Stockholm School of Economics Spring 2017
Department of Finance B.Sc. Thesis

Can Investors Benefit from Using a Simple Fundamental-Based Stock Selection Strategy?

Hugo Karlsson* & Simon Lindström**

Bachelor Thesis in Finance
Stockholm School of Economics
May 2017

Abstract

This paper examines Piotroski's (2000) fundamental-based F-Score strategy, on the Stockholm Stock Exchange between 1996 and 2017, to investigate: (1) if the strategy can identify future over- and underperformers, and (2) if this information constitute a market inefficiency over the most recognized common risk factors. We find that the strategy is overall successful, with an average annual return of 17 percent and a mean market-adjusted return of 8 percent. Moreover, several tests indicate that the strategy can separate future over- and underperformers, which is strongest for stocks with high book-to-market values. In contrast to previous research, the benefits of the strategy are not limited to small- and medium sized firms, which reduces concerns about feasibility. Importantly, the ability to identify future over- and underperformers are intact after adjusting returns, using Fama and French's (1993) three factor model. Thus, since risk fails to explain a majority of the high returns, our findings indicate the existence of long-term market mispricing.

Keywords: Piotroski's F-Score, Efficient Markets, Abnormal Returns, Fundamental Analysis, Stock Selection, Market Mispricing

Tutor: Irina Zviadadze, Assistant Professor, Department of Finance, Stockholm School of Economics

Acknowledgements: Special thanks to our tutor, Irina Zviadadze, for her valuable insight and support.

* 23352@student.hhs.se
** 23386@student.hhs.se

Table of contents

1. Introduction

Since the rise of equity markets, investors have explored stock selection strategies that can "beat the market". The prevalence of investor outperforming the market is a well-documented phenomenon with two dominating regimes of explanation; the risk-based view, that higher return is due to higher risk (Fama and French, 1992), and mispricing view, where the market incorrectly incorporates all available information (La Porta et al. 1997). Investors' returns are first tested against a model of the risk-based view, then the significance of the residual, the difference of investors' actual return and expected return, is computed to determine if this abnormal return constitutes mispricing. Theoretically, investors cannot consistently yield abnormal returns since all available information should be available to investors and therefore reflected in the price (Fama, 1970). Therefore, a significant abnormal return could constitute a mispricing, contradicting market efficiency. Simultaneously, the advancements of the asset pricing theory, with inclusion of more risk-based anomalies, have created models with increased explanatory power that have decreased the risk of falsely conclude abnormal returns. Still, several studies show that investors can earn abnormal returns and that these anomalies persist over time (e.g. Rosenberg et al. 1985).

Piotroski (2000) assents the existence of strategies to exploit market mispricing. He devises the F-Score as the aggregate of nine simple fundamental-based bools to evaluate firms and examines the strategy on stocks with high book-to-market ratios, value stocks, since their valuation more heavily relies on the financial statements. His findings suggest that the strategy can separate value stocks winners and losers, as well as shift the return distribution. In addition, Piotroski shows that through buying stocks with a high F-Score and shorting low F-Score, investors can earn high risk-adjusted returns, contradicting the semi-strong form of market efficiency. Furthermore, studies on other markets (e.g. Aggarwal and Gupta, 2009) supports the existence of market mispricing and show similar findings to Piotroski's.

The background of this paper is the increased accessibility of stock screeners, which has lowered the barrier for the individual investor to apply stock selection strategies (e.g. Piotroski F-Score) in the process of "beating the market". Thus, our work is highly relevant for individual investors since the paper test the usefulness of these strategies. In this paper, we search to evaluate one of these strategies, the F-Score strategy, on the Swedish Market and more specifically test the results against both risk-adjusting and mispricing explanations. The main research question is:

Can an investor, using the simple fundamental-based F-Score strategy (Piotroski, 2000), identify future over- and underperformers, and if so, does this information constitute a market inefficiency over the most well recognized common risk factors?

In relation to prior research on the F-Score domain, our paper contributes in several ways. Firstly, we give special focus on aspects where the F-score has been criticised, as with risk-adjustments and abnormal returns, rather than questing the fundamental rational that has been the focus of several other papers. Secondly, we test a more practical application that eases the assumption of investing first when companies by law must have released its annual report. Most Swedish firms report annual statements often before February, which means that around half of the return reflect the following year's F-Score, creating a huge mistiming. Thirdly, the paper provides concrete findings of the Swedish market where previous research is limited, both in how the researchers relate to risk and statistical methods. Lastly, our paper uses recent data with most of the sample after the release of Piotroski's paper and other research on the F-Score domain. This is especially interesting due to Schwert (2003), who argues that documented anomalies may rather change form or impact than disappear. To sum up, our paper contributes to previous research by both testing the robustness of previous research, as well as investigating new perspectives and improvements in the methodology.

The sample consists of all stocks on the Stockholm Stock Exchange for the period January 1996 to February 2017. Tests related to the strategy's ability to identify future over- and underperformers are complemented with tests of abnormal returns, book-to-market ratios, firm size and time. The results of our paper suggest that there is a strong relationship between F-Score and future returns. Furthermore, several tests indicate that the strategy can separate over- and underperformers and that this ability is strongest within the highest book-to-market quintile. The strategy provides a respectable average annual return of 17 percent and a mean market-adjusted return of 8 percent. The overall findings are in line with previous studies, however, in contrast to Piotroski's (2000) results, the benefits of the strategy are not limited to small- and medium sized firms. Importantly, the strategy's ability to find over- and underperformers persist when the returns are adjusted for risk, using the Fama and French three factor model (1993). Moreover, the abnormal returns are significant, especially for value stocks, and is not sensitive to time.

To conclude, our findings support that an investor, using a simple fundamental-based stock selection strategy, can obtain abnormal returns and it thus exemplifies a failure of the semi-strong form of market efficiency. Furthermore, we show the existence of long-term market mispricing, in line with the Piotroski's F-Score (2000) and several previous research on stock selection strategies.

Our paper is structured as follows: first we present the relevant theoretical framework. In section three, the F-Score strategy and Piotroski's findings are discussed in detail. Section four describes our methodology regarding sample, calculations and definitions. In the fifth section, we present and discuss the results. Lastly, we discuss the conclusions and implications of the findings.

2. Theoretical Framework

This section describes the theoretical framework related to Piotroski's (2000) F-Score strategy. The Efficient Market Hypothesis is the most relevant theoretical base related to the performance of fundamental-based investing. We also discuss Asset Pricing Theory, which becomes important when trying to explain return in terms of common risk factors.

2.1 The Efficient Market Hypothesis

This paper investigates a simple fundamental-based investment strategy with the goal of shifting the distributions of actual returns. Fama (1965) argues that this is not possible in an efficient market since, on the average, perfect competition will make information on intrinsic value to be reflected "instantaneously" in actual prices. The information could be related to either actual or expected changes that impacts the firm's prospects. Therefore, stocks always trade at their fair value and can be neither over- nor undervalued.

This constitutes the theoretical framework Efficient Market Hypothesis [EMH] (Fama, 1970), that defines three different forms of efficient markets. Below we define, explain and relate these different forms of efficiency and discuss potential anomalies.

2.1.1 Three Forms of EMH

An efficient market should fully reflect all available information. Fama (1970) tests the market efficiency regarding three natures of information: weak, semi-strong and strong. According to the weak form of efficiency, trading based on historical prices (e.g. technical analysis) should not generate abnormal returns. Semi-strong efficiency adds other public information such as earnings and dividend announcements. From this perspective, Piotroski's fundamental-based strategy cannot reveal arbitrage opportunities. Lastly, the strong form of the EMH includes all available information including, but not limited to, private information.

Fama (1970) finds significant support for both the weak and the semi-strong form of market efficiency, while he shows a couple of exceptions from the strong form of efficiency. Nevertheless, he argues that the strong form should be used as a benchmark for discussing deviation rather than representing the correct picture of reality. Furthermore, the EMH predicts a variance in traders' success due to future positive information announcements. However, these announcements must be impossible to predict, otherwise they would already have been incorporated in the price.

2.1.2 Market Anomalies

Efficient Market Hypothesis is one of the most recognised theories in financial economics, but researchers have had problem replicating it in practice. Such contradiction is known as anomalies and EMH argues that they will disappear in the long run as they get discovered by the market. Nonetheless, prior research, as the likes of Hawawini and Keim (2000), have found example of observed anomalies to be continuous. Furthermore, Schwert (2003) argues that documented anomalies may rather change form or impact than necessarily disappear.

Several theories have been made to why anomalies arise and persist. Anomalies can be related to time variant return patterns e.g. stocks prices moving more on Fridays compared to Mondays, known as the weekend effect (e.g. French, 1980). Returns can also be related to fundamental value, in which an anomaly is observed through cross-sectional return patterns. An example is Banz (1981) who introduced the size effect, that smaller firms outperforms larger firms. Fama and French (1992) focus on the risk-based explanation and argue that the size effect is due the increased risk. This will be discussed more in detail in the following sub-section.

2.2 Asset Pricing Models

Asset pricing models are used for estimating expected return to test the performance of a stock selection strategy, like Piotroski's (2000) F-Score strategy. The expected return is what an estimated model predicts the return of the security to be, given certain characteristics e.g. risk. The model is meant to represent an efficient market (e.g. Fama, 1970). The difference between expected and actual return is the abnormal return.

One possible pitfall in the calculation is that testing abnormal returns per definition requires the researcher to test two hypotheses at the same time (Cuthbertson, 1996). On the one hand, it tests whether the market is inefficient in the sense that it has created abnormal returns. However, it also tests if the market model represents an efficient market. Consequently, concluding that an investment strategy created significant abnormal return may be incorrect, since another possible explanation is that the asset pricing model is inadequate. This is commonly referred to as the "joint hypothesis problem".

To summarize, the choice of model is crucial because of the sensitivity of the results to the chosen model. In the following sub-sections, we will describe the most recognised models and factors used in determining the expected return.

2.2.1 The Capital Asset Pricing Model

Markowitz (1952) started to analyze portfolio theory using the mean returns and the variance of assets. Under the assumption that investors are risk averse, he concludes that they will always choose the less risky asset, given the same expected return. The Markowitz Portfolio Theory was the predecessor of the Capital Asset Pricing Model [CAPM] which independently developed by Sharpe (1964), Linter (1965) and Mossin (1966). In the CAPM there is an important distinction made between two types of risks: systematic and idiosyncratic. Systematic risk concerns the whole market, while idiosyncratic risk relates to firm specific events. Consequently, only idiosyncratic risk can be diversified away by holding several assets, since events between firms are assumed to not be correlated. The CAPM formula can be summarized as:

$$R_{it} = R_{ft} + \beta_i\left(R_{Mt} - R_{ft}\right) \quad [1]$$

The systematic risk is captured by $\beta_i(R_{Mt} - R_{ft})$ where the coefficient β_i is the security's sensitivity to changes in the return of the market and $R_{Mt} - R_{ft}$ describes the market risk premium, difference in return between the market portfolio and the risk-free rate. In addition to systematic risk the expected return of the security R_{it} is compensated for time, opportunity cost of the risk-free asset, represented by the term R_{ft}.

Investors will thus only be compensated for systematic risk, since it is the highest possible return and variance trade-off that can be found using the completely diversified market portfolio and leverage. As a result, investors can choose how to construct their portfolio after their risk aversion. For example, investors willing to take minimal risk invest a substantial proportion in the risk-free asset and less in the market portfolio.

The CAPM has also received plenty of criticism for its simplicity, assumptions and lack of ability to fully imitate an efficient market. For example, Roll (1977) argues that a stock market index, which is widely used as the market portfolio estimate in the CAPM, is insufficient and the market portfolio would need to include all instruments with market value e.g. commodities. Moreover, Fama and French (1993) argues that several other factors, e.g. size and value, are necessary to explain the return. Fama (2004) concludes that "the problems are serious enough to invalidate most applications of the CAPM."

2.2.2 The Fama and French's Three-Factor Model

Following the critique of CAPM's practical usefulness, research has, since the release of the theory in mid '60s, been focusing on finding additional factors that can explain return. The term multifactor model, which include multiple common risk factors, was introduced in the Arbitrage Pricing Theory, by Ross (1976). However, Ross does not present specific factors, instead he argues that each investor should choose which and how many factors, they consider relevant for their analysis.

Graham and Dodd (1940) coined the term "value investing" in which accounting numbers are used to calculate a firm's intrinsic value (true value) to find under- and overvalued stocks. In a semi-strong form of efficient market this would be impossible since the information is available for everyone. Nonetheless, variables related to value investing have been shown to explain returns (Basu, 1977). Rosenberg et al. (1985) argues for a "value effect;" that stocks with high book-to-market value (so-called value stocks) have historically outperformed stocks with low book-to-market value (growth stocks). Many following studies confirm these findings including De Bondt and Thaler (1987), and Fama and French (1992).

Fama and French (1993) further argues the value effect is risk-based and therefore add it as an additional risk factor to the CAPM. Chen and Zhang (1998) support their claim and argue that value effect is a consequence of high B/M firms being characterized by high financial leverage, uncertainty regarding future earnings and financial distress. La Porta et al. (1997), on the other hand, gives an alternative explanation to the difference in performance between value and growth stocks. He argues that investor's expectations about future growth is too optimistic, leading to systematic mispricing. Piotroski (2000) agrees with the latter explanation and argues that the systematic mispricing is due to the market being too slow in incorporating value-relevant information about value firms.

Banz (1981) argues for a "small firm effect" which also suggest a mispricing. He finds that common stocks of small firms have, on average, higher risk-adjusted returns than larger firms. The effect is distinguishable for very small firms but he does not find a significant difference between medium- and large-sized firms. Whether the size effect is a risk factor or related to other size characteristics are debated. Other papers argue that the size effect is a liquidity effect, since larger firms have higher trading activity and investors are willing to pay a premium to easier liquidate their position without impacting the stock price (e.g. Chordia et al 1998). Fama and French (1992) agree that liquidity is one cause of the size effect, however, argue there are several other important effects related to size. The size and value effects are the base of Fama and French's (1993) three-factor model, one of the most recognized extensions of the CAPM, which can be seen below:

$$R_{it} = R_{ft} + \beta_{mi}(R_{Mt} - R_{ft}) + \beta_{si}SMB \ + \beta_{hi}\text{HML} \qquad [2]$$

The expected return is, in addition to the CAPM (see function [1]), explained by the factors Small-Minus-Big (size effect) and High-Minus-Low (value effect). To calculate the factors, the authors create portfolios based on book-to-market and size break-points. They find that the model has higher explanatory value and conclude that CAPM fails to capture important risk factors and thereby also fail to explain the return, leading to mispricing.

Multifactor models, e.g. Fama and French's (1993) three factor model have, however, been criticised among some researchers. Lo and MacKinlay (1990) argue that in hindsight; you can always find and show deviations from CAPM through data snooping, in which researchers look for statistical significance without an initial hypothesis. Moreover, multifactor models have yet to alone resolve CAPM deviations (MacKinlay, 1995).

2.2.3 Additional Factors

Apart from the two of the most recognised models within asset pricing theory, explained above, there is still several other influential and recognised effects. Jegadeesh and Titman (1993) examine that a stock that have performed well (poorly) will continue on that trend and through going long (short) an investor can earn abnormal return. The tendency that stocks previous performance is a leading indicator of its future performance is known as the momentum effect and is a contradiction of the weak form of market efficiency. However, they argue that the momentum effect is limited to around twelve months and if an investor hold the investment for another year then around half of the return is lost. Jegadeesh and Titman (1993) argue that the momentum effect is not related to systematic risk and instead they argue that it is due to behavioural causes and delayed price reactions to firm-specific information. Carhart (1997) incorporates the momentum effect (PR1YR) in Fama and French's (1993) three factor model, and finds that this model, see below, has improved explanatory power of expected return.

$$R_{it} = R_{ft} + \beta_{mi}\left(R_{Mt} - R_{ft}\right) + \beta_{si}SMB + \beta_{hi}\mathrm{HML} + \beta_{pi}PR1YR \qquad [3]$$

Amihud (2002) examines another potential risk factor in illiquidity risk premium. He argues that if investors expect illiquidity, they will require a higher return to compensate for difficulties in selling without affecting the price. However, if the illiquidity is unexpected, it has a negative effect on the return.

Fama and French (2015) further extend their own three factor model. They presented two additional factors: profitability (robust-minus-weak) and investment (conservative-minus-aggressive). The profitability factor forecasts that stocks with high operating profitability outperform and the investment factor forecasts that firms investing more conservatively have higher returns. The extended model is summarized below:

$$R_{it} = R_{ft} + \beta_{mi}\left(R_{Mt} - R_{ft}\right) + \beta_{si}SMB + \beta_{hi}\mathrm{HML} + \beta_{ri}RMW + \beta_{ci}CMA \qquad [4]$$

However, despite increasing the number of factors, it still has drawbacks. According to Fama and French (2015), the five-factor model fail to capture the low average returns on small stocks whose returns behave like those of firms that invest a lot despite low profitability. Furthermore, the model has been criticized for not including the momentum factor.

In conclusion, opinions differ about how many and which factors to use and since more is lowers simplicity and usability. Ericsson and Karlsson (2004) test 15 different observable factors and try a total of 32768 model combinations on forty years of U.S. stock data. They only find robust evidence for the three factors from Fama and French (1993) while the fourth most supported factor: momentum, is sensitive to sample selection.

As previously discussed, there is a risk of data-mining and -snooping when including additional factors (e.g. Lo and MacKinlay, 1990). Some of which is the risk of overfitting problem where the model describes the random error in the sample instead of the underlying relationship. Moreover, a new factor is supposed to add a dimension but if it is already reflected in another factor then it causes lack of robustness. Lastly, the extent of behavioural finance have been heavily discussed recently, questioning the CAPM and other asset pricing models on the assumption of rational investors (e.g. Dempsey, 2013 & Akerlof and Shiller, 2009). However, that discussion is beyond the scope of this paper.

3. Previous Literature

Piotroski (2000) believes the market is too pessimistic in the pricing of value stocks and therefore designs a strategy to exploit that stand-point. First will be a description of the F-Score, followed by Piotroski's method, results and concluding by reviewing follow-up studies.

3.1 The F-Score Strategy

Piotroski's (2000) focuses on the stocks within the highest quintile of book-to-market, so-called value stocks. He argues that value stocks benefit more from fundamental analysis than growth stocks, the lowest quintile of book-to-market, due to growth stocks valuation is based on long-term forecasts while value stocks are more short-term oriented. Thus, annual reports are more useful in the valuation of value stocks. Piotroski incorporates prior research on the quintile to the choice of signals and its interpretations. E.g. Chen and Zhang (1998) argue that value stocks are more likely to be financially distressed and increased leverage therefor signals a decreased chance of meeting future obligation. Piotroski therefore incorporates change in leverage. He also states that investors find value stocks more uninteresting and, consequently, they are not as followed by analysts, making it more likely that exploitable mispricing exists.

Piotroski (2000) therefore creates nine simple fundamental-based signals (see table 1) to assess firms' financial situation. The signals are binary since the strategy should be simple and feasible to implement in practice. Piotroski then defines the F-Score as the sum of the nine signals with a higher score representing better expected return development.

The nine signals are divided in to three categories: profitability, leverage/liquidity and operating efficiency. Profitability assess the ability and long-term sustainability to generate future cash flow. Leverage/liquidity focus on the capital structure and capacity to meet obligations. Lastly, operating efficiency consists of the two components in DuPont that describes the margin and how efficient the firm uses its capital.

Piotroski (2000) conducts several statistical tests if the F-Score can identify over- and underperforming firms. He tests the difference between high F-Score firms; and first low F-Score firms, and then towards all value firms. He rejects his null hypothesis for both cases at the 1 percent significance level, concluding that F-Score has an ability to identify over- and underperforming firms.

Although Piotroski (2000) argues that most the signals are new, there exist several overlaps with other known investment strategies. Since these mispricing's are known and Piotroski's F-Score can identify over- and underperformers, the strategy must be based on continuous errors in the pricing of the stocks.

Table 1
Summary of the Nine F-Score Signals

Category	Name	Description	F-Score
Profitability	ROA	Return on assets	1 if positive 0 if negative
Profitability	CFO	Cash flow from operations	1 if positive 0 if negative
Profitability	ΔROA	Change in return on assets	1 if positive 0 if negative
Profitability	ACCRUAL	CFO/Assets - ROA	1 if positive 0 if negative
Leverage/liquidity	ΔLEVER	Change in long-term debt	0 if positive 1 if negative
Leverage/liquidity	ΔLIQUID	Change in current ratio	1 if positive 0 if negative
Leverage/liquidity	EQ_{OFFER}	Issue of common equity	0 if positive 1 if negative
Operating efficiency	ΔMARGIN	Change in gross margin	1 if positive 0 if negative
Operating efficiency	ΔTURN	Change in asset turnover ratio	1 if positive 0 if negative

The table summarizes the nine accounting-based variables in Piotroski's (2000) F-Score strategy. The signals are binary and can thus have a value of 1 (good) or 0 (bad). The sum of the signals result in an aggregate score called the F-score that can have a value 0 to 9, which is supposed to evaluate the financial situation of a firm. Piotroski's strategy is to buy high F-Score firms (8 or 9) and short low F-Score firms (0 or 1), with a holding period of one year.

3.2 The Methodology and Results of Piotroski's Study

To test for continuous error, Piotroski's (2000) collects a sample consisting firms on the U.S. stock market over 21 years, between 1976 and 1996, gathered from the database Compustat. Following will explain how Piotroski calculates the F-Score and returns.

Piotroski firstly calculates the book-to-market ratio at each firm's fiscal year end during the past year, divides them in to quintiles to only continue using firms from the highest B/M quintile. He then calculates the F-score based on the following fiscal year's financial statements and four months after the fiscal period end he long (short) firms with high (low) F-Score, an F-Score of 8 or 9 (0 or 1). The investment takes place four months after the fiscal period end because the U.S. law prohibits firms to further delay the release of the annual report. He argues that the four-month assumption reflects the practical implementation of when the information must be public. After firms are included in the portfolio they are hold for twelve months, but since the portfolio is equal-weighted, portfolio size can differ between periods.

Take as an example, a company with the calendar year as their fiscal year and assume they are in the highest B/M quintile based on the numbers at the end of 1980. The F-Score will be calculated based on the information for the next fiscal year (1981/01/01 – 1981/12/31). If the firm is considered to either have a high (8 or 9) or low F-Score (0 or 1), it is added to the portfolio. The investment is done four months after the fiscal year end, 1982/05/01, and liquidated one year later, 1983/04/30.

Piotroski calculates the return if the stock is longed (shorted) as: the sell price (buy price) divided by the buy price (sell price) minus one. If a stock delists he assumes the return to be zero, since buyouts and bankruptcies are assumed to balance out each other. Piotroski finds that high and low F-Score firms respectively represent approximately 10 percent and 3 percent of the entire sample. Moreover, the stock selection strategy yields an average annual return of 23 percent over the sample period, which is 7.5 percentage points higher than for the whole B/M quintile. He concludes that using the strategy shifts the entire return distribution to the right, which is important since the sample distribution have a negative median market-adjusted return.

One potential concern with the conclusion is whether the returns are realizable in practice, due to the strategy's reliance on liquidity and ability to short small firms. He tests in respect to size, share price and share turnover terciles and find enough significance to conclude that F-score is realisable. In addition, Piotroski tests whether previously recognised anomalies, e.g. momentum, accruals and equity offerings, can explain the difference in returns but does not find significant increased robustness. Furthermore, Piotroski argues that the investment strategy is also not time-dependent.

One of the major criticism of Piotroski (2000) is the lack of test with a risk-based explanation. Instead, he provides three arguments to reject concerns related to risk. Firstly, he argues that different B/M and size characteristics among different F-Scores are unlikely to explain such a significant difference in returns. Secondly, he tests the relationship between F-Scores and return on assets the next year. The results show that the high F-Score firms show best future financial performance and thus, an explanation with regard financial distress, in line with e.g. Chen and Zhang (1998), is improbable. Finally, using historical performance signals, he concludes that the firms with highest subsequent returns have lowest financial and operating risk before the investment.

To sum up, Piotroski (2000) finds that a simple fundamental-based stock selection strategy, the F-Score, can identify over- and underperformers within the value stocks segment. He believes that the market is too pessimistic in their view of value stocks, but whether his results are due to market inefficiency or an undetected anomaly is left unanswered.

3.3 Follow-Up Research

The follow-up research mainly consists of three types: a continuation study by Piotroski and So, replications studies and a new variant of the F-Score.

Piotroski and So (2012) focus on finding an explanation to the value effect, in terms of the risk versus mispricing view. They investigate whether there exists congruence between a firm's expectations implied by current price and the strength of the fundamentals, measured by the F-Score. They conclude, and further strengthens the results of Piotroski (2000), that the underlying reason to the abnormal returns are systematic expectation errors and the existence of market mispricing.

There is an extensive amount of replication studies on different markets, including emerging markets (Hyde, 2013), the U.K. market (Rathjens and Schellhove, 2011) and the Indian market (Aggarwal and Gupta, 2009). They all support F-Score strategy ability to identify over- and underperformers and most papers also check their results against recognised risk factors. However, Rathjens and Schellhove (2011) find that the nine signals work better for growth stocks than for value stocks on the U.K. market.

The most cited follow-up study of Piotroski (2000) is Mohanram (2005) who presents an alternative variant of the F-Score strategy. He defines eight binary signals and calls the aggregate sum G-Score. The strategy differs from Piotroski's in two ways. Firstly, signals are quite different since the strategy is designed for growth stocks. Secondly, several of the signals are benchmarked against industry averages, making them more complicated to calculate. In line with Piotroski, Mohanram finds evidence that supports the mispricing view rather than the risk-based explanation. Nevertheless, the returns of the strategy do, compared to Piotroski, more heavily rely on shorting stocks. Potential criticism of the G-score therefore concerns whether the abnormal returns are realizable.

4. Data and Methodology

The following section presents thoroughly this paper's methodology and how it differs from Piotroski (2000). In other words, we explain definitions, calculations and potential selection biases.

4.1 Sample Selection

The investment period is set from 1 January 1996 to 31 March 2017, which means that we consider publicly available F-Scores until the end of March 2016. Data are collected from the Thomson Reuters Datastream (see appendix 2 for a full list of all used variables). The selection criteria are stocks on the Stockholm Stock Exchange. The clear majority of stocks on the Swedish Market is thus included, to avoid potential bias if the sample only contains of companies with certain characteristics. The choice of Sweden as the market of interest has primarily two explanations. Firstly, our assessment is that previous research within the area F-Score on the Swedish Market are weak both in terms of the statistical methods and risk-adjusting. Secondly, it is the market we have knowledge and interest in.

The choice to drag the end date to March is to increase the sample size, as an overwhelming majority of all firms report the information needed to compute the F-Score, during the first three months of the year. The investment period is therefore approximately 21 years, the same number as Piotroski (2000). The high number of years enables to assess the strategy regardless of the macroeconomic conditions. The period includes extreme financial climates the Dot-Com Crash in 2000 and the Financial Crisis of 2007-2008. However, in terms of abnormal returns, the market condition should be reflected in the market risk factor. Moreover, the period includes years both before and after the release of Piotroski's paper in 2000, which is interesting from the perspective of Schwert's (2003) argument that anomalies change form once they become public.

4.2 The Strategy

In many respects, we search to replicate the paper from Piotroski (2000). However, giving special focus on aspects where the F-score has been criticised, as with risk-adjusting and abnormal returns, rather than questing the fundamental rational that has been the focus of several other papers, e.g. how differences in accounting standards impact the results. Some problems arise when we attempt to replicate the F-Score computation made of Piotroski, as he uses another data collector, Compustat, with some differences in how they define variables. The choice of using Datastream were due to Compustat only relatively recently added data for the European market.

Rathjens and Schellhove (2011) examine the U.K. market with data provided by Datastream and identify two differences in the F-Score computation. Firstly, to correct for differences in "Cash Flow from Operations", they use the variable "Funds from Operations" and deduct change in working capital. Secondly, Datastream does not provide a variable on whether a firm has issued equity. As an approximation Rathjens and Schellhove use two criteria; "Common Shares Outstanding" must have changed and "Net Proceeds from Sale/Issue of Common & Preferred" must be positive. We decide to use these definitions due to using Datastream and since we find the changes to also be applicable on the Swedish market. For more detailed definitions of all F-Score variables, see table 2. The sum of the nine binary signals result in an aggregate score (F-Score) which the following function describes:

$$F_{SCORE} = F_{ROA} + F_{\Delta ROA} + F_{CFO} + F_{ACCRUAL} + F_{\Delta MARGIN} + F_{\Delta TURN} + F_{\Delta LEVER} + F_{\Delta LIQUID} + F_{EQ_{OFFER}} \quad [5]$$

For a firm to be included, all the necessary information to compute the F-Score must be available. Here is a potential bias since companies with certain characteristics are more likely lack the necessary data, such as low market capitalization (small firms) or gross margin for banking firms. To minimize the number of drops we use variables with high number of observations from Thomson Reuters Worldscope via Datastream.

In contrast to Piotroski (2000), we do not solely apply the strategy for the highest B/M quintile, since previous studies, e.g. Rathjens and Schellhove (2011), show that the strategy could work better for other quintiles. Therefore, we choose to analyse all B/M and size quintile. This also increases the comparability with previous studies within the area. The classification of quintiles is made according to Piotroski's (2000) methodology, where B/M and size quintiles at this year (t) are based on the book and market capitalization values at the fiscal period end, the year before (t-1).

In previous research, mainly two different definitions have been used for high and low F-Score firms. On the one hand, Piotroski (2000) has a extensive sample and therefore define high (low) F-Score as 8 or 9 (0 or 1). Follow-up studies have increased the classification to include F-scores of 2 and 7, making the breakpoints between 0 to 2 and 7 to 9. The advantage with more narrowly defined F-scores ranges is that it better represents the ability of the strategy to identify over- and underperformers. However, it decreases the number of observations, which may cause a lack of robustness. For these reasons and the characteristics of the Swedish Stock Exchange, we have chosen to use the wider range of 0 to 2 and 7 to 9.

Table 2
Definition of the Nine F-Score Signals

NAME	DEFINITION
ROA	(Net Income before Extraordinary Items/Preferred Dividends$_t$) / Total Assets$_{t-1}$ $F_{ROA,t}$=1 if > 0 and $F_{ROA,t}$=0 if < 0
CFO	Funds from Operations$_t$ - Increase/decrease in Working Capital$_t$ $F_{CFO,t}$=1 if > 0 and $F_{CFO,t}$=0 if < 0
ΔROA	$ROA_t - ROA_{t-1}$ $F_{\Delta ROA,t}$=1 if > 0 and $F_{\Delta ROA,t}$=0 if < 0
ACCRUAL	CFO_t / Total Assets$_{t-1}$ - ROA_t $F_{ACCRUAL,t}$=1 if > 0 and $F_{ACCRUAL,t}$=0 if < 0
ΔLEVER	Long Term Debt$_{t-1}$ / (0.5*Total Assets$_{t-1}$ + 0.5 * Total Assets$_{t-2}$) – Long Term Debt$_t$ / (0.5*Total Assets$_t$ + 0.5 * Total Assets$_{t-1}$) – $F_{\Delta LEVER,t}$=1 if > 0 and $F_{\Delta LEVER,t}$=0 if < 0
ΔLIQUID	Current Ratio$_t$-Current Ratio$_{t-1}$ $F_{\Delta LIQUID,t}$=1 if > 0 and $F_{\Delta LIQUID,t}$=0 if < 0
EQ_{OFFER}	Common Shares Outstanding$_t$-Common Shares Outstanding$_{t-1}$+ + Net Proceeds from Sale/Issue of Common & Preferred$_t$ $F_{EQ_{OFFER},t}$=1 if = 0 and $F_{EQ_{OFFER},t}$=0 if ≠ 0
ΔMARGIN	Gross Profit Margin$_t$-Gross Profit Margin$_{t-1}$ $F_{\Delta MARGIN,t}$=1 if > 0 and $F_{\Delta MARGIN,t}$=0 if < 0
ΔTURN	Total Asset Turnover$_t$-Total Asset Turnover$_{t-1}$ $F_{\Delta TURN,t}$=1 if > 0 and $F_{\Delta TURN,t}$=0 if < 0

The table gives a detailed description of the computation of the F-Score for each firm and year. The F-Score strategy consists of nine accounting-based signals, which can have a value of 1 (good) or 0 (bad). The data were retrieved from Thomson Reuters Datastream and for more detailed information regarding variables and ID numbers, see appendix 2.

4.3 Calculation and Assumptions of Returns

As we make our calculations and investment in hindsight, it is essential that we try to replicate the most realistic scenario possible for an investor to avoid a bias. Therefore, we must ensure that the information, for which the investment is based on, is available at the time of an investment. In contrast to most of the previous research, including Piotroski (2000), we do not use the simplified assumption that the investment takes place when a company by law must have released its annual report. In the case of the Swedish Market this would have been seven months from the fiscal period end. However, that assumption would be unrealistic since the quarterly earnings report for the last quarter, which includes all necessary information to calculate the F-Score, normally is released already within three months after the fiscal year end. The seven months' investment assumption would thus mean, once companies the year after announce their new financial numbers early spring, the return for the next half-year would reflect the new F-Score. To increase the more practical and actual F-score analysis between two reporting dates, we use the variable provided by Datastream called "Earnings per Share Report Date Fiscal Year End", making it possible to match the investment with the exact date of the quarterly earnings report release. Since an investor also knows, often weeks in advance, when the new year's financial statements are released; the stock is held until the day before the next year's report release.

To calculate actual returns we use Datastream's variable "Total Return Index" since it adjusts for changes in price that is not directly related a change in valuation from the market e.g. dividends and splits. In practice, it means that dividends are reinvested which is a reasonable assumption to make. Moreover, we use closing prices which is essential due to lack of information regarding when during the day, the report is released. Even if we had that information it would not be appropriate to use something else than closing prices, since it takes time to compute the F-Score and the aim of the strategy is to exploit a one-year market mispricing, which should not depend on precision-based intraday trading. The exact computation of the return in a long position is described in the function below. If the position is short, it applies the inverse relationship between the total return indexes.

$$Return = \frac{Total\ Return\ Index_{Report\ Date-1,\ year\ t+1}}{Total\ Return\ Index_{Report\ Date,\ year\ t}} - 1 \qquad [6]$$

OMXSPI is chosen as the proxy for the market portfolio because it is a value-weighted all-share index that includes all stocks on the Stockholm Stock Exchange. To calculate the market-adjusted return for a stock, we match the return of the stock [6] with the return of the market for the same period [7].

$$Market\ Return\ = \frac{OMXSPI_{Report\ Date-1,\ year\ t+1}}{OMXSPI_{Report\ Date,\ year\ t}} - 1 \qquad [7]$$

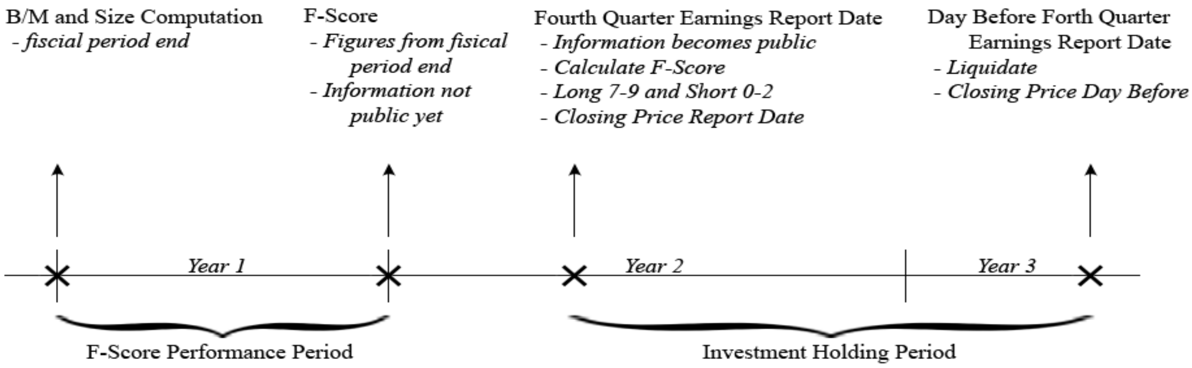
$$Market\ Adjusted\ Return\ = Return - Market\ Return \qquad [8]$$

We follow Piotroski's (2000) simplified assumption that if a firm delists; its return is zero during the period, with the argument that it does not necessarily mean that they are in trouble. On the one hand, a firm may delist involuntary; due to bankruptcy or inability to meet financial covenants set by the stock exchange. On the other hand, it may have beneficial reasons, for example through a buyout from a private equity firm which is associated with positive premium effects on the share price. This argumentation has however been criticized, since there is a greater chance for low F-Score firms, that is subject of shorting, to be delisted for involuntary reasons. Thus, there are differing ways the assumption can impact our results.

Another assumption made is that if the total return index of a firm is the same between the two dates, the observation is not included. The reason is that our sample partly consists of untradeable stocks with a special purpose over a limited life time, e.g. redemption shares. These stocks have a fixed total return index and their returns are therefore zero. Furthermore, they are issued by firms that already have a regular, publicly traded share, but still uses the same underlying financials, making them misleading. There is a risk that other stocks drop out because of the made assumption because total return index uses two decimals, but the effect is minimal.

Lastly, like Piotroski (2000) suggests, the F-Score portfolio is equally weighted among all holdings, which in practice means that an investor before each year decides a fixed amount that will be invested in high and low F-Scores. Thus, the portfolio changes with the mean return of all stocks within the portfolio. An equally-weighted portfolio enables increased diversification and thus less exposure to movements in certain stocks. On the other hand, it creates difficulties in the implementation of the strategy, especially because an investor cannot tell how many investment opportunities that will arise during the year. However, the trade-off is necessary, since eventual rebalancing would conflict the crucial assumption about a passive holding period of twelve months.

Figure 1
Example of the Strategy in Practice for one Stock



The figure shows an example of a firm with the calendar year as their fiscal period. The potential investment in year 2 is based on the numbers from the fiscal period end (year 1), which is reported in the fourth quarter earnings report. The investment is liquidated the day before the next year's report release.

4.4 Asset Pricing Models and Abnormal Return

As discussed in the section "2.2 Asset Pricing Models" one need an asset pricing model or other kind of benchmark to calculate if returns are abnormal. We also discussed some of the critique directed to the CAPM model e.g. its simplicity, and for that reason it is appropriate to include additional factors. Since the strategy tries to identify over- and underperformers within the value stock segment; Fama and French's three factor model is a natural starting point since it includes a factor on book-to-market ratios. Moreover, there are several studies on the Swedish market that concludes a high explanatory power of the model, regardless of market conditions (e.g. Kilsgård and Wittorf, 2010).

However, one can argue the need of additional factors to improve it as benchmark. E.g. Kobelyatskyy and Fulgentiusson (2011) show that the momentum factor is significant and adds explanatory value on the Swedish market. On the other hand, Ericsson and Karlsson (2004) argue that momentum is sensitive to sample selection and does not improve explanatory value over the three-factor model. Furthermore, it is questionable if the effect is related to risk or other more behavioural and irrational causes. As a result, we have chosen to use the Fama and French's (1993) three factor model as our regression model, described in equation 9.

$$R_{it} - R_{ft} = \alpha_i + \beta_{mi}(R_{Mt} - R_{ft}) + \beta_{si}SMB + \beta_{hi}\text{HML} + \varepsilon_{it} \quad [9]$$

When calculating the factors HML and SMB, Fama and French (1993) divide all stock in to breakpoints based on size (market capitalization) and book-to-market ratios. Size is divided in to two equally large groups, called small and big, while book-to-market is divided in to three groups, with the 30 and 70 percentiles as breakpoints. In contrast to Fama and French we construct the portfolios at the end of July each year because the Swedish regulation is more generous and allows the annual report to be published up to seven months after the fiscal year end. The size factor is therefore based on the market capitalization at July 31 same year. The B/M ratios for the year (t) are the book value of equity for the fiscal end the year before (t-1) and market capitalization at the end of December the same year (t-1). The different breakpoints result in six different portfolios for the period August 1 (year t) to July 31 the year after (year t+1). The portfolio construction logic is described in Figure 2.

Figure 2
The Construction of Portfolios to Calculate the HML and SMB Factors

		B/M		
		Low (30%)	Neutral (40%)	High (30%)
Size	Small (50%)	S/L	S/N	S/H
	Big (50%)	B/L	B/N	B/H

Description of how the six portfolios are constructed based on book-to-market (B/M) and size breakpoints.

In line with Fama and French, we drop firms with a negative book-to-market value. Furthermore, for a company to be included, the data for book value and market capitalization to construct the portfolios must be available. Since report dates differ both between firms and years within the same firm; we need daily factors and returns and match them with the holding period. Since we use closing prices, the daily return of an individual stock is calculated as follows:

$$r_{i,d} = \frac{Total\ Return\ Index_d}{Total\ Return\ Index_{d-1}} - 1 \qquad [10]$$

The individual return of the six portfolios is the value-weighted return of all stocks within that portfolio. Value-weighting (w) is how large the firm's market capitalization is in relation to the capitalization of the whole individual portfolio. The weights are based on the market capitalization the 31st of July each year. Thus, the return in each of the six portfolios is calculated as:

$$Return_{Portfolio} = \sum w_i * r_i \qquad [11]$$

The main portfolios of interest is Small, Big, High and Low. To calculate the return of these we take the simple average of the underlying portfolios included in that breakpoint. An example for the portfolio Small is find below.

$$Return_{Small} = \frac{r_{Small/Low} + +r_{Small/Neutral} + r_{Small/High}}{3} \qquad [12]$$

The period over which we observe the factors in the regression must match with the returns of the investment. However, since the investment dates and horizons differ from firm to firm; we cannot rely on yearly factors, but daily returns. However, daily factors can create compounding error. To correct for this, one must remember that HML and SMB, per definition, is the difference in return between two portfolios. Therefore, one avoids compounding error through calculating the return of the individual portfolio and then match the returns of these portfolios with the specific holding period in question. We therefore focus on the individual main portfolios: small, big, high and low for which we create indexes with 1995-12-31 as base to 2017-03-31. At this date, the index starts with a value of 1 and changes with that individual portfolio's return each trading day. The method is similar to how the market return factor is calculated. This means that for each investment in a stock the matching SMB and HML factors can be calculated as follows:

$$Portfolio\ Index_{D=n} = 1 * (1 + r_{1996-01-01}) * (1 + r_{1996-01-02}) \dots * (1 + r_n) \quad [13]$$

$$SMB = \frac{Small\ Portfolio\ Index_{\ Report\ Date-1,\ year\ t+1}}{Small\ Portfolio\ Index_{\ Report\ Date,\ year\ t}} - \frac{Big\ Portfolio\ Index_{\ Report\ Date-1,\ year\ t+1}}{Big\ Portfolio\ Index_{\ Report\ Date,\ year\ t}} \quad [14]$$

$$HML = \frac{High\ Portfolio\ Index_{\ Report\ Date-1,\ year\ t+1}}{High\ Portfolio\ Index_{\ Report\ Date,\ year\ t}} - \frac{Low\ Portfolio\ Index_{\ Report\ Date-1,\ year\ t+1}}{Low\ Portfolio\ Index_{\ Report\ Date,\ year\ t}} \quad [15]$$

As earlier described, the benchmark for the market portfolio return is the OMXSPI index since it includes all stocks on the Stockholm Stock Exchange. The risk-free rate is the 3 months Treasury bill rates provided by Riksbanken. These rates change daily but have annual compounding. We approximate the daily risk-free return using the average of 252 trading days in a year (see equation 16). As a result, we can create a daily index for the Treasury bill, with the same logic as equation 13. The excess market factor for a given observation is therefore:

$$Daily\ Treasury\ Bill\ Rate = (Yearly\ Treasury\ Bill\ Rate + 1)^{\frac{1}{252}} - 1 \quad [16]$$

$$R_M - R_f = \frac{OMXSPI_{Report\ Date-1,\ year\ t+1}}{OMXSPI_{Report\ Date,\ year\ t}} - \frac{Treasury\ Bill\ Index_{Report\ Date-1,\ year\ t+1}}{Treasury\ Bill\ Index_{\ Report\ Date,\ year\ t}} \quad [17]$$

The Fama and French three factor model will often be used in one of two ways when testing if the strategy has created abnormal returns. Firstly, testing will be made from the perspective that the F-Score strategy is a portfolio and thus regress the annual excess portfolio returns with respect to the three factors. This implies that the number of observations will be the same as the amount of years where we invest (21), which can be considered as low. Nonetheless, the assessment is that this is the most relevant scenario since the strategy itself is about a one-year holding period and what happens between every investment entry and liquidation date is not the main focus. The goal of the regression is to test if the intercept is significant since it implies that the strategy has created significant abnormal. However, as discussed earlier, we also test how well the asset pricing model describes the returns and therefore the significance of individual betas and the r-squared are highly relevant. Our alternative hypotheses with the regression is that the F-Score stock selection strategy can create abnormal returns (H_1: $\alpha_i \neq 0$).

Secondly, we use the Fama and French's (1993) three factor model to predict abnormal return of every individual stock each year. The methodology is in line with Fama and French, in dividing all stocks into 25 portfolios based on book-to-market and size quintiles (see section 4.2 The Strategy for methodology). For each of the 25 portfolios' return, we run a regression whose model is used to find the expected return for the stocks within that specific quintile. The intuition is that using quintiles of proven factors, each portfolio has a unique risk characteristics and work as a relatively accurate predictor of the expected return of the stocks within the portfolio. Thus, it increases the robustness of the test described in the paragraph above since it accounts for individual stocks' exposure to risks related to size and book-to-market. Although the methodology is simplified and we make tests based on a prediction which decreases the robustness, our alternatives are few since each stock have a maximum of 21 return observations and in many cases even fewer. Thus, we must find similarities between stocks and group them based on risk characteristics. Moreover, the test is a complement to previously described tests.

5. Results

In simplifying terms, our research question tries to answer if the F-Score strategy works well as identifier of future stock performance and if so, why that is the case. Thus, we have chosen to divide the results in to two parts: raw returns and abnormal returns. Raw returns focus on the first part of the research question: how well F-Score predict future returns and thus can identify over- and underperformers. The next part, abnormal returns, focus on the explanation behind the strategy's performance i.e. whether the risk-based or mispricing view is the underlying cause.

5.1 Raw Returns

In this section, we focus on the strategy from a more general point of view and how the strategy performs, regarding to book-to-market, size and time. It thus also adds the element about when the strategy is working. The main goal in this section is to investigate the strategy's ability to find future over- and underperformers.

5.1.1 Descriptive Statistics

Table 3 and 4 treat the main descriptive statistics for the nine signals and the F-Score in relation to return. The total amount of observations with all necessary data is 5069. Since the amount of active stocks on the Stockholm Stock Exchange have been stable between 250 and 350 per year, the number 5069 indicate that a clear majority have had the necessary data.

Table 3
Descriptive Statistics of the Nine F-Score Signals

Variable	Proportion with positive signal
ROA	70.40%
CFO	76.30%
ΔROA	52.70%
ACCRUAL	64.80%
ΔLEVER	35.80%
ΔLIQUID	47.90%
EQ_{OFFER}	69.50%
ΔMARGIN	53.80%
ΔTURN	52.00%

The data consist of all stocks on the Stockholm Stock Exchange from January 1996 to March 2017. We compute the F-Score for each firm and year using the methodology summarized in table 2. The F-Score strategy consists of nine accounting-based signals, which can have a value of 1 (good) or 0 (bad). The table shows the proportion of the signals that have a value of one, indicating a positive signal for a firm's financial situation.

Table 4
Descriptive Statistics of the F-Score Strategy

F- SCORE	Mean	Std. Dev.	Min	Max	NR
0	-37.12%	0.4643	-86.00%	44.70%	6
1	9.82%	0.7417	-94.10%	201.40%	68
2	-2.98%	0.81	-94.90%	426.70%	187
3	10.11%	0.8847	-97.00%	888.00%	483
4	13.32%	0.6885	-100.00%	677.00%	825
5	16.95%	0.699	-99.00%	898.70%	1203
6	16.27%	0.5651	-97.30%	605.00%	1200
7	17.43%	0.54	-87.70%	579.80%	731
8	25.47%	0.5606	-66.00%	492.40%	303
9	21.04%	0.4021	-50.20%	141.20%	63
Low	-0.43%	0.788447	-94.86%	426.74%	261
High	19.86%	0.539654	-87.66%	579.84%	1097

The table presents return means and standard deviation for every F-Score. At the far right the number of firms in our sample with a specific F-Score is shown. The data consist of all stocks on the Stockholm Stock Exchange from January 1996 to March 2017. Piotroski's (2000) F-Score is supposed to measure a firm's financial position and the higher score the better. We compute the F-Score for each firm and year using the methodology summarized in table 2. The return of a stock is calculated using the closing price of the day the fourth quarter earnings report was released and the closing price the day before next year's report release (see function 6). This means that the holding period differs across firms and years, but is approximately one year.

The F-Score distribution resembles of a normal distribution with most of the observations, almost 90 percent, concentrated around F-Score between 3 and 7. Furthermore, one can observe that low F-Score firms are significantly fewer than high F-Score firms, with 5.15 and 21.64 percent respectively. The distribution is in line with Piotroski's results except for the proportion of firms with 0 in F-Score, which is considerably lower for our sample.

Importantly, we observe a positive relationship between F-score and returns. The relationship is not perfect, in the sense that the mean return increases for each F-Score, and most eye-catching is the almost 10 percent mean-return for F-scores of 1. However, when we cluster the values in low, medium and high F-score firms, the score work as qualified predictor of future stock returns. The mean difference in return between high and low firms is 20.29 percent (19.86 and -0.43 respectively), which is the first indicator that the strategy can separate out- and underperformers. Moreover, the relationship is quite monotonic as a change in F-Score not has a linear relationship with return, but instead changes the return at different rates across the distribution. Nonetheless, since the strategy takes the starting point in buying high and shorting low F-score firms, our results indicate that most of the portfolio returns derive from high F-Scores.

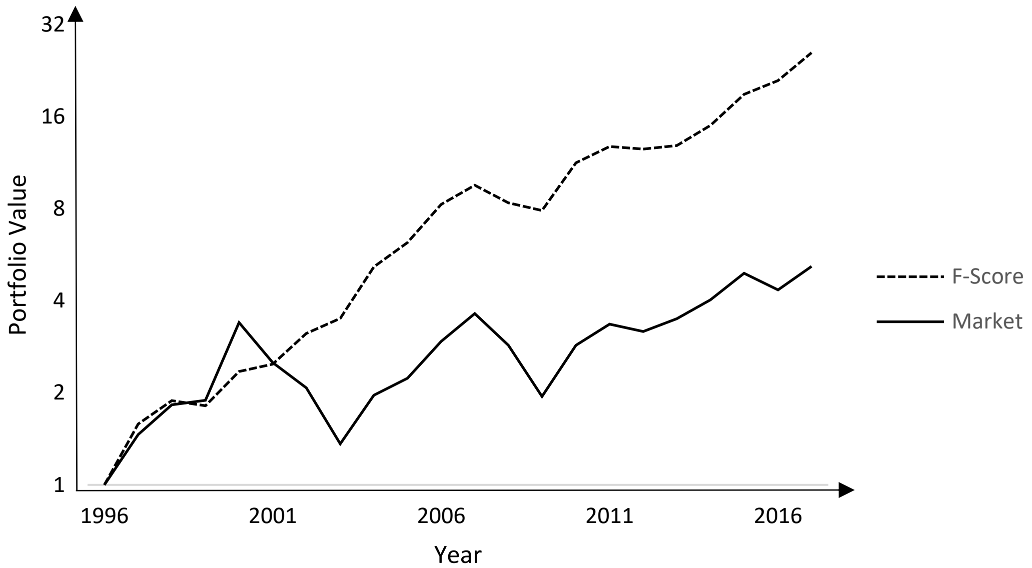
5.1.2 Market-Adjusted Returns over Time

Figure 2 describes the average development of the F-Score and market portfolio returns during our sample period. This is interesting from the perspective to see how returns relate to time and market conditions. The F-Score portfolio is equally weighted among all holdings and thus changes with the average return of all positions each year.

The graph indicates that if an investor started to follow the F-Score strategy at January 1996, his holdings would be worth approximately 25 times more at February 2017. This implies a mean annual return of 17 percent and an 8 percent mean market-adjusted return. The difference in performance, between the portfolio and the market, is noticeable to the advantage of the F-Score portfolio, if only just slightly before the new millennium.

In terms of raw returns, the portfolio handled the dot-com crisis well despite the overall negative market trend. During the financial crisis of 2007-2008 the return of the F-Score portfolio was negative, yet not to such an extent that the exposure to systematic risk seems to be what explains the return. The strong years have mainly occurred in the near future of the two crises, during a time when the market in general has been in a recovering phase.

Figure 3
Return of the F-Score and Market portfolio over Time



The graph shows the total return of the F-score and the market portfolio (OMXSPI) between 1996 and 2017. Both portfolios start with a value of one. The portfolio value (y-axis) has a logarithmic scale to mitigate the visual effect of compounding. The data consist of all stocks on the Stockholm Stock Exchange. Piotroski's (2000) F-Score is supposed to measure a firm's financial position and the higher score the better. The F-Score is computed for each firm and year using the methodology summarized in table 2. We define high (low) F-Score firms as firms with an F-Score of 7-9 (0-2). The investment strategy is to buy high F-Score firms and short low F-Score firms. The portfolio is equally-weighted between all stocks. We calculate the returns using the closing price of the day the fourth quarter earnings report was released and the closing price the day before next year's report release (see function 6). This means that the holding period differs across firms and years, but is approximately one year.

In terms of statistical power the usage of a graph is weak, but we have chosen to include it for its ability to visually summarize the returns over time. Furthermore, high market-adjusted returns are not necessarily an indication of success since it may be due to risk. Thus, more tests regarding how the strategy relates to time will be presented in the section “5.2 Abnormal Returns”.

5.1.3 Returns Conditioned on Book-to-Market

In contrast to Piotroski (2000), we do not solely evaluate the performance of the F-Score strategy within the highest book-to-market quintile. The raw returns conditioned on F-Score and book-to-market quintiles are summarized in table 5, where we also make t-tests of mean-return differences between high and low F-Score firms, as well as high and all F-Score firms. These tests contribute to conclusions about if the strategy successfully differentiates between over- and underperformers. It is then interesting to make comparisons across different quintiles, specially to set our results in relation to Piotroski’s findings.

According to Chen and Zhang (1998) high B/M firms are in general poor performing firms and from the perspective that the F-Score is supposed to evaluate the financial position of a firm, it is reasonable to assume that the F-Scores for value stocks are lower than for the average firm. However, our results show that proportionally the number of low F-Score firms are the highest for growth stocks (see appendix 1a). Growth stocks also have proportionally fewest observations with high F-Scores. Furthermore, Chen and Zhang argue that the average high B/M firm is in general smaller than the average firm on the market, which our sample also indicates (see appendix 3). This is an explanation to why the number of observations are lowest for the highest book-to-market quintiles, since small firms are more likely to lack all necessary data. A third observation is the historical phenomena that value stocks have outperformed growth stocks (e.g. high-minus-low in Fama and French’s three factor model) is supported in our results, since the average return of the highest quintile is almost three percent points higher than for the lowest quintile.

The observation in section “5.1.1 Descriptive Statistics” that the F-Score distribution resembles a normal distribution does not change when we condition returns based on book-to-market quintiles. Moreover, as we group F-Scores in low, neutral and high firms there is no problem for robustness because of potential low numbers of observations. The lowest amount of observations for a specific grouping is 36, for low F-Score firms in the highest B/M quintile. However, for specific F-Scores, the mean return can be affected by extreme values. For example, the average return in the second B/M quintile for an F-Score of 0 is -51.58 %, which derives from one single observation. But the robustness of specific F-Score is not of interest in this paper.

The results of the t-tests, between the mean returns of high and low F-Score firms, suggest that the strategy can separate future over- and underperformers across different B/M quintiles. The one-tailed t-test is significant at the one percent level for both value and growth stocks, with a mean return difference of 59.37 percent and 24.06 percent, respectively. Furthermore, the mean return difference between high and low F-Scores are positive across all quintiles. Nonetheless, biggest difference is found within the highest B/M quintile.

Table 5
Returns Conditioned on B/M

	Low		**B/M Quintiles**		*High*	
F - SCORE	1	2	3	4	5	All
0	-20.64%	-51.58%	-54.43%		-21.00%	-37.12%
1	3.73%	23.53%	2.47%	50.64%	-15.29%	9.82%
2	-16.46%	12.59%	14.07%	2.26%	-34.76%	-2.98%
3	11.08%	11.74%	18.01%	-4.83%	11.32%	10.11%
4	9.45%	16.89%	14.50%	9.11%	16.21%	13.32%
5	16.20%	20.69%	18.25%	14.83%	13.72%	16.95%
6	20.18%	20.49%	9.21%	18.09%	12.63%	16.27%
7	12.77%	16.26%	17.97%	19.47%	22.75%	17.43%
8	19.54%	22.96%	19.82%	21.44%	47.84%	25.47%
9	12.72%	10.18%	18.82%	26.14%	35.08%	21.04%
All	**12.83%**	**18.32%**	**15.23%**	**14.08%**	**15.67%**	**15.28%**
High	14.02%	17.64%	18.51%	20.63%	31.61%	
Low	-10.05%	13.27%	7.50%	10.80%	-27.76%	
High - Low	24.06%	4.37%	11.01%	9.83%	59.37%	
t-statistic	3.4290***	0.4052	1.3694*	1.1749	3.9868***	
High - All	1.19%	-0.68%	3.28%	6.90%	15.41%	
t-statistic	0.203	-0.1425	0.8531	1.7512**	2.6261***	

The table shows how return varies across different F-scores and book-to-market quintiles. Moreover, the table presents one-tailed t-tests between the return of high F-Score firms and low/all F-Score firms. The significance levels 10 %, 5 % and 1 % are shown with *, ** and ***. The data consist of all stocks on the Stockholm Stock Exchange from January 1996 to March 2017. Piotroski’s (2000) F-Score is supposed to measure a firm’s financial position and the higher score the better. The F-Score is computed for each firm and year using the methodology summarized in table 2. We define high (low) F-Score firms as firms with an F-Score of 7-9 (0-2). Book-to-market quintiles are calculated for year t using the book value of equity and market capitalization at each firm’s fiscal period end, year t-1. We calculate the returns using the closing price of the day the fourth quarter earnings report was released and the closing price the day before next year’s report release (see function 6). This means that the holding period differs across firms and years, but is approximately one year. The number of observations within each combination of F-Score and B/M quintile can be found in appendix 1, panel A.

Importantly, however, the difference between high and all F-Score firms is only significant at the one percent level for the highest quintile. In this quintile, the mean return difference between high and all firms is 15.41 percent, while for growth stocks the same difference is, surprisingly, negative. A negative difference between high and all do also apply to the second B/M quintile. Undoubtedly, the overall results strongly support that the strategy works best for value stocks, which is along the lines of Piotroski's argumentation. In line with his findings, we reject the null hypothesis between high-low and high-all at the one percent significance level.

The general observation in the descriptive statistics suggest that the clear majority of the returns derive from high F-Score firms and the mean return of low F-Score firms, which were recommended to be shorted, was almost zero. However, when returns are conditioned on B/M, the strategy is accurate, in the two highest quintiles, at finding substantial negative return for low F-Scores as well. Thus, in these quintiles, the low F-Scores, on average, contribute more substantially to the return of the F-Score portfolio. As a result, it raises some concerns regarding the possibility to short stocks, especially since high book-to-market firms are smaller in general.

5.1.4 Returns Conditioned on Firm Size

Next we look at returns, relative to firm size, which is important from the perspective of risk, size effect, and to test the feasibility of the strategy, as smaller firms are characterized by lower trading liquidity and generally fewer possibilities of shorting. The methodology is identical to the previous section and the results are presented in table 6.

The similarities of a normal distribution for F-scores still exists across different quintiles. However, for the largest companies the tails are smaller, indicating more financial stability. Furthermore, the linear relationship between returns and F-Scores are most evident in the two highest quintiles, while the mean return distribution is more random for the smaller firms. In our sample, the portfolio of firms in the lowest size quintile, on average, do not outperform the portfolio of firms in the highest size quintile. However, when calculating the Fama and Frenchs (1993) small-minus-big factor, we find support for the size effect in our sample period between 1996 and 2017. The most reasonable explanation to this incongruence is due to the numbers in table 6 is equal-weighted average, while small-minus-big factor is value-weighted.

Table 6
Returns Conditioned on Size

	Low		**Size Quintiles**		*High*	
F - SCORE	1	2	3	4	5	All
0	-21.00%	44.68%	-61.60%			-37.12%
1	12.38%	12.26%	15.32%	-20.11%		9.82%
2	-5.93%	-2.37%	5.13%	-3.61%	-15.75%	-2.98%
3	-0.47%	15.84%	20.04%	4.34%	14.66%	10.11%
4	16.12%	19.65%	13.16%	5.59%	12.49%	13.32%
5	20.30%	17.53%	18.94%	16.67%	12.40%	16.95%
6	14.05%	12.65%	19.75%	16.87%	16.11%	16.27%
7	13.74%	24.31%	11.75%	22.55%	14.42%	17.43%
8	41.98%	21.06%	23.60%	27.74%	20.17%	25.47%
9	29.22%	19.81%	19.26%	15.71%	23.10%	21.04%
All	**13.34%**	**16.45%**	**16.68%**	**15.13%**	**14.39%**	**15.28%**
High	24.71%	23.28%	15.43%	23.51%	16.59%	
Low	-1.42%	1.82%	3.91%	-8.63%	-15.75%	
High - Low	26.13%	21.46%	11.52%	32.14%	32.34%	
t-statistic	2.4538***	2.1192**	1.3381*	2.8949***	3.3189***	
High - All	11.37%	6.84%	-1.25%	8.38%	2.20%	
t-statistic	1.3125*	1.1194	-0.2687	2.456***	0.8767	

The table shows how return varies across different F-scores and size quintiles. Moreover, the table presents one-tailed t-tests between the return of high F-Score firms and low/all F-Score firms. The significance levels 10 %, 5 % and 1 % are shown with *, ** and ***. The data consist of all stocks on the Stockholm Stock Exchange from January 1996 to March 2017. Piotroski's (2000) F-Score is supposed to measure a firm's financial position and the higher score the better. The F-Score is computed for each firm and year using the methodology summarized in table 2. We define high (low) F-Score firms as firms with an F-Score of 7-9 (0-2). Size quintiles are calculated for year t using market capitalization at each firm's fiscal period end, year t-1. We calculate the returns using the closing price of the day the fourth quarter earnings report was released and the closing price the day before next year's report release (see function 6). This means that the holding period differs across firms and years, but is approximately one year. The number of observations within each combination of F-Score and size quintile can be found in appendix 1, panel B.

The one-tailed t-tests of mean return differences, between high and low F-Score firms, show significance at the one percent level for the first, fourth and fifth quintiles (mean return differences of 26.13 %, 32.14 % and 32.34 %, respectively). The second and third quintile also show significance, but at the five and ten percent levels. The fact that all quintiles show some statistical significance indicate that the strategy, in contrast to B/M, is not specifically designed for firms of a particular size quintile. Which can also be observed in the highest t-statistics being lower for size than B/M quintiles. The high-all tests further strengthen this conclusion, where only the fourth quintile shows significance (one percent), with a mean return difference of 8.38 percent.

Although Piotroski makes size classifications based on terciles, we choose to use the same break-points as for book-to-market for consistency and comparability reasons, being the same methodology between tests and relative to Fama and French (1993). Nevertheless, contrary to us, Piotroski does not find any significance in his tests for large firms, but finds one percent significance for small and medium-sized firms in both the high-low and the high-all tests. Our findings suggest that the strategy ability to find underperformers is most pronounced for the two highest size classification (average mean return of -8.63 and -15.75 percent). As a result, potential concerns regarding shorting abilities are mitigated, since it is reasonable to assume that most of the stocks within the two highest quintiles can be shorted.

5.2 Abnormal Returns

This section tries to explain the raw return and answer why the F-Score strategy is successful, both from the perspective of risk-based and market mispricing view. Therefore, we test if the strategy's ability to identify future over- and underperformers are intact after adjusting returns for risk. As previously stated, we give special focus on aspects where the F-score has been criticised, as with risk-adjusting and abnormal returns.

5.2.1 Three-Factor Model Regression

To evaluate whether Piotroski's F-Score portfolio has created abnormal returns we regress the F-Score portfolio against Fama and French's three factor model. The choice to use portfolio returns instead of individual stock returns is to increase the robustness. Since the data becomes less noisy, the model can create a more fitting line. Based on 21 yearly observations, the investment period, the regression has an r-squared and adjusted r-squared of 80.20 and 76.70 percent, respectively. This indicates that the factors can explain a large majority of the returns of the F-Score portfolio.

All the coefficients, including the intercept, are significant at the one percent level. The market risk premium beta is relatively low and broadly follows the results from the subsection "5.1.2 Market-Adjusted Returns over Time", which show that the portfolio returns are positively correlated with the market, but not particularly vulnerable to bear markets. The coefficients of SMB and HML are positive and therefore show that the portfolio is exposed to the size as well as value effect. Interestingly, our results show high explanatory value of the risk-based view, which contradict the arguments Piotroski makes of not needing to test his returns to the most recognised common risk factors.

Table 7
Three Factor Model Regression for the F-Score Portfolio

	Constant	Excess Market Return	SMB	HML
Coefficient	0.0984	0.5766	0.3523	0.3519
Std. Err.	0.0205	0.0831	0.1199	0.0826
t-statistic	4.79***	6.94***	2.94***	4.26***
R-squared	0.8018			
Adj. R-squared	0.7668			
Number of obs.	21			
Root MSE	0.08643			

The table summarizes the regression between the dependent variable excess return of the F-Score portfolio and the independent variables in Fama and French's (1993) three factor model: excess market return, HML and SMB. The significance levels 10 %, 5 % and 1 % are shown with *, ** and ***. The data consist of all stocks on the Stockholm Stock Exchange from January 1996 to March 2017. Piotroski's (2000) F-Score is supposed to measure a firm's financial position and the higher score the better. The F-Score is computed for each firm and year using the methodology summarized in table 2. We define high (low) F-Score firms as firms with an F-Score of 7-9 (0-2). The investment strategy is to buy high F-Score firms and short low F-Score firms. The portfolio is equally-weighted between all stocks. We calculate the returns using the closing price of the day the fourth quarter earnings report was released and the closing price the day before next year's report release (see function 6). The OMXSPI index are used as market proxy and the risk-free rate is the 3 months Treasury bill rates. The additional Fama and French (1993) factors are calculated using the methodology described in their paper. Since the reporting dates differ across firms the factors are matched with the holding period (see functions 13-17). The regression is based on a yearly average of the factors for the observations within the F-Score portfolio.

However, the most striking finding is the unexplained return and the significant alpha of 9.84 percent, which means that the null hypothesis can be rejected at the one percent level. To conclude, the common risk factors cannot fully explain the high returns of Piotroski's F-Score strategy on the Swedish stock market between 1996 and 2017, leading to the creation of high significant abnormal return.

5.2.2 Abnormal Returns Conditioned on Book-to-Market

As earlier discussed in the theoretical framework, conclusions related to abnormal returns are to some degree dependent on the asset pricing model. To increase the robustness of our initial findings, we therefore make additional tests related to abnormal returns, with regard to the risk characteristics of individual stocks. Further testing the strategy's capability to identify future over-and underperformers when the returns are risk-adjusted. By conditioning abnormal returns on book-to-market, the results become more comparable with Piotroski; whose sample only consists of value stocks. Table 8 summarizes the findings of abnormal returns conditioned on B/M quintiles. In each B/M quintile, the sum of the abnormal returns is almost exactly zero percent, due to B/M quintiles being used as break-points in the classification of the 25 portfolios.

Table 8
Abnormal Returns Conditioned on B/M

	Low		B/M Quintiles		*High*	
F - SCORE	1	2	3	4	5	All
0	-34.09%	-57.41%	-57.41%		-12.26%	-45.91%
1	-10.44%	-1.60%	-20.30%	30.57%	-17.71%	-6.55%
2	-22.37%	-4.19%	-2.01%	0.97%	-36.21%	-11.35%
3	1.33%	-7.56%	-0.19%	-8.52%	-5.49%	-3.55%
4	0.83%	-2.19%	-0.14%	-5.77%	1.31%	-1.23%
5	-0.91%	1.61%	3.66%	2.63%	0.12%	1.52%
6	4.41%	4.30%	-4.60%	0.33%	-4.38%	0.30%
7	3.86%	-1.58%	2.73%	2.29%	4.37%	2.14%
8	1.66%	0.51%	1.28%	5.88%	29.88%	7.14%
9	-23.87%	-11.25%	3.31%	-13.58%	7.88%	-5.50%
All	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
High	2.57%	-1.58%	2.39%	2.54%	12.74%	
Low	-18.79%	-4.78%	-10.42%	6.19%	-29.15%	
High - Low	21.36%	3.20%	12.81%	-3.65%	41.90%	
t-statistic	3.4662***	0.3138	1.9148**	-0.5217	3.4092***	
High - All	2.57%	-1.58%	2.39%	2.54%	12.74%	
t-statistic	0.4866	-0.3581	0.6888	0.7562	2.5375***	

This table shows how abnormal return varies across different F-scores and book-to-market quintiles. Moreover, the table presents one-tailed t-tests between the abnormal return of high F-Score firms and low/all F-Score firms. The significance levels 10 %, 5 % and 1 % are shown with *, ** and ***. The data consist of all stocks on the Stockholm Stock Exchange from January 1996 to March 2017. Piotroski's (2000) F-Score is supposed to measure a firm's financial position and the higher score the better. The F-Score is computed for each firm and year using the methodology summarized in table 2. We define high (low) F-Score firms as firms with an F-Score of 7-9 (0-2). Book-to-market quintiles are calculated for year t using the book value of equity and market capitalization at each firm's fiscal period end, year t-1. Size quintiles for year t are computed using market capitalization at each firm's fiscal period end, year t-1. We divide portfolio based on B/M and size quintiles, resulting in 25 portfolios. For each one of these portfolios we run a Fama and French's three factor model regression to predict the abnormal return. We calculate the returns using the closing price of the day the fourth quarter earnings report was released and the closing price the day before next year's report release (see function 6). The OMXSPI index are used as market proxy and the risk-free rate is the 3 months Treasury bill rates. The additional Fama and French (1993) factors are calculated using the methodology described in their paper. Since the reporting dates differ across firms the factors are matched with the holding period (see functions 13-17).

The results are overall similar to raw returns conditioned on book-to-market, where the t-tests on the difference between high and low F-Score firms are significant at 1 percent for the first and fifth quintiles (21.36 percent and 41.90 percent, respectively). In addition, only the fifth quintile shows one percent significance from the t-tests on the difference between high and all F-Score firms. The findings thus suggest the strategy performs best for value stocks, also when adjusted for common risk factors. Moreover, the F-Score strategy still can find over- and underperformers.

However, the results for abnormal return differ in some crucial aspects from the results for raw returns. Firstly, the low F-scores are more heavily what drives the return, indicating that the strategy, after all, is quite dependent on the shorting ability. Secondly, the positive relationship between F-Score and the mean returns is less eye-catching. For example, the mean abnormal return for F-Scores of 9, is negative in three out of the five quintiles, which indicates that the highest F-Score are exposed to higher risk. Lastly, the mean abnormal returns for high (low) F-Scores firms are positive (negative) in four out of five quintiles, which questions whether the F-Score strategy can be applied on any other book-to-market quintile than value stocks.

5.2.3 Abnormal Returns over Time

From a theoretical point of view, the abnormal returns should not be affected by the market condition since it is already reflected in the market risk premium component. Yet, one can still observe other patterns, e.g. Schwert's (2003) argument that anomalies change form after they are public rather than disappearing.

The results in table 9 is a continuation of the discussion on section "5.1.2 Market-Adjusted Returns over Time". The table shows, first and foremost, that the mean abnormal return difference between high and low F-Scores firms are positive for 16 out of 21 years. Whereof the positive difference is significant, with five percent significance, in 9 of the years. The high (low) portfolio have positive (negative) abnormal returns in 13 (14) years. This indicates that Piotroski's F-Score is rather time-invariant, which is important for an investor with a limited holding period. Secondly, it should be said that because the observations are fewer for low F-Score firms, their means can be very extreme (e.g. 1996 and 2003). The general observation is the same as in the previous subsection; from the perspective of abnormal returns, the strategy is better at finding underperformers than overperformers. Thirdly, the introduction of the strategy in year 2000 does not seem to have affected the results negatively, instead the abnormal returns are greatest and most significant in the subsequent years. To conclude, the findings suggest that abnormal returns have not been limited to a specific period.

Table 9
Abnormal Returns over Time

Year	High	Low	High - Low	t-statistic
1996	12.25%	-102.36%	114.61%	3.7872***
1997	4.00%	-	4.00%	0.4119
1998	-6.36%	-36.27%	29.91%	2.1515**
1999	-11.06%	65.76%	-76.82%	-2.2551
2000	12.98%	-	12.98%	1.7619**
2001	14.22%	-57.45%	71.67%	6.8125***
2002	-1.00%	-51.20%	50.20%	4.9408***
2003	4.25%	100.84%	-96.59%	-2.1295
2004	-0.19%	-31.56%	31.36%	2.6803***
2005	-0.21%	-8.94%	8.73%	0.6577
2006	5.67%	-15.80%	21.47%	0.7600
2007	3.46%	14.44%	-10.98%	-0.8712
2008	9.49%	1.78%	7.72%	0.6898
2009	6.25%	-10.14%	16.39%	0.7204
2010	2.61%	-18.53%	21.13%	1.9204**
2011	-1.76%	-10.76%	9.00%	0.9328
2012	-2.49%	-0.24%	-2.25%	-0.1910
2013	-0.28%	-18.72%	18.44%	1.1227
2014	13.37%	-36.70%	50.07%	5.099***
2015	4.84%	-14.86%	19.70%	1.7825**
2016	4.01%	10.31%	-6.30%	-0.3928

The table shows how abnormal returns for the high and low F-Scores develop over time. The tests to the right are one-tailed t-test between the return of high and low F-Score firms each year. For two years (1997 & 2000) the one-tailed t-tests are made between the return of the high F-Score firms and zero due to lack of observations. The significance levels 10 %, 5 % and 1 % are shown with *, ** and ***. The data consist of all stocks on the Stockholm Stock Exchange from January 1996 to March 2017. Piotroski's (2000) F-Score is supposed to measure a firm's financial position and the higher score the better. The F-Score is computed for each firm and year using the methodology summarized in table 2. We define high (low) F-Score firms as firms with an F-Score of 7-9 (0-2). Book-to-market quintiles are calculated for year t using the book value of equity and market capitalization at each firm's fiscal period end, year t-1. Size quintiles for year t are computed using market capitalization at each firm's fiscal period end, year t-1. We divide portfolio based on B/M and size quintiles, resulting in 25 portfolios. For each one of these portfolios we run a Fama and French's three factor model regression to predict the abnormal return. We calculate the returns using the closing price of the day the fourth quarter earnings report was released and the closing price the day before next year's report release (see function 6). The OMXSPI index are used as market proxy and the risk-free rate is the 3 months Treasury bill rates. The additional Fama and French (1993) factors are calculated using the methodology described in their paper. Since the reporting dates differ across firms the factors are matched with the holding period (see functions 13-17).

6. Conclusion and Implications

This paper examines Piotroski's (2000) F-Score strategy, on the Stockholm Stock Exchange between January 1996 and March 2017, with the goal to investigate the strategy's ability to identify future over- and underperformers, and try to explain the reason behind the performance in terms of risk-based and mispricing view. From a theoretical perspective, we conclude that stock selection strategies, like Piotroski's, should not create abnormal returns since it contradicts the semi-strong form of efficient markets. From previous research, on different market and time periods, we find consensus on that the F-Score strategy is useful for an investor, due to the existence of market mispricing. A clear majority concludes that this effect is most pronounced for stocks with high book-to-market values, so-called value stocks.

In this paper, we find a positive relationship between F-Score and future return, indicating that the strategy can identify future over- and underperformers. This ability is highly significant when firms are grouped into low, medium and high F-scores, especially in the value stock quintile. Furthermore, we find that the ability is not limited to small- and medium sized firms, in contrast to previous research, which reduces concerns about feasibility. Importantly, the ability to identify future over- and underperformers prevails after adjusting returns for the most recognized risk factors, and yet again; value stocks show highest significance. Thus, since a substantial amount of the returns remain unexplained after adjusting for risk, using Fama and French's (1993) three factor model, the results indicate the existence of market mispricing. Lastly, the abnormal returns have not been concentrated to any specific year, but instead been relatively persistent, indicating long-term market mispricing.

The research and conclusion described above contributes to previous research on the Piotroski's (2000) F-score in three ways. Firstly, our findings give further support and robustness to previous studies, concluding the F-Score as a successful stock selection strategy. Secondly, the improved methodology provides a more practical and realistic picture of how an investor can take advantage of the F-Score strategy. Thirdly, giving special focus on aspects where the F-score has been criticised. By adding the risk-adjusted perspective, which many papers lack, we show evidence related to the causality between F-Score and common risk factors.

Looking at the bigger picture, our results indicate the effectiveness for an investor of a simple fundamental-based stock selection strategy. Furthermore, the inability to explain the portfolio returns on a risk-adjusted basis exemplify a failure of the semi-strong form of market efficiency, supporting the existence of market mispricing.

Suggestions of future research are to focus on differences between markets e.g. accounting standards, market structures and, investors' view on value and growth stocks. For example, Rathjens and Schellhove's (2011) results indicate that the F-Score strategy performs better for growth stocks than for value stocks, on the U.K. stock market. Identifying the drivers of the disparity enables increased explanation and understanding of the F-Score as well as market mispricing.

As mentioned earlier, one key aspect of our paper is the increased practical application in the methodology and it would thus be interesting to test when during the one year holding period that the market incorporates the information (F-Score), creating the abnormal return. Furthermore, how this change would impact the results of previous studies on other markets, e.g. is any market worse/better at incorporating the information? Another improvement related methodology would be to avoid the assumption of zero return when delisting, through finding a fitting variable.

Lastly, the choice to only include the most recognised risk factors of Fama and French's (1993) three factor model is a limitation of our method, since conclusions about market mispricing and abnormal returns are residuals of the chosen model (joint hypothesis problem). One can argue, it is quite unlikely that additional risk factors would explain the whole alpha of 9.84 percent and thus change our overall conclusion, but it would add more explanatory elements. Proposedly, such an increase should include factors related to momentum, traded liquidity and, Fama and French´s (2015) five-factor model.

References

Aggarwal, Navdeep, and Mohit Gupta. "Do high book-to-market stocks offer returns to fundamental analysis in India?." *Decision (0304-0941)* 36.2 (2009).

Akerlof, George A., and Robert J. Shiller. "Animal spirits: How human psychology drives the economy, and why it matters for global capitalism." *Princeton: Princeton University Press* (2009).

Amihud, Yakov. "Illiquidity and stock returns: cross-section and time-series effects." *Journal of financial markets* 5.1 (2002): 31-56.

Banz, Rolf W. "The relationship between return and market value of common stocks." *Journal of financial economics* 9.1 (1981): 3-18.

Basu, Sanjoy. "Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis." *The journal of Finance* 32.3 (1977): 663-682.

Brennan, Michael J., Tarun Chordia, and Avanidhar Subrahmanyam. "Alternative factor specifications, security characteristics, and the cross-section of expected stock returns." *Journal of Financial Economics* 49.3 (1998): 345-373.

Carhart, Mark M. "On persistence in mutual fund performance." *The Journal of finance* 52.1 (1997): 57-82.

Chen, Nai-fu, and Feng Zhang. "Risk and return of value stocks." *The Journal of Business* 71.4 (1998): 501-535.

Cuthbertson, Keith. "The expectations hypothesis of the term structure: The UK interbank market." *The Economic Journal* (1996): 578-592.

De Bondt, Werner FM, and Richard H. Thaler. "Further evidence on investor overreaction and stock market seasonality." *Journal of finance* (1987): 557-581.

Dempsey, Mike. "The capital asset pricing model (CAPM): the history of a failed revolutionary idea in finance?." *Abacus* 49.S1 (2013): 7-23.

Ericsson, Johan, and Karlsson, Sune, "Choosing Factors in a Multifactor Asset Pricing Model: A Bayesian Approach" *SSE/EFI Working Paper Series in Economics and Finance* No. 524 (2004), Stockholm School of Economics

Fama, Eugene F. "Efficient capital markets: A review of theory and empirical work." *The journal of Finance* 25.2 (1970): 383-417.

Fama, Eugene F. "The behavior of stock-market prices." *The journal of Business* 38.1 (1965): 34-105.

Fama, Eugene F., and Kenneth R. French. "A five-factor asset pricing model" *Journal of Financial* 116.1 (2015): 1-22.

Fama, Eugene F., and Kenneth R. French. "Common risk factors in the returns on stocks and bonds." *Journal of financial economics* 33.1 (1993): 3-56.

Fama, Eugene F., and Kenneth R. French. "The capital asset pricing model: Theory and evidence." *The Journal of Economic Perspectives* 18.3 (2004): 25-46.

Fama, Eugene F., and Kenneth R. French. "The cross-section of expected stock returns." *the Journal of Finance* 47.2 (1992): 427-465.

French, Kenneth R. "Stock returns and the weekend effect." *Journal of financial economics* 8.1 (1980): 55-69.

Graham, Benjamin, and David Dodd. "Security Analysis (The Classic 1940 Second Edition)." *New York: McGraw-Hill Education* (1940).

Hawawini, Gabriel, and Donald B. Keim. "The cross section of common stock returns: A review of the evidence and some new findings." *Security Market Imperfections in Worldwide Equity Markets* 3 (2000).

Hyde, Charles E. "An emerging markets analysis of the Piotroski F score." *Macquarie University* (2013).

Jegadeesh, Narasimhan, and Sheridan Titman. "Returns to buying winners and selling losers: Implications for stock market efficiency." *The Journal of finance* 48.1 (1993): 65-91.

Kilsgård, David, and Wittorf, Filip, “The Fama and French Three-Factor Model – Evidence from the Swedish Stock Market“, *Lund’s University* (2010)

Kobelyatskyy, Mykhaylo, and Fulgentiusson, Henrik, “Momentum – Trendspotting in the Swedish Stock Market“, *Lund’s University* (2011)

Lintner, John. "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets." *The review of economics and statistics* (1965): 13-37.

Lo, Andrew W., and A. Craig MacKinlay. "Data-snooping biases in tests of financial asset pricing models." *Review of financial studies* 3.3 (1990): 431-467.

MacKinlay, A. Craig. "Multifactor models do not explain deviations from the CAPM." *Journal of Financial Economics* 38.1 (1995): 3-28.

Markowitz, Harry. "Portfolio selection." *The journal of finance* 7.1 (1952): 77-91.

Mohanram, Partha S. "Separating winners from losers among lowbook-to-market stocks using financial statement analysis." *Review of accounting studies* 10.2 (2005): 133-170.

Mossin, Jan. "Equilibrium in a capital asset market." *Econometrica: Journal of the econometric society* (1966): 768-783.

Myers, Stewart C., and Nicholas S. Majluf. "Corporate financing and investment decisions when firms have information that investors do not have." *Journal of financial economics* 13.2 (1984): 187-221.

Piotroski, Joseph D. "Value investing: The use of historical financial statement information to separate winners from losers." *Journal of Accounting Research* (2000): 1-41.

Piotroski, Joseph D., and Eric C. So. "Identifying expectation errors in value/glamour strategies: A fundamental analysis approach." *Review of Financial Studies* 25.9 (2012): 2841-2875.

Porta, Rafael La, et al. "Good news for value stocks: Further evidence on market efficiency." *The Journal of Finance* 52.2 (1997): 859-874.

Rathjens, Hauke, and Schellhove, Hendrik, “Simple Financial Analysis and Abnormal Stock Returns - Analysis of Piotroski's Investment Strategy”, *Stockholm School of Economics* (2011)

Riksbanken, “Interest & exchange rates”, www.Riksbanken.se, retrieved 2017-04-03

Roll, Richard. "A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory." *Journal of financial economics* 4.2 (1977): 129-176.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein. "Persuasive evidence of market inefficiency." *The Journal of Portfolio Management* 11.3 (1985): 9-16.

Ross, Stephen A. "The arbitrage theory of capital asset pricing." *Journal of economic theory* 13.3 (1976): 341-360.

Schwert, G. William. "Anomalies and market efficiency." *Handbook of the Economics of Finance* 1 (2003): 939-974.

Sharpe, William F. "Capital asset prices: A theory of market equilibrium under conditions of risk." *The journal of finance* 19.3 (1964): 425-442.

Thomson Reuters Datastream, retrieved 2017-04-03

Appendices

Appendix 1

Panel A - Number of Observations across F-Scores

	Low		B/M Quintiles		*High*	
F - SCORE	1	2	3	4	5	Nr
0	2	1	2		1	6
1	29	9	12	9	9	68
2	59	41	28	42	17	187
3	125	87	93	75	103	483
4	164	189	175	156	141	825
5	226	263	274	222	218	1203
6	251	294	264	195	196	1200
7	148	162	196	129	96	731
8	35	60	76	82	50	303
9	6	13	18	15	11	63
Nr	1045	1119	1138	925	842	5069

Panel B - Number of Observations across F-Scores and Size Quintiles

	Low		Size Quintiles		*High*	
F - SCORE	1	2	3	4	5	Nr
0	1	1	4			6
1	22	20	19	7		68
2	63	60	37	16	11	187
3	142	111	101	70	59	483
4	165	158	197	171	134	825
5	205	226	245	255	272	1203
6	159	214	292	254	281	1200
7	62	136	178	169	186	731
8	37	44	76	66	80	303
9	9	12	9	15	18	63
Nr	865	982	1158	1023	1041	5069

Appendix 1a) and 1b) show the number of observations across different F-Scores and Book-to-market/Size quintiles. The data consist of all stocks on the Stockholm Stock Exchange, from January 1996 to February 2017. Piotroski's (2000) F-Score is supposed to measure a firm's financial position and the higher score, the better. We compute the F-Score for each firm and year using the methodology summarized in table 2. We define high (low) F-Score firms as firms with an F-Score of 7-9 (0-2). Book-to-market quintiles are calculated for year t using the book value of equity and market capitalization at each firm's fiscal period end, year t-1. Size quintiles for year t are calculated using market capitalization at each firm's fiscal period end, year t-1.

Appendix 2
Underlying Variables

Variable ID	Variable Name
WC02999	Total Assets
WC03251	Long Term Debt
WC08106	Current Ratio
WC05301	Common Shares Outstanding
WC08306	Gross Profit Margin
WC08401	Total Asset Turnover
WC05491	Book Value – Outstanding Shares – Fiscal
WC08002	Market Capitalization Fiscal Period End
WC04251	Net Proceeds from Sale/Issue of Common & Preferred
WC01551	Net Income before Extraordinary Items/Preferred Dividends
WC04201	Funds from Operations
WC05905	Earnings per Share Report Date Fiscal Year End
RI	Total Return Index
P	Price Adjusted
PI	Price Index
WC04900	Increase/decrease in Working Capital

The table summarizes all variables used in the computation of the F-Score. All the data were retrieved from Thomson Reuters Datastream.

Appendix 3
Number of Observations across the 25 Portfolios

		Low		B/M Quintiles		*High*	
		1	2	3	4	5	Nr
Low	1	141	145	165	186	228	865
	2	213	166	184	204	215	982
Size Quintiles	3	242	232	263	213	208	1158
	4	250	307	218	156	92	1023
High	5	199	269	308	166	99	1041
	Nr	1045	1119	1138	925	842	5069

The table shows the numbers of observations in the 25 portfolios, with break-points based on size and B/M quintiles. The data consist of all stocks on the Stockholm Stock Exchange, from January 1996 to February 2017. Book-to-market quintiles are calculated for year t using the book value of equity and market capitalization at each firm's fiscal period end, year t-1. Size quintiles for year t are calculated using market capitalization at each firm's fiscal period end, year t-1.