Utility of Piotroski F-Score for predicting Growth-Stock Returns

Jan-Hendrik Markus Mohr
Münzenberger Straße 10
60389 Frankfurt

Tel.: 0172 162 89 51
Mail: jhmmohr@googlemail.com

1 Introduction

Over the course of the last decades, the analysis of structural reasons for equity out- or underperformance has been a widely discusses academic topic. New explanatory factors, such as accruals (Sloan, 1996), were established and former explanatory factors lost some of their predictive power, as Fama and French (2003) show in the case of beta.

One of the more recent explanatory factors is the F-Score (Piotroski, 2000), which has strong practical utility in separating winners from losers in the value segment of the market.

In this paper, I provide evidence on the utility of F-Score in the growth segment of the market.

In section two, I review literature on the structural differences of value and growth stocks. Additionally, I explain the methodology of Piotroski and how F-Score is constructed. Next, I take a look at earlier attempts to differentiate winners from losers in the growth sub-sample of the market.

In section three, I explain my research tools and establish the necessary tests to provide evidence for my hypotheses. Also, I formulate my hypotheses and the related null hypotheses.

In section four, I present the empirical results from my study and discuss the results. I find a strong market-adjusted return difference between high and low F-Score stocks. However, I provide additional data that raises concern on whether a strategy that buys high F-Score and shorts low F-Score stocks will be successful in asset management practice.

In the last section, I sum up my findings and introduce possible practical implementation tools to seize the apparent return difference between high and low F-Score growth stocks.

2 Literature Review

This paper examines the utility of fundamental analysis for separating over- and underperforming “growth” (e.g. low ratio of past financial data vs. current market cap) stocks. Unlike for "value" (e.g. high ratio of past financial data vs. current market cap) stocks, not much research has been conducted on the nature of return differences within the field of growth stocks. Among the many fundamental indicators I chose "F-Score" (Piotroski, 2000) to test its application in identifying return differences within the universe of eurozone growth stocks.

2.1 Systematic Differences in Value and Growth Stock Returns

After Basu (1977, 1983) and Rosenberg, Reid and Lanstein (1985) conducted early research on the systematic outperformance of value stocks over growth stocks, the explanation for this return difference has been the topic of discussions among academics and practitioners alike.

Fama and French, who incorporated their initial findings into the widely-quoted 3-factor model (Fama and French, 1993), explain the structural outperformance of value stocks with the embedded risk of these stocks. This value return premium would therefore be a fair compensation for the risk an investor has to bear when holding value stocks. Vice versa, growth stocks allegedly underperform due to the negative risk premium that expresses the less risky nature of these stocks. Indeed, Fama and French observe a higher degree of financial distress at high book-to-market firms which could be an indicator for higher risk of these companies.

A competing explanation for the value premium is provided by Lakonishok, Shleifer and Vishny (1994), who cite the possibility of excess extrapolation of historic trends as a possible reason why value stocks are priced too low and growth stocks are priced too high. In addition, La Porta (1996) finds that investors' expectations about future growth are too extreme. Analysts appear to *over*-extrapolate prior fundamental performance into the future and hence exaggerate the expected growth rates. Since earnings expectations might play a role in the (low) high pricing of (value) growth stocks, too extreme earnings expectations would result in (under-) overvaluation and thereby explain the return characteristics of both group of stocks. La Porta finds evidence for this thesis as stocks with high earnings growth expectations underperform stocks with low earnings expectation on average by 20,9% in the 1982-1991 timeframe. This excess return is still significant after controlling for size and industry groups. Moreover, given the generally rising equity markets during the time, La Porta's findings provide an interesting contrast to De Bondt and Thaler (1987) who show that value strategies tend to outperform in declining equity markets.

While neither of both explanations have so far been falsified, an extensive body of empirical findings from around the world supports the basic assumption that there is, in fact, a structural difference in the return pattern between value and growth stocks. International evidence for a value-premium / growth-discount is given by Chan, Hamao and Lakonishok (1991) for Japan; Fama and French (1998) for 13 international markets and Brouwer, van der Put and Veld (1996) for France, Netherlands, Germany and UK. In a more recent paper Artmann, Finter and Kempf (2010) find statistically significant indications for a systematic pricing difference between German value and growth stocks. All of these findings suggest that there is in fact a systematic return difference in stock returns that is linked to valuation.

2.2 Joseph Piotroski's "F-Score" as an Indicator to Differentiate Value-Stocks

After 2000, numerous authors have tested different methods of improving the returns of stocks in the value as well as in the growth universe. Understandably, much research has been done on indicators that can be applied to select the best value stocks since there is, at first sight, more practical utility for this type of information. In fact, while long equity portfolios of "especially attractive" value stocks can easily be built in a real-life setting, shorting "especially unattractive" growth stocks demands higher investor sophistication and is constrained by practical issues such as stock borrowing availability or borrowing costs.

As observed by Piotroski (2000), value stocks show skewed returns: while the mean market-adjusted return of value stocks is significantly positive, the median market-adjusted return of value stocks is negative. Hence, the average value stock performs below market; however few value stocks show very significant outperformance. Piotroski therefore constructs a fundamental indicator that aims at separating the few winning value stocks from the majority of value stocks that lose due to their distressed nature (Fama and French, 1993). A similar result can be found in the research paper of Grantham (2010), where the structural return pattern of value stocks is explained. The authors finds a general long-term outperformance of value stocks that can nevertheless be severely impacted during times of harsh economic deterioration. For example, Grantham finds that value stocks performed vastly below market during the Great Depression and also during the post-2007 financial crisis. This effect is explained by the solid financial strength that was needed for corporate survival during the time- an attribute value stocks rarely possess.

2.2.1 Construction of F-Score

The construction of F-Score is systematically different from other multi-variable fundamental indicators. One of the most prominent of these fundamental indicators is 'Z-Score' (Altman, 1964), which shows statistically significant results in predicting bankruptcy of a company. After publication, Z-Score has gained great attention among practitioners in fields like risk intermediation, ratings or security analysis. While the inputs to Z-Score show a high overlap with F-Score, the latter does not anchor to specific values in the companies' fundamentals. Instead, F-Score considers a) in what directions the fundamentals of a company are trending and b) whether general financial health conditions are met (i.e. "positive RoA: yes/no"; "equity issuance: yes/no"; "positive accruals yes/no" etc.). F-Score consist of nine binary variables that can be clustered into three dimensions of company health: profitability, balance sheet health and operating efficiency.

Profitability

Piotroski argues that positive *Return on Assets ("RoA")* and *Cash-Flow from Operations / Assets ("CFO"),* both trailing one year, are signs for a firm's ability to generate funds internally, which can be seen as a positive in an environment where most of the sampled value stocks are distressed. Bother are assigned 1 in the model if positive and 0 if negative.

In addition, Piotroski considers the year-over-year *change in Return on Assets ("ΔRoA")* as "suggestive of an improvement in the firm's underlying ability to generate positive future cash-flows" (Piotroski, 2000, p. 7). Again, 1 gets assigned if the trend is positive, 0 if the trend is negative.

The fourth profitability variable is the *Relation between CFO and RoA ("Accrual").* Sloan (1996) provides evidence that positive accruals could be indicative for a) lower subsequent profitability and also b) aggressive management of earnings. This variable assumes the value 1 if CFO > RoA and 0 if CFO < RoA.

Balance Sheet Health

Within the generally distressed value sample, an increase (decrease) in leverage (liquidity) is expected to result in more financial risk (Piotroski, 2000, p.7). Therefore, Piotroski assigns a 1 (0) to *year-over-year decrease (increase) in long-term debt / assets ("ΔLever").* In terms of liquidity, Piotroski uses the change of a firm's current ratio (*"ΔLiquid")* as a proxy for the ability to service currently maturing liabilities. A positive year-over-year change in liquidity gets assigned a 1, a negative change a 0. A third variable in the financing dimension of the F-Score is the monitoring of equity issuance. An increase in share count might indicate severe financial distress at a value firm, since equity issuance does come along with high cost of capital a firm has to accept when its share price is low. Managers are likely to only accept this cost of capital when they truly have no other financing choice (Piotroski, 2000, p.8). Therefore, Piotroski assigns companies with an increase in share count (year over year) a 0 and all other companies a 1.

Operating Efficiency

In this third dimension of company fundamental health, Piotroski evaluates the *Changes in gross margin and changes in asset turnover ("ΔMargin" and "ΔTurn").* Both are part of the Return on Asset decomposition and hence have a causal influence on RoA. Piotroski assigns to a positive trend in both variables a 1 and to a negative trend a 0.

2.2.2 Empirical Testing of F-Score

Piotroski analyzes the value stocks (defined as top 20% in terms of book-to-market) within the COMPUSTAT database in the 1976-1996 timeframe. To identify the predictive power of F-Score, Piotroski separates "low" (0-1) F-Score companies from "high" (8-9) F-Score companies and compares the subsequent one and two year returns of both groups.

Piotroski observes an average difference in mean-return between high and low F-Score value stocks of 23%. Additionally, 50% of high F-Score stocks outperform the naïve value portfolio (indicating still skewed returns), while low F-Score firms underperform in 68% of the time. Similar return patterns were identified for the two-year holding period. Piotroski observed the strongest hedge returns between high and low F-Score firms within the sub-sample of the smallest and under-followed value stocks.

2.3 Attempts to Differentiate Growth Stocks

The inflation and subsequent breakdown of the "New Economy"-bubble caused academic interest in the structural differences between successful and unsuccessful growth stocks. Post-2000, the poor performance of prior high-flying stocks was not only caused by shrinking valuations but also by unprecedented accounting scandals like WorldCom, Nortel Networks and Enron. Jensen (2004) argues that managers are more inclined to apply aggressive accounting when their companies' stock price is valuated excessively. Coming along with a stronger capital market focus and the heavy use of stock options as incentives, managers feel great pressures to meet capital market expectations and therefore engage in earnings management activities to sustain high stock valuations. Those activities are suspected to cause significant agency costs for investors in highly-valuated companies.
Jensen's view supports a thesis that growth stocks do not only underperform due to too optimistic cash-flow growth expectations. In fact, Jensen provides reasons to suspect that the Fama and French (1992) concept of high quality growth / low-quality value stocks might be faulted. Beneish and Nichols (2009) establish an indicator ("O-Score") that aims at identifying the companies within the growth stock universe that engage in overly aggressive accounting. The authors therefore construct a tool-kit to utilize Jensen's agency cost theory by ex-ante identifying the growth stocks with the highest agency costs. Basis of the O-Score are five components: abnormally high sales growth, high accruals, acquisition activity, equity issuance and the so-called PROBM measure. PROBM is a scoring systems developed by Beneish (1999) to evaluate a company's accounting integrity by considering financial ratios that could indicate aggressive accounting. Most notably, there is a high overlap between the inputs to O-Score and the composition of F-Score. However, F-Score emphasizes financial health while O-Score stresses the role of accruals in a company's financial statements.
Beneish and Nichols test the predictive ability of O-Score with 27,000 firms between 1993 and 2004. The companies with the highest O-Score show large market-adjusted abnormal returns between -22% and -25% in the one-year trading period after formation. O-Score keeps its high explanatory power even after controlling for possible other explaining variables (e.g. accruals, size effect, momentum etc.). In addition, the authors show that high O-Score companies are significantly more likely to restate earnings than low O-Score companies.

Although Beneish and Nichols research shows very convincing empirical evidence that O-Score is indeed a good tool to separate winning from losing growth stocks, O-Score

focuses mainly on the accounting integrity of a given company. Financial strength and the level of distress are only evaluated indirectly. However, this negligence could possibly leave out important information since it is not clear whether *only* aggressive accounting is the factor that allows separating winning from losing growth stocks.
In an effort to apply the general idea of F-Score to growth stocks, Mohanram (2005) constructs an indicator ("G-Score") that is supposed to better reflect the underlying fundamental situation of growth companies. The author combines traditional fundamental data like earnings, accruals and cash-flows with more growth-related fundamental figures such as intensity of R&D or growth stability. His main argument in doing so is to respect the context in which fundamental analysis is being conducted. According to Mohanram, the generally positive performance history of growth stocks attracts great investor interest, especially among sophisticated investors and analysts. Hence, plain fundamental analysis as in F-Score is allegedly not as effective since the information dissecting is faster in growth stocks than in (underfollowed) value stocks. A vital difference to F-Score is the extensive use of industry-benchmarks to calculate the components of G-Score. E.g., not the absolute direction of change in RoA is measured (like in F-Score), but rather the relative change in RoA in comparison to an industry benchmark.
In empirical testing, G-Score seems to indicate differences in growth stock returns very reliably. From 1978-2001, backtesting of a strategy that buys the growth stocks with the highest G-Score and shorts the growth stocks with the lowest G-Score would have generated size-adjusted returns of greater 20% annually. This result also holds up when controlled for other suspected explanatory factors of stock returns. Most notably, a significant part of the return difference between high and low G-Score firms stems from identifying the underperforming companies (contributing 17,5% to the hedge return). This indicates that a proper utilization of G-Score requires access to (economically viable) shorting of low G-Score stocks.
Mohanram observes the highest predictive ability of G-Score within the largest and most widely followed sub-segment of the growth stock universe. This is an interesting contrast to Piotroski's (2000) findings that fundamental analysis bears most fruit in a slow information dissecting environment.
Piotroski (2004) provides further empirical evidence on fundamental analysis and discusses the results shown by Mohanram. Regarding the importance of context in fundamental analysis, he agrees with Mohanram that there are systematic differences between growth and value stocks in the application of fundamental research. Most

notably, fundamental research seems to work best for small, underfollowed value stocks and for large, well-followed growth stocks. Moreover, Piotroski finds significant evidence that fundamental analysis explains stock returns most reliably when the fundamental data shows a contradicting picture to the pricing (i.e. value or growth) of a given stock:

"Effectively, financial signals confirming the expectations that are likely already imbedded in price are assimilated into price quickly, while contrarian signals are (generally) discounted until future confirmatory news is received. As a result, historical good news for value firms is a tradable opportunity, and vice-versa for trading opportunities conditional on bad news." (Piotroski, 2004, p. 23-24)

However, Piotroski does not agree with Mohanram's assessment regarding the need for a modified indicator. In fact, Piotroski shows that F-Score can also be used to separate growth stocks. Piotroski finds that F-Score is able to separate winning from losing stocks across all book-to-market portfolios. In contrast to Mohanram's findings, F-Score shows the greatest hedge return (i.e. buying high F-Score companies vs. shorting low F-Score companies) within the growth sub-section of the market.

2.4 Motivation for Research

My aim is to contribute empirical evidence regarding the ongoing debate whether or not fundamental analysis is dependent on the valuation context of the sampled stocks. To achieve this, I am going to backtest a strategy that buys high F-Score growth stocks and shorts low F-Score growth stocks. In my analysis I focus on the eurozone equity market from 1999-2010. There are two dimensions why this could provide interesting input to the ongoing debate: a) few empirical findings exist around strategies that evaluate the power of fundamental analysis within the eurozone growth stock universe and b) the 1999-2010 timeframe provides an interesting backtesting environment since it covers two large bear markets and two large bull markets.

Since I focus on growth stocks, one might ask why not to test the G-Score that is allegedly designed to evaluate growth stocks. There are three primary reasons why I stick to testing the traditional F-Score and do not use the more sophisticated G-Score. Firstly, G-Score is vastly more critical to design since it requires thorough construction of industry benchmarks which could be reason to data mining issues (Piotroski, 2005). Secondly, G-Score relies on relative industry data which could be misleading when an entire industry is in good/bad shape. This was the case in the deflation of the post-

2000 tech bubble or in the housing-related boom/bust before/after 2007. Lastly, F-Score's simplistic and intuitive construction underlines the indicator's possible application as an asset management tool. In fact, should F-Score prove to project future returns within asset classes, the opportunity opens for practitioners to offer products that capture these returns.

3 Research Design

3.1 Sample Selection

My gross database consists of all fundamental and return data from the eurozone stock market from 1999-2010. I receive this data from the data providing firm MFIE Capital bvba., which operates the www.value-investing.eu service.. After having received the data, I checked a sample of 50 companies as to whether or not the data is a fair representation of originally reported financial figures. I found out that the data from MFIE is of great quality and in each case shows the financial figures as they are in reality.

Subsequently, I clear the gross sample for the following factors:

Firstly, I exclude all companies that do not have sufficient financial data to calculate P/B; F-Score etc. because of incomplete reporting or unavailability of data.

Secondly, I remove all utilities, financial intermediaries, real estate investment companies and investment companies as their reporting is structured very differently than the reporting in the service or industrial sector. This approach is congruent with earlier analysis of stock returns, such as in Fama and French (1992).

Thirdly, I remove all companies with a daily trading volume < EUR 10.000. This number is somewhat arbitrary as minimum trading volumes strongly depend on the size of an asset pool that aims to invest in a given strategy. I use EUR 10.000/day as a proxy, since this is the minimum trading volume that the smallest reasonable institutional equity mandate (around EUR 10 Mio. assets under management; policy for most custodians as minimum to set up a single fund mandate) would need to set up a 1% position during 20 days of trading. Embedded within this line of reasoning is the assumption that a single market participant can buy or sell 50% of the daily trading volume at a maximum.

Next, I sort the entire sample into P/B-quintiles and extracted the 20% with the highest

P/B-ratio. Similar how Piotroski (2000) refers to the highest 20% of B/M-stocks as "value stocks", I will here refer to the highest 20% of P/B-stocks as "growth stocks". The number of stocks in the growth sample ranges from a minimum of 152 stocks (in 1999) to a maximum of 409 (in 2007).

Finally, I calculate returns for each stock using a one-year holding period. Since the data is collected by June 30 of each year (when all fiscal year reporting is public), I measure the buy-and-hold return from June 30 of year t_0 to June 30 of year t_1 by applying the formula

$$Ret_1 = (P_{t1} + Div_1)/\ P_{t0} - 1$$

with:

Ret_n: Return between t_0 and t_n

P_{tn}: Price in t_n

Div_n: Dividends paid between t_0 and t_n

I consistently use the mid-price between ask and bid for a stock at the specific date t_0 and t_1. When a stock is delisted after t_0, I take the price for which the delisting took place as the t_1 price. I expect delisting do occur on average in the middle of the period (after six months). For delisted stocks, this would result in an anchoring bias towards the market-return as those stocks would, on average, only contribute six months of performance data to the analysis. If anything, my handling of delistings would smooth any market-adjusted returns and I suspect no adverse effect from setting delisting price = t_1.
In addition, I calculate the market-adjusted return for each stock by subtracting the mean growth stock return of all stocks in the t_0 to t_1 period from a single stock's performance. I weight all growth stocks equally when calculating the mean return.
The result is the out- or underperformance of a given growth stock in comparison to all growth stocks in a certain period.

3.2 Test Design

In my testing, I want to generate empirical evidence whether or not a market-neutral strategy of buying growth stocks with a high F-Score and shorting growth-stocks with a low F-Score yields satisfactory returns. Moreover, I want to evaluate the significance and robustness of these hedge returns and control for possible alternative sources for cross-sectional variation of growth stock returns. Therefore, I apply two different sets of tests: Firstly, I construct portfolios of high and low F-Score growth stocks. I refer to companies with a F-Score of 0-3 as “low F-Score” and to companies with a F-Score of 7-9 as “high F-Score”. This is different from Piotroski (2000), as he referred to 0-1 F-Score stocks as “low F-Score” and to 8-9 F-Score stocks as “high F-Score”. I deviate from this approach to arrive at a larger sub-sample and to be independent from rare outliners. My eurozone sample is smaller than the US sample from Piotroski and it also covers a shorter time frame and I hence apply a broader definition of “high” and “low” F-Scores.
I compare the market-adjusted returns for a one-year holding period for both portfolios and calculate the hedge return for the strategy by subtracting the low F-Score market-adjusted portfolio return from the high F-Score market-adjusted portfolio return. Portfolios are readjusted once a year, always at June 30th.
For simplicity reasons, I do not consider trading costs, slippage or taxes in this analysis.

In a second step, I test whether the results are significantly different from zero at a meaningful confidence level. I do so by applying a variety of t-tests at different confidence levels.

The results from this analysis will provide evidence on whether I can confirm or falsify the following hypotheses:

Hypothesis 1

The market adjusted return of the high F-Score sub-sample is >0.
Null: The market adjusted return of the high F-Score sub-sample is <=0.

Hypothesis 2

The market adjusted return of the low F-Score sample is <0.
Null: The market adjusted return of the low F-Score sample is >=0.

Hypothesis 3

The return of buying high F-Score stocks and shorting low F-Score stocks is >0.
Null: The return of buying high F-Score stocks and shorting low F-Score stocks is <=0.

After testing my hypotheses, I take a look at some practical considerations regarding the hedge portfolio by discussing statistical features around operational feasibility of each sub-portfolio. Especially market capitalizations and trading volumes of the sampled stocks are of interest since they allow a conclusion on the practical implementation of the strategy.

In addition to the evaluation of hedge returns, I am going to control for possible other factors that could explain the returns in the growth-segment of the eurozone equity market. To do so, I build a multifactor regression that consists of the explanatory factors size, P/B, momentum, accruals, equity offerings and F-Score. My model closely matches the model used by Piotroski (2000, p. 22):

$$MA\_Ret_i = \alpha + \beta_1 \log(MV_i) + \beta_2 \log(P/B_i) + \beta_3 MOM_i + \beta_4 ACC_i + \beta_5 EQOFFER_i + \beta_6 FSCORE_i + \epsilon$$

with:

MA_Ret_i: market adjusted return for a given stock at the time of portfolio formation

MV_i: market value of a given stock at the time of portfolio formation

P/B_i: price-to-book ratio of a given stock at the time of portfolio formation

MOM_i: prior 12 months stock return of a given stock at the time of portfolio formation

ACC_i: (Earnings – Cash-Flow)/Assets for a given stock at the time of portfolio formation

$EQOFFER_i$: equity issuance in the prior 12 months before portfolio formation results in a dummy variable 1, no issuance or buy-backs result in a dummy variable of 0.

$FSCORE_i$: F-Score at the time of portfolio formation.

ϵ_i: error term

The factors in this regression are based on widely-quoted research. Indeed, size effect and P/B are components of the original three-factor model (Fama and French, 1992).

Momentum has been established as another strong explanatory factor of stock returns (Chan, Jegadeesh, Lakonishok, 1996) and can be found as the extension to Fama and French's work in the four-factor-model of Carhart (1996). Sloan (1996) shows how accruals are a strong explanatory factor for stock returns. Loughran and Ritter (1995) found similar explanatory utility for equity offerings.

Unlike Fama and French (1992) and Carhart (1996), neither Piotroski nor I include beta as a factor in his model as recent evidence (Fama and French, 2003) suggests that the relationship between market return and beta is not clear.

While the general methodology of my multifactor regression is the same as in Piotroski (2000), I decide to use more recent research when constructing the momentum factor. Hancock (2010) shows how 12 months momentum is best suited to predict subsequent stock returns. Therefore, I use 12 months momentum as opposed to 6 months momentum as in Piotroski (2000, p. 23).

In my testing, I first run a regression over the entire growth sample with the factors size, P/B, momentum, accrual and equity offerings. In a second regression, I include F-Score as a sixth factor. I formulate two hypotheses:

Hypothesis 4

The explanatory ability of the factor model increases significantly when F-Score is included as a sixth factor.

Null: The explanatory ability of the factor model does not increase when F-Score is included as a sixth factor.

Hypothesis 5

The factor B_6 for F-Score is significantly greater than 0.

Null: The factor B_6 for F-Score is 0.

Finding evidence that allows the falsification of the null hypotheses adds meaningfully to the general question whether or not F-Score qualifies as an explanatory variable in the growth stock universe.

4 Empirical Results

4.1 Evidence from the Returns of a Market-Neutral Strategy in Growth-Stocks

Table 1 shows evidence on how a market-neutral strategy of buying high F-Score growth stocks and shorting low F-Score growth stocks performed in the 1999-2010 time intervals. We see that buying high F-Score companies yields an average market-adjusted mean return of 10,74% per year. Notably, the average market-adjusted mean return of the high F-Score stocks is more than 4% points higher than the average median of those stocks. This leads us to believe that the mean might be influenced by a few very positive years, which appears to be the case in 1999. However, there are two reasons to assume that the very positive market-adjusted return of the high F-Score growth stocks is not due to outliner years. Firstly, the median market-adjusted return is, albeit lower than the mean, still much greater than zero. Secondly, the market-adjusted return for high F-Score growth stocks is positive in each year between 1999 and 2010.
I further applied different t-tests on the results that show that the 10,74% annual market-adjusted mean return is significant at the 99% confidence level. Moreover, the t-tests for each single year show that the market-adjusted returns are statistically significant in eight out of twelve years. In these eight years, we observe significant results at the 95% confidence level in five years, at the 99% confidence level in two years and at the 99,9% level in one year. Hence, my findings allow me to falsify the null hypothesis to the corresponding hypothesis 1.

Next, I evaluate the market-adjusted returns of the low F-Score sample. Over the 12 years starting in 1999, the low F-Score stocks show average market-adjusted mean returns of -13,82% per year. In fact, this closely matches the equivalent average market-adjusted median returns, which are -13,62%. The low F-Score portfolio show negative market-adjusted mean returns in ten out of eleven years. T-testing of the return data shows that the -13,82% mean market-adjusted return is significant at the 99,9% confidence level. Annual market-adjusted returns are significant in seven out of twelve years, with confidence levels of 95% (in 2000 and 2009), 99% (in 2007 and 2008) and 99,9% (in 2001, 2002 and 2006). Therefore, my data supports the rejection of the null hypothesis that the market adjusted return of the low F-Score sample is >=0.

Finally, I consider the market-neutral hedge returns of the strategy by analyzing the high F-Score minus low F-Score returns. Returns of this strategy would be achieved, if an

investor were to buy a portfolio of EUR X high F-Score stocks and short an equivalent amount EUR X of low F-Score stocks. Result would be a market-neutral portfolio with the market-adjusted return of both portfolio-"sides" as the exclusive source of return.
Over the twelve-year time frame, the market-neutral strategy shows an annual return of 24,57%, which is significant at the 99,9% confidence level. The hedge-return was positive in each but one year and in this single year it did not show a large negative return. All these findings provide evidence that the null hypothesis 3 can be rejected.

Concluding this first set of tests, I find convincing evidence that a market-neutral asset management strategy of buying high F-Score and shorting low F-Score growth stocks offers attractive returns. This finding confirms earlier research by Piotroski (2004) who states that F-Score does not lose its predictive ability when applied to growth (instead of value) stocks. However, my results provide some counter-evidence to Mohanram (2005) who provided evidence that fundamental analysis is strongly context-dependent and that F-Score loses its predictive ability once applied outside the original value stock universe.

4.2 Operational Feasibility of the Hedge Portfolio

Although the data above provides striking evidence for the utility of a strategy that buys high F-Score stocks and shorts low F-Score stocks, the number of stocks in each portfolio could be a reason for criticism. In fact, within the high F-Score portfolio, the number of stocks ranges from 31 (1999) to 111 (2005). In the low F-Score portfolio, the number of stocks is a minimum of 19 in 1999 and a maximum of 81 in 2009. One might argue that this relatively low sample size could make the strategy vulnerable in a practical set-up.
The reason for the changes in the number of stocks in each portfolio is two-fold. The first major driver behind the number of stocks in each portfolio is the total number of stocks in the Eurozone stock universe, as we derive our sub-sample from this universe. As the number of listed companies increases, so does the number of growth stocks among which we select our high / low F-Score sub-samples. Secondly, general trends in the economic climate of the Eurozone influence the distribution of F-Score in the growth sample. When the economy is booming, we might suspect a higher number of high F-Score stocks and vice versa. We see empirical evidence for this as the ratio between the number of high and the number of low F-Score companies is high in times of economic well-being (1999 and 2004-2006) and low in times of economic depression (2001-2002 and 2009-2010). This could result in a change of portfolio diversification, as portfolio managers would be very

diversified on the long side (in an economic boom) and very diversified on the short side (during a recession). Possible effects on return and especially on risk of the portfolios are not investigated further in this paper but provide a good starting point for further research.

However, the absolute number in each sub-portfolio is very relevant from a practical perspective. Most institutional managers are constrained by diversification rules that oftentimes limit exposure to one security to levels that are normally below 5% (BVI information circular, 2008). This means that portfolio managers would need at least 20 securities to chose from on each the long and the short side of the portfolio. In our 1999-2000 sample, there seems to be a limited number of shares to chose from, especially on the short side. Adding to this fact are natural constrains to shorting, such as borrowing costs, availability of shares to borrow etc. While tests for significance show highly significant market-adjusted returns even for the small portfolio sizes, practical considerations might prove that the short strategy is hard to implement in a real-life setting. In this paper, I do not evaluate borrowing constrains or economic feasibility of shorting the sampled stocks. However, this is another promising starting point for further research.

Adding to the aforementioned discussion around the number of available securities in each portfolio, it seems reasonable to take a closer look at the characteristics of each the high F-Score portfolio and the low F-Score portfolio. My focus is on the median market capitalization and the median trading volume in the sampled stocks. Information on both parameters sheds further light on the practicability of implementing a market-neutral strategy using F-Score in the growth stock universe. I use median market capitalization and median trading volume, respectively, since these provide useful indicators whether or not the stocks in the sample are implementable. The means of both indicators do not offer much information for practical purposes because each mean can be upward biased by very large and heavily traded companies within the sample. In fact, empirical data from my sample shows exact this effect, as mean market capitalizations and trading volumes vastly exceed medians in all years.
In table 2, I list the median market capitalizations for each sample from year 1999 to year 2000.The empirical finding is that in every year (with the exception of 1999), the high F-Score sample shows a higher median market capitalization than the entire growth sample. Therefore, an investor of high F-Score companies would be biased towards larger

capitalization stocks.
A second finding is the median market capitalization of low F-Score companies. Indeed, the data shows that the low F-Score companies tend to be much smaller in market capitalization than a) the entire growth stock universe and b) the respective high F-Score companies. A possible conclusion of these findings is the fact that a market-neutral strategy of buying high F-Score companies and shorting low F-Score companies would include a size bias, which might explain some part of the return difference.
Further research could center on the logic relationship between size and F-Score, i.e. whether large (small) size causes high (low) F-Scores or whether high (low) F-Scores cause a company to perform well (bad) in terms of market capitalization. The first causal relationship would support the thesis that size is a predictor for distress and that larger companies are less likely to be in a distressed state (Altman, 1964). Such research could also aim at explaining the relatively high fluctuations in median market capitalization year over year. A possible starting point for research could be to test whether F-Scores at companies have any embedded auto-correlation or to evaluate what causes the median market capitalizations to fluctuate so widely.
However, this topic needs further research attention and the apparent size / F-Score correlations in my findings motivate research in that field.
I do cover part of the size / F-Score correlation question in the next part of this paper though, as size is one of the explanatory factors in the multifactor regression I construct.

From an implementation perspective, the findings in the median market capitalization data illustrate a constraint for investors to implement the short-side of a market-neutral portfolio. In fact, engaging in companies with a median market cap of less than EUR 100.000.000 (as shown in the sample in six out of twelve years) could be challenging for larger institutional investors. This issue is intensified by two more aspects. Firstly, the availability of borrowing shares of small companies seems questionable and secondly do agency costs increase as company size decreases (Schwert, 1983).

To further evaluate this matter, table 3 shows the median trading volumes for each sample. Median trading volume provides another data point on the practical utility of running a market-neutral strategy of buying high F-Score growth stocks and shorting low F-Score growth stocks. Similar to the findings regarding median market share, high F-Score companies seem to trade with much higher liquidity than a) the average growth stock and b) the average low F-Score stock.

Frankly, the average median trading volume for high F-Score companies should be sufficient even for larger institutional investors to build a relevant position in a given stock. I derive this conclusion from my earlier assumption, that one single investor can buy up to 50% of an average daily trading volume when entering or exiting a stock position. Therefore, even a position of several million EUR could be set up in high F-Score growth stocks in a short time. Nevertheless, the results from the low F-Score sample show a different picture. In this case, median trading volumes are very small and seem unlikely to be interesting for larger institutional investors. Again, this finding supports the earlier evidence that buying high F-Score stocks should be available even for larger institutional investors but shorting the respective low F-Score stocks is prone with implementation issues.

4.3 Controlling for other Sources of Cross-Sectional Variation in Returns

Although our basic test of the market-neutral portfolio shows very strong and significant outperformance (underperformance) of high (low) F-Score growth stocks, this could still be due to embedded correlation between F-Score and other explanatory factors. As explained in the Research Design section, I control for five other return patterns, namely size, P/B, momentum, accruals and equity offerings. As pointed out in Piotroski (2000), accruals and equity offerings are correlated with F-Score since they are both parts of F-Score. Additionally it makes intuitive sense for stock price momentum to be correlated with F-Score which displays some kind of 'fundamental momentum'. Additionally, the low information dissecting attributes of certain equity markets are seen as explanation for momentum strategies (Chan, Jegadeesh, Lakonishok, 1996) and F-Score (Piotroski, 2000) to work. Hence, it is important to test whether F-Score has additional explanatory utility over the aforementioned factors. I do so in first testing a multivariate regression consisting of five variables and then adding F-Score as a sixth variable. I observe the change in R^2 and test whether the change is significant at the confidence level of 99%. Next, I evaluate the ß-factor for F-Score and test whether it is significantly different from zero at the 99% confidence level.

Table 4 shows the regression estimates for the five variables and the applicable test statistics. The literature opinion that P/B, accrual and equity offerings have significant negative ß-factors is confirmed. Moreover, the very significant positive ß-factor for the momentum variable supports the view that a high momentum is predicator for high market-adjusted returns. Unlike the established literature opinion though, size effect does not show significant explanatory utility at the 95% or higher confidence level. This is a noticeable data point for follow-up research on the size / F-Score correlation.

Next, I add the variable F-Score as a sixth factor and run the regression again. Results are shown in table 5.

We now see that size effect loses even more explanatory utility and that the ß-factor for equity offerings is not any more significant at the 95% or higher confidence level.
By adding F-Score as a sixth variable, R^2 increases by 0,0123. Note that the initial R^2 is fairly low and that adding F-Score increases R^2 by a third- albeit from a low level. I apply a F-test to evaluate the significance of the incremental R^2-change by adding F-Score to the regression. At the 99% confidence level, the critical F-test result is 6,6421 (1 and 3496 as degrees of freedom). However, the empirical F-test value for the incremental increase in R^2 is much larger at 45,40. Therefore, I can reject null hypotheses 4.
We can also see that the ß-factor for F-Score is significantly positive, which is a reason to falsify null hypotheses 5.
In comparison to Piotroski (2000), I observe a very similar ß-factor for F-Score (0,034) as did Piotroski (0,025 to 0,03). These results do also have very similar t-statistics.

From my regression findings I conclude that F-Score adds significantly to the explanatory utility of a five-factor multivariate regression that aims at explaining market adjusted returns. Moreover, the respective ß-factor for F-Score is positive and significantly different from zero which leads me to believe that increasing F-Scores relate to increasing market-adjusted returns.

5 Conclusion

I can sum up my findings as follows:

a) separating growth stocks by applying F-Score seems to be a promising strategy. In constructing a market-neutral portfolio, buying high F-Score and shorting low F-Score growth stocks seems to yield a positive return.

b) When applying a multivariate regression to explain market-adjusted returns and variance of returns within the growth segment of the market, adding F-Score as an additional explaining factor adds significantly to the explanatory power of a multivariate regression.

c) Albeit evidence shows that shorting low F-Score growth stocks is an attractive strategy, I expect several implementation problems coming along with such a strategy.

I conclude by illustrating several ways how to seize the discussed return differences and how to cope with practical implementation issues.
The most straightforward way to take advantage of the apparent trading opportunity within the growth stock universe is the rebuild the tested market-neutral strategy in a real-life setting. Given that shorting stock is very much restricted for many investors, a hedge fund is the appropriate vehicle to run such an equity long/short strategy (Kaiser, 2004). However, I suspect this strategy to perform lower than the empirical results might indicate for two reasons: a) trading costs and borrowing costs were not included in my backtest but would incur for a hedge fund and b) availability to short stocks is likely to be restricted. A possible alternative is a long-only strategy of buying high F-Score growth stocks in a vehicle such as a mutual fund. Given that the high F-Score growth stocks show very consistent market-adjusted performance in my testing, I suspect a long-only strategy to outperform its benchmark. Nevertheless, leaving out the short side of an asset management strategy is likely to be not optimal, as my testing suggests attractive sources for alpha in shorting low F-Score growth stocks. A 130/30 strategy seems to be the dominant strategy in the trade-off dilemma between seizing the returns of the short portfolio and the restrictions of shorting (Johnson et al., 2007). Such a strategy buys high F-Score growth stocks for 100% of the portfolio and borrows an additional 30% of equity to purchase even more of such stocks. To offset the increased market exposure, an equivalent of 30% of equity is sold short in low F-Score growth stocks. Thereby, the 130/30 strategy would enhance the returns of the 100% high F-Score portfolio by adding the hedge return of the 30/30 overlay. This strategy adds value, if the hedge return of the 30/30 part exceeds the borrowing costs. I suspect this to be the case in a real-life setting as the testing of my hedge strategy suggests a hedge return many times higher than the margin borrowing costs at the time. The practical advantage of a 130/30 strategy lies at the low required number of stocks sold short. If an asset manager aims at limiting individual position sizes to, for instance, 2% each, the manager would need only 15 stocks on the short side to run a 130/30 strategy. This seems to be much more practical than the required 50 short positions if the manager were to run a 100/100 market-neutral strategy.

Further research could provide evidence on the past results of such a strategy and test whether the likely outperformance can be explained by other explanatory factors.

6 References

Altman, Edward I., 1968, Financial Ratios, Discriminant Analysis and Prediction of Corporate Bankruptcy, *Journal of Finance* 23, 589-609.

Artman, Sabine, Philipp Finter, and Alexander Kempf, 2010, Determinants of Expected Stock Returns: Large Sample Evidence from the German Market, mimeo.

Basu, Sanjoy, 1977, Investment Performance of Common Stocks in Relation to their Price Earnings Ratios: A Test of the Efficient Market Hypothesis, *Journal of Finance* 32, 662-682.

Basu, Sanjoy, 1983, The Relationship Between Earnings' Yield, Market Value and Return for NYSE Common Stocks: Further Evidence, *Journal of Financial Economics* 12, 129-156.

Beneish, Messod D., 1999, The Detection of Earnings Manipulation, mimeo.

Beneish, Messod D., D. Craig Nichols, 2009, Identifying Overvalued Equity, mimeo.

Brouwer, Iwan, Jeroen van der Put, and Chris Veld, 1996, Contrarian Investment Strategies in a European Context, mimeo.

BVI, 2008, Investment Fonds bieten Inflationsschutz, information circular.

Campbell, John Y., Robert J. Shiller, 2011, Valuation Ratios and the Long-Run Stock Market Outlook: An Update, *Cowles Foundation*, Discussion Paper 1295.

Carhart, Mark M., 1997, On Persistence in Mutual Fund Performance, *Journal of Finance* 52, 57-82.

Chan, Louis K.C., Yashui Hamao, Josef Lakonishok, 1991, Fundamentals and Stock Returns in Japan, *Journal of Finance* 46, 1739-1764.

Chan, Louis K.C., Narasimhan Jegadeesh, and Josef Lakonishok, 1996, *Journal of Finance* 51, 1681-1713.

Chan, Louis K.C., Josef Lakonishok, 2002, Value and Growth Investing: A Review and Update, mimeo.

DeBondt, Werner F.M., Richard H. Thaler, 1987, Further Evidence on Investor Overreaction and Stock Market Seasonality, *Journal of Finance* 42, 557-581.

Dubinsky, Andrew, 2006, Value Investing Retrospective, mimeo.

Fama, Eugene F., Kenneth R. French, 1992, The Cross-Section of Expected Stock Returns, *Journal of Finance* 47, 427-465.

Fama, Eugene F., Kenneth R. French, 1993, Common Risk Factors in the Returns of Stocks and Bonds, *Journal of Financial Economics* 33, 3-56.

Fama, Eugene F., Kenneth R. French, 1996, Multifactor Explanations of Asset Pricing Anomalies, *Journal of Finance* 51, 55-84.

Fama, Eugene F., Kenneth R. French, 1998, Value vs. Growth: The International Evidence, Journal of Finance 53, 1975-1999.

Fama, Eugene F., Kenneth R. French, The CAPM: Theory and Evidence, mimeo.

Grantham, Jeremy, 2010, On the Potential Disadvantages of Graham and Dodd-Type Investing, *GMO Letters to the Investment Committee* 16.

Hancock, Tom, 2010, Momentum- A Contrarian Case for Following the Herd, *GMO White Paper*.

Jensen, Michael C., 2004, Agency Cost of Overvalued Equity, mimeo.

Johnson, Gordon, Shannon Ericson, and Vikram Srimurthy, 2007, An Empirical Analysis of 130/30 Strategies: Domestic and International 130/30 Strategies Add Value Over Long-Only Strategies, *Journal of Alternative Investments* 10, 31-42.

Kaiser, Dieter G., 2004, Hedgefonds. Entmystifizierung einer Anlageklasse. Strukturen – Chancen – Risiken, *Gabler Verlag*, Wiesbaden

La Porta, Rafael, 1996, Expectations and Cross-Section of Stock Returns, *Journal of Finance* 51, 1715-1742.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, Contrarian Investment, Extrapolation, and Risk, *Journal of Finance* 49, 1541-1578.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, 1985, Persuasive Evidence of Market Inefficiency, *Journal of Portfolio Management* 11, 9-17.

Lee, Charles M.C., Bhaskaran Swaminathan, 1999, Prime Momentum and Trading Volume, mimeo.

Loughran, Tim, Jay R. Ritter, 1995, The New Issues Puzzle, *Journal of Finance* 50, 23-51.

Mohanram, Partha S., 2005, Separating Winners from Losers among Low Book-to-Market Stocks using Financial Statement Analysis, *Review of Accounting Studies* 10, 133-170.

Piotroski, Joseph D., Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers, *Journal of Accounting Research* 38, 1-41.

Piotroski, Joseph D., 2004, Further Evidence on the Relation between Historical Changes in Financial Condition, Future Stock Returns and the Value/Glamour Effect, working paper.

Piotroski, Joseph D., 2005, Discussion of "Separating Winners from Losers among Low Book-to-Market Stocks using Financial Statement Analysis, *Review of Accounting Studies* 10, 171-184.

Schwert, G. William, 1983, Size and Stock Returns, and other Empirical Regularities, *Journal of Financial Economics* 12, 3-12.

Sloan, Richard G., 1996, Do Stock Prices Fully Reflect Information in Accruals and Cash Flows About Future Earnings?, The *Accounting Review* 71, 289-315.

7 Tables

Table 1

Portrayed are market-adjusted results of the high F-Score and low F-Score sample. Additionally, the hedge return of combining both strategies can be seen in column four. **Bold**, <u>underlined</u> *and* <u>double-underlined</u> *results indicate significant results at the 5%, 1% and 0,1% level, respectively.*

	high F-Score Mkt Adj Returns	low F-Score Mkt Adj Returns	high-low difference	number of observations high / low
1999	**0,4348**	-0,1904	<u>0,6252</u>	31 / 19
2000	0,1177	**-0,1826**	**0,3003**	33 / 26
2001	<u>0,1449</u>	<u>-0,2376</u>	<u>0,3825</u>	43 / 45
2002	**0,1271**	<u>-0,1896</u>	<u>0,3166</u>	32/ 49
2003	0,0319	-0,1482	**0,1800**	48 / 32
2004	**0,0765**	-0,0324	**0,1089**	90 / 34
2005	<u>0,0102</u>	0,0340	-0,0238	111 / 35
2006	0,0449	<u>-0,2673</u>	<u>0,3122</u>	96 / 56
2007	**0,0535**	<u>-0,1200</u>	<u>0,1735</u>	100 / 73
2008	<u>0,0717</u>	<u>-0,1019</u>	<u>0,1736</u>	83 / 68
2009	0,1045	**-0,1242**	**0,2287**	35 / 81
2010	**0,0717**	-0,0986	**0,1703**	55 / 66
Mean	<u>0,1074</u>	<u>-0,1382</u>	<u>0,2457</u>	
sample sd	0,1106	0,0843	0,1616	
Median	0,0741	-0,1362	0,2044	
t-stat	3,3656	-5,6772	5,2669	

Table 2

Portrayed are median market capitalizations for the entire sample and both sub-samples.
All numbers are as of June 30 in each year.
All numbers in Euro.

	Growth sample	High F-Score	Low F-Score
1999	79.985.000	96.880.000	177.035.000
2000	1.500.080.000	2.546.250.000	551.875.650
2001	877.075.000	1.711.215.000	244.300.000
2002	390.440.000	521.080.000	161.975.600
2003	308.210.000	1.147.829.000	63.412.400
2004	238.220.750	584.765.000	48.698.000
2005	182.505.000	428.060.000	28.444.000
2006	227.848.400	433.530.000	50.219.700
2007	405.006.100	729.411.300	133.787.200
2008	348.542.300	1.473.410.000	77.750.000
2009	245.056.950	784.639.600	101.850.800
2010	227.555.500	557.351.500	68.867.100

Table 3

Portrayed are median trading volumes of the 30 days prior to June 30 of each year. All numbers in Euro.

	Growth sample	High F-Score	Low F-Score
1999	36.296	37.671	88.509
2000	832.990	2.454.427	241.470
2001	373.739	251.155	203.140
2002	120.820	67.945	102.980
2003	132.185	277.700	86.638
2004	109.885	253.354	55.210
2005	133.643	436.430	59.685
2006	254.596	675.348	41.029
2007	457.002	756.727	62.902
2008	245.984	3.043.674	56.930
2009	124.716	318.500	84.898
2010	100.493	646.266	31.989

Table 4

Portrayed are the results from a multivariate regression of five factors that are modeled to explain the market-adjusted returns over my entire growth sample. The time period considered are the twelve 12 months periods from June 30 1999 to June 30 2011. EQOFFER is a dummy variable that takes the value 1 if new equity was issued in the twelve months prior to June 30 of each year, and 0 otherwise.

	Coefficients	*Standard error*	*t-stat*	*p-value*
Intercept	0,0522	0,0569	0,9171	0,3592
log MV	0,0147	0,0089	1,6473	0,0996
log P/B	-0,2787	0,0303	-9,1967	0,0000
MOM	0,0547	0,0086	6,3505	0,0000
ACC	-0,1288	0,0329	-3,9100	0,0001
EQOFFER	-0,0402	0,0166	-2,4202	0,0156

R^2=	0,0395
adj. R^2=	0,0381
degrees of freedom:	3500

Table 5

Portrayed are the results from a multivariate regression of six factors that are modeled to explain the market-adjusted returns over my entire growth sample. Unlike in table 4, F-Score is now included as a factor ("FSCORE").
The time period considered are the twelve 12 months periods from June 30 1999 to June 30 2011.
EQOFFER is a dummy variable that takes the value 1 if new equity was issued in the twelve months prior to June 30 of each year, and 0 otherwise.

	Coefficients	*Standard error*	*t-stat*	*p-value*
Intercept	-0,0320	0,0579	-0,5534	0,5800
log MV	-0,0039	0,0093	-0,4245	0,6713
log P/B	-0,2547	0,0303	-8,3996	0,0000
MOM	0,0434	0,0087	4,9749	0,0000
ACC	-0,1148	0,0328	-3,5016	0,0005
EQOFFER	-0,0166	0,0169	-0,9853	0,3246
FSCORE	0,0340	0,0051	6,7380	0,0000
R^2=	0,0518			
adj. R^2=	0,0502			
degrees of freedom:	3500			