Financial Asset Pricing Models", *Review of Financial Studies*, *3*(3), 431–467.

Lukac, L. P. and Brorsen, B. W. (1990). "A Comprehensive Test of Futures Market Disequilibrium", *Financial Review*, *25*(4), 593–622.

Lukac, L. P., Brorsen, B. W., and Irwin, S. H. (1988). "A Test of Futures Market Disequilibrium Using Twelve Different Technical Trading Systems", *Applied Economics*, *20*(5), 623–639.

Mandelbrot, B. (1963). "The Variation of Certain Speculative Prices", *Journal of Business*, *36*(4), 394–419.

Memmel, C. (2003). "Performance Hypothesis Testing with the Sharpe Ratio", *Finance Letters*, *1*, 21–23.

Menkhoff, L. and Taylor, M. P. (2007). "The Obstinate Passion of Foreign Exchange Professionals: Technical Analysis", *Journal of Economic Literature*, *45*(4), 936–972.

Metghalchi, M., Marcucci, J., and Chang, Y.-H. (2012). "Are Moving Average Trading Rules Profitable? Evidence from the European Stock Markets", *Applied Economics*, *44*(12), 1539–1559.

Miller, R. G. (1974). "The Jackknife - a Review", *Biometrika*, *61*(1), 1–15.

Moskowitz, T. J., Ooi, Y. H., and Pedersen, L. H. (2012). "Time Series Momentum", *Journal of Financial Economics*, *104*(2), 228–250.

Park, C.-H. and Irwin, S. H. (2007). "What Do We Know About the Profitability of Technical Analysis?", *Journal of Economic Surveys*, *21*(4), 786–826.

Pätäri, E. and Vilska, M. (2014). "Performance of Moving Average Trading Strategies Over Varying Stock Market Conditions: The Finnish Evidence", *Applied Economics*, *46*(24), 2851–2872.

Patton, A., Politis, D. N., and White, H. (2009). "Correction to "Automatic Block-Length Selection for the Dependent Bootstrap" by D. Politis and H. White", *Econometric Reviews*, *28*(4), 372–375.

Politis, D. N. and Romano, J. P. (1994). "The Stationary Bootstrap", *Journal of the American Statistical Association*, *89*(428), 1303–1313.

Politis, D. N. and White, H. (2004). "Automatic Block-Length Selection for the Dependent Bootstrap", *Econometric Reviews*, *23*(1), 53–70.

Rhea, R. (1932). *The Dow Theory: An Explanation of its Development and an Attempt to Define its Usefulness as an Aid in Speculation.* Reprint in 1993 by Fraser Publishing Company.

Rossi, B. and Inoue, A. (2012). "Out-of-Sample Forecast Tests Robust to the Choice of Window Size", *Journal of Business & Economic Statistics*, *30*(3), 432–453.

Samuelson, P. A. (1965). "Proof That Properly Anticipated Prices Fluctuate Randomly", *Industrial Management Review*, *6*(2), 41–49.

Sharpe, W. F. (1966). "Mutual Fund Performance", *Journal of Business*, *39*(1), 119–138.

Sharpe, W. F. (1994). "The Sharpe Ratio", *Journal of Portfolio Management*, *21*(1), 49–58.

Shin, S. and Soydemir, G. (2010). "Exchange-Traded Funds, Persistence in Tracking Errors and Information Dissemination", *Journal of Multinational Financial Management*, *20*(4), 214–234.

Sortino, F. A. and Price, L. N. (1994). "Performance Measurement in a Downside Risk Framework", *Journal of Investing*, *3*(3), 59–64.

Sullivan, R., Timmermann, A., and White, H. (1999). "Data-Snooping, Technical Trading Rule Performance, and the Bootstrap", *Journal of Finance*, *54*(5), 1647–1691.

Van Horne, J. C. and Parker, G. G. (1967). "The Random-Walk Theory: An Empirical Test", *Financial Analysts Journal*, 87–92.

Van Horne, J. C. and Parker, G. G. (1968). "Technical Trading Rules: A Comment", *Financial Analysts Journal*, 128–132.

Welch, I. and Goyal, A. (2008). "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction", *Review of Financial Studies*, *21*(4), 1455–1508.

White, H. (2000). "A Reality Check for Data Snooping", *Econometrica*, *68*(5), 1097–1126.

Zakamulin, V. (2014). "The Real-Life Performance of Market Timing with Moving Average and Time-Series Momentum Rules", *Journal of Asset Management*, *15*(4), 261–278.

Zakamulin, V. (2015). "A Comprehensive Book at the Empirical Performance of Moving Average Trading Strategies", Working paper, University of Agder.

Zakamulin, V. (2017). *Market Timing with Moving Averages: The Anatomy and Performance of Trading Rules*. Palgrave Macmillan, UK (forthcoming).

Appendices

Appendix I

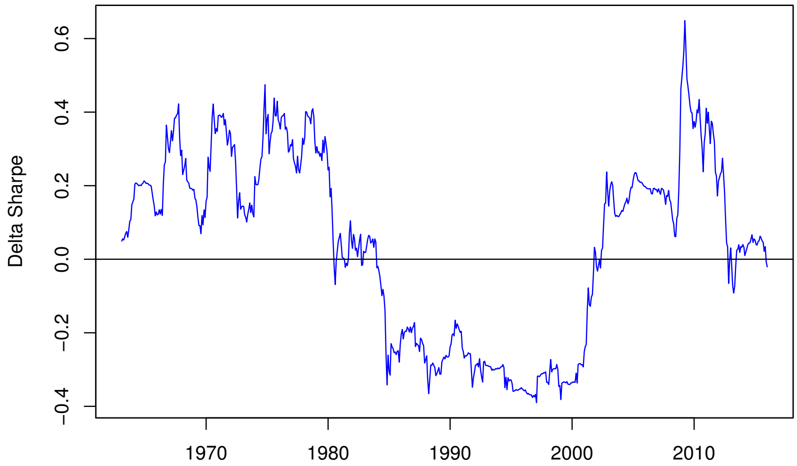
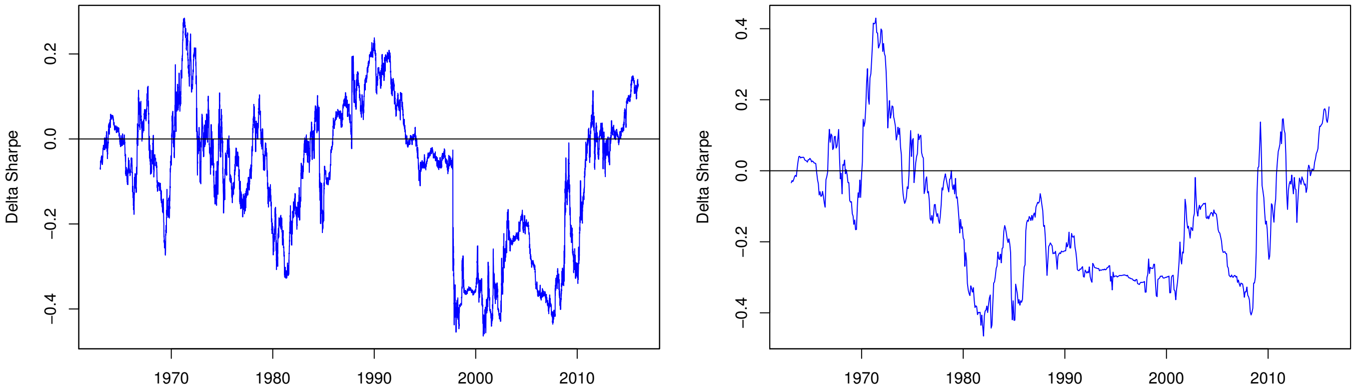


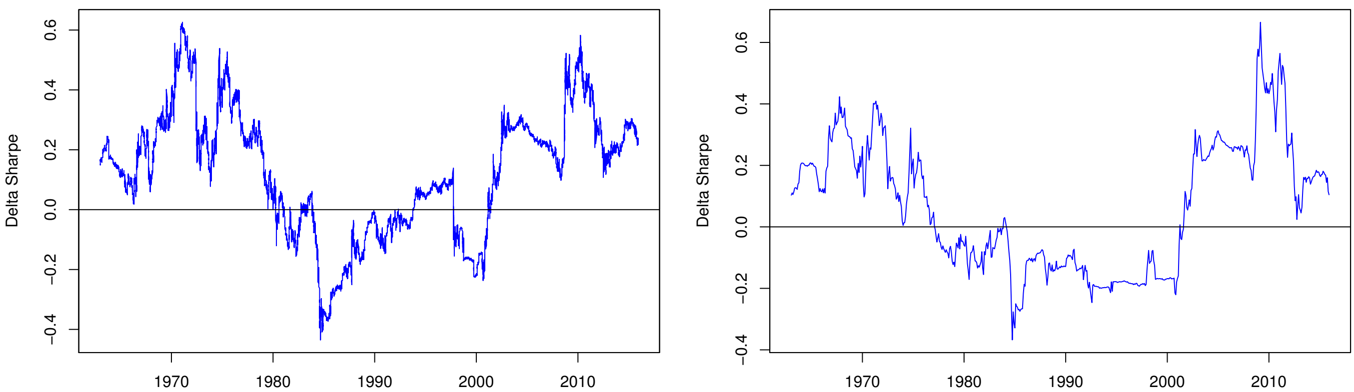
Figure 9: Rolling 10-year outperformance of monthly trading on the S&P Compostie index over the period January 1953 to December 2015. The results are simulated out-of-sample using the **COMBI** rule for the rolling estimation scheme. Delta Sharpe denotes the difference between the Sharpe ratio of the active strategy and the passive benchmark and is calculated $\Delta = SR_{MT} - SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. The solid black line gives where $SR_{MT} = SR_{BH}$.



(a) Daily

(b) Monthly

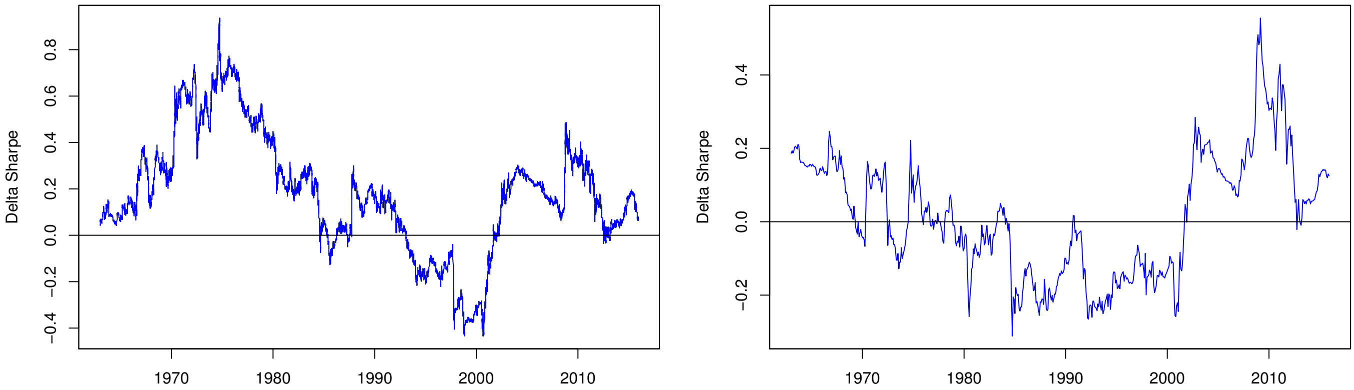
Figure 10: Rolling 10-year outperformance in trading on the DJIA index over the period January 1953 to December 2015. The results are simulated out-of-sample using the **COMBI** rule for the rolling estimation scheme. Delta Sharpe denotes the difference between the Sharpe ratio of the active strategy and the passive benchmark and is calculated $\Delta = SR_{MT} - SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. The solid black line gives where $SR_{MT} = SR_{BH}$.



(a) Daily

(b) Monthly

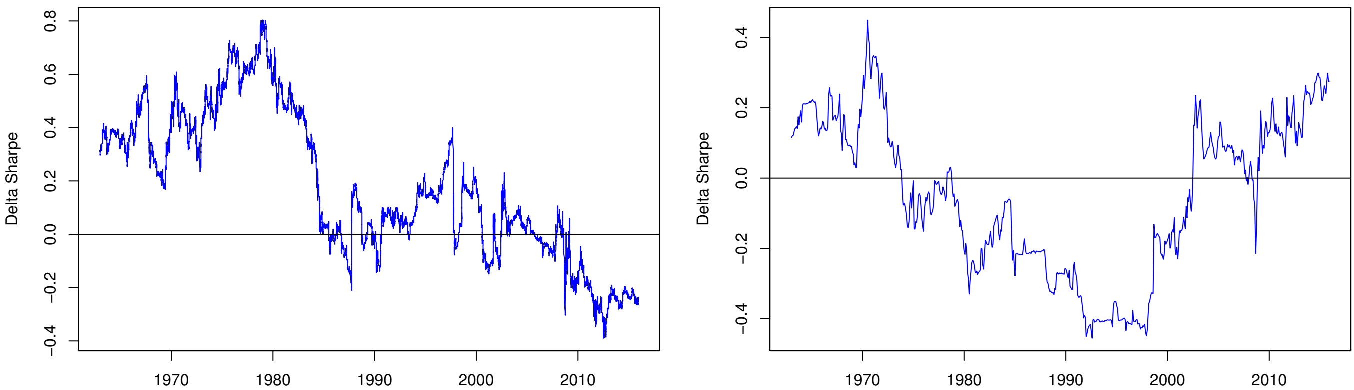
Figure 11: Rolling 10-year outperformance in trading large stocks over the period January 1953 to December 2015. The results are simulated out-of-sample using the **COMBI** rule for the rolling estimation scheme. Delta Sharpe denotes the difference between the Sharpe ratio of the active strategy and the passive benchmark and is calculated $\Delta = SR_{MT} - SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. The solid black line gives where $SR_{MT} = SR_{BH}$.



(a) Daily

(b) Monthly

Figure 12: Rolling 10-year outperformance in trading growth stocks over the period January 1953 to December 2015. The results are simulated out-of-sample using the **COMBI** rule for the rolling estimation scheme. Delta Sharpe denotes the difference between the Sharpe ratio of the active strategy and the passive benchmark and is calculated $\Delta = SR_{MT} - SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. The solid black line gives where $SR_{MT} = SR_{BH}$.

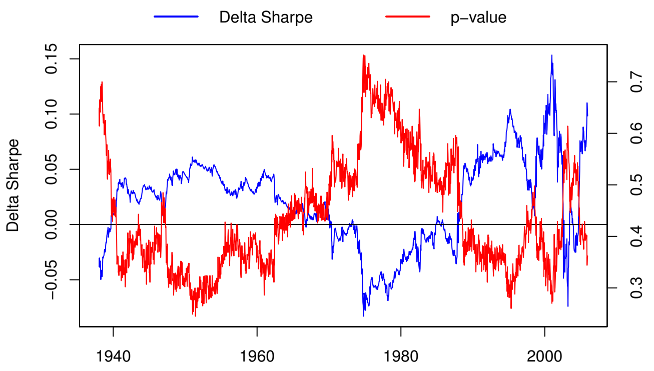


(a) Daily

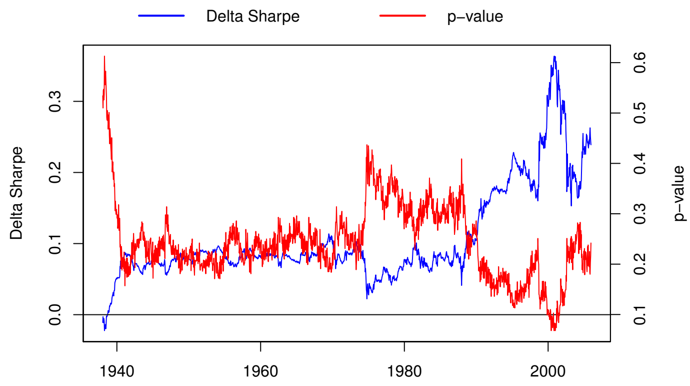
(b) Monthly

Figure 13: Rolling 10-year outperformance in trading value stocks over the period January 1953 to December 2015. The results are simulated out-of-sample using the **COMBI** rule for the rolling estimation scheme. Delta Sharpe denotes the difference between the Sharpe ratio of the active strategy and the passive benchmark and is calculated $\Delta = SR_{MT} - SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. The solid black line gives where $SR_{MT} = SR_{BH}$.

Appendix II

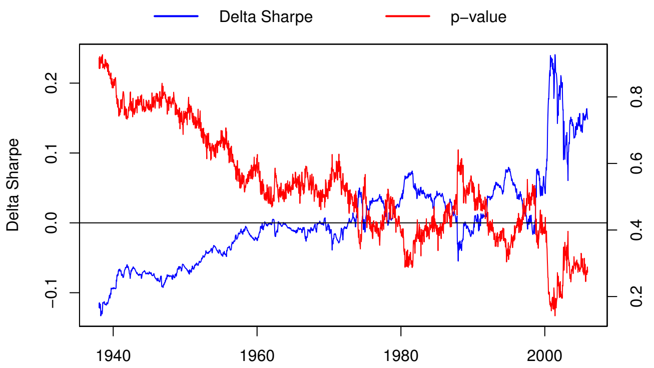


(a) Rolling

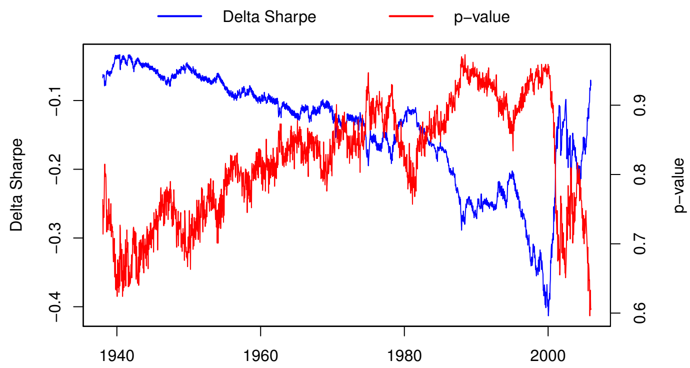


(b) Expanding

Figure 14: S&P Composite daily: Delta Sharpe and p-value for different choices of split points from January 1938 to Desember 2005. The results are simulated out-of-sample using the **COMBI** rule for both rolling and expanding estimation schemes. Delta Sharpe denotes the difference between the Sharpe ratio of the active strategy and the passive benchmark and is calculated $\Delta = SR_{MT} - SR_{BH}$. For each Sharpe ratio we test the null hypothesis $H_0 : SR_{MT} \leq SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. The solid black line gives where $SR_{MT} = SR_{BH}$.

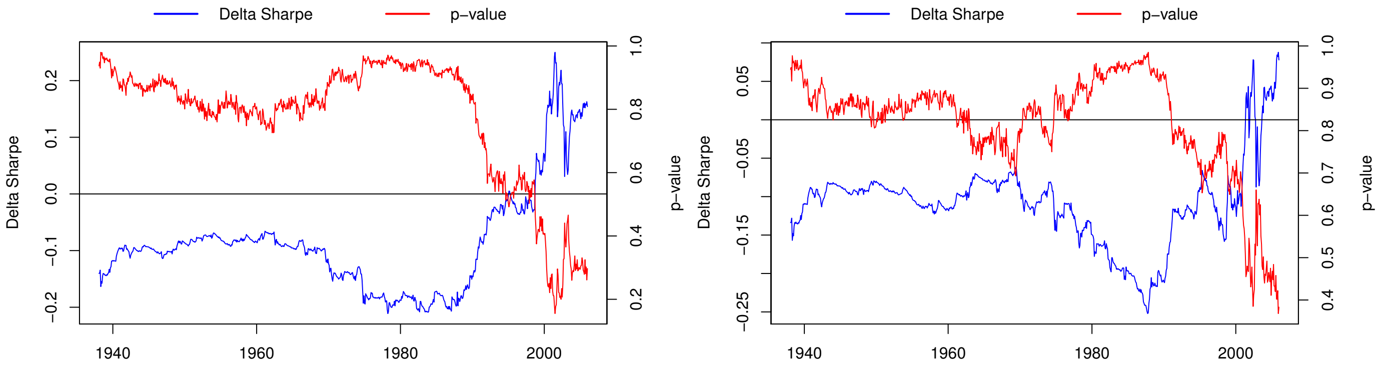


(a) Rolling



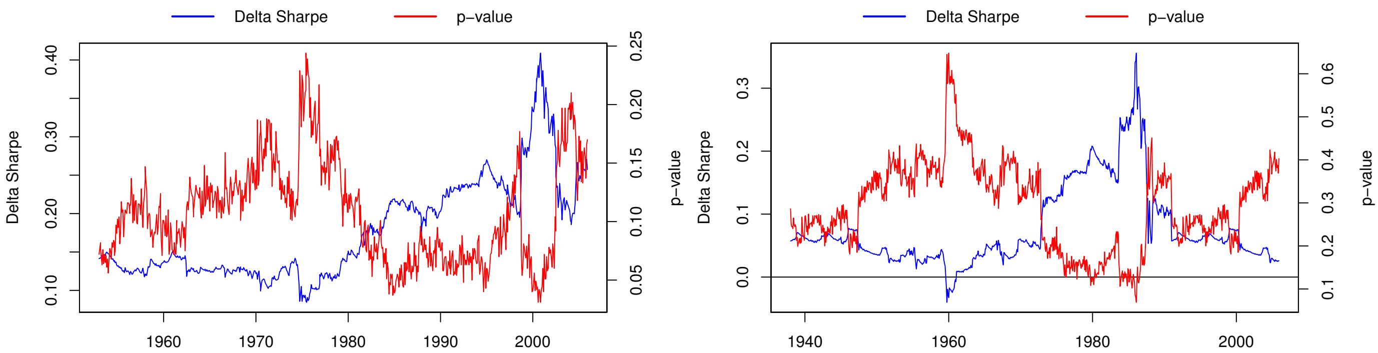
(b) Expanding

Figure 15: DJIA daily: Delta Sharpe and p-value for different choices of split points from January 1938 to Desember 2005. The results are simulated out-of-sample using the **COMBI** rule for both rolling and expanding estimation schemes. Delta Sharpe denotes the difference between the Sharpe ratio of the active strategy and the passive benchmark and is calculated $\Delta = SR_{MT} - SR_{BH}$. For each Sharpe ratio we test the null hypothesis $H_0 : SR_{MT} \leq SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. The solid black line gives where $SR_{MT} = SR_{BH}$.



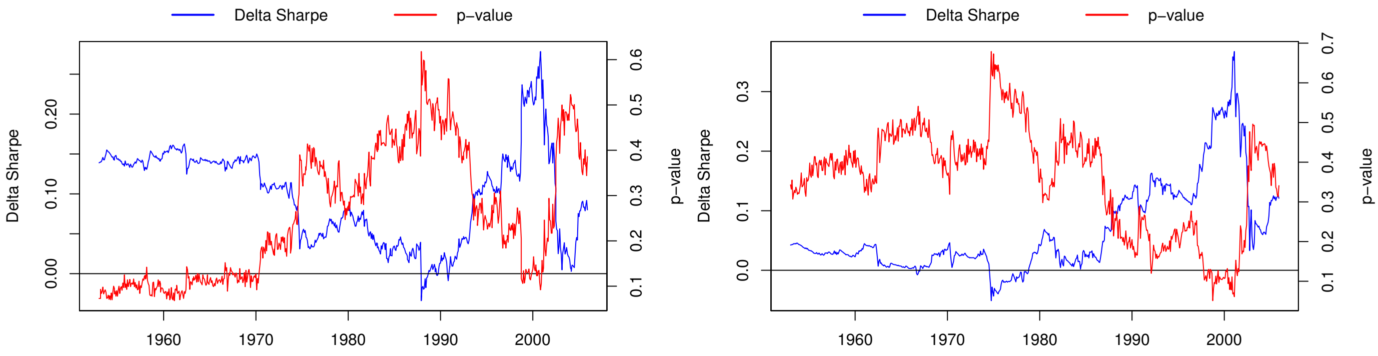
(a) Rolling (b) Expanding

Figure 16: DJIA monthly: Delta Sharpe and p-value for different choices of split points from January 1938 to Desember 2005. The results are simulated out-of-sample using the **COMBI** rule for both rolling and expanding estimation schemes. Delta Sharpe denotes the difference between the Sharpe ratio of the active strategy and the passive benchmark and is calculated $\Delta = SR_{MT} - SR_{BH}$. For each Sharpe ratio we test the null hypothesis $H_0 : SR_{MT} \leq SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. The solid black line gives where $SR_{MT} = SR_{BH}$.



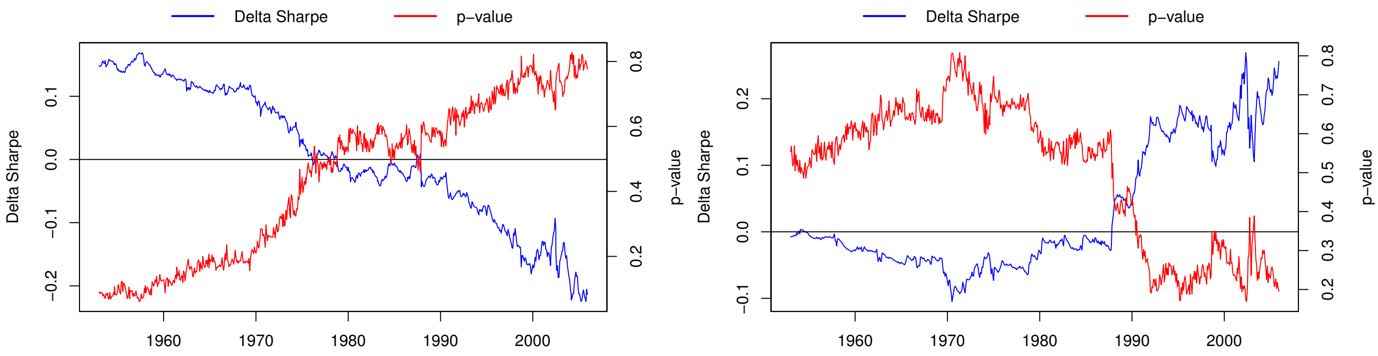
(a) Daily (b) Monthly

Figure 17: Large stocks: Delta Sharpe and p-value for different choices of split points from January 1953 to Desember 2005. The results are simulated out-of-sample using the **COMBI** rule for the rolling estimation scheme. Delta Sharpe denotes the difference between the Sharpe ratio of the active strategy and the passive benchmark and is calculated $\Delta = SR_{MT} - SR_{BH}$. For each Sharpe ratio we test the null hypothesis $H_0 : SR_{MT} \leq SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. The solid black line gives where $SR_{MT} = SR_{BH}$.



(a) Daily (b) Monthly

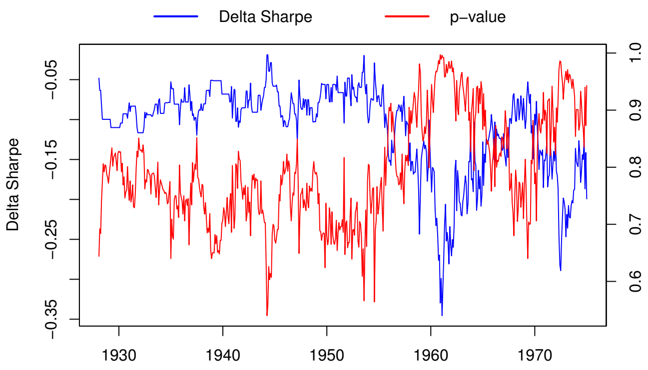
Figure 18: Growth stocks: Delta Sharpe and p-value for different choices of split points from January 1953 to Desember 2005. The results are simulated out-of-sample using the **COMBI** rule for the rolling estimation scheme. Delta Sharpe denotes the difference between the Sharpe ratio of the active strategy and the passive benchmark and is calculated $\Delta = SR_{MT} - SR_{BH}$. For each Sharpe ratio we test the null hypothesis $H_0 : SR_{MT} \leq SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. The solid black line gives where $SR_{MT} = SR_{BH}$.



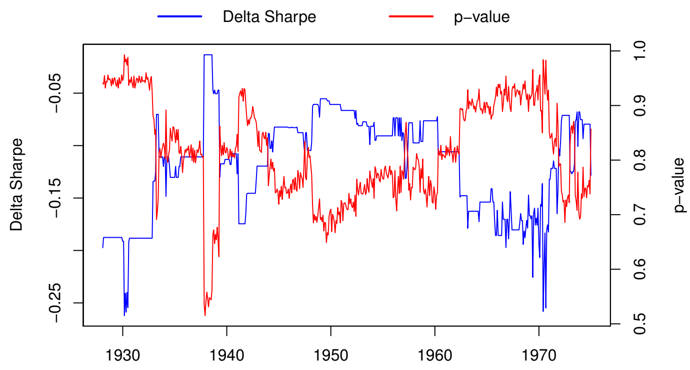
(a) Daily (b) Monthly

Figure 19: Value stocks: Delta Sharpe and p-value for different choices of split points from January 1953 to Desember 2005. The results are simulated out-of-sample using the **COMBI** rule for the rolling estimation scheme. Delta Sharpe denotes the difference between the Sharpe ratio of the active strategy and the passive benchmark and is calculated $\Delta = SR_{MT} - SR_{BH}$. For each Sharpe ratio we test the null hypothesis $H_0 : SR_{MT} \leq SR_{BH}$, where SR_{MT} is the Sharpe ratio of the active strategy and SR_{BH} is the Sharpe ratio of the passive benchmark. The solid black line gives where $SR_{MT} = SR_{BH}$.

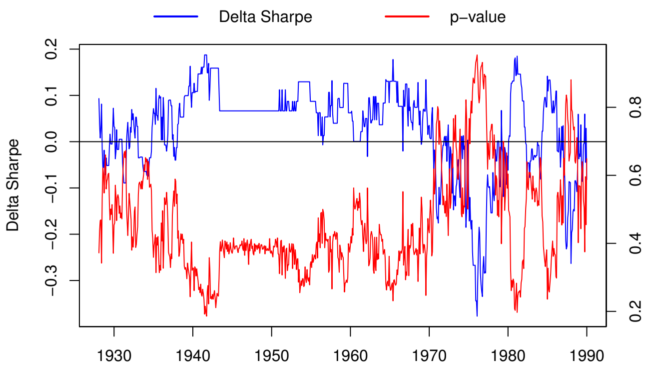
Appendix III



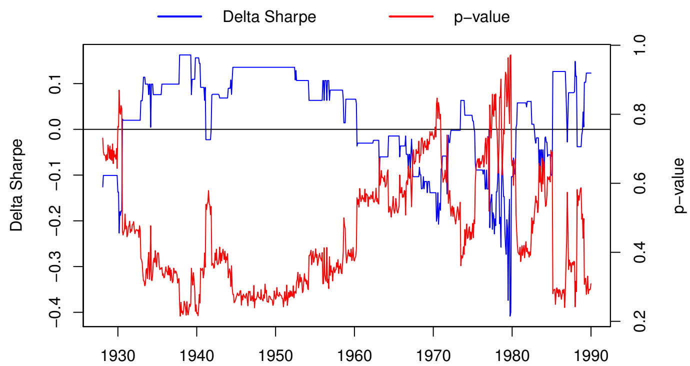
(a) Rolling, 1985-2015



(b) Expanding, 1985-2015



(c) Rolling, 2000-2015



(d) Expanding, 2000-2015

Figure 20: DJIA Monthly: Delta Sharpe and p-value for out-of-sample simulations using the **COMBI** rule, when the out-of-sample window is fixed. The plots shows the performance in the fixed window for different choices of in-sample periods. For plots (a) and (b) the fixed window is set from January 1985 to December 2015, and for plots (c) and (d) the fixed window is set from January 2000 to December 2015. The dates on the x-axis represents the choice of start date of the in-sample period.

Appendix IV

Reflection Note

Håvard Løseth Modell

To complete our Master degree in Financial Economics, we wanted to find a topic in which we could utilize the quantitative skills and financial theory we have adopted throughout our studies. We therefore chose to look closer at the profitability of trading stocks using different market timing trading rules. The goal of the thesis then became to thoroughly investigate the performance of stock market timing using strategies based on moving averages and time-series momentum. We did so by extensive tests of different strategies on data of the US stock market in the historical period from 1928 to the end of 2015.

We found little to no statistically significant evidence that the active market timing strategies hold outperforming capabilities over the passive buy-and-hold. This result also persist across different indices and stock portfolios as well as for both monthly and daily data frequency. We further found that the performance is highly non-uniform as well as dependent on the choice of split point between in- and out-of-sample periods and also the length of the in-sample period. This led us to conclude that timing the US stock market using moving averages and time-series momentum rules does not outperform the passive buy-and-hold strategy. We also argue that previous studies finding market timing strategies to "work" contain a data mining bias leading to flawed results.

Our study in itself, with how to trade, and different market timing strategies might not be very relevant when looking at internationalization. But if we take a step back, and look slightly broader, including stock market trading and the financial market, it becomes very relevant. The US stock market is not only a reflection of the American economy, but is influenced by factors from economies around the world. In the same way, economics and the financial markets in the rest of the world is largely dependent on the economic superpower America. This has become very evident with the different financial crises we have had throughout the years. Take for instance the crises of 2007/2008. Bad mortgage loans in the US created a worldwide financial crises, affecting people around the whole world. We also have the internationalization impact of the internet when it comes to stock trading. With an internet access you can easily be situated in Spain, trading equities in the US, Japan and Norway at the same time. This again shows how stock trading and the financial markets no longer is restricted by country boarders.

Innovation is an important factor in a sector where computers and technology plays such an important role. In trading, we have seen a wast increase in the use of computers and trading algorithms over recent years. Almost all major investment banks and hedge funds today use computers and trading robots to some extent. Recently we have also seen that several large institutional investors lay off stock traders, replacing them with robots programmed with sophisticated trading algorithms. The possible performance of such artificial intelligence has also been proven by the historical returns of the Medallion fund of the quantitative hedge fund Renaissance Technologies. With an average annual return of over 40% since 1990, it has beaten the market every year. With historical success and the possibilities coming from new technology and increased computer power, it is no doubt that robots and artificial intelligence has reserved a place in the future of stock trading, as well as in the financial markets in general.

With great power comes great responsibility, and with asset under management ranging from billions to trillions of dollars, many institutional investors today has great power. The focus on responsible investments has increased in recent years, and today there are several funds focusing on making only sustainable investments in e.g. renewable energy and climate friendly technology. Still, some argue that the only responsibility of an investment bank is to maximize their customers wealth within the boundaries of the law. It therefore also becomes a responsibility for the government and law-makers. They should regulate the market in a fashion that protects both the interests of the people and the future of this planet. As greed and unlawful actions always will be a problem in a world driven by money, governing bodies also has a responsibility to see to it that this laws are adhered to.

Appendix V

Reflection Note

Lars Magnus Lynngård

In this study we have investigated the performance of trend following strategies, specifically moving average and momentum trading rules. The aim of the study is to examine stock markets and portfolios to see if a simple and active trading strategy can outperform the buy-and-hold strategy. We have tested the strategies on the stock indices; S&P Composite and DJIA, and stock portfolios; small-cap, large-cap, growth and value. The main findings can be summarized as follows. When simulating the strategies in a real-life setting we are generally not able to find active strategies that outperform the passive (B&H). However, when tested on small stocks, we found that the active strategy outperformed the passive in the period from 1955 to 2000, but not thereafter.

The thesis is completed using knowledge and skills attained throughout the education from courses like, econometrics, computational finance, advanced econometrics, corporate finance, finance theory, etc. Specifically, in advanced econometrics and computational we learned how to write an academic paper which in turn prepared us for the format the thesis is written in. In computational finance we also learned the tools necessary to complete the data programming in R, the main software used for completing the thesis. The topic studied in this thesis is quite specific and narrow. As a consequence, the following discussion/reflection will be broadened to technical analysis, trading and portfolio management in general.

In order to examplify how interconnected the financial markets are today, consider that the 500 largest companies in the US, measured by the S&P 500, earns 50% of their revenues abroad. In practice, this means that the US economy (and its companies) is not only dependent on domestic variables like growth, employment, interest rates etc., but also on the countries where the revenues comes from (e.g. China, Europe, India). This dependency and global interconnected markets came to light by the recent global financial crisis starting in 2007. Even though the crisis was connected to the US housing market, the consequences were severe worldwide. For example, all three private commercial banks in Iceland (which was considered robust) defaulted. It is now 10 years since the global financial crisis emerged, the world has become even more interconnected, which in turn makes me believe that the next global financial crisis may even be more severe then the last. However, we have seen in recent years that populist politicians in Europe and US (Trump) are on the rise with anti-globalization rhetoric. Trump have explicitly said that he wants to "close" the borders and bring American factories abroad back to the United States. If this trend continues we might see a decline or reversion in globalization. Time will tell.

The financial industry is currently undergoing massive changes. Specifically, we see almost every day new capabilities and improvements from AI (artificial intelligence) and robots (algorithms). This have several implications. First, many jobs will become redundant when algorithms can do it 100x faster and for "free". Recently, large financial institutions have announced that robots and algorithms have taken over and will continue to take over jobs previously completed by humans. This includes capital management, customer service, etc. However, as some jobs disappear and becomes redundant, new jobs and positions are needed. And second, not only are robots and algorithms cheaper, they are also much faster. In trading this is referred to as high frequency trading (HFT). High frequency trading refers to the process of computers trading millions times a day (mainly against other computers/algorithms) gaining small profits each time. The algorithms have become incredibly sophisticated, and before you and I are able to open a news article, the algorithm have already interpreted the meaning and magnitude of the news, and placed a trade. As a consequence, humans are not able to compete against the algorithms, and the algorithms are getting better and faster each day. In addition, from almost anywhere in the world it is now possible to trade assets worldwide, and not only stocks, but also bonds, currencies, commodities (e.g. gold, oil, beef, coffee) and real-estate (REIT). Each day, new derivatives and trading products (ETFs and ETNs) are launched which allows investors to trade or make a bet about anything. And more importantly, you no longer have to call (and pay) your broker. Nowadays most people can trade through online retail brokerage firms faster and cheaper then previously. In summary, the financial industry have changed a lot and will continue to change.

There is an ongoing debate about the importance of responsible investing. According to Fiona Reynolds, director of UN's responsible investment department, over 63 trillions USD of capital is invested according to UN's principle of responsible investments and EGS

(Environmental, Social and Governance). A number that is increasing by the day. In order to help investors (small and large) to invest more responsibly, Morningstar provide sustainability ratings based on EGS for a wide arrange of different funds. This allows investors to not only invest in funds that fits the individual risk-reward level, but also the individuals values and core principles.