Vaasan yliopisto
UNIVERSITY OF VAASA

Kasper Koski

Value in Fundamental Stock Screening

F-Score Investment Strategy Performance and Additional Fundamental Analysis in the US Equity Market

School of Accounting and Finance
Master's thesis in Finance
Master's Degree Programme in Finance

Vaasa 2020

UNIVERSITY OF VAASA
School of Accounting and Finance
Author: Kasper Koski
Title of the Thesis: Value in Fundamental Stock Screening : F-Score Investment Strategy Performance and Additional Fundamental Analysis in the US Equity Market
Degree: Master of Science in Economics and Business Administration
Programme: Master's Degree Programme in Finance
Supervisor: Klaus Grobys
Year: 2020 **Pages:** 77

ABSTRACT:

The objective in this study is to assess the performance of an F-Score based trading strategy in the US equity market and analyze whether the strategy can be improved by excluding companies that have unreliable earnings figures and increased default risk, proxied by M-Score and Z-Score, respectively. Piotroski (2000) argues that by screening high book-to-market (HBM) companies with the F-Score, an aggregate of nine fundamental signals indicating financial strength, one can earn significant abnormal returns. Furthermore, Piotroski (2000) claims that the spread between long and short position returns is the biggest when low distress risk companies are used. Moreover, Beneish (2013) finds that companies, which may have manipulated their earnings figures tend to earn lower returns. To identify manipulators, Beneish (2013) uses M-Score (Beneish 1999), also calculated from financial statement data.

In this study, companies are first ranked based on their fiscal year-end book-to-market ratios after which, the top tercile (high book-to-market companies) is taken into further analysis. Next, the HBM companies are ranked with F-Score so that financially sound (weak) companies are assigned to a long (short) portfolio. In stages three and four, the F-Score portfolios are screened with M-score and Z-Score so that companies with high manipulation probability and inflated default risk are excluded. Portfolio formations are carried out in June and the positions are held for one year, after which the ranking is repeated. The first portfolio formations are in June 1999 and the last in June 2016.

Based on the analysis conducted for S&P 500 constituent companies, an F-Score based trading strategy generates (positive) abnormal returns over the sample period, but only for the long leg. Moreover, the long position returns seem to be mainly driven by the underlying performance of the HBM portfolio. However, high F-Score companies seem to be more profitable than their low F-Score counterparts. Additionally, by using M-Score to identify companies that may have managed their earnings and excluding them, the risk-adjusted performance of the long portfolios can be improved compared to an F-Score-only strategy. The exclusion of possible earnings manipulators also decreases the returns of the short portfolios, but not enough to reach acceptable statistical significance. Additional Z-Score screening on the other hand seems to be inefficient for both long and short portfolios. The results suggest that an F-Score based financial strength analysis is, to some extent, useful also when only large companies are analyzed. Moreover, the results indicate that additional fundamental analysis that considers the quality of reported earnings can be beneficial when implementing a value strategy.

KEYWORDS: Value Investing, Fundamental Analysis, Earnings Management, F-Score, M-Score

VAASAN YLIOPISTO
Laskentatoimen ja rahoituksen yksikkö
Tekijä: Kasper Koski
Tutkielman nimi: Value in Fundamental Stock Screening : F-Score Investment Strategy Performance and Additional Fundamental Analysis in the US Equity Market
Tutkinto: Kauppatieteiden maisteri
Oppiaine: Master's Degree Programme in Finance
Työn ohjaaja: Klaus Grobys
Vuosi: 2020 **Sivumäärä:** 77

TIIVISTELMÄ:

Tutkielman tarkoituksena on arvioida F-Score perusteisen arvosijoitusstrategian toimivuutta Yhdysvaltojen osakemarkkinoilla sekä analysoida, voiko strategian tuottoja parantaa käyttämällä ainoastaan yhtiöitä, joiden raportoidut tuottoluvut ovat luotettavia ja joilla on alhainen konkurssiriski. Piotroski (2000) esittää, että pisteyttämällä korkean B/M-luvun yhtiöt yhdeksän fundamenttimuuttujan F-Scorella, sijoittajan on mahdollista erotella tulevaisuudessa parhaiten pärjäävät yhtiöt heikosti menestyvistä. Piotroski (2000) myös huomauttaa, että pitkän ja lyhyen position tuottoero on suurin alhaisen konkurssiriskin yhtiöitä käytettäessä. Beneish (2013) puolestaan toteaa, että yhtiöt, jotka ovat mahdollisesti vääristäneet raportoituja tuottolukujaan, ansaitsevat alhaisempia tuottoja tulevaisuudessa. Mahdollisten tulosmanipuloijien tunnistamiseen Beneish (2013) käyttää niin ikään tilinpäätöstiedosta laskettua M-Scorea (Beneish 1999).

Tutkielmassa käytetyn sijoitusstrategian vaiheet voidaan esittää seuraavasti. Ensimmäisessä vaiheessa yhtiöt järjestetään tilinpäätöstiedosta lasketun B/M-luvun perusteella suuruusjärjestykseen, jonka jälkeen suurimpien B/M-lukujen yhtiöiden tertiili otetaan lisäkäsittelyyn. Toisessa vaiheessa korkean B/M-luvun yhtiöt järjestetään F-Scoren perusteella portfolioihin siten, että fundamenteiltaan vahvat (korkea F-Score) yhtiöt muodostavat ostoposition ja heikot yhtiöt (matala F-Score) myyntiposition. Kolmannessa vaiheessa muodostetuista portfolioista poistetaan M-Scoren perusteella epäluotettavien tuloslukujen yhtiöt. Viimeisessä vaiheessa portfolioiden osakkeet rajoitetaan Z-Scoren osoittamana alhaisen konkurssiriskin yhtiöihin. Portfoliot muodostetaan kesäkuun ensimmäisenä kaupankäyntipäivänä ja positiot pidetään muuttumattomana yhden vuoden ajan, jonka jälkeen edellä mainitut vaiheet toistetaan. Ensimmäiset portfoliot muodostetaan kesäkuussa 1999 ja viimeiset kesäkuussa 2016.

Tutkielman tulosten perusteella F-Scoreen perustuva sijoitusstrategia on tuottava S&P 500-indeksin osakkeille vuosien 1999 ja 2017 välillä. Epänormaalit tuotot kuitenkin rajoittuvat ostopositioon ja johtuvat suurilta osin pohjana olevien korkean B/M-luvun osakkeiden yleisesti hyvistä tuotoista. Korkean F-Scoren yhtiöiden voidaan myös todeta olevan kannattavampia alhaisen F-Scoren yhtiöihin verrattuna. Tulosten perusteella voidaan lisäksi todeta, että pitkän position riskikorjattuja tuottoja voidaan edelleen parantaa poistamalla M-Scoren avulla epäluotettavien tuottolukujen yhtiöt. Mainittujen yhtiöiden poistaminen myös alentaa lyhyen position tuottoja, mutta ei riittävästi saavuttaakseen tilastollisen merkitsevyyden. Portfolioiden osakkeiden rajaaminen alhaisen konkurssiriskin yhtiöihin ei puolestaan näytä muuttavan osto- ja myyntiportfolioiden riskikorjattuja tuottoja. Tulosten valossa fundamenttianalyysiin perustuva taloudellisen aseman ja tilinpäätöslukujen laadun arviointi perinteisen B/M-lukuun pohjautuvan arvosijoitusstrategian toteutuksessa on kuitenkin hyödyllistä myös silloin, kun strategian pohjana käytetään ainoastaan suuria yrityksiä, joskin F-Scoren erottelukyky näyttää olevan heikompi isoille yhtiöille.

AVAINSANAT: Arvosijoittaminen, Fundamenttianalyysi, Tulosmanipulaatio, F-Score, M-Score

Contents

Figures

Tables

1 Introduction

The Efficient Market Hypothesis (Fama 1970) states that investors cannot earn abnormal returns by buying undervalued assets or by selling overvalued assets as market prices already reflect all the available information. In other words, if new information arises, that information is immediately incorporated into market prices. Thus, fundamental analysis should be inefficient in return prediction according to the EMH. However, during the recent decades, it has become clear that financial markets do not function as efficiently as proposed by the EMH since a variety of different pricing inefficiencies or anomalies have been found to consistently violate the underlying assumptions of EMH.

Stock prices rarely reflect the company's actual fundamental value. Therefore, by buying (selling) stocks that have high (low) fundamental value and low (high) market price, investors can earn better average returns. Stock selection strategies that concentrate in finding assets which market prices are significantly lower than their intrinsic value, can be referred as value investing, which initially dates back to ideas of Benjamin Graham (1934). (Bodie Kane & Marcus 2014:655.)

Value investing has been a popular topic among researchers and also lays the foundation for Piotroski's (2000) study. He suggests that with fundamental analysis, it is possible find companies that have the best and worst future prospects. According to Piotroski (2000), high book-to-market companies tend to be fundamentally weaker in general, so the ability to find the best performers from a pool of poorly performing companies, can be especially useful for value investors.

Piotroski (2000) claims that by ranking high book-to-market companies with F-Score, an aggregate of nine binary (value of 1 or 0) performance signals indicating fundamental strength, significantly higher returns can be achieved. The nine F-Score signals can be separated into three main categories that measure the company's profitability, liquidity/leverage/source of funds and operating efficiency. According to Piotroski (2000), companies that have an aggregate F-score of 9-8 can be viewed as financially strong whereas companies with F-Score 0-2 can be viewed as financially weak. Moreover, he implements a strategy that takes a long (short) position to high (low) F-score companies. Piotroski (2000) argues that this strategy yields an average annual return of 23% on a market-adjusted basis.

In addition to financial strength evaluation, fundamental analysis can be applied to multiple other contexts such as earnings quality assessment. Beneish (1999) uses financial ratios to examine the financial characteristics of earnings manipulators. Based on the common factors between fraudulent companies, he computes a manipulation probability metric known as the M-Score. In a later study, Beneish, Lee and Nichols (2013) find that the M-Score has power also in return prediction as companies with high manipulation probability tend to earn significantly lower returns than their low M-Score counterparts. Moreover, earnings management has been found to be relatively common within public companies. For example, based on a conducted survey of 169 Chief Financial Officers, Dichev, Graham, Harvey and Rajgopal (2013:1) report that "*about 20% of firms manage earnings to mispresent economic performance*".

From a practical point of view, fundamental analysis is highly accessible for investors as it in most cases relies on very basic calculus and ratio analysis. Thus, fundamental analysis is relatively easy to implement also for individual investors that seek to make better investment decisions.

1.1 Purpose and Structure of the Study

The purpose of the study is to examine value investing in the framework of Piotroski (2000) as the aim is to determine, whether the F-Score based investing strategy can generate abnormal returns among blue-chip companies in the US stock market. Moreover, the study investigates whether the F-Score strategy can be enhanced with additional fundamental analysis using the Beneish M-Score and the Altman's Z-Score.

As mentioned in the introduction, earnings management is a common issue among public companies. Therefore, the assessment of earnings quality together with financial strength analysis could be very beneficial. Although the F-Score considers accrual-based earnings management, as per Sloan (1996), as one of the nine fundamental signals, M-Score could capture this aspect of fundamental strength better. Beneish (2013) suggests that the M-Score provides more information than accruals alone though the two seem to be positively correlated with each other. Beneish (2013) also reports that the M-score has significant power in return prediction.

Thus, this study analyzes if the F-Score's ability to separate winners from losers can be enhanced by excluding companies that have high earnings manipulation probability and hence, lower quality earnings figures. Additionally, the aim is to examine whether F-Score strategy returns increase if only low-distress companies are used. Piotroski (2000) argues that companies with low default risk tend to earn higher returns in general. Moreover, he suggests that F-Score tends to be especially powerful in detecting the worst performers among low distress risk companies and, thus, increasing the return spread between long and short F-Score portfolios.

The structure of the thesis is constructed as follows. After the introduction, the theoretical framework will be provided by first reviewing the prior literature around value investing and fundamental analysis. After the literature review, the most central ideas behind modern financial theory and asset pricing are presented in Chapter 3. The fourth chapter will present the Efficient Market Hypothesis and clarifies few of the well-known deviations from the hypothesized market efficiency.

Chapter five of the thesis presents the main concepts of fundamental analysis and asset valuation. Especially F-Score, M-Score and Z-Score are investigated in detail due to their key role in this thesis. The examination of the aforesaid fundamental metrics also ensures a smooth transition to the empirical part of the thesis. The empirical part consists of the description of data and methodology in chapter 6, followed by the main empirical analysis in chapter 7. Concluding remarks are provided in chapter 8, which is also the last section of this thesis.

1.2 Hypothesis

The research hypotheses in this thesis are structured to four different pairs of null and alternative hypotheses. The first two hypotheses sets consider the effectiveness of the F-Score trading strategy. Moreover, the hypotheses are constructed to reflect the abnormal returns estimated with the Fama-French (2015) Five-Factor model. That is, the first hypotheses pair considers the return difference between high and low F-Score portfolios:

H1,0: *High F-Score portfolios do not generate higher abnormal returns than low F-Score portfolios*
H1,1: *High F-Score portfolios do generate higher abnormal returns than low F-Score portfolios*

The second set of hypothesis reflects the effectiveness of the F-Score based strategy compared to a benchmark portfolio. In this study, the benchmark portfolio is considered to be the high book-to-market portfolio without F-, M- or Z-Score screening. The second set of hypotheses read as follows:

H2,0: *Abnormal returns of high (low) F-Score portfolios are not higher (lower) than the benchmark's*
H2,1: *Abnormal returns of high (low) F-Score portfolios are higher (lower) than the benchmark's*

The third hypotheses pair focuses on the M-Score's ability to increase the F-Score returns. If the exclusion of possible earnings manipulators has a positive impact on the F-Score strategy, it should increase the long portfolio returns and decrease the returns of a short portfolio. Thus, the third hypotheses are as follows:

H3,0: *M-Score screening does not increase (decrease) the abnormal returns of long (short) portfolios constructed with the F-Score*
H3,1: *M-Score screening does increase (decrease) the abnormal returns of long (short) portfolios constructed with the F-Score*

The last pair reflects an assumption that using financially less distressed companies based on the Z-Score, improves F-Score's ability to separate future winners from losers. The assumption is based on Piotroski (2000), who claims that F-Score has limited power among high-distress companies. The last hypotheses are presented below as:

H4,0: *Z-Score screening does not increase (decrease) the abnormal returns of long (short) portfolios constructed with the FM-Score*
H4,1: *Z-Score screening does increase (decrease) the abnormal returns of long (short) portfolios constructed with the FM-Score*

1.3 Contribution

The thesis aims to contribute to existing literature by shedding light on the possible usefulness of additional fundamental screening when implementing an F-Score based value strategy. That is, the study analyzes whether the M-Score as a proxy for earnings quality can be used as a complementary tool for an F-Score strategy.

Additionally, the thesis will extend prior literature regarding F-Score by providing recent results. The company-specific financial statement data covers years from 1997 to 2015 and the stock price series cover years between 1999 and 2017. The first portfolio formations are carried out in the first trading day of June 1999 and the positions are then held for one year. This one year buy and hold cycle is repeated until 2016 as the last holding period ends at the last trading day of May 2017.

2 Literature Review

This chapter discusses the previous studies that are in the scope of this thesis. That is, this chapter together with the introduction gives insight on how this thesis is positioned in relation to other researches.

The presented literature review is divided into two subchapters that clarify separately the aspects of value investing and fundamental analysis from an academic point of view. As per the scope of this thesis, the latter subchapter considers fundamental analysis studies that mainly revolve around the F-Score and its implications.

2.1 Value Effect Research

Generally, the value effect or value premium, means high book-to-market companies' historical tendency to outperform their low book-to-market counterparts. HBM companies are usually referred to as value stocks, while the latter are known as growth stocks. Although B/M-ratio[1] is probably the most used ratio when classifying value and growth stocks, it is not the only one as ratios such as E/M[2] or EBITDA/EV[3] are also often used. The value effect has been a popular topic in finance research due to its persistency over the past decades. For example, Statman (1980) and Rosenberg, Reid and Lanstein (1985) document pricing inefficiencies in the US equity market from 1960s to 1980s, as they find that high book-to-market ratio is positively associated with future stock returns.

Furthermore, consistent evidence is found by Fama and French (1992) as they argue that market beta (i.e. stock's sensitivity to market risk) alone is not able to capture changes

---

[1] *Book Value of Equity/Market Capitalization*
[2] *Earnings/Market Capitalization*
[3] *Earnings before interest, taxes, depreciation and amortization/Enterprise Value*

in stock returns, as should happen according to the Capita Asset Pricing Model[4] (Sharpe 1964, Litner 1965, Mossin 1966). Moreover, they find that the highest book-to-market decile portfolio yielded an average positive return of 1,63% per month over a sample period from July 1963 to December 1990. Oppositely, the lowest B/M decile portfolio returned only 0,64% on average. Additionally, stock returns tend to decrease as company size increases (Fama & French 1992:446,449-451) as also proposed earlier by Banz (1981). The unanimous findings regarding value premium in latter half of the 20th century led to the Fama-French (1993) three factor asset pricing model, which together with CAPM's market beta uses factors for value and firm size to capture the variation in stock returns.

The value effect has also been recognized globally outside the US markets. For example, Chan, Hamao and Lakonishok (1991) document positive value returns in the Japanese market. Moreover, Fama and French (2012) examine value jointly with size (Banz 1981) and momentum[5] (Jegadeesh & Titman 1993) premiums. In their study, Fama and French (2012) report significant value premiums in all investigated markets. That is; Europe, Japan, Asia-Pacific and North America. Consistent with earlier studies, the returns tend to be smaller for larger companies.

Further international evidence is also provided by Asness, Moskowitz and Pedersen (2013) who study value and momentum in different asset classes and regions. They find that value and momentum premiums exist both in equities but also in different asset classes such as currencies and commodities. They document that the value premiums are positively correlated across different asset classes and markets. However, value premiums tend to be negatively correlated with momentum returns. Moreover, Asness et al. (2013) suggest that combining value and momentum strategies increases the Sharpe-ratio compared to individual strategies. That is, the risk-adjusted performance. Benefits

---

[4] CAPM and Fama-French asset pricing models are examined in more detail in Chapter 3: *Return and Risk*
[5] Momentum refers to an anomaly, where stocks that have recently performed well (poorly), continue their good (poor) performance in the intermediate future (3-12 months).

of value-momentum combinations are also documented by Leivo (2012) in the Finnish equity market and by Grobys and Huhta-Halkola (2019) in the Nordic region.

Though value premiums by themselves are widely documented, the underlying drivers of these returns are not completely clear. As with many financial market anomalies, the explanations are divided into risk-based, and investor behavior-based explanations. For example, Fama and French (1992,1996) suggest that the superior performance of value stocks compared to growth stocks mirror the deteriorating fundamentals, such as leverage and distress risk of these companies. In other words, the increased returns reflect the risk that is also increased by the poorer prospects of these companies. Alternative explanation is given by Asness et al. (2013) who propose funding liquidity risk as a partial reason for the value premiums.

Others such as Chan and Lakonishok (2004), however, propose that the returns of value strategies are not due to higher fundamental risk, but investor behavior. That is, investors are over-optimistic about the growth potential of growth stocks, causing an undervaluation of high book-to-market companies. Moreover, the value premiums are subsequently realized when the mispricing is later corrected by the market.

The value-growth anomaly has also been investigated more recently by Piotroski and So (2012) in the US market. They report that the returns to value-growth strategies can be explained by errors in market expectations. That is, when the B/M ratio does not reflect the actual financial strength of the stock, proxied by F-Score. Walkshäusl (2017) confirms the findings of Piotroski and So (2012) in the European market, suggesting that the findings are not dependent on the analyzed region. Moreover, it is explained that the high (low) value (growth) returns tend to be driven by companies with strong (weak) underlying fundamental strength (Walkshäusl 2017:867-868).

2.2 Fundamental Analysis and F-Score Research

Like value investing, fundamental analysis research has been a popular topic among practitioners and academics. Fundamental analysis can be used to evaluate companies' financial strength metrics such as profitability or leverage. This information can be then used to evaluate the prospects of the company.

For example, Altman (1968) finds that financial statement information can be used to predict bankruptcy. Moreover, he introduces a model known as the Z-Score[6], which uses five different variables to assess whether a company is in a risk of becoming default. According to Altman (1968), the model predicted correctly up to 90% of the bankruptcies. An alternative well-known bankruptcy prediction model is the O-Score proposed by Ohlson (1980), who uses significantly higher number of observations in his study and suggests that O-Score has a better bankruptcy predictability than the Z-Score especially over an intermediate time horizon.

F-Score however, an aggregate nine binary financial strength signals, was created to distinguish between good and bad value stocks. Moreover, Piotroski (2000) finds that the return spread between good and bad companies increases if the portfolios are screened with Z-Score so that only companies with low distress risk are used. Since the original publication by Piotroski (2000), F-Score's ability to separate winner stocks from losers and its implications in different contexts have been widely analyzed due to its strong performance and relatively easy application. Fama and French (2006) report that the F-Score is capable of predicting future stock returns as it captures information about future profitability, which is positively associated with stock returns. Furthermore, they confirm that high accruals are negatively associated with future profitability and subsequent stock returns like originally proposed by Sloan (1996).

[6] The specific composition of F-, M- and Z-Scores are presented in Chapter 5: *Fundamental Analysis*

In his study, Sloan (1996) uses financial statement information of US companies over a sample period from 1962 to 1991 to investigate whether the level of accruals and cash as components of earnings predict stock performance. According to Sloan (1996), higher positive accruals lead to deterioration future profitability and stock returns. For this reason, Piotroski (2000) uses accruals as one of the nine fundamental variables in the F-Score as it is suggested that value companies may be more prone to earnings management through accruals. Earnings management is also studied by Beneish (1999), who introduces the M-Score to detect earnings manipulation. M-Score has been successfully used for example to detect the accounting fraud of Enron in the early 2000s. In a later study, Beneish (2013) suggests that the M-score provides more information about future stock returns than accruals alone, despite statistically significant positive correlation between the two.

In addition to a traditional F-Score strategy that uses high B/M companies as the stock base, F-Score has been successfully combined with other strategies. For example, Tikkanen and Äijö (2018) examine whether the F-Score can be used with other value strategies in the European equity market. Specifically, these are B/M, E/M, D/M, EBIT/EV, EBITDA and Novy-Marx strategies. They find that all high F-Score portfolios generated positive abnormal returns (alphas). It is also reported that low F-Score portfolios performed worse than the corresponding benchmark strategies, though statistically significant negative alphas are only found for B/M, E/M and D/M strategies. The highest (lowest) annual alpha of 7,44% (-10,50%) is generated by screening the EBITDA/EV (E/M) portfolio. Additionally, Tikkanen and Äijö (2018:503) find that the use of F-Score increases (decreases) the Sharpe and Sortino ratios of high (low) portfolios. Though the generated returns tend to decrease as company size increases, F-Score screening can be viewed profitable also for bigger stocks.

Another combination strategy study is made by Turtle and Wang (2017). They report that although high F-Score companies tend to outperform the low F-Score companies, the effect is even more significant when the portfolios are double-sorted with momentum.

That is, the long (short) portfolios include previous winners (losers) that also have strong (weak) fundamentals. Turtle and Wang (2017:135) report that on a raw return basis this long-short strategy generates a positive return of 5,2% per quarter. The authors point out that the evidence does not support the risk-based explanations for F-Score returns as proposed by Fama and French (2006). Moreover, Turtle and Wang (2017:138) suggest that the mispricing is more likely to be driven by investors’ underreaction to information especially during periods when the general market sentiment is high.

3 Return and Risk

In this chapter, the concepts of risk and return of an asset are clarified in order to provide the first set of theoretical ground for the thesis. In the first subchapter, the general ideas of return and risk are reviewed after which, the focus shifts to the most relevant and dominating asset pricing models in financial literature[7]: The Capital Asset Pricing model and the Fama-French factor models.

3.1 Return and Risk

Generally, risk and return of an asset are assumed to comove. In other words, securities that provide a high expected return also bear higher level of risk. Oppositely, securities with lower expected return are less exposed to different risk factors. The risk of a stock can be divided into two separate components: Systematic risk and non-systematic risk. The systematic component of risk affects all securities equally through changes in the financial markets such as changes in business cycles and changes in interest rates. Thus, systematic risk is often referred as market risk. Non-systematic risk, however, affects only specific securities. This firm-specific risk results from changes in a company's own operations or from changes in the industry the company is operating in. (Bodie, Kane & Marcus 2014:206.)

One key difference between the two components of risk is how they can be reduced with diversification. Markowitz (1952) proposes that by selecting a variety of securities from different industries into a portfolio, the total level of risk can be reduced without causing a proportional decrease in the portfolio's expected return. This technique, however,

[7] An alternative often used asset pricing model is the Carhart (1997) 4F model, which extends the Fama-French (1993) 3F model with a momentum factor. This model however, is limited outside the scope of this thesis, since the abnormal returns in the empirical part are estimated with the Fama-French 5F model (Fama & French 2015).

reduces only the firm-specific risk level, while the market component remains the same. Hence, the two risk components can be further referred as diversifiable- and non-diversifiable risk, respectively. The effect of diversification on the two risk components are illustrated in the following figure 1, where the vertical axis illustrates the portfolio risk in terms of volatility (standard deviation of returns) and horizontal axis demonstrates the number of stocks in the portfolio:

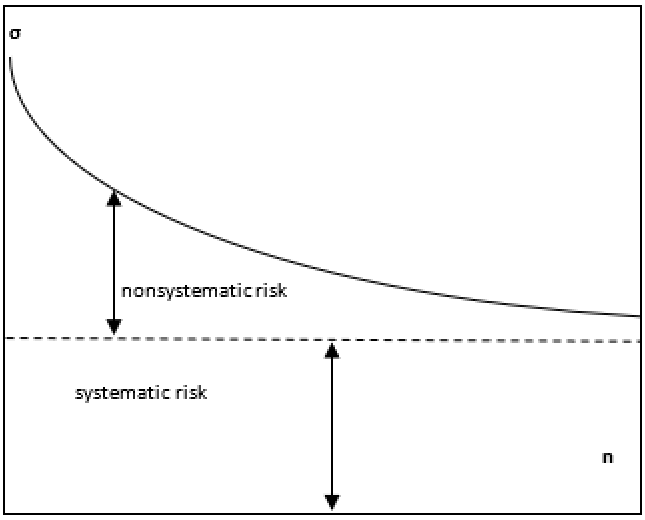

Figure 1. Components of risk and the power of diversification (Bodie et al. 2014:207)

As can be observed from the above figure, the level of systematic risk remains the same, while the level of non-systematic risk decreases as the number of stocks in the portfolio increases. It is also important to note the rate of risk reduction is not linear: at first, when new stocks are added to a portfolio, the risk exposure reduces rapidly, but the rate of decent slows down as more stocks are added into a portfolio. (Bodie et al 2014:207.)

3.2 The Capital Asset Pricing Model

Since the publications by Sharpe (1964), Litner (1965) and Mossin (1966) the Capital Asset Pricing Model (later CAPM) has been one of the most important pieces of modern asset pricing theory. The model considers the relationship between the expected return of an asset and its exposure to market risk. Moreover, the relationship between an asset's expected return and the market risk can be presented as in the following equation 1 (Bodie et al. 2014:298):

$$E(r_i) = r_f + \beta_i\,(E(r_m) - r_f), \tag{1}$$

where $E(r_i)$ is the expected return of stock *i,* r_f is the risk-free rate, $(E(r_m)- r_f)$ is the market risk premium and β_i is the market sensitivity of the stock *i*. Since financial markets are highly complex, CAPM possesses a set of underlying assumptions to simplify this complexity. Bodie et al. (2014:304) list these assumptions as:

1. Investors are rational mean variance optimizers
2. Investors have identical planning period
3. Investors have identical expectations
4. There are only publicly held and traded assets without short selling restrictions
5. Investors may lend or borrow at a risk-free rate
6. There are no transaction costs or taxes
7. All information is publicly available

As might be clear, such a strict set of assumptions does not mirror real financial markets that precisely as investors are not completely rational and nor are they identical. Moreover, trading of assets causes costs and short positions are not always possible. However, the model helps to understand the relationship between risk and expected return.

According to the model, the expected return of stock *i* depends on the risk-free rate and on the risk premium that increases simultaneously as the market risk of stock *i*, denoted as Beta, increases. The relationship between the expected return and Beta is further illustrated in the following figure 2:

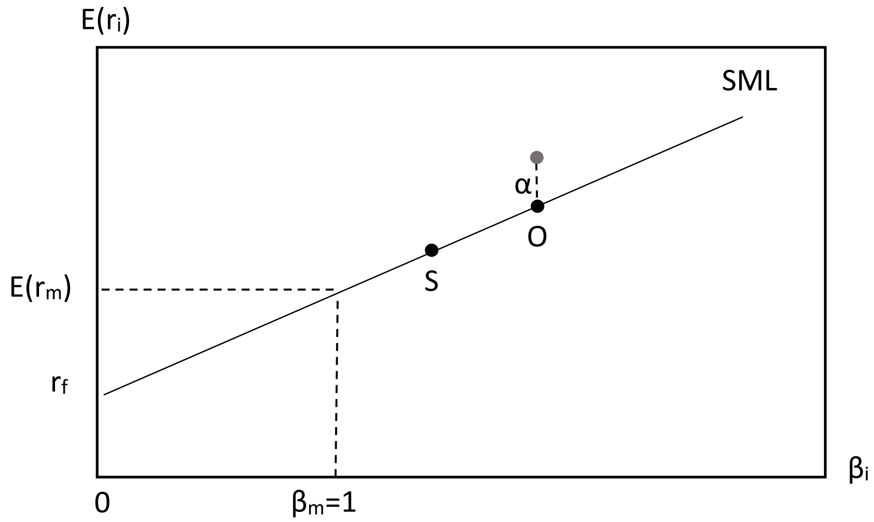

Figure 2 SML and a positive-alpha stock (Bodie et al. 2014:299)

In the figure, the X-axis presents the market risk, while the expected return is presented on the Y-axis. In market equilibrium, all stocks are on the same linear line, known as the Security Market Line (SML). In the graph, there are two stocks *S* and *O*, of which *S* is a stock with lower expected return and risk. Should the realized return of a stock *O* deviate positively (negatively) from the prediction of CAPM, it would mean that this underpriced (overpriced) stock is above (below) the SML. Moreover, the deviation or the abnormal return is denoted as Alpha.

3.3 The Fama-French Factor Models

Despite being widely used, CAPM has limited power to explain stock returns as it only considers market risk exposure. To tackle the restrictions of the Capital Asset Pricing Model, Eugene Fama and Kenneth French introduced the Fama French three-factor model (later FF3 model). Later, the FF3 model was enhanced by adding two new explanatory factors. Thus, the FF3 model is evolved to a five-factor model and later to a six-factor model. These models are presented next in the following subchapters.

3.3.1 Fama-French Three-Factor Model

According to Fama and French (1993) the FF3 model has significantly better ability to explain stock returns than CAPM. In addition to a market risk factor, they construct two additional risk factors: SMB and HML. SMB, or Small-minus-Big, takes into account the historically good performance of small companies compared to big companies. Similarly, HML considers the historically good performance of value companies (indicated by high Book-to-Market ratio) compared to growth companies (indicated by low Book-to-Market ratio). The FF3 asset pricing model is presented as follows:

$$\mathrm{E}(\mathrm{r}i) - \mathrm{r}f = \alpha i + bi[\mathrm{E}(rm) - rf] + si\mathrm{E}[SMB] + hi\mathrm{E}[HML], \tag{2}$$

where ri and $\mathrm{r}f$ are the return of stock i and the risk-free rate, respectively. The market factor *[E(rm)-rf]* is the expected return of a broad market portfolio less the risk-free rate. Size factor *E[SMB]* is the expected return difference between small and big companies, while the value factor *E[HML]* is the expected return difference between value and growth companies. Coefficients *bi*, *si* and *hi* describe the return sensitivity of stock *i* to market, size and value factors respectively. Lastly, *αi* is the intercept term, indicating the possible abnormal return that is not explained by the factors. (Bodie et al. 2014:426-428.)

3.3.2 Fama-French Five-Factor Model

To enhance the FF3 model, Fama and French (2015) add two new variables into the old model, creating a new model known as Fama French Five-Factor model (Later FF5). The new variables are called profitability factor RMW, and investment factor CMA. Furthermore, RMW describes the return difference between companies that have robust profitability and weak profitability, whereas the CMA is the return difference between conservative and aggressive investment firms (Fama & French 2015:3). The FF5 asset pricing model can be presented as:

$$E(ri) - rf = \alpha i + bi[E(rm) - rf] + siE[SMB] + hiE[HML] + \text{riE[RMW]} + \text{piE[CMA]}. \quad (3)$$

According to Fama and French (2015) the new FF5 model explains returns better than the original FF3 model. However, adding the two new variables, the model can explain also return changes that were previously captured by the FF3 model's HML factor. Thus, a four-factor model that excludes the HML factor, has similar power compared to the FF5 model. In other words, the FF3 model's HML factor becomes redundant as the two new variables RMW and CMA are added (Fama & French 2015,2017).

3.3.3 Fama-French Six-Factor Model

The most recent model asset pricing model introduced by Fama and French (2018) is the Fama-French Six-Factor model. The model uses the same five factors as the FF5 model; Mkt, SMB, HML, RMW and CMA, but adds a momentum factor UMD, which stands for *up minus down*. That is, the factor considers the performance difference of portfolios constructed on recent winners (upward performance) minus recent losers (downward performance).

In their study, Fama and French (2018) examine multiple different factor model combinations in order to shed light to the relationship between the individual factors and models constructed on them. In the case of the FF6 model, they find that in a general sense, by adding the momentum factor to the FF5 model, the power of the model increases. However, it is pointed out that the results change when different factor compositions are used. For example, the best six-factor model uses the market (Mkt) and size (SMB) factors accompanied with value (HML), cash profitability (RMWc), investment (CMA) and momentum factor (UMD). Moreover, the latter four factors are constructed on small stock return spreads. However, the authors argue that factors which use both big and small stocks, could work equally well though the results differ in their study. (Fama & French 2018:235,238,247-248).

In addition to being a widely used estimation model for abnormal returns, the Fama-French factor models give valuable insight about the risk exposure characteristics of the investigated portfolio. For example, portfolio returns' positive and significant loading on the market and the SMB factor would indicate that the portfolio returns co-move with the market returns and the long-side of SMB. That is, small stocks. Moreover, a negative and significant loading on the RMW factor would indicate co-movement with the returns of companies that have low profitability. Lastly, positive loading on the CMA factor would indicate that the companies in the analyzed portfolio are conservative investors that in other words, have low asset growth.

4 Financial Market Efficiency

One of the central ideas in financial theory is that financial markets are assumed to functions in an efficient manner. That is, the first part of this chapter clarifies the aspects of the efficient market hypothesis proposed by Fama (1970).

However, since financial markets do not always function efficiently for example due to irrational investor behavior, the chapter also presents some of the most documented financial market anomalies and the possible reasons for these inefficiencies. This examination helps to understand the difference and relationship between theory and actual financial markets.

4.1 The Efficient Market Hypothesis

The Efficient Market Hypothesis (hereafter EMH) was introduced by Fama (1970) and its fundamental assumption is that all relevant and available information is incorporated immediately into stock prices. Thus, investors cannot make abnormal returns by buying (selling) undervalued (overvalued) securities. EMH is can be connected to the concept of Random Walk, which describes stock prices' tendency to randomly vary over time. That is, as new information comes to financial markets unpredictably and this information is immediately incorporated to equity prices, the subsequent price movements are also unpredictable and *random* (Malkiel 2003).

According to Fama (1970:383), there are three different levels of market efficiency: weak form, semi-strong form and strong form efficiency. The three levels of efficiency are stacked in a way that the second level of efficiency cannot be fulfilled if the first level is not also fulfilled. Thus, in a case of strong form efficiency, weak-form and semi-strong form efficiency also exists. The relationship between the different levels of efficiency and their connection to information availability is illustrated in the next figure:

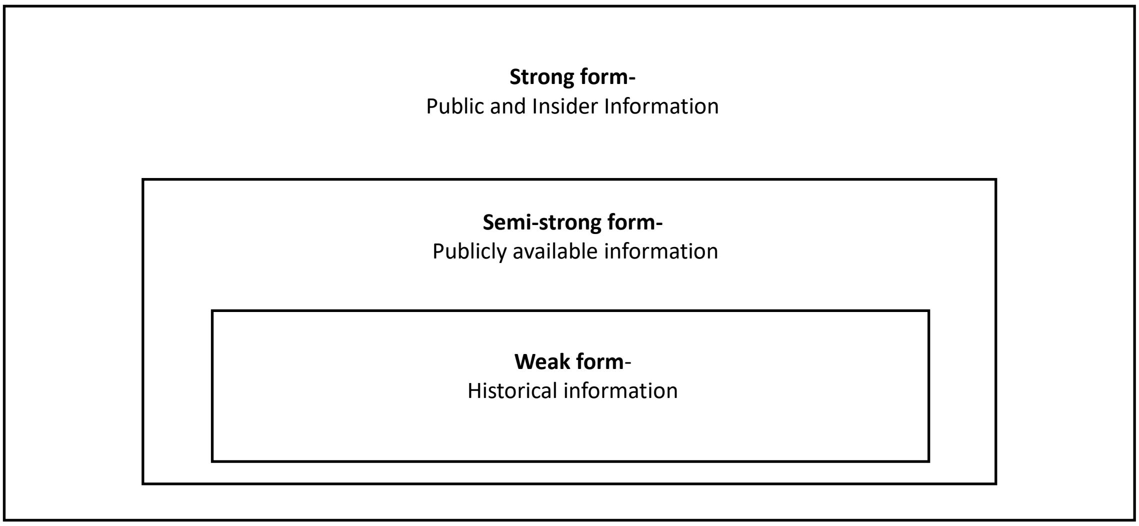

Figure 3 The Levels of Market Efficiency (Nikkinen et al. 2002:84)

In case of weak-form efficiency, market prices should reflect historical information (Fama 1970). This means that investors cannot use technical analysis to achieve higher returns. In other words, past market data, such as trading volume and historical price information cannot be exploited for superior returns. According to Fama (1970) semi-strong market efficiency means that market prices reflect not just historical information, but also all publicly available information. Thus, investors should not be able to earn higher returns by analyzing, for example, financial statements. This assumption suggests that fundamental analysis is ineffective in return prediction, as all the reported information is immediately mirrored in the assets' market prices.

The last and the strictest level of market efficiency is the strong form efficiency, which means that stock prices reflect historical information, publicly available information as well as insider information. For example, a board member of a company could use privileged information of his or her company to achieve superior returns. In the case of strong form efficiency, this information would also be reflected in the market prices. However, Fama (1970) suggests that the strong form efficiency is so strict that it does not reflect real financial markets very well.

Should the EMH hold, investors should not earn above average returns on the long run. However, multiple different financial market anomalies have been found to consistently violate the theoretical market efficiency, proposed by EMH. The next subchapter presents few of the most documented deviations from the market efficiency.

4.2 Deviations from Market Efficiency

Broadly speaking, financial market anomalies can be divided into three different categories based on their nature of occurrence. Fundamental anomalies are based on discrepancies between reported information and market prices. The value effect can be viewed as an example of a fundamental anomaly. Technical anomalies on the other hand, arise from past market data, such as historical price movements. Lastly, calendar anomalies refer to inefficiencies that occur during specific months or days. (Pompian 2011:15-16.)

When considering technical anomalies, the most prominent inefficiency is the momentum effect, which is also probably the most persistent and documented anomaly overall. Specifically, momentum refers to stocks' tendency to continue their past recent performance. In other words, stocks with the best (poorest) recent performance tend to continue their good (poor) performance up to a year. The momentum effect was first documented by Jegadeesh and Titman (1993), though a similar effect has been documented earlier by Levy (1967).

Specifically, Jegadeesh and Titman (1993) investigate US companies over a sample period from 1965 to 1989. In total, they set up 16 different strategies that have quarterly varying holding- and preceding performance measurement periods. That is, the analyzed periods are 3, 6, 9 and 12 months. In general, the implemented strategy can be referred to as J-month/K-month strategy, which means that companies are assigned into decile portfolios according to their performance over the preceding J-months. Companies that have performed the best are assigned to a long portfolio, whereas the worst performers are assigned to a short portfolio. These positions are the held for the next K-months.

Jegadeesh and Titman (1993:69) report that the highest returns are generated by a 12/3-strategy that ranks the stocks based on their performance over the past 12 months and then holds the positions for three months. Specifically, the strategy yields 1,31% per month. Furthermore, the returns increase to 1,49% per month if a week is skipped before portfolio formation and the holding period. It is reported that the momentum strategies perform well within intermediate horizons, but the returns tend to decrease after a year. After the publication of Jegadeesh and Titman (1993), the momentum effect has been widely analyzed, and it has been found to be persistent regardless the markets (Fama & French 2012) and asset classes (Asness et al. 2013).

Due to its persistency, the possible explanations for momentum returns have been also widely analyzed along with the different implications. Though the mechanism of the returns has not been conclusively reported, most studies suggest that the returns are due to irrational investor behavior. For example, Jegadeesh and Titman (1993) themselves suggest that the returns are due to investors' over- and underreaction to information. Moreover, Daniel, Hirshleifer and Subrahmanyam (1998) suggest that investors are unable to assess information correctly and they tend to be overconfident, which can cause short term overreactions to information. Thus, driving the momentum returns.

When it comes to calendar anomalies, one of the most persistent is the January Effect, which refers stocks' tendency to generate significantly higher returns in January than during other months (Pompian 2011:16). As with many anomalies the effect tends to be especially strong with smaller stocks. The link between January effect and size-effect is also documented by Keim (1983) as he discovers that small stocks tend to outperform bigger ones especially in the beginning of January. It is often proposed that January effect is driven by investors' tendency to sell poorly performed stocks at the end of each year to claim tax benefits (Martikainen 1990:121) . Alternatively, Ritter and Chopra (1989) suggest that institutional investors rebalance their portfolios in December by selling riskier assets to present less risky assets in their balance sheets. Subsequently, these stocks rise in value during January due to repurchases. The magnitude of the January effect also seems to be a good performance proxy for the rest of the year, as companies which have the highest returns in January tend to overperform those that have the lowest returns in January (Cooper, McConnell & Ovtchinnikov 2006). That is, the inefficient performance from February to December can be referred as the *other* January Effect.

This chapter together with the preceding chapters have clarified both the theoretical framework of financial markets as well as known deviations from these assumptions. Though the presentation is not conclusive, it is important to note that the discussion between theory and the real world is never-ending. For the reader, the key is to understand that when a new anomaly is documented, it causes changes and improvements to the theoretical models accordingly. As investors and assets are not identical, it is impossible to create a model that captures every aspect of the financial markets accurately. However, the theoretical models and their development reflect what is known so far. In the following chapter, the focus shifts from theoretical framework to hands-on fundamental analysis, followed by the empirical part of the paper.

5 Fundamental Analysis

In general, fundamental analysis can be viewed as a method where the intrinsic value of a stock or another equivalent asset is determined based on the company's *fundamentals.* In other words, the value of a stock is derived from reported accounting information such as earnings, which can be used for future forecasts. For example, a fundamental analyst can use income statement ratios to determine the profitability of a company or calculate metrics using a company's balance sheet to evaluate if the company has a healthy amount of debt or not. That information is then used when evaluating the prospects of that company. (Bodie et al. 2014:356.)

As stated in chapter 4, investors should not be able to earn abnormal returns using fundamental analysis, as it violates the assumptions of the semi-strong market efficiency. However, the usefulness of fundamental analysis in return prediction and in various more specific functions has been documented by multiple researchers over the past decades. This chapter presents the three most essential fundamental analysis measures regarding this study: these being the Piotroski F-Score, the Beneish M-Score and the Altman's Z-Score. At first, however, few basic stock valuation models are presented as these models are inextricably linked to the concept of fundamental analysis. Moreover, it helps to clarify the difference between a stock's intrinsic value and its market value.

5.1 Stock Valuation

This section presents three different stock valuation methods, namely: dividend discount model, discounted cash flow model and relative valuation. As can be observed from the models' names, the approach in the first two absolute valuation models is to *discount* future cash flows to present value.

5.1.1 Discounted Cash Flow Models

The dividend discount model (DDM) assumes that the current value of a stock is the sum of future dividends into perpetuity. Hence, DDM can be presented in an equation as follows (Bodie et al. 2014:596):

$$V_0 = \frac{D_1}{(1+k)^1} + \frac{D_2}{(1+k)^2} + \frac{D_3}{(1+k)^3} + \cdots + \frac{D_t}{(1+k)^t} \tag{4}$$

where *V* is the value of the stock at time 0. *D* is the expected dividend of the stock at time *t* and *k* is the required rate of return. Since DDM requires dividends to be forecasted for every year into the future, it can be simplified by adding an expected dividend growth rate into the equation. This variation of the model is known as the Gordon Model or the constant growth DDM. The equation for the Gordon model is presented below as (Bodie et al. 2014:597):

$$V_0 = \frac{D_1}{(k-g)} \tag{5}$$

where *g* is the expected growth rate of the dividend. In the model, it is expected that the dividend has a steady growth rate. The DDMs imply that the value (V) of the stock in question increases, if the dividends (D) increase over time, the required rate of return (k) of the stock is lowered or, if the expected dividend growth rate (g) increases. However, in order to the DDM work properly, it assumes that *k* is always higher than *g*.

An alternative approach to DDMs is the Discounted Free Cash Flow Model (hereafter referred as DFCM). It can be used to value any company, but it is particularly useful for example in cases where dividends are not paid, thus making DDMs inapplicable. Additionally, DFCM can provide more information compared to DDMs it uses data beyond dividends. Where DDMs consider dividends as the sole form of cash flow, DFCM considers cash flow as the amount of after-tax capital that is generated by the company's operations and left for equity holders after capital expenditures (CAPEX) and depreciation. According to DFCM, the value of a firm can be obtained as (Bodie et al. 2014:618):

$$P_0 = \sum_{t=1}^{\infty} \frac{FCF_t}{(1 + WACC)^t} \tag{6}$$

where *P* is the value of the company at time 0, *FCF* is the free cash flow and *WACC* is the weighted average cost of capital. The free cash flow seen in the numerator and the weighted average cost of capital seen as the denominator of the equation above, can be calculated as:

$$FCF = EBIT\,(1 - t_c) + \text{Depriciation} - CAPEX - \text{Increase in net WC} \tag{7}$$

where, *EBIT* is the earnings before interest and taxes, *t* is the corporate tax rate, *WC* is working capital. Moreover, WACC is computed as:

$$WACC = r_D \frac{D}{V} + r_E \frac{E}{V} \tag{8}$$

where the lower-case *r*s indicate the cost of debt and cost of equity respectively, D/V is the share of debt and E/V the share of equity in the company. When the company value P is obtained using the equations, the value of individual stock is simply P divided by the number of shares outstanding. As may be clear from the formulas, the discounted cash flow models rely heavily on different estimations and forecasts. In other words, the calculated value of the stock may increase significantly if the investor performing the analysis, for example, overestimates the dividend growth rate in the Gordon model. This problem is not present in the relative valuation models, which are presented next.

5.1.2 Relative Valuation

The main idea in the relative valuation is to compare a stock's different price multiples to its peers or to the industry average. Most commonly used price multiples are Price-to-Earnings (P/E), Price-to-Book (P/B)[8] and Price-to-Sales (P/S) ratios. In each ratio, price refers to the current market price of the stock, whereas the denominators refer to earnings, book value and sales per share, respectively. As each multiple has the current market price of the stock as the numerator, it gives insight to how the market is valuing the stock. Thus, the multiples are free from investor-specific assumptions that are present in the absolute valuation models.

The P/E ratio mirrors the market's beliefs of the stock's growth prospects. The ratio increases as the current market prices increases in relation to its current earnings, which implies that the market is expecting earnings growth in the future. Usually, riskier companies have lower P/E ratios, mirroring the lower growth opportunities (Bodie et al. 2014:612). The P/S ratio has similar implications as the P/E ratio as it mirrors the expected sales growth. P/S ratio can be used as an alternative for the P/E ratio for companies that do not have earnings, such as start-up companies.

Though the P/E ratio can be used to differentiate value companies from growth companies as growth companies have higher P/E ratios, the more often used multiple for this is the P/B ratio. The ratio divides the current market price per share with the book value per share. Lower P/B ratio could indicate that the market is undervaluing the company. On the other hand, a low market price and thus lower P/B, could imply that something is wrong with the company. However, it is usually considered that if a company has a P/B less than 1, the stock is a potential investment.

[8] In finance literature, Book-to-Market ratio is often used to distinguish between value and growth stocks. B/M ratio is the reverse of the P/B as it compares the book value of equity to the stocks market capitalization. Thus, higher B/M would indicate higher value whereas, low B/M stocks are considered as growth stocks. B/M ratio is also used in this thesis.

Though the different price multiples can provide useful information about the stock's valuation, the multiples should be used with caution as the average ratios vary significantly between industries. Usually analysts use a combination of different multiples and mainly use them to evaluate companies to their peers within the same industry. Next, the focus in this chapter shifts to more detailed examination of value signals, earnings quality and financial strength.

5.2 Piotroski F-Score

Value investors' main objective is to detect mispriced companies, which market price is lower than their intrinsic value. Joseph Piotroski created the F-Score as an extension to the basic Book-to-Market trading strategy in 2000. In general, the F-Score is a screening tool for a value investor: As the B/M strategy relies solely on the B/M ratio when ranking the stocks, the F-Score aims to detect the best value stocks from the group of high BM companies. Thus, F-Score can be classified as quality, rather than value detection tool.

Specifically, the F-Score is a combination of nine different signals, each worth one or zero depending on whether the respective criterion is met or not. Hence, the aggregate F-Score varies between zero and nine: Companies with F-Score of 9-8 are considered high value and financially healthy, whereas companies with an aggregate F-Score of 0-2 are considered low value and financially unhealthy. (Piotroski 2000). The complete F-Score can be presented as:

$$\text{F-Score} = ROA + \Delta ROA + CFO + ACCRUAL + \Delta LEVER + EQOFFER + \Delta LIQUID + \Delta MARGIN + \Delta TURN \quad (9)$$

Piotroski (2000) divides the nine different criteria into three different categories, which measure the company's: (i) profitability, (ii) leverage/liquidity/source of funds as well as (iii) operating efficiency. Profitability variables in the F-Score measure the company's ability to generate funds through its internal operations. Higher ability to generate earnings would subsequently lead to higher returns in the future. Variables ROA, ΔROA and CFO measure this ability to generate funds, while the fourth profitability variable ACCRUAL signals if the company has managed its earnings through positive accruals. According to Sloan (1996), accrual-based earnings inflation predicts poor future performance. The four profitability variables are presented by equations 10-13. (In the equations, *t* denotes value at the latest fiscal year end):

$$\mathrm{ROA} = \frac{\mathrm{Net\ income\ before\ extraordinary\ items_t}}{\mathrm{Total\ assets_t}}, equals\ 1\ if\ ROA > 0, \tag{10}$$

otherwise 0

$$\Delta\mathrm{ROA} = \mathrm{ROA_t} - \mathrm{ROA_{t-1}}, equals\ 1\ if\ \Delta ROA > 0\ , otherwise\ 0 \tag{11}$$

$$\mathrm{CFO} = \frac{\mathrm{Cash\ flow\ from\ operations_t}}{\mathrm{Total\ assets_t}}, equals\ 1\ if\ CFO > 0, otherwise\ 0 \tag{12}$$

$$\mathrm{ACCRUAL} = \mathrm{CFO_t} - \mathrm{ROA_t}, equals\ 1\ if\ CFO_t > ROA_t, otherwise\ 0 \tag{13}$$

Company's leverage and liquidity are measured through three different variables: ΔLEVER, EQOFFER and ΔLIQUID. The first variable ΔLEVER measures the change in the company's long-term debt level. According to Piotroski (2000:8), increase in long-term debt may indicate higher financial risk. Thus, decrease in long-term debt level is considered as a good signal. Similarly, if the company has issued new shares (EQOFFER), this may indicate that the company cannot generate required funds internally and thus is required to rely on external funding. Therefore, if the company did not issue new equity, EQOFFER has a value of 1. Variable ΔLIQUID measures the company's ability meet its current debt obligations through the change in the company's current ratio. The three variables are presented as follows:

$$\Delta\mathrm{LEVER} = \frac{\mathrm{LTD_t} - \mathrm{LTD_{t-1}}}{\mathrm{Average\ total\ assets}}, \mathrm{equals\ 1\ if\ \Delta LEVER < 0, otherwise\ 0} \tag{14}$$

$$\mathrm{EQOFFER} = \mathrm{Shares\ outstand}_{t} - \mathrm{Shares\ outstand}_{t-1},\ \text{equals 1 if EQOFFER} = 0\text{, otherwise 0} \tag{15}$$

$$\Delta\mathrm{LIQUID} = \mathrm{Current\ ratio}_{t} - \mathrm{Current\ ratio}_{t-1}, \text{equals 1 if } \Delta\mathrm{LIQUID} > 0\text{, otherwise 0} \tag{16}$$

The last two components of the F-Score, ΔMARGIN and ΔTURN measure the company's operating efficiency. ΔMARGIN measures the change in the company's gross-margin. Positive change in gross-margin can be viewed as a good signal, as it indicates that the company has been able to reduce the direct costs (COGS) of its sales. Alternatively, improved gross-margin can be due to increased price of the company's products, which increases revenue. ΔTURN measures the change in the company's asset turnover. Piotroski (2000:9) clarifies that improved asset turnover is considered as a good signal as it implies that the company is generating more (same) revenue with the same (fewer) amount of assets. The two variables can be presented as:

$$\Delta\mathrm{MARGIN} = \mathrm{Gross\ marg\%}_{t} - \mathrm{Gross\ marg\%}_{t-1}, \text{equals 1 if } \Delta\mathrm{MARGIN} > 0\text{, otherwise 0} \tag{17}$$

$$\Delta\mathrm{TURN} = \frac{\mathrm{Total\ Sales}_{t}}{\mathrm{Total\ assets}_{t}} - \frac{\mathrm{Total\ Sales}_{t-1}}{\mathrm{Total\ assets}_{t-1}}\text{, equals 1 if } \Delta\mathrm{TURN} > 0\text{, otherwise 0} \tag{18}$$

In his study, Piotroski (2000) constructs the strategy so that companies are first ranked into quantiles based on their book-to-market ratios at the fiscal year-end. In the second step, the top quantile (the highest BM-ratios) of companies is re-ranked into portfolios with the F-Score. Finally, long position is taken to the portfolio that includes the high F-Score companies (F-Score = 8,9), while the portfolio that includes low F-Score companies

(F-Score=0,1) is shorted. Piotroski (2000) clarifies that this strategy yields 23% annually during the sample period from 1976 to 1996.

5.3 Beneish M-Score

Unlike the F-Score, M-Score is not a value-investing tool per se. The M-Score was introduced by Messod Beneish in 1999 and it is used to detect companies that have high probability to become earnings manipulators in the future. Moreover, the score consists eight different variables calculated from the company's financial statements. In the original study, Beneish (1999) examines companies that have been flagged as accounting manipulators by US authorities to find out the common factors between those companies. In total, the sample consists of 74 companies flagged as manipulators. Beneish (1999:25) clarifies that companies, which had manipulated their earnings, had significantly higher sales growth. Specifically, there is a 25-percentage point difference in growth medians as the median for manipulators is 34,4% compared to non-manipulators' 9,4%. Companies that had manipulated their earnings were also smaller in terms of total assets and/or sales, less profitable and had more debt (Beneish 1999:25). Specifically, the Beneish model is calculated as:

$$\text{M-Score} = -4.84 + 0.92 \times DSRI + 0.528 \times GMI + 0.404 \times AQI + 0.892 \times SGI + 0.115 \times DEPI - 0.172 \times SGAI + 4.679 \times TATA - 0.327 \times LVGI. \quad (19)$$

According to Beneish (1999:26), seven of the eight variables used in the model are constructed as indexes, which improves the model's ability to detect abnormalities between consecutive years. The first variable in the model is Days' sales in receivables index (DSRI), which measures the change in receivables to sales ratios between years t and t-1. According to Beneish (1999:26), significant and abnormal increase in DSRI could indicate that the company has overstated its revenues. Specifically, DSRI is calculated as:

$$DSRI = (\text{Net Receivables}_t / \text{Sales}_t) / (\text{Net Receivables}_{t-1} / \text{Sales}_{t-1}). \tag{20}$$

Next variable in the model is gross margin index (GMI), which indicates the change in the company's profitability between year t and the prior year. If the gross margin of a company has deteriorated from the previous year, GMI will have a value of >1. Furthermore, decreasing margin is expected to increase the possibility of earnings manipulation as it would indicate poorer future prospects (Beneish 1999:26.). GMI can be presented as:

$$GMI = [(\text{Sales}_{t-1} - \text{COGS}_{t-1}) / \text{Sales}_{t-1}] / [(\text{Sales}_t - \text{COGS}_t) / \text{Sales}_t]. \tag{21}$$

The quality of the company's assets is measured with asset quality index (AQI). According to Beneish (1999:26), *"Asset quality in a given year is the ratio of noncurrent assets other than PP&E to total assets and it measures the proportion of total assets for which future benefits are less certain".* If the value of AQI is > 1, it might indicate that the company has increased its cost deferral, meaning that occurred costs are booked as assets and will be expensed on a later period. A decreased asset quality and increase in asset realization risk can increase the company's probability to engage in earnings manipulation. The equation for AQI is as follows:

$$\begin{aligned} AQI = {} & [1 - (\text{Current Assets}_t + \text{PP\&E}_t + \text{Securities}_t) / \text{Total Assets}_t] / \\ & [1 - ((\text{Current Assets}_{t-1} + \text{PP\&E}_{t-1} + \text{Securities}_{t-1}) / \text{Total Assets}_{t-1})]. \end{aligned} \tag{22}$$

The fourth variable (SGI) measures the sales growth between the given year and the preceding one. Beneish (1999:27) points out that due to the characteristics of growth companies, they are more likely to engage earnings manipulation compared to their non-growth counterparts. Growth companies have incentive to keep growing and meet the earnings expectations as their stock price could be negatively affected, if investors were to think that the growth has slowed down. Thus, high sales growth would indicate higher likelihood of earnings manipulation. The equation for SGI is presented as:

$$SGI = Sales_t / Sales_{t-1.} \quad (23)$$

DEPI, or the depreciation index, measures the change in depreciation rates. If the depreciation rate is smaller than in the year before, this change might imply that the company has artificially increased its earnings. In other words, if the useful life of an asset has been increased, the corresponding expense (depreciation) will decrease. Thus, inflating the company's earnings (Beneish 1999:27.) DEPI is calculated as:

$$DEPI = (Depreciation_{t-1} / (PP\&E_{t-1} + Depreciation_{t-1})) / (Depreciation_t / (PP\&E_t + Depreciation_t)). \quad (24)$$

According to Lev and Thiagarajan (1993:196), abnormal increase in sales, general and administrative (SGA) expenses in relation to sales can be interpreted as a negative signal as it might be a result of inadequate cost management. Moreover, since the amount of such cost are usually relatively fixed, highly increased SGA costs in relation to sales can also indicate that the company's management seeks to increase sales aggressively (e.g. with marketing). Therefore, Beneish (1999:28) assumes that an increase in SGA to sales ratio would increase the probability of earnings management. Specifically, the sales, general and administrative expenses index is computed as:

$$SGAI = (SG\&A\ Expense_t / Sales_t) / (SG\&A\ Expense_{t-1} / Sales_{t-1}) \quad (25)$$

The last index formulated variable in the model is leverage index LVGI, which measures the changes in the company's debt to assets ratio. Beneish (1999:28) clarifies that more leveraged companies have higher risk to meet the demands of their debt covenants. Thus, increased leverage would indicate higher likelihood to manipulate. The equation for LVGI is further illustrated in the equation 26:

$$\text{LVGI} = [(\text{Current Liabilities}_t + \text{Total Long Term Debt}_t) / \text{Total Assets}_t] / [(\text{Current Liabilities}_{t-1} + \text{Total Long Term Debt}_{t-1}) / \text{Total Assets}_{t-1}] \quad (26)$$

The only non-index variable in the model is TATA that is used to capture earnings manipulation through positive accruals. In other words, the variable measures if accounting earnings are driven by actual cash profits or not. If earnings are not supported by cash, this will implicate that the company is implementing "creative" accounting choices (Beneish 1999:28). The formula for TATA is as follows [9]:

$$\text{TATA} = [\text{Current Assets-Cash})\text{-}(\text{Current Liabilities-Current Maturities of LTD -Income Tax Payable})\text{-Depreciation and Amortization}] / \text{Total Assets} \quad (27)$$

Beneish (1999:28-30) finds that five of the eight variables result statistically significant coefficients: DSRI, GMI, AQI, SGI and TATA. Therefore, it can be stated the likelihood of earnings manipulation increases, if the company has disproportionate increases in receivables, sales and accruals. Moreover, the probability of manipulation also increases if the asset quality or gross-margin of a company has decreased. However, Beneish (1999:30) points out that the variables regarding leverage, depreciation and SGA expenses do not show statistical significance as they might be related to earnings management, which cannot be classified as manipulation. For example, changes in depreciation rate can also be viewed as normal, non-fraudulent, managerial behavior.

It is important to recognize that since the M-score is a probabilistic model, it cannot detect manipulators with a 100% accuracy. The accuracy to which the model can identify manipulation is a tradeoff between two types of errors: Types 1 and 2. Type 1 error means that the company is classified as non-manipulator although it in reality is manipulating its earnings. Oppositely, type 2 error means that the company is not manipulating

[9] For clarity, the equation for TATA is now presented as in Beneish (1999:27). However, later on this thesis TATA is calculated as: [*TATA=(Net income before extraordinary items-CF from operations)/Total assets*] as proposed by Beneish et al. (2013:76).

its earnings but is classified as a manipulator by the model. From an investor's point of view, the cost of Type 1 error is much higher than for Type 2, since by wrongly classifying a manipulator as non-manipulator, an investor could suffer significant losses if he or she has a long position on such stock. Type 2 error is less costly since investors have many other alternative stocks to choose from. In other words, "missing" an investment opportunity has little impact for an investor compared to a situation, where that opportunity is used to invest in a fraudulent company. Thus, it is argued that from investors' point of view a cutoff value of -1,78 (manipulator if, M-score > -1,78) gives good results, when classifying companies. (Beneish 1999:30-33.)

In their study, Beneish et al. (2013) investigate whether the M-score has power in return prediction instead of fraud detection. They find that companies that which have been flagged as manipulator earn significantly lower returns. Specifically, the average one year ahead returns for flagged companies are -7,5% whereas non-manipulators earned 9,9pps more at 2,4%. They also document that during the sample period from 1993 to 2010, a book to market strategy generated an average annual return of 8,0%. However, when the B/M strategy was enhanced with the M-Score, the strategy generated an annual return of 13,7%. Moreover, Beneish et al. (2013:63-65) claims that returns generated by the M-score are driven by accruals, though the two are significantly correlated. As presented in section 5.2, accruals are used in the F-Score to detect possible earnings management. The findings of Beneish et al. (2013) suggest that M-score may have superior power compared F-score's ACCRUAL variable to predict future returns.

5.4 Altman's Z-Score

Originally created by Edward I. Altman in the late 1960s, the Altman Z-Score is a fundamental measure, which is used to evaluate the level of distress or bankruptcy risk of a company. Z-Score relies on five different ratios and their multiples to detect if the company is heading towards bankruptcy or not. Specifically, the score is computed as:

$$\text{Z-Score} = 1.2X1 + 1.4X2 + 3.3X3 + 0.6X4 + 0.99X5, \tag{28}$$

where

X1= working capital scaled by total assets

X2 = retained earnings scaled by total assets

X3 = EBIT scaled by total assets

X4 = market value of equity scaled by book value of total liabilities

X5 = sales scaled by total assets

According to Altman (1968:595), X1, the first of the five ratios, seems to be the best signal to evaluate whether the company in question has liquidity problems as it captures the possible deterioration of liquid assets through working capital, which is calculated as current assets less current liabilities.

The second ratio (X2) is constructed to clarify the profitability of a company over time. Should the company under investigation be relatively young, its retained earnings are small as there has not been time to cumulate earnings like an older company would have. Thus, new and young companies are more prone to face financial distress than older ones. Similar to X2, the third ratio X3, can be also viewed as a profitability signal as it clarifies how the company's assets are generating profit. (Altman 1968:595).

The fourth ratio (X4) in the equations introduces the market value aspect of the equation. According to Altman (1968:595), the ratio measures the level of insolvency risk as it tells: *"how much the firm's assets can decline in value before the liabilities exceed the assets and the firm becomes insolvent"*. Lastly, the final ratio X5 is used to capture how the company's assets generate sales. Generally, a Z-score higher than 2,99 is considered as a "green-zone", indicating a little risk of default. Oppositely, a Z-score less than 1,81 would indicate a high probability of default in the future. Moreover, Z-score between the two values indicates increased, but not immediate risk, thus it can be referred as the "grey zone".

6 Data and Methodology

This chapter clarifies the data structure used in this thesis and explains the research methodology by giving a detailed description of the implemented trading strategy. The used methodology and later the presentation of results mimics the style used by Tikkanen and Äijö (2018). In the last section of this chapter, limitations to this study are explained.

6.1 Data

The data for this thesis is obtained from the Thomson Reuters Datastream through the School of Accounting and Finance at University of Vaasa (Fama-French risk factors are obtained from the Kenneth French data library). The data consists financial statement information of S&P500 constituent companies from 1997 to 2016 and stock price data that covers years between 1999 and 2017. The obtained financial statement information consists values at the fiscal year-end whereas the stock price information has a monthly frequency. Risk-free rate is the one-month US T-bill rate similar to the Fama-French factors. All monetary figures are in United States Dollars (USD/$).

Companies that lack required financial statement data to compute B/M-ratio, F-Score, M-Score or Z-Score are excluded from the sample. Moreover, as is a common practice in financial literature, financial companies and REITs are also removed from the sample due to different interpretation of the financial statements of these companies. Lastly, companies with negative book-to-market ratio are also excluded from the sample.

The total number of firm-year observations in the dataset is 4509, which yields a final sample of 1503 high book-to-market (HBM) firm-year observations during the sample period. Due to data availability, the calculated portfolio returns are exposed to survivorship bias. That is, the sample consists of only companies that are included in the index at last year of the sample (limitations are further clarified in a separate section in this chapter). For that reason, the benchmark index used in the return presentations is a modified version of the S&P 500 index, which is calculated from the data used in this study. This decision ensures that also the “market” returns are exposed to the same positive return deviation.

6.2 Trading Strategy

The implemented trading strategy has multiple steps. The first steps follow the trading strategy proposed by Piotroski (2000). Thus, companies are first ranked based on their fiscal year-end book-to-market ratios. In his study, Piotroski (2000) ranks the companies into five different portfolios based on the B/M ratios after which, the portfolio with the highest B/M-ratios (HBM) are taken into further analysis. Due to significantly smaller sample size in this study, tercile cutoffs are used to include more companies into the HBM portfolio. The aggregate Piotroski F-Score is then calculated for the companies in the HBM portfolio according to formula 9, which was covered in detail in chapter 5.2:

$$\text{F-Score} = ROA + \Delta ROA + CFO + ACCRUAL + \Delta LEVER + EQOFFER + \Delta LIQUID + \Delta MARGIN + \Delta TURN \quad (9)$$

F-Score of 9-7 are assigned into a long portfolio, whereas companies with low f-score (0-3) are assigned into a short portfolio.

In order to enhance F-Score’s ability to separate winner stocks from losers, additional fundamental screens are applied using the Beneish M-Score and Altman’s Z-Score. Specifically, the resulting F-Score portfolios are screened with the M-score, so that companies which have high manipulation probability (indicated by M-score >-1,78[10]) are excluded from the F-Score ranked portfolios. M-Score was defined in chapter 5.3 as:

$$\text{M-Score} = -4.84 + 0.92 \times DSRI + 0.528 \times GMI + 0.404 \times AQI + 0.892 \times SGI + 0.115 \times DEPI - 0.172 \times SGAI + 4.679 \times TATA - 0.327 \times LVGI \tag{19}$$

It is important to point out that in his study, Beneish (2013) assigns the flagged companies (M>-1,78) into a short portfolio. In this study, however, the flagged companies are excluded from the portfolios. This decision has a similar reasoning as the type 1 and 2 errors presented in chapter 5.3. That is, since investors have many stocks to choose from, losing an investment opportunity is less costly than a situation where that opportunity is used to buy or sell a stock with unreliable earnings figures.

Lastly, the portfolios are screened with Altman's Z-score. According to Piotroski (2000), the highest return spread between high and low F-Score companies are generated when only low default risk companies are used. According to chapter 5.4, Z-Score was computed as:

$$\text{Z-Score} = 1.2X1 + 1.4X2 + 3.3X3 + 0.6X4 + 0.99X5 \tag{28}$$

After the Z-Score application, the portfolios only consist of companies that are in the top Z-score tercile of that year. In other words, the final F-Score based portfolios are "cleaned" from companies that may have poor earnings quality or have an increased default risk exposure as indicated by M- and Z-Scores, respectively. To avoid the lookahead-bias, portfolio formations are carried out each year in the first trading day of June. As proposed for example by Piotroski (2000) and Chan et al. (1991) such time between the fiscal year end and the portfolio formation ensures information availability for investors. All portfolios are equally weighted.

[10] In the following sections of this study, "high" M-Score portfolio refers to a value of M-Score<-1,78, not to the actual value. In these cases, "high" is for clarity used interchangeably with "long" (position).

6.3 Risk-Adjusted Performance Measurement

The risk adjusted performance of the portfolios will be analyzed using Sharpe (1966) and Sortino (1991) ratios. Sharpe ratio, also known as the reward-to-volatility ratio, is computed as:

$$S_p = \frac{R_p - R_f}{\sigma_p} \tag{29}$$

where, R_p is the portfolio return, R_f = the risk-free rate, and σ_p= standard deviation of the portfolio's monthly excess returns over the risk-free rate. The statistical difference between Sharpe ratios between different portfolios are tested with Memmel (2003) test, which is a corrected version of the original Jobson and Korkie (1981) test. The z-test statistic is calculated as:

$$z = \frac{\sigma_a\, \mu_b - \sigma_b\, \mu_a}{\sqrt{\theta}} \tag{30}$$

where, σ_a and σ_b are the standard deviations of returns of portfolios *a* and *b* respectively, μ_a and μ_b are the mean returns of portfolios *a* and *b* respectively, and θ is the asymptotic variance which is obtained as:

$$\theta = \frac{1}{T}\left[2\sigma_a^2\sigma_b^2 - 2\sigma_a\sigma_b\sigma_{ab} + \frac{1}{2}\mu_a^2\sigma_b^2 + \frac{1}{2}\mu_b^2\sigma_a^2 - \frac{\mu_a\mu_b}{\sigma_a\sigma_b}\sigma_{ab}^2\right] \tag{31}$$

where, *T* is the number of observations, σ_{ab} is the covariance of returns of portfolios *a* and *b.* As mentioned, Sortino ratios are calculated for each portfolio alongside with the Sharpe ratio. Sortino ratio can be viewed as a modification of Shape ratio: Where Sharpe ratio considers the overall standard deviation of returns as a measure of risk, Sortino ratio (SR) uses only the root-mean-square deviation from MAR, or the minimum acceptable return, which in this case is the risk-free rate. In other words, only negative returns or the downside risk is considered. Hence, the Sortino Ratio does not penalize high positive returns that in the case of Sharpe ratio, would increase volatility. Sortino ratio is calculated as:

$$\mathrm{SR_p} = \frac{\mathrm{R_p - MAR}}{\sqrt{\frac{1}{\mathrm{n}}\sum_{\mathrm{R_p<MAR}}\left(\mathrm{R_p - MAR}\right)^2}} \tag{32}$$

Abnormal returns are measured using the Fama-French (2015) five-factor model, which was more thoroughly examined in chapter 3. The model gives insight to the risk exposure characteristics of the returns of different portfolios. In the Fama-French regressions, Newey and West's (1987) standard errors are used in order to mitigate the impact of heteroscedasticity. The FF5 regression model reads as follows:

$$(\mathrm{r}i - \mathrm{r}f) = \alpha i + bi[(rm) - rf] + si[SMB] + hi[HML] + \mathrm{ri[RMW]} + \mathrm{ci[CMA]} \tag{33}$$

6.4 Limitations

This section discusses the limitations to the study. As briefly explained in the data section of this thesis, due to data availability, the results are exposed to survivorship bias. This means that only companies are present in the last year of the data are included to the analysis, which may cause the results to deviate positively as only companies that *survived* from the start of the to the end of the sample period are analyzed. That is, the sample size increases gradually from the beginning of the sample period. More robust results would be obtained by using data that each year has all the constituent companies. To counteract the effects of the survivorship-bias, market returns are calculated from the whole dataset. This ensures that the market returns are also exposed to the same bias. Actual market returns would have been lower than those presented in this study.

Additionally, this study does not account for transaction cost nor taxes that arise from changing the companies in the portfolios. Moreover, the portfolio sizes become smaller as more fundamental screens are applied. The small number of companies in the portfolios affects especially the triple-screened FMZ portfolios that hold only three stocks on average. However, this study is not exposed to liquidity issues as only large companies are analyzed.

7 Empirical Analysis

This section of the thesis covers the empirical results. The structure of the empirical analysis is as follows: at first, descriptive statistics are provided to better understand the data structure and the characteristics of the analyzed companies. In the second part, portfolio returns are presented in terms of compound annual growth rates, year-by-year raw returns and Fama-French five-factor estimated abnormal returns. In the third section of the empirical analysis, the focus is on the risk-adjusted performance of the implemented trading strategies. Moreover, the third section reports the risk exposure characteristics of the portfolios according to the Fama-French factor loadings and Sharpe and Sortino ratios as the portfolio performance measures.

The last part of the empirical analysis discusses the obtained results and links them to existing literature to explain the possible reasons behind them. Moreover, the discussion section also considers the research hypotheses that were presented in chapter 1.2.

7.1 Descriptive Statistics

Descriptive statistics of the used data set are illustrated in table 1, where panel A provides the figures for the HBM companies and panel B for the whole data set. The table presents the size characteristics in terms of market capitalization and total assets in millions of USD. Additionally, the average F-, M- and Z-scores are provided along with the F-Score signals that were presented in chapter 5.

As can be observed from the table, the average book-to-market ratio for an HBM company is 0,664 which is almost twice as high when compared to an average company in the whole data set. However, the average book-to-market ratios in both panels are not very high, which could be due to the generally higher valuation of blue-chip stocks compared to smaller stocks. Market capitalization for non-HBM companies is on average higher than the corresponding value for an HBM company. This is however contradicted by the higher average total assets of HBM companies, subsequently leading to higher mean book-to-market ratio.

TABLE 1 Descriptive statistics of company size, FMZ-Scores, and the F-Score signals. Panel A reports the values for the HBM companies, whereas Panel B illustrates the corresponding values for the whole data set.

PANEL A: HBM stocks	**Mean**	**Median**	**St.Dev**	***n***
MktCAP (M$)	16772	7894	33303	*1503*
TOTAL ASSETS (M$)	27595	13277	60989	*1503*
B/M-RATIO	0,664	0,591	0,302	*1503*
F-SCORE	5,426	5,000	1,646	*1503*
M-SCORE	-2,624	-2,642	1,658	*1503*
Z-SCORE	2,599	2,183	1,957	*1503*
CFO	0,880	1,000	0,326	*1503*
ACCRUAL	0,917	1,000	0,276	*1503*
DLEVER	0,444	0,000	0,497	*1503*
DLIQUID	0,491	0,000	0,500	*1503*
DMARGIN	0,486	0,000	0,500	*1503*
DROA	0,499	0,000	0,500	*1503*
DTURN	0,476	0,000	0,500	*1503*
EQOFFER	0,355	0,000	0,479	*1503*
ROA	0,878	1,000	0,328	*1503*
PANEL B: All stocks				
MktCAP (M$)	27173	9574	54410	*4509*
TOTAL ASSETS (M$)	22439	8237	53134	*4509*
B/M-RATIO	0,382	0,322	0,280	*4509*
F-SCORE	5,688	6,000	1,602	*4509*
M-SCORE	-2,601	-2,618	3,922	*4509*
Z-SCORE	5,122	3,586	7,042	*4509*
CFO	0,915	1,000	0,279	*4509*
ACCRUAL	0,904	1,000	0,294	*4509*
DLEVER	0,417	0,000	0,493	*4509*
DLIQUID	0,493	0,000	0,500	*4509*
DMARGIN	0,547	1,000	0,498	*4509*
DROA	0,544	1,000	0,498	*4509*
DTURN	0,493	0,000	0,500	*4509*
EQOFFER	0,462	0,000	0,499	*4509*
ROA	0,914	1,000	0,281	*4509*

When examining the average FMZ-scores, it can be determined that the average (median) F-score 5,426 (5,000) of an HBM company is slightly lower than the corresponding F-score 5,688 (6,000) of an average company. The lower average Z-Score 2,599 (2,183) indicates that HBM companies have increased default risk compared to an overall average of 5,122 (3,586). Based on the M-score, companies cannot be generally considered to be prone to earnings manipulation as the threshold value indicating aggressive earnings management is -1,78 as explained in chapter five.

Since each of the F-score signal can have either value of 1 or 0, the threshold for the median is 0,5, which separates positive and negative signals. Thus, an average value of <0,5 would result a median of 0,000. Based on the median values of the F-score signals, it can be observed that HBM companies are in worse financial condition, which is consistent with Piotroski (2000) and Fama and French (1995).

For example, the variables DROA and DMARGIN suggest that on average, the profitability and operating efficiency of an HBM company deteriorates from the year before. Oppositely for the whole data set, these variables have a median of 1, suggesting improving profitability and operating efficiency. However, the signals are in some cases slightly contradictory to each other. For example, the variable ROA for HBM and non-HBM companies are 0,878 and 0,914 respectively, signaling that all companies are profitable on average.

Consistent evidence with Piotroski (2000:14) can be also observed in the case of ACCRUAL, DLIQUID and DLEVER. The high mean ACCRUAL suggests that on average companies are not prone to creative accounting, thus indicating more reliable reported figures. The low means of DLIQUID however, indicate that companies’ ability to handle their payment obligations decrease from the year before. Moreover, at low mean of DLEVER signals that the level of debt increases on average from the year before.

Table 2 provides the distributions of the different F-score values as well as the number of companies that are flagged as possible manipulators based on M-score.

TABLE 2 Distribution of F-Score and M-Score

F-SCORE	***n=1503***
9	*19*
8	*130*
7	*273*
6	*302*
5	*361*
4	*221*
3	*128*
2	*47*
1	*21*
0	*1*
M-SCORE	***n=1503***
Flagged (M>-1,78)	*107*
Not Flagged	*1396*

The table shows that most companies in the HBM portfolio have an aggregate F-Score between five and seven, five being the largest group. This distribution is consistent with Piotroski (2000) as the F-Score values are more concentrated in the high values. This concentration results short portfolios to be smaller than the corresponding long portfolios in the trading strategy, since the qualification of a stock to a portfolio is based on absolute values rather than on relative proportions.

The bottom part of table 2 presents the distribution of the M-score. In total, 107 firm-year observations are flagged as potential manipulators. Beneish et al. (2013) shows that in his study, 17,4% of the sample are flagged companies. In this case, the proportion is ten percentage points lower at 7,66%. However, the smaller proportion of flagged companies is consistent with the findings of Beneish et al. (2013:66) as the proportion tend to decrease as size, proxied by market value, increases.

7.2 Portfolio Returns

This section of the empirical part presents the returns of the implemented trading strategy. As presented in chapter 5, the HBM companies are assigned into long and short portfolios based on their aggregate F-score. Specifically, companies with an aggregate F-score of 9-7 (3-0) are assigned into a long (short) portfolio. Portfolio formations are carried out each year in the first trading day of June. In the following tables, headers "F-Score, FM-Score and FMZ-score" have the following interpretation: *F-Score* refers to a F-score-only strategy, *FM-Score* refers to a strategy where companies with M-score>-1,78 are excluded. *FMZ-score* refers to a strategy where companies with M-score>-1,78 along with companies that are not included in the top Z-Score tercile of that year are excluded.

Table 2 provides the compound annual growth rates for each portfolio in Panel A. Panel B reports the aggregate number of companies (specific portfolio sizes by year are provided in the appendix). Panels C reports the Fama-French 5F alphas, that is, the abnormal returns. In the table, "hi" refers to a long portfolio of a corresponding trading strategy, whereas "lo" refers to a short portfolio.

According to the figures in Panel A, the compound annual growth rates for a F-score long (short) portfolio are higher (lower) than the corresponding CAGRs for the market and HBM portfolios. Specifically, the long F-Score portfolio outperforms the HBM and market portfolios by 2,02% and 9,08% in terms of CAGR. Additionally, the short F-Score portfolio earns lower returns than the HBM portfolio but exceeds the market return.

From the table it can be observed that the return spread increases as companies with high manipulation probability are excluded. That is, the exclusion increases the growth rate for the long portfolio and decreases the growth rate of the short portfolio. Moreover, the high FM-portfolio earns the highest growth rate at 16,88%. In terms of CAGR, the high FM-portfolio outperforms the HBM portfolio by 2,80% and the market by 9,86%. The application of the Z-Score screen decreases the CAGRs for both, the long and short portfolios. Furthermore, the spread between high and low FMZ-portfolios is the highest at 9,58%. However, as can be determined from the figures in Panel B of the table, aggregate number of companies in the portfolios drastically decreases after the application of the Z-Score screen. For example, the average number of companies in the low FMZ-portfolio is only 3[11], whereas the corresponding averages for the low F- and FM-portfolios are 10 and 9 respectively.

TABLE 3 Portfolio returns. Panel A reports the compound annual growth rates for each portfolio, Panel B the aggregate number of companies and Panel C the annualized Fama-French 5F alphas. In the Fama-French 5F regressions monthly return series (216 observations) are used. T-statistics are reported in the parenthesis. *,** and *** illustrate statistical significance at the 10%, 5% and 1% levels, respectively.

Portfolio Returns	**F-Score**	**FM-Score**	**FMZ-Score**
PANEL A: Annual Returns (CAGR)			
Hi	16,10%	16,88%	14,42%
Lo	12,37%	11,23%	4,84%
HBM	14,08%	14,08%	14,08%
Market	7,02%	7,02%	7,02%
Hi-Lo	3,73%	5,65%	9,58%
Hi-HBM	2,02%	2,80%	0,34%
Hi-Market	9,08%	9,86%	7,40%
PANEL B: Number of Companies			
Hi	425	408	133
Lo	185	157	59
HBM	1503	1503	1503
PANEL C: Fama-French 5F Alphas			
Hi	**5,37%****	**6,08%****	3,06%
t-stat	(2,280)	(2,334)	(0,769)
Lo	4,93%	1,64%	1,23%
t-stat	(1,218)	(0,438)	(0,204)
HBM	**5,06%****	**5,06%****	**5,06%****
t-stat	(2,544)	(2,544)	(2,544)

Similar patterns can be observed from panel C, which reports the abnormal returns estimated with the FF5 model. When considering the abnormal returns of the long leg, significant alphas are generated only by the high F- and FM- portfolios. Moreover, the highest alpha is generated by the high FM-portfolio at 6,08% per annum with a t-statistic

[11] Specific portfolio sizes by year are provided in the Appendix 1

of 2,334 (p<0,05). Additionally, the HBM portfolio itself yields positive abnormal returns at almost 5,06% per year suggesting that the performance of the long portfolios is mostly driven by the underlying performance of the HBM portfolio. The abnormal returns for the F- and FM-score long portfolios have similar magnitude as reported by Tikkanen and Äijö (2018) in the European stock market. They find that for an HBM portfolio that is enhanced with the F-score the annual abnormal return is 6,22% (p<0,01).

However, when examining the short portfolios, none of the strategies generate significant abnormal returns suggesting inconsistent performance from year-to-year. The raw returns of the portfolios are further illustrated in figures 4 and 5, which show the year-by-year returns for the long and short portfolios against the market portfolio. In the figures, years refer to the end of the holding period. As can be observed from figure 4, the long portfolios consistently outperform the market portfolio in raw return basis.

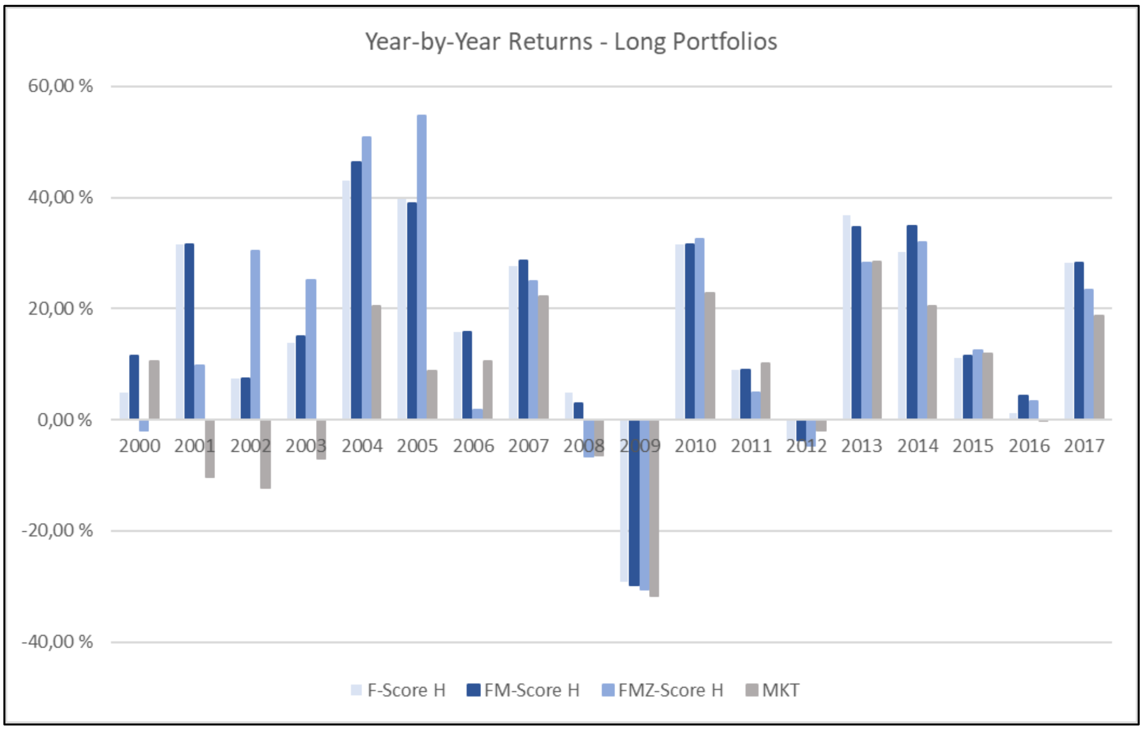

FIGURE 4 Year-by-Year returns for the long portfolios and the market portfolio.

Though the returns are higher for the long portfolios throughout the sample period, the spreads in returns are higher during the period preceding the most recent financial crisis in 2008. The reason for this could be that after the financial crisis, more attention has been paid to companies' financial position, which reduces the possible pricing errors and subsequently lowers the returns.

Figure 5 presents the year-by-year raw returns for the short portfolios and for the market portfolio. As evident, similar consistency in return patters cannot be observed for the short portfolios as with the long portfolios. The returns are prone to extremely high variation from year to year. In many years, the short portfolios earned higher returns than the market and in some years, even higher than the long portfolios. The worst draw-downs were years 2000, 2004 and 2010 when the returns were approximately 50%. Along with the compound annual growth rates and FF5 alphas presented in table 3, the return patterns presented in figure 5 suggest that F-score and additional M- and Z-Score screens have limited power to predict future returns for the short portfolios. The next section of the empirical part discusses the risk-adjusted performance of the portfolios.

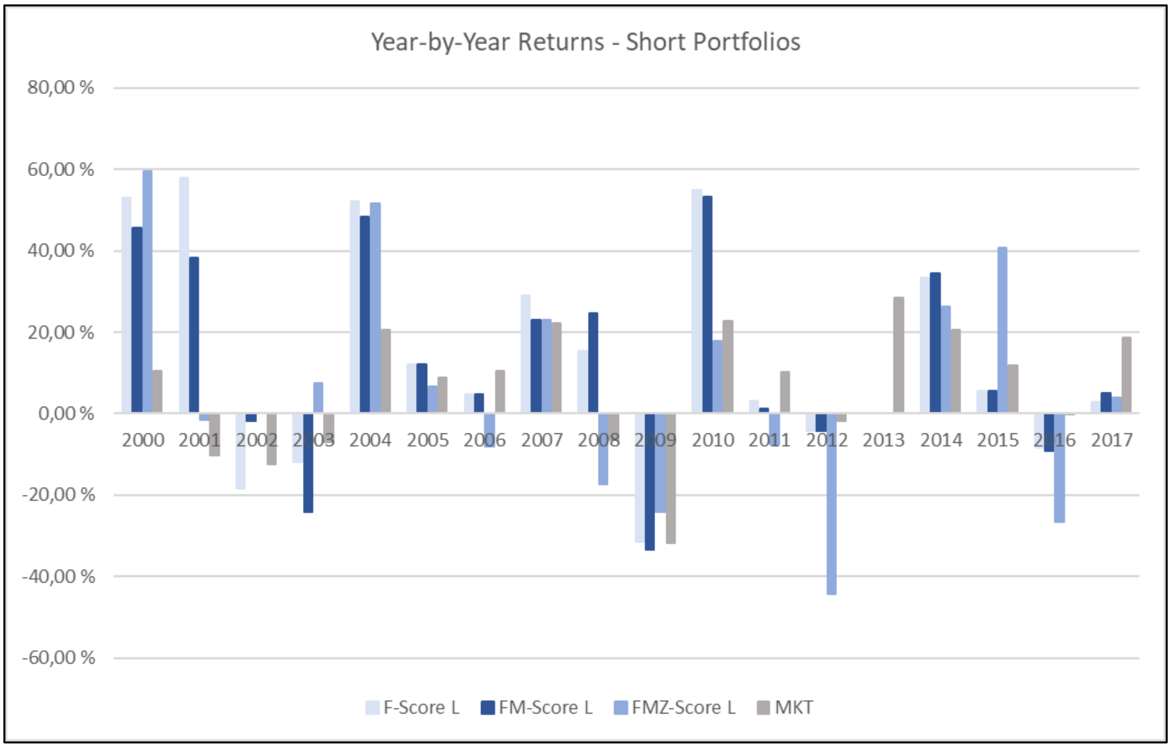

FIGURE 5 Year-by-Year returns for the short portfolios and the market portfolio

7.3 Risk-Adjusted Performance

This section of the empirical analysis presents the risk-adjusted performance and risk exposure characteristics of the investigated portfolios. At first, the Fama-French 5F model regression results are provided after which, the portfolio performance measures are presented. Namely, Sharpe and Sortino ratios. Table 4 reports the FF5 regression results, where panel A presents the factor loadings for the long portfolios and panel B the corresponding figures for the short portfolios.

TABLE 4 Fama-French 5F Loadings. Coefficients for the FF5 factors are estimated using the FF5 model with monthly return series (216 observations). The intercept terms (alphas) have been annualized for presentation purposes. T-statistics are provided in the parenthesis. *,** and *** illustrate statistical significance at the 10%,5% and 1% levels, respectively.

Fama-French 5F Loadings		**F-Score**	**FM-Score**	**FMZ-Score**
Panel A: High F-Score Portfolios				
	Alpha (annual)	**5,37%****	**6,08%****	3,06%
		(2,280)	(2,334)	(0,769)
	Market	**0,982*****	**1,001*****	**1,203*****
		(16,711)	(16,194)	(9,107)
	SMB	**0,231*****	**0,194****	0,197
		(3,071)	(2,119)	(0,866)
	HML	**0,235****	**0,277****	**0,382****
		(2,054)	(2,304)	(2,102)
	RMW	**0,396*****	**0,429*****	**0,736*****
		(4,676)	(4,414)	(3,293)
	CMA	0,178	0,131	-0,062
		(1,276)	(0,927)	(-0,197)
Adj.R-sqr.		*72,74%*	*71,62%*	*51,42%*
PANEL B: Low F-Score Portfolios				
	Alpha (annual)	4,93%	1,64%	1,23%
		(1,218)	(0,438)	(0,204)
	Market	**1,264*****	**1,338*****	**1,011*****
		(17,899)	(17,213)	(9,163)
	SMB	0,085	0,191	0,183
		(0,568)	(1,181)	(1,181)
	HML	0,229	0,233	**-0,458****
		(1,303)	(1,642)	(-2,405)
	RMW	-0,088	0,129	**-0,511****
		(-0,436)	(0,925)	(-2,422)
	CMA	0,268	0,357	**0,849****
		(1,007)	(1,577)	(2,110)
Adj.R-sqr.		*63,35%*	*63,60%*	*43,61%*

Both high and low portfolios have highly significant loading on the market factor, but the coefficients are slightly higher for the short portfolios suggesting a higher market risk exposure for the short portfolios. The long portfolios have also positive and highly significant loadings on the RMW factor, ranging between 0,396 and 0,736. Thus, high F-score companies tend to have more robust profitability than their low F-Score counterparts, whose loadings on the RMW factor are negative for the F- and FMZ-portfolios, and statistically insignificant for the F- and FM-portfolios. Similar findings are also reported by Tikkanen and Äijö (2018), who also report positive and highly significant RMW loadings for the high F-Score portfolios, but negative or insignificant loadings for the short portfolios.

Also consistent with the findings of Tikkanen and Äijö (2018), the long portfolios have statistically significant and positive loadings on the size factor SMB and value factor HML. The positive coefficients for the two aforesaid factors indicate exposure to the long leg of these factors. In the case of SMB, this would mean that the companies share the risk characteristics of smaller companies. Moreover, the returns are partly explained by the higher book-to-market of these companies as indicated by the positive coefficient of the HML factor.

For the short portfolios, the risk characteristics are different and have a less systematic pattern when compared to the long portfolios. For the F-score and FM-score portfolios, the market factor has the only statistically significant coefficient. However, the loadings on the HML and CMA factors are positive and almost significant at the 10% level ($p<0,12$). In the case of CMA, this weakly implies that companies whose financial health is compromised (as indicated by the low F-score) have low asset growth.

For the low FMZ-portfolio, the results differ slightly from the other two short portfolios although the market and CMA loadings are similar. For example, the HML coefficient is negative (-0,458) and statistically significant at 5% level, meaning that companies in that portfolio share the risk characteristics of large companies. Moreover, the RMW factor's coefficient is -0,511 (p<0,05) implying significantly weaker profitability.

The adjusted R-squares are relatively high, ranging from 72,72% to 43,61%. The adjusted R-squares are however much smaller for the FMZ-score portfolios at 51,42% and 43,61% as for the other four portfolios the lowest adjusted R-square is 63,35%. The low R-squares of the FMZ-score portfolios can be due to the small number of companies in those portfolios, causing higher return variation. The adjusted R-square values are consistent with e.g. Äijö and Tikkanen (2018) and suggest that the FF5 model explains relatively well the variation in the portfolio returns. Moreover, the adjusted R-squares are higher for the long portfolios. However, as the long F-score and FM-score portfolios have significant alphas, some of the variation in returns is left unexplained.

The risk-adjusted portfolio performance is further demonstrated in table 5's panels A and B, which report the annualized Sortino and Sharpe ratios for the examined portfolios, respectively. As can be observed from the table, the Sortino ratio is the highest for the FM-score strategy at 1,498 when compared to the F-Score strategy (1,002) and FMZ-strategy (0,737). Additionally, the spreads are the highest for the FM-Score strategy. Specifically, the FM-score's Sortino ratio is 1,057 higher than the market's and 0,554 higher than the Sortino ratio of the HBM portfolio. The Sortino Ratio for the low FMZ-Score is the smallest of the three at 0,202 but it does not make up the poorer value of the of the high portfolio (0,737).

When analyzing the Sharpe ratios, similar patterns emerge as the high FM-Score portfolio is the best performing with a Sharpe ratio of 0,908 followed by F-Score (0,890) and FMZ-score (0,619). Also, the application of the M-score screen decreases the Sharpe ratio of the low F-Score portfolio, thus increasing the spread between high and low portfolios. Again, the lowest Sharpe ratio can be found from the low FMZ-portfolio, but it is overcome by the poor Sharpe ratio of the long portfolio. According to the Sharpe ratios and the spread test, the high F-Score and FM-Score outperform the low portfolios and the market. However, the difference between the two high portfolios and the HBM portfolio does not reach statistical significance.

TABLE 5 Annualized Sortino and Sharpe ratios. The table also presents the statistical significance of the Sharpe spread, based on the Memmel (2003) version of the Jobson-Korkie (1981) test. Z-statistics are provided in the parenthesis. *,** and *** illustrate statistical significance at the 10%, 5% and 1% levels, respectively.

	F-Score	FM-Score	FMZ-Score
PANEL A: Sortino Ratios			
Hi	1,002	1,498	0,737
Lo	0,486	0,457	0,202
HBM	0,944	0,944	0,944
Market	0,441	0,441	0,441
Hi-Lo	0,516	1,041	0,535
Hi-HBM	0,058	0,554	-0,207
Hi-Market	0,560	1,057	0,295
PANEL B: Sharpe Ratios			
Hi	0,890	0,908	0,619
Lo	0,532	0,487	0,244
HBM	0,860	0,860	0,860
Market	0,464	0,464	0,464
Hi-Lo	**0,359****	**0,421****	0,375
z-stat	(2,058)	(2,511)	(1,408)
Hi-HBM	0,030	0,480	-0,241
z-stat	(0,292)	(0,481)	(-1,403)
Hi-Market	**0,426*****	**0,444*****	0,154
z-stat	(2,676)	(2,714)	(0,743)

According to the long portfolios' Sharpe ratios, it can be determined that though the FM-Score strategy is the best performing, the difference to the other two long portfolios are smaller than with the Sortino ratio. This smaller difference implies that the Sharpe ratio of the FM-score portfolio is penalized for having higher returns, which causes volatility to increase. This increase in volatility subsequently decreases the Sharpe ratio. As the Sortino ratio only considers downside volatility as risk, the high returns of the FM-score strategy do not affect the Sortino ratio.

7.4 Discussion

The preceding sections of this chapter have presented the results of an F-Score based trading strategy and the effect of additional fundamental screening. This part of the empirical analysis focuses on recapping the presented results and providing explanations by analyzing findings of other researchers. Moreover, the section considers the research hypotheses set in chapter 1.2.

7.4.1 F-Score Performance

As presented in chapter 1.2, the first alternative hypothesis considers the effectiveness of an F-Score screening. Furthermore, the objective of the first hypotheses pair is to determine whether the F-Score can separate future winners from losers by assigning high and low F-Score stocks into different portfolios. Moreover, the second alternative hypothesis assess the return difference between the high F-Score portfolio and HBM portfolio. Specifically, the hypotheses 1.1 and 2.1 read as follows:

H1,1: *High F-Score portfolios do generate higher abnormal returns than low F-Score portfolios*
H2,1: *Abnormal returns of high (low) F-Score portfolios are higher (lower) than the benchmark*

According to Fama-French 5F alphas presented in panel C of table 3 and in table 4, the long portfolio alphas exceed the corresponding short portfolio alphas. Specifically, the alpha of a high F-Score portfolio is 5,37% ($p<0,05$) compared to a low F-Score portfolio alpha of 4,93% ($p>0,1$). The difference in performance is also supported by the Sharpe and Sortino ratios (table 5) as the long portfolios have better risk-adjusted performance compared to short portfolios, as indicated by the higher ratios. However, the spread in returns tends to be driven by the long portfolio performance as only the long portfolio alphas are statistically significant at an acceptable level. This leads to an interpretation, where the first null hypothesis is rejected, but the effect of the F-Score screening is not as strong as it would be in a situation where the short portfolio alphas are negative and statistically significant. Moreover, the higher returns of the F-Score long portfolio seems to be for the most parts driven by the overall performance of the HBM portfolio. Though the abnormal returns and Sharpe and Sortino ratios are higher for the F-Score portfolio, the difference is marginal. In terms of the Fama-French 5F alphas, the annual difference is only 31 basis points when compared to the HBM alpha.

The weaker separative effect of the F-Score may be conditional on company size as proposed in the original study by Piotroski (2000). Piotroski (2000:20-23) suggests, that F-Score's ability to separate winner stocks from losers is decreased among larger companies. Moreover, it is suggested that the differentiation ability of the F-Score is smaller among companies that have high analyst coverage, though the effect is not isolated to companies with no analyst following. In this study, the analyzed companies are without exception large and have high analyst coverage, which can drive the smaller F-Score returns.

However, Tikkanen and Äijö (2018:504-505) contradict the claims of Piotroski (2000) as they find that the F-Score is a beneficial screening tool also among big European companies. Though they report that smaller companies generate better returns in general, the returns are high also for big stocks. For example, when the F-Score is used to screen high B/M companies the CAGR for a high F-Score-size portfolio is 19,90% compared to 17,85% of a small F-Score-size portfolio. Additionally, it is reported that the low F-Score portfolio CAGRs increase in four out of six different value strategies when applied to large companies instead of small stocks. Oppositely for high F-Score portfolios, the B/M-based strategy is the only one that has higher CAGR among large companies (Tikkanen & Äijö 2018:505). That is, the relatively weaker performance of the short portfolios compared long portfolios in this study can be partly driven by the large stock sizes.

When assessing the positive performance of the high F-Score companies, profitability seems to be a highly contributing factor. According to the FF5 factor loadings in table 4, the profitability factor of high F-Score portfolio is positive and highly significant at 0,396. For the low F-Score portfolio, the RMW loading is negative and insignificant, suggesting that F-Score can separate profitable stocks from others.

Profitability and its positive relation to future returns have been well documented. That is, Piotroski (2000) also uses profitability as one of the three main categories in the F-Score. For example, Haugen and Baker (1996:419) provide evidence that profitable companies, measured by asset turnover, ROE, ROA and profit margin, have higher expected returns compared to less profitable companies. Furthermore, Fama and French (2006) suggest that profitability together with high B/M ratio is positively associated with future stock returns.

7.4.2 Additional Fundamental Screening

In addition to an F-Score-only strategy, this study has investigated the usefulness of other fundamental metrics to enhance the F-Score's performance. That is, the hypothesis 3.1 and 4.1 reflect the M-score's and Z-Score's ability to increase (decrease) the high (low) F-Score portfolio returns. The hypotheses were set as follows:

H3,1: *M-Score screening does increase (decrease) the abnormal returns of long (short) portfolios constructed on with the F-Score*
H4,1: *Z-Score screening does increase (decrease) the abnormal returns of long (short) portfolios constructed on with the FM-Score*

According to the reported results, M-Score screening seems to have a positive impact on the F-Score portfolios. As reported in table 3, the high FM-Score portfolio has the highest compound annual growth rate (16,88%) and the highest annual FF5 alpha of all at 6,08% ($p<0,05$). These results suggest that the using the M-Score as a proxy for earnings quality, the F-Score's abnormal returns can be increased, which leads to a rejection of the H3.0 hypothesis. Moreover, the application of the M-Score screen also tends to decrease the low F-Score portfolio returns. Even though the low FM-Score portfolio does not reach statistically significant levels, the long portfolio results indicate that the M-Score can act as a complementary tool when implementing an F-Score strategy. Furthermore, the results are supported by the risk-adjusted performance signals as the high FM-Score portfolio has the highest Sharpe (0,908) and Sortino (1,498) ratios, indicated by table 5.

Though the combination of the F-Score and M-Score has not been covered in literature to the best of the author's knowledge, the results are aligned with other preceding studies regarding earnings management and stock returns. As presented earlier, Beneish (2013) finds that high M-Score companies tend to earn lower returns. For example, it is reported that the average return of the highest B/M decile portfolio of non-flagged companies is 6,4%, whereas the flagged companies in the highest B/M decile returned a negative 11,7%. Although, the effect is less prominent among low B/M companies the return difference is significant on average. Moreover, the difference between returns is slightly stronger among smaller companies, but also significant among large companies (Beneish 2013:66). In this study, the higher returns of the FM-Score long portfolio compared to a F-Score only portfolio suggest that companies with high manipulation probability do earn lower returns.

Though not analyzed precisely in this study, the M-Score's ability to increase the returns may by driven by its ability to detect earnings management better than accruals alone. As mentioned, Sloan (1996) reports that companies which have managed their earnings through high positive accruals have poorer prospects and thus Piotroski (2000) uses accruals as one of the nine fundamental signals. Beneish (2013:57) however, points out that the M-Score performs well among companies which have low accruals. In other words, among companies which seem to have higher earnings quality based on accruals. According to table 1 of this study, it can be observed that the analyzed companies on average have low accruals as indicated by an accrual variable mean (median) of 0,917

(1,000). That is, the M-Score may detect information beyond the F-Score's accrual variable.

The last hypothesis of this study assesses the usefulness of the Z-Score together with the F-Score. Piotroski (2000) claims that the F-Scores ability to detect future losers is especially good when applied to low distress risk companies, which subsequently results the greatest return spread between long and short portfolios. Furthermore, Piotroski (2000:32) presents that the results are in line with Dichev (1998) who argues that low default risk tends to lead to higher stock returns. This claim, however, cannot be confirmed by the results obtained in this study. Though the spread in CAGRs between high and low FMZ-Score portfolios is the biggest, the difference in returns is not supported by the FF5 alphas nor by the performance signals in table 5. Furthermore, the FMZ-Score portfolios do not have significant alphas and the risk-adjusted performance of the long portfolio loses to the HBM portfolio. Thus, Z-Score screening does not seem to have positive impact on the F-Score strategy.

The discrepancies between results regarding the effect of the Z-Score screening in this study and Piotroski (2000) could be due to differences in analyzed companies. Piotroski (2000) uses much higher number of companies, both small and large. In this study, however, only large companies are analyzed. Moreover, the analyzed companies are constituents of the S&P 500, which means that generally the companies are top-tier performers. That is, the actual level of distress risk of these companies can be significantly lower, although some of them are indicated as "high" default risk based on the unconditional Z-Score screen. Should the actual default risk of these companies be very high, there is a great possibility that such company would be dropped from the index. Lastly, the application of the Z-Score screen makes the portfolios significantly smaller. Before the Z-Score screen the long portfolios hold around 23 stocks and the short portfolios around 10, whereas the long (short) portfolios hold only 7 (3) stocks on average after the screen. Thus, the FMZ-Score portfolio results should be assessed with much higher caution. However, based on the conducted analysis the null hypothesis H4.0 is accepted.

The relationship between book-to-market effect and distress risk is also investigated by Griffin and Lemmon (2002). They suggest that HBM stocks overperform glamour stocks especially among high distress companies, indicated by Ohlson's O-Score. However, the magnitude of the value premiums decreases as the level of distress decreases. Though, Griffin and Lemmon (2002) use O-Score for distress risk measurement, they point out that the results are robust also with other measures such as Altman's Z-Score. Piotroski (2000) also explains that although the return spread between high and low F-Score portfolios is the largest among low distress companies, the highest long portfolio returns are generated among medium distress companies. Due to the conflicting observations regarding the impact of distress risk on value strategies, further research around the topic is needed to clarify the relationship between bankruptcy probabilities and F-Score returns.

When analyzing the obtained results as a whole, few prominent patterns emerge. Firstly, the returns of each strategy are driven by the long leg of the strategy. Secondly, the long leg performance seems to be driven by the underlying HBM portfolio and profitability. Although both the long F-Score and FM-Score portfolios generate higher alphas than the HBM portfolio, the difference is marginal for the F-Score-only portfolio. For the FM-Score portfolio however, the effect is stronger when compared to the HBM portfolio. However, all the long portfolios exceed the market portfolio returns. Due to the data limitation presented in chapter 6.1 and 6.4, however, the positive performance of the long portfolios should be treated with appropriate caution as the returns may positively deviate compared to data set, where each year has all the index companies of that specific year.

From a practical point of view, the good long portfolio performance can be seen as promising due to limits of short selling. That is, the implementation of a long position is easier for individual and institutional investors than setting up a short position. Stambaugh, Yu and Tuan (2012:290) explain that short positions are not taken by investors for example due to transaction costs. Moreover, institutional investors such as mutual funds in some cases take only long positions due to regulations. Thus, the results provide possibilities for investors that tend to prefer long-only strategies when investing.

8 Conclusions

According to the semi-strong form of market efficiency, investors should not earn above average returns using fundamental analysis as stock prices already reflect all the publicly available information (Fama 1970). Despite the assumptions of the Efficient Market Hypothesis, value stocks have historically outperformed growth stocks. However, these high book-to-market companies tend to be financially weaker in general. Piotroski (2000) suggests that by ranking HBM stocks with the F-Score, a composite of nine fundamental strength signals, it is possible to find HBM companies that have the best and worst prospects. Furthermore, Piotroski (2000) reports that the F-Score strategy generates an annual market adjusted return of 23%. Additionally, Beneish (2013) argues that companies that have managed their earnings, tend to earn lower returns. To detect possible earnings manipulation, Beneish (2013) uses a manipulation detection model, known as M-Score (Beneish 1999).

This study has assessed the usefulness of fundamental analysis in the US equity market, in a framework proposed by Piotroski (2000) and Beneish (2013). Specifically, using financial statement information of S&P 500 constituent companies from 1997 to 2016 and monthly closing prices from 1999 to 2017, the study implements a F-Score based trading strategy. That is, companies are ranked each year based on their fiscal year-end book-to-market ratios, of which the highest B/M tercile is taken into further analysis. In the second stage, the HBM companies are assigned into long and short portfolios based on their F-Scores. That is, fundamentally strong companies (F-Score 7-9) are assigned to a long portfolio, whereas financially weak companies (F-Score 3-0) are assigned to a short portfolio. In the third stage, companies which have high earnings manipulation probability (M-Score>-1,78) are excluded from the portfolios. In the last stage, the companies in the portfolios are further limited to companies, which are included in the lowest default risk tercile of that year, indicated by Z-Score (Altman 1968).

The obtained results suggest that a F-Score based trading strategy does generate abnormal returns, but only for the long leg. Moreover, the magnitude of the long portfolio alphas is only marginally higher than the HBM portfolio alpha. Consistent with Piotroski (2000), the F-Score's ability to separate winners from losers seems to be weaker when applied to large companies, though it is useful to some extent. According to the Fama French 5F loadings, however, the high F-Score companies tend to be more profitable than low F-Score companies.

However, the use of M-Score to exclude companies with lower quality earnings, seems to have a positive impact on the F-Score strategy. The exclusion both increases the annual long portfolio alpha from 5,37% to 6,08% ($p<0,05$) and decreases the short portfolio alpha from 4,93% to 1,64%. The short portfolio alphas however are not statistically significant. In addition to the Fama-French 5F alpha, the application of the M-Score screen increases the Sharpe and Sortino Ratios of the high F-Score portfolio, leading to the interpretation that M-Score can be used as a complementary tool for a F-Score strategy. On the other hand, using only low distress risk companies based on Z-Score, the FM-Score returns do not increase further.

Though the results provide evidence regarding the beneficial relationship between F-Score and M-Score, further research around the subject is needed to confirm that the results are not isolated to the used sample or market. In this study, although wider partitions are used for example with the B/M ratio, the portfolio sizes become increasingly small when more fundamental screens are applied, especially when considering the short portfolios. Moreover, this study does not address the relationship between the specific variables that are used to compute F-Score and M-Score. Lastly, due to the conflicting results in prior research and in this study, the relationship between distress risk and value premiums should be examined further. However, from an investor's point of view, the results suggest that earnings quality is something that should be addressed, when implementing a fundamental analysis-based trading strategy such as an F-Score strategy.

References

Asness, C., Moskowitz T., & Pedersen L. (2013). Value and Momentum Everywhere. *The Journal of Finance,* 68(3), 929–985.

Altman, E. (1968). Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy. *Journal of Finance,* 23(4), 589-609.

Banz, R.W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics,* 9, 3–18.

Beneish, M. (1999). The Detection of Earnings Manipulation. *Financial Analysts Journal*, 55(5), 24-36.

Beneish, M., Lee C., & Nichols, D. (2013). Earnings Manipulation and Expected Returns. *Financial Analysts Journal*, 69(2), 57-81.

Bodie Z., Kane A., & Marcus A. (2014). Investments. New York: McGraw-Hill Education. 10.ed. 1016 p. ISBN 987-0-07-786167-4.

Carhart, M.M. (1997). On persistence in mutual fund performance. *The Journal of Finance,* 52, 57-82.

Chan, L.K., Hamao, Y., & Lakonishok J. (1991). Fundamentals and stock returns in Japan. *Journal of Finance,* 46, 1739-1789.

Chan, L., & Lakonishok, J. (2004). Value and Growth Investing: Review and Update. *Financial Analysts Journal,* 60(1): 71–86

Cooper, M.J, McConnell, J.J., & Ovtchinnikov A.V. (2006). The other January Effect. *Journal of Financial Economics*, 82(2), 315-341.

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor Psychology and Security Market Under- and Overreaction. *The Journal of Finance,* 53(6), 1839-1885.

Dichev, I. (1998). Is the risk of bankruptcy a systematic risk? *Journal of Finance,* 53, 1131-1147.

Dichev, I., Graham, J., & Rajgopal, M. (2013). Earnings Quality: Evidence from the Field. *Journal of Accounting and Economics*, 56, 1-33.

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

Fama, E., & French, K. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance,* 47(2), 427–465.

Fama, E., & French, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, 3-56.

Fama, E., & French, K. (2006). Profitability, investment and average returns. *Journal of Financial Economics*, 82, 491-518.

Fama, E., & French, K. (2012). Size, Value, and Momentum in International Stock Returns. *Journal of Financial Economics,* 105, 457–472.

Fama, E., & French, K. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1-22.

Fama, E. & French, K.(2017). International tests of a five-factor asset pricing model. *Journal of Financial Economics,* 128(2), 234-252.

Fama, E. & French, K.(2018). Choosing factors. *Journal of Financial Economics,* 123, 441-463.

Graham, B. & Dodd, D.(1934). Security Analysis. New York: McGraw Hill Book Co*,* 735 p, ISBN 0-07-144820-9.

Griffin, J.M., & Lemmon, M.L. (2002). Book-to-Market Equity. Distress Risk, and Stock Returns. *The Journal of Finance,* 57(5), 2317-2336.

Grobys, K., & Huhta-Halkola, T. (2019). Combining value and momentum: Evidence from the Nordic equity market. *Applied Economics,* 51, 2872-2884.

Haugen, R. A., & Baker, N.L. (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics,* 41, 401-439.

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

Jobson, J., & Korkie, B. (1981). Performance Hypothesis Testing with the Sharpe and Treynor Measures. *The Journal of Finance,* 36(4), 889-908.

Keim, D. (1983). Size-related anomalies and stock return seasonality: Further empirical evidence. *Journal of Financial Economics,* 12(1), 13-32.

Leivo, T. (2012). Combining value and momentum indicators in varying stock market conditions: The Finnish evidence. *Review of Accounting and Finance*, 11, 400–447.

Lev, B., & Thiagarajan, R. (1993). Fundamental information analysis. *Journal of Accounting research,* 31(2), 190-215.

Levy, R. (1967). Relative strength as a criterion for investment selection. *Journal of Finance,* 22(4), 595-610.

Lintner, J. (1965a). The valuation of risk assets and selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics,* 47(1), 13-37.

Lintner, J. (1965b). Security prices, risk, and maximal gains from diversification. *The Journal of Finance*, 20(4), 587-615.

Malkiel, B. (2003). The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspecives*, 17(1), 59-82.

Markowitz, H. (1952). Portfolio Selection. *Journal of Finance*, 7(1), 77-91.

Martikainen, T. (1995). *Arvopaperit.* 1.ed. Juva: WSOY. 153 p. ISBN 951-0-20735-7.

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

Newey, W., & West, K. (1987). A Simple Semidefinite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica,* 55(3), 703–708.

Nikkinen, J., Rothovius, T., & Sahlström, P. (2002). *Arvopaperisijoittaminen*. 1.ed. Vantaa: Werner Söderström Osakeyhtiö. 244 p. ISBN 951-0-26627-2.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18(1), 109-131.

Piotroski J. (2000). Value investing: The use of Historical Financial Statement Information to Separate Winners from Losers. *Journal of Accounting Research*, 38(3), 1-41.

Piotroski, J., & So, E. (2012). Identifying Expectation Errors in Value/Glamour Strategies: A Fundamental Analysis Approach. *Review of Financial Studies,* 25, 2841–2875.

Pompian, M. (2011). *Wiley Finance Ser: Behavioral finance and wealth management: How to build optimal portfolios that account for investor biases.* 2.ed. New Jersey: John Wiley & Sons, Inc. 352 p. ISBN 978-1-118-18227-7.

Ritter, J.R., & Chopra, N. (1989). Portfolio Rebalancing and the Turn-of-the-Year Effect. *Journal of Finance,* 44(1), 149-166.

Rosenberg, B., Reid, K., & Lanstein, R. (1985). Persuasive evidence of market inefficiency, *Journal of Portfolio Management,* 11, 9-17

Sharpe, W. (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *Journal of Finance*, 19(3), 425-442.

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

Sloan, R. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings, *Accounting Review*, 71(3), 289-315.

Sortino, F., & van der Meer, R. (1991). Downside risk. *The Journal of Portfolio Management* 17(4), 27–31 Appendices.

Stambaugh, R., Yu, J., & Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics,* 104(2), 288–302.

Stattman, D. (1980). Book values and stock returns. *The Chicago MBA: A Journal of Selected Papers,* 4, 25-45.

Tikkanen, J., & Äijö J., (2018). Does the F-score improve the performance of different value investment strategies in Europe?. *Journal of Asset Management,* 19, 495-506.

Turtle, H., & Wang, K. (2017). The Value in Fundamental Accounting Information. *The Journal of Financial Research,* 40(1), 113–140.

Walkshäusl, C. (2017). Expectation Errors in European Value-Growth Strategies. *Review of Finance*, 21(2), 845–870.

Appendix 1. Portfolio Sizes by Year

	F-Score		FM-Score		FMZ-Score		
	hi	lo	hi	lo	hi	lo	HBM
1999	8	13	7	12	1	3	60
2000	19	7	19	7	5	2	64
2001	14	7	14	3	3	0	65
2002	12	14	12	9	4	4	69
2003	21	19	19	19	7	10	75
2004	24	7	23	7	6	4	79
2005	34	7	33	7	10	3	79
2006	30	4	29	3	8	3	81
2007	32	9	31	7	9	4	83
2008	20	9	18	7	6	5	84
2009	20	14	20	14	9	4	89
2010	14	16	14	9	3	2	94
2011	35	6	35	6	12	1	95
2012	32	0	31	0	11	0	93
2013	23	18	20	16	7	5	98
2014	37	8	35	8	12	3	97
2015	32	11	30	9	13	2	100
2016	18	16	18	14	7	4	98
Total	425	185	408	157	133	59	1503
mean	24	10	23	9	7	3	84
median	22	9	20	8	7	3	84