Price is What You Pay, Value is What You Get

- Dissecting the Quality Anomaly in US Equity Returns

Master Thesis, First Year
Master in Economics
Filip Düsing
Erik Ivarsson
Supervisor: Birger Nilsson

Abstract

The purpose of the thesis relates to the Quality anomaly observed in the US equity market, where stocks with Quality characteristics tend to outperform and have higher risk adjusted returns. By dissecting the Quality anomaly, the thesis aims to analyze the drivers of the over performance of Quality and investigate the presence of a systematic Quality premium. From previous research, three areas have been identified as theoretical gaps – the magnitude of selection bias, how quality performs during different market conditions and if Quality has explanatory power in a cross sectional setting. By forming an aggregated, zero-investment Quality portfolio, regress it on traditional factor models, analyze condition Beta and perform a Fama Macbeth Cross Sectional Regression, the thesis aspires to address these gaps. Four main conclusions were brought to light;

(i) The lack of a coherent definition of the Quality factor impose selection bias;

(ii) There are tendencies towards flight to Quality – the risk-adjusted returns in excess of the market is mainly generated in down markets and over longer periods;

(iii) The presence of a Quality premium is observed. When regressed on multifactor models, the Quality portfolio generates monthly significant alpha in between 0.431 - 0.549 %. Furthermore, the Quality portfolio loads significantly negative on market Beta, and tendencies are observed on significant negative factor loadings on SMB and HML. Thus, traditional factor models cannot explain the Quality premium.

(iv) The premium appears to be caused by systematic errors rather than exposure to a systematic source of risk, as the Quality anomaly becomes evident first during longer time periods or during crises. In a CSR, Quality cannot be rejected to improve the model, but as such, it is also concluded not to be a compensation for carrying a systematic risk premium.

Table of contents

1 Introduction

The first chapter of the essay provides a brief background to the evolvement of factor investing, from both an academic as well as an institutional perspective. This is followed by a literature review, which puts the Quality Anomaly in context, in terms of estimation techniques and previous research. Subsequently, a short review of the anomaly is discussed, followed by identification of theoretical gap, our research questions and the purpose of this essay.

1.1 Background

Equity risk premium is a multi-facetted expression, but it is often used in the context of the framework presented by Ross (1976), who stated that expected return of a financial asset can be modeled as a function of several sources of risk. In this spirit, empirical finance researchers have tried to uncover and determine common characteristics among stocks that exhibit abnormally high risk adjusted returns. Analogous to the progress within academia, a trend has emerged amongst institutional and retail investors; the inflows to systematic equity factor risk premium strategies, mainly through Exchange Traded Funds (ETFs), has exhibited a growth of 31 % CAGR during the last five years (Blackrock, 2017).

Equity risk premium is as mentioned often described in the context of the framework presented by Ross (1976), Arbitrage Pricing Theory (APT). APT differs from CAPM in the sense that expected return of a financial asset can be modeled as a function of several sources of risk, not only the market factor, Beta[1]. Hence, in a general sense, a factor can be thought of as a specific trait, important for explaining an asset's risk and return characteristics. However, the field is broad and can also incorporate strategies such as put/call writing strategies, volatility roll down, carry trades, total return swaps and repurchase agreements (Bank for International Settlements, 2017).

Since the framework of CAPM was outlined in the 1950's and 1960's, one of the core concepts has been diversification – investors are compensated merely for holding market risk and not idiosyncratic risk (Markowitz 1952, 1959; Treynor 1961; Sharpe 1964; Lintner 1965; and Mossin 1966). Even if diversification is a cornerstone in the theoretical framework, the idea that a market portfolio (a capitalization weighted portfolio consisting of all available stocks) provides the highest risk adjusted

[1] However, Ross does not state what these factors should be, but the sources of risk (and thus the drivers of asset returns) is often assumed to be modeled as either macroeconomic dynamics or theoretical equity market indices.

return, is not uncontroversial. Fama (1976) as well as Haugen, Nardin and Baker (1996) showed that a cap weighted index will be mean variance efficient only when considered in the context of four main assumptions. The first is that all investors have homogenous expectations regarding risk and expected return for all securities. Second, there should be no constraint in terms of short-selling. Furthermore, the returns from any investor's portfolio should not be exposed to taxes. Lastly, the investment opportunity set, i.e the universe of tradable stocks, should be restricted to the stocks included in the cap-weighted index. Even in Wilshire 5000 (the most comprehensive equity index in the US), the fourth condition is violated (Haugen, Nardin, Baker, 1996). Neither is it realistic that the other assumptions are fulfilled. Absent above mentioned assumptions, even the most comprehensive cap-weighted portfolios engage positions within the efficient set. This finding implies that there is no practical way of implementing a "truly" mean variance efficient portfolio, using a broad index. This might be one of the explanatory sources as to why factor investing has increased in popularity throughout the years.

Fama et. al. (1969) and Fama (1991) are famous for outlining the Efficient Market Hypothesis, which has been central in testing to what extent stock market returns can be forecasted. There is a quite broad consensus amongst contemporary financial economist that markets are hard to predict, and any achievement of doing so is a result of mere chance (Bogle, 2009; Malkiel, 1995, 2012; Gruber, 1996; Barras et al, 2010; Berk and Binsbergen, 2012; Kosowski et al, 2006; Wermers, 2003; Jones and Wermers, 2011; Kinnel, 2010; Arnott, Berkin, and Ye, 2000; Ibbotson and Kaplan, 2000).

In this context, the chase for alpha and the employment of active management investment strategies, appear to be a futile endeavor. This, combined with the empirical arguments against the idea that cap-weighted indices are mean variance efficient, leaves investors with a confusing setup. However, when the returns of active managers that consequently beat their respective benchmark is dissected, an interesting conclusion comes to light. Fama and French (2010) as well as Ang, Goetzmann and Schaefer (2009) showed that a majority of successful mutual funds tend to be exposed to well documented risk factors in the equity markets. In a similar note, Mok, Bender and Hammond (2013) found that about 50 % of the excess returns of mutual funds could be explained by the Fama French Three Factor Model. Thus, this implies that there is something systematic in the alpha generated by these managers. One of the conclusions in Mok et al is that some of this alpha is not a stochastic residual, but rather a systematic, extractable, dimension of equity return that can be harvested by using systematic factor strategies.

Lately, this line of thinking has been widely adopted by the industry. BlackRock, the world's largest asset manager, earns a significant portion of their revenue from passive investing solutions and Exchange Traded Funds (ETFs) (BlackRock, 2017). Assets allocated to ETFs amounts to approximately USD 2 885 Bn, corresponding to 16 % of total assets under management in the US. Furthermore, the growth rate for ETFs during the last five years has been more than double that of other investment vehicles, aggregated (ETFGI, 2017).

The increased demand from investors combined with a vast contribution within academia on equity risk premium, has enabled access to low cost ETFs in the equity markets[2]. These investment strategies are mainly constructed by using a broad parent index, but assigning different member weightings than in the cap weighted index. The most popular strategies base their member weightings on factors such as low Volatility, Dividend, Value, Quality, Size, Growth and Momentum rather than market capitalization.

The implementation is often determined by using fundamental metrics as proxies for various risk premiums. One such factor, that has been more frequently and widely adopted, is Quality. In a general context, Quality screens aim to capture the premium of companies that have stable business models as well as growing profits, low leverage and solid cash flows (MSCI, 2013). Thus, the Quality strategy is often communicated as a way of mimicking the returns of investment guru's such as Warren Buffet or Benjamin Graham. However, there is no coherent way of defining Quality, and even if various Quality screens have performed well in the past, there are different opinions as to why they have outperformed the market and if they will continue to do so going forward.

This essay aims at dissecting the Quality Anomaly, and investigating potential sources of its historical outperformance. To fulfil the purpose, the essay is divided in three sub areas: investigating selection bias, the performance of Quality and Quality as a systematic risk factor. These three areas are identified as gaps from previous research. Before these are described more in detail, a section of previous research is presented, which puts the Quality Anomaly in context.

[2] The cost for strategy ETF:s issued by iShares (BlackRock) is between 5 – 75 basis points annually, and the vast majority has net expenses of 20 basis points (iShares, 2017).

1.2 Literature Review and Previous Research

1.2.1 Factor Investing

CAPM can be viewed as the first factor model for asset prices. It uses only the market Beta as a factor, important for explaining asset returns (Markowitz 1952, 1959; Treynor 1961; Sharpe 1964; Lintner 1965; Mossin 1966; and Black 1972). Since the CAPM framework was outlined, a lot of research has been conducted in the spirit of Ross' APT (1976). Subsequently, a lot of evidence has been put forward to strengthen the existence of certain risk premiums in the equity markets, most of them starting from a point where they see CAPM as an obsolete model that lacks explanatory power. Before these academic efforts are descried more in detail, an outline of previous research regarding factor methodology is presented.

Connor (1995) stated that there are three main categories of factors: macroeconomic, statistical, and fundamental. Macroeconomic factors can for instance be surprises in PMI, changes in various business cycle variables, inflation and changes in the yield curve. (See also Chen, Ross, and Roll (1986)).

Statistical factor models on the other hand takes advantage of various statistical estimation techniques, such as principal components analysis. The principal components analysis method selects a linear combination of asset returns which contribute with the highest variance (Egloff, Leippold, Wu 2010; Litterman, Scheinkman, 1991; Stock, Watson 1999; Johnson & Wichern 2009). The principal component analysis was developed by Pearson (1901) and Hotelling (1933), but one of the more cited modern reference is Jolliffe (2002). There is a vast spectrum of methods concerning estimation techniques, and except from principal component analysis, some of the more frequently used techniques are panel regressions, Bayesian models and latent factor models, to mention a few (Miller, 2006).

Fundamental factors aim to capture certain characteristics among stocks, and to be a proxy of traits that are not directly observable. Fundamental factors have been thoroughly studied since the framework of APT was outlined, as a part of the field of academic asset pricing. Among the first to describe the prominence of individual stock traits as an explanatory variable of stock returns was Fama and Macbeth (1973). They could not reject the hypothesis that no measure of risk, in addition to Beta, systematically affected expected returns. They created a framework for testing for various risk premiums in the equity market, the so called Fama Macbeth Regressions. Due to this contribution, the foundation for many empirical papers regarding equity risk premiums was outlined.

One of the most cited and well known academic efforts in the field of fundamental factors originates from the research of Eugene Fama and Kenneth French in the early 1990s. Fama and French (1992, 1993) put forward a model explaining US equity market returns with three factors: the "market" (defined as in the traditional CAPM), the size factor (a sort based on large capitalization stocks versus small capitalization stocks) and the value factor (high book-to-market value of equity versus low book-to-market value of equity). Fama and French concluded that market Beta has a low explanatory power in terms of explaining the cross-sectional variation in the returns of stocks and bonds, implying that the traditional CAPM framework is insufficient as an explanatory model for the drivers of asset returns. Throughout the past decades, empirical finance researchers have studied a multitude of auxiliary stock traits, ranging from cash flow-, income statement- and balance sheet metrics, in order to unfold new fundamentally based factors in the equity markets.

1.2.2 The Quality Anomaly

Quality, as an investment strategy, seeks to capture the excess returns of companies that are efficient in an operational sense, are stable in terms of earnings and cash flows, have low leverage, are highly profitable and associated with low operational risk. Even if this definition is quite vague, these traits and this line of thinking have been popular in the active investment industry for decades. But as a more quantitative phenomenon, Quality is a rather new occurrence. As a dimension in factor investing, Quality was popularized first by Asness, Frazzini, and Pedersen (2013), but it is still not consistently defined. This indicates that the risk sources behind a Quality screen has not yet been uncovered and documented thoroughly.

Like other stock market anomalies, the Quality anomaly has been identified from empirical tests of the Capital Asset Pricing Model of Sharpe (1964), Lintner (1965) and Black (1972), and later the multifactor models of Fama and French (1992) and Carhart (1997). The evidence from an empirical stand point indicates that portfolios sorted on Quality metrics such as profitability, earnings quality and safety have produced higher risk-adjusted returns relative to the market portfolio. However, the size of the premium is not coherent; it varies depending on which metrics are used, the time period investigated, the geographical market or stock sample examined as well as the asset pricing model used to measure portfolio risk. This leaves a somewhat disintegrated picture of the Quality anomaly.

One of the first published Quality screens goes back to Benjamin Graham and the book "The

Intelligent Investor" from 1949. In chapter 14, Graham outlines a screen of combined financial metrics (Graham, 1949). Graham's strategy was built on the premise that undervalued and underappreciated companies, that meet some given criteria, should be subject to higher expected returns. Graham considered metrics such as debt ratios, earnings stability, past earnings and dividend growth to be as important as valuation metrics such as price-to-earnings and price-to-book ratios.

Another publication of Graham and Dodd (Graham, Dodd, 1934, page 351, Security Analysis) outlines a more rigor definition of Quality. This line of thinking was adapted by Graham's disciple, the famous Warren Buffet. Buffet's investment company Berkshire Hathaway has realized a Sharpe-ratio of 0.76, higher than any other stock or mutual fund with a history of more than 30 years, and Berkshire has consecutively produced significant alpha compared to the CAPM (Asness et al, 2013).

Another impactful individual, who share a similar investment philosophy as Buffet, Dodd and Graham, is Peter Lynch. His efforts at the Magellan Fund and Fidelity Investments have made him an investment guru in the asset management industry. The common denominator is that they all invest in stocks that exhibit the Quality characteristics discussed above. Due to their high returns in excess of the stock market, their strategies have been subject to dissection by both practitioners as well as academia.

Some of the early research on statistical relationships of earnings and stock performance links back to Foster (1977); Watts and Leftwich (1977); Albrecht et al. (1977); Beaver (1970); and Griffin (1977), indicating that profitable stocks tend to outperform broad benchmarks. In the 1980's, Graham and Dodd's Earnings Quality measure was re-introduced into the academic sphere, as a descriptive characteristic of earnings for academic researchers (O'Glove, 1987; Lev 1989), and thus forming a more coherent view on quality as a definition.

Sloan (1996) was one of the first to validate the excess returns to high earnings quality stocks, where accruals proxy for earnings quality. Other examples of studies of this type are Lev and Sougiannis (1996), who evaluate how investor responds to increased or decreased earnings and accruals. This is was also described by Landsman et al. (2008). Bender and Nielsen (2013) and Kozlov and Petajisto (2013) both reconfirm the accruals effect, but for the the 2000s. They find similar results, namely that the accruals anomaly persisted throughout the time period examined in generating positive alpha. Leippold and Lohre (2010) concluded that in 22 of the 26 markets they examined, the accruals constituted an evident anomaly.

In a similar manor, Huang (2009) finds that firms with stable cash flows tend to outperform, measured as volatility of cash flow. The paper argued that cash flow volatility provides a better measure of the overall riskiness of a company than accruals.

Dichev (1998), Griffin and Lemmon (2002), Vassalou and Xing (2004) and Campbell et al. (2008) all found that financial distress, as defined by trailing financial ratios, on average is associated with lower returns. The result applies for different markets and during diverse time periods. This was also investigated by George and Hwang (2010), by studying low leverage companies. The authors showed that there is a significant return premium in companies with low leverage, and when the results were put in a risk-adjusted return context, the results became even more clear. Another effort in this space was made by Penman, Richardson and Tuna (2007). They separated the book-to-market ratio into an asset and a leverage component. From this, they concluded that the leverage component of the book-to-market ratio negatively predicts to stock returns.

Other modern endeavors within this field, generated in the spirit of Graham's screen, have been made by Pitroski (2000) and Greenblatt (2005). Pitroski (2000) proposed an investment screen, based on 9 financial metrics, and the screen outperformed and produced significant alpha. The investment strategy bought expected winners and shorted expected losers, by filtering on financial metrics, and it generated 23 % annual return between 1976 and 1996. The strategy was robust over time. The screen has been revisited by many practitioners since it was outlined, and the screen has continued to perform well (Hyde, 2015).

Greenblatt published a book in 2005, "The Little Book that Beats the Stock Market", in which he outlined a stock screen, called "The Magic Formula" (Greenblatt, 2005). A dissection of the strategy was done by Novy-Marx in 2013, where the results indicated that the investment strategy produced significant results (alpha of 2,8 % yearly) (Novy-Marx, 2013). Before this, Novy-Marx (2012) identified a proxy for profitability that was concluded to be closely correlated with average return. The sample at hand was the US stock market, and the time period spanned from 1963 to 2010. Stocks with high profitability characteristics produced alpha of 1,44 % per annum, and the return could not be explained by CAPM, Fama French Three Factor Model or Carhart's Four Factor Model. Moreover, the alphas were significant over time.

Dechow, Ge, and Schrand (2010) made a comprehensive review of how earnings persistence, accruals, earnings smoothness and loss avoidance affect stock prices. When weighing in on more than 300 papers (many originating from the accounting field) they concluded the following. First, abnormal accruals tend to have positive persistence. Second, investors appear to recognize the distinction between High and Low Quality firms (such as abnormal accruals and normal accruals), but they do not fully incorporate the implications into price. Third, the literature is inconsistent regarding the causality from observing a quality company (in terms of fundamentals) to the consequences for future period earnings. Some papers find that quality is predictable, whereas others do not.

A paper that merged fundamental metrics with a multi factor asset model was outlined by Chen, Novy-Marx and Zhang (2011). They included ROE in an alternative three factor model, with the aim to improve the explanation the cross-sectional variation stock returns. They showed that a long-short ROE factor earned a statistically significant average return of 0,71 % per month from 1972 to 2010, thus confirming that ROE can serve as a proxy for profitability and earnings quality.

Moreover, Novy-Marx (2014) finds that gross profitability performs relatively better than quality strategies such as Graham's quality, especially among large-cap US stocks. They also concluded that profitability has approximately the same explanatory power as book-to-market in explaining the cross section of average stock returns.

Aharoni, Grundy, and Zeng (2013) document a somewhat weaker but still statistically reliable relation between firm investment and average return. (See also, Haugen and Baker 1996; Cohen, Gompers, and Vuolteenaho 2002; Fairfield, Whisenant, and Yohn 2003; Titman, Wei, and Xie 2004; Fama and French 2008, 2014.). A more extensive paper, published by Asness, Frazzini, and Pedersen (2013) indicates that high quality companies, measured by three categories (profitability, stable growth, and high payout ratio) has significantly higher risk-adjusted returns than the market portfolio.

In 2013, a paper by Frazzini, Kabiller, and Pedersen was published, called Buffet's Alpha. Buffet's investment company Birkshire Hathaway has consequently produced significant alpha, but the alpha was found to be insignificant when controlling for exposures to other factors than Beta. This spurred an interesting discussion regarding to what extent Buffet's Alpha could be extracted using factors in the equity market (Frazzini, Kabiller, Pedersen, 2013).

In 2014, Fama French reiterated their factor approach from the 90s, in a paper called "A Five Factor

Asset Pricing Model". They extend the original three factor model with two additional measures – RMW and CMA. RMW stands for Robust Minus Weak, and is a sort based on the robustness of profits and earnings power. CMA stands for Conservative Minus Aggressive, hence sorting stocks based on the level of investments made by the company. Their results indicate that the model provides a good description of asset returns; however, the results were not statistically significant (Fama, French 2014).

Asness et al. (2014) found that portfolios sorted on profitability, safety and earnings quality have generated statistically significant alphas, both globally as well as in the US. The average profitability premium in the US over the period from 1956 to 2012 was 0,4 % per month, and the Four Factor Model alpha amounted to 0,53 % per month.

Despite the pervasiveness of quality investment strategies, not much research has attempted to explain why quality outperforms. In theory, it is reasonable that quality stocks should command higher prices. As Asness et al. (2014) point out, investors should be willing to pay a higher price for companies with quality characteristics, as these companies tend to have either higher expected cash flows or lower volatility in cash flows- Due to this effect, this kind of companies would not necessarily imply higher risk-adjusted returns, since they are priced higher.

To sum up previous research efforts, there is a clear indication that a quality screen, in various constellations, exhibit higher risk adjusted returns than the market, and that return patterns cannot be explained by the CAPM. The main challenge seems to be how to define the quality factor consistently and objectively by uncovering reliable proxies for the sources that drive the return of Quality stocks. As with other market anomalies, this premium may exist due to a variety of aspects, and the critiques range from insufficient risk models, measurement errors, data mining effects and overfitting, to behavioral biases (survivorship bias, home country bias, familiarity bias and selection bias to name some) and institutional constraints, including restrictions on short selling, tax effects et cetera (Davis, 2001). This is more thoroughly discussed in the Method section.

Exhibit 1 summarizes eight of the most widely studied factors.

Exhibit 1: Well documented Systematic Factors from the Academic Research

Factor	Explanation
Value	Captures excess returns to stocks that have low prices relative to their fundamental value
Low Size	Captures excess returns of smaller firms (by market capitalization) relative to their larger counterparts
Momentum	Reflects excess returns to stocks with stronger past performance
Low Volatility	Captures excess returns to stocks with lower than average volatility, Beta, and/or
Dividend Yield	Captures excess returns to stocks that have higher-than-average dividend yields
Quality	Captures excess returns of stocks that are characterized by low debt, stable earnings growth, and other "quality" metrics
Growth	Captures companies that have high historical sales and EPS growth, and high expected growth in EPS.
Liquidity	Captures companies that have low liquidity in their tradable assets

Companies that exhibits one or more of the characteristics outlined above, also tend to exhibit certain patterns in stock returns on average. These patterns cannot always be explained by the CAPM. In 2008, Fama and French (2008) dissected some of the most frequently used factors to conclude that anomalies can be linked to factors, in a cross sectional setting. Thus, some of these traits are considered to constitute anomalies that are left unexplained by the traditional framework of Modern Portfolio Theory.

1.3 Theoretical Gap and Research Questions

As described above, previous research indicates that quality screens generate higher risk adjusted returns, but there is not yet a clear, quantitative definition of quality as a factor. Neither has previous research been able to produce coherent results as to why quality characteristics tend to outperform the market. There is no clear explanation to why this phenomenon occurs; is it due to errors in investor expectations, or could there be a systematic, non-diversifiable source of risk, that investors demand higher returns for? Quality metrics seem to constitute a proxy for some source of risk, due to the excess returns generated over long time periods, but there is no broadly accepted explanation today. The most prevailing explanation is a "residual" remark; the reason why quality has outperformed historically is due to errors in expectations. Thus, the theoretical gap that this essay focus on is the lack of explanations of the quality anomaly.

Therefore, the scope of the thesis is to investigate the following areas:

Magnitude of Selection Bias: Create zero investment portfolios for single metrics, frequently used in the literature, to analyze the magnitude of the selection bias. Another aspect of the selection bias will also be analyzed, namely that two Quality portfolios are constructed, one based on previous research and one screen based on Svenska Handelsbanken (SHB) selection of metrics, and how different combinations of metrics affect the portfolio.

The Performance of Quality: To enhance the understanding the Quality anomaly, the thesis analyzes how Quality performs during different market conditions, as well as outlining if there exist a premium is in terms of alpha, in a multifactor setting.

The Quality Factor: Can a zero-investment portfolio, which is long High Quality stocks and short Low Quality stocks, help explain a larger proportion of the cross-sectional variation in equity returns, thus stating a systematic source of risk, or are the returns linked to a systematic error among investors?

1.4 Purpose

The purpose of the thesis is therefore to investigate why stocks with Quality characteristics tend to outperform and have higher risk adjusted returns. By dissecting the Quality anomaly, this essay aims to investigate the presence of a systematic Quality premium.

2 Theory

The chapter covers the main theories and academic concepts within the field of financial economics which relates to market efficiency and asset pricing models. The theories presented will serve as the basis for the thesis, providing the background and foundation essential to analyzing the Quality anomaly.

2.1 The Capital Asset Pricing Model

The Capital Asset Pricing Model (CAPM) is considered the foundation of asset pricing theory and constitutes the framework of Modern Portfolio Theory. It was developed by joint contributions from Markowitz (1952, 1959), Treynor (1961), Sharpe (1964), Lintner (1965), Mossin (1966) and Black (1972). The basis of the CAPM relies on the idea that there are two types of risk from which returns are generated – Systematic risk and unsystematic risk. Systematic risk can be viewed as market risks and hence cannot be diversified away, while unsystematic risk, so called idiosyncratic risk, is specific risk to an individual asset and thus uncorrelated to market movements. The unsystematic risk can be eliminated by diversification, why investors should only be compensated for carrying systematic risk. The CAPM provides a framework for measuring the systematic risk as a function between expected return and exposure to the market (Beta).

$$r_{it} = r_f + \beta_{im}(r_{mt} - r_f)$$

Where $\mathrm{r_{it}}$ is the return of any asset *i* during time *t*, r_f is the risk free rate, β_{im} is the sensitivity of asset *i* to the market return, r_{mt}. Usually, this equation is determined by regression analysis. If $(r_{mt} - r_f)$ is defined as a vector of excess market returns, **X**, and r_{it} is defined as a vector of asset i:s returns, **Y**, OLS can be used to estimate the β_{im}, by $(\mathbf{X'X})^{-1}\ \mathbf{X'Y}$. β_{im} can also be estimated using GLS if autocorrelation or heteroscedasticity is observed. Then an intertemporal homoscedastic covariance matrix $\boldsymbol{\Psi}_t$ can be used to estimate β_{im}. Markowitz proves that under certain assumptions, the β_{im} is defined as:

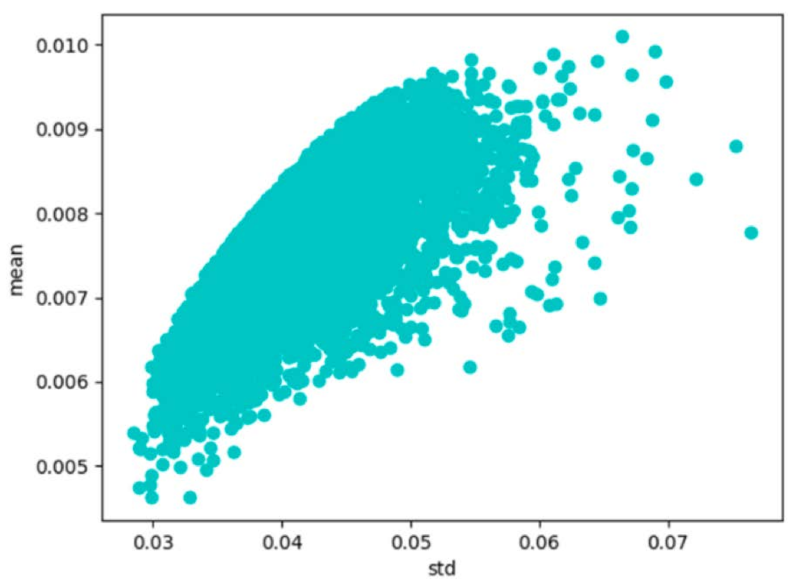
$$\beta_{im} = \frac{Cov\ (r_i, r_m)}{Var\ (r_m)}$$

2.2 The Mean Variance Framework

The Mean Variance Framework, popularly referred to as Modern Portfolio Theory, MPT, states that an investor wants to maximize the return for any given level of risk. MPT assumes that investors are risk averse, implying that an investor facing two portfolios with similar return characteristics but different risk levels, will chose the less risk one.

By this logic, an investor face a trade-off in terms of risk and return, and an investor will only increase the level of risk if she is compensated by higher expected returns. The relationship will be the same for all investors, but investors will evaluate the trade-off differently as each investor's utility function differs in terms of risk aversion. Consequently, the framework outlines a concept referred to as the efficient frontier, showing the combination of all available assets, which in turn shows the efficient set – the portfolios with the highest return for every level of risk. (Markowitz, 1952)

The efficient set shown below:



2.3 Arbitrage Pricing Theory

As an alternative to the Mean Variance Asset Pricing Model proposed by Sharpe (1964), Lintner (1965) and Treynor (1961), Arbitrage Pricing Theory (APT) was introduced by Ross in 1976. In the Mean Variance Model, the linear relation between return and risk (Beta) is used in order to price assets. The APT builds on CAPM's ability to price risky assets, but relies on the argument that the return of assets is driven by various macro-economic factors rather than only the exposure to the market factor. Ross (1976) asserts that there are an infinite number of factors, both macro-economic and firm-specific, on which the assets expected return depend on. The APT formula is depicted below:

$$r = r_f + \beta_1 F_1 + \beta_2 F_2 + \cdots + \beta_n F_n$$

Where r is the rate of return, r_f the risk-free rate, β_i represents the sensitivity of asset i in relation to a factor, and F_i is the systematic factor.

The APT is far less restrictive in terms of its assumptions than the CAPM, and leaves more room for the investor to customize and develop a model for a specific asset's return. However, this implies that the investor must in turn identify each of the factors used in every specific asset, which is no trivial matter. In the CAPM, which can be viewed as a simplified model of the APT, only one factor needs to be considered: the risk of a particular asset relative to the market.

2.4 The Efficient Market Hypothesis and Systematic Risk versus Systematic Errors

In general, there are two main camps in the debate over what drives factor returns—one based on the view that markets are efficient and that factors reflect "systematic" sources of risk, and one based on the view that investors either exhibit behavioral biases or are subject to different constraints. These constraints can be such as time horizons, investable geography, ability to use leverage or tax effects (Ross, 1976).

2.4.1 Systematic Risk

In accordance with the APT (Ross, 1976) and the Efficient Market Hypotheses (Fama et al 1969), systematic risk refers to the risk attached to the factors; since these factors provide the stock its traits and the risk to these traits cannot be diversified away given an efficient market and rational investors, thus making the sources of risk 'systematic'. A premium in terms of excess return should therefore be earned for companies carrying systematic risk. Systematic risk could for instance be found in exposure macroeconomic factors, such as growth and inflation, in factors such as Value, Size and Momentum, since these are sensitive to shocks in the economy and thus must carry a return premium for the investor. Another example of systematic risk can be found in the small cap premium, which is believed to be due to exposure to stocks characterized by low liquidity (Liu, 2006), low transparency (Zhang, 2006) and are more likely to be distressed (Chan and Chen, 1991; Dichev, 1998).

2.4.2 Systematic Errors

Another explanation for the excess return earned by factors is due to investors' systematic errors. One suggestion, which can be derived from literature in behavioral finance, asserts that investors exhibit biases, such as chasing winners, preferring familiar investments and overconfidence, which can explain the observed factor anomalies in the market. Another camp within those who argue for systematic errors as explanation for factor anomalies, suggests that even though investors behave rational, they can be subject to different constraints. Investor constraints and frictions from regulatory and industry practices are argued to affect factor performances. For instance, studies have shown that low volatility stocks earn a premium over time horizons stretching beyond 10 years, while most investors prefer a much shorter time horizon and stocks with high liquidity. Therefore, an investor with a longer time horizon should earn higher returns, a premium, for carrying the horizon risk.

2.5 Factor Models

2.5.1 Fama and French Three Factor Model

In 1993, Eugene Fama and Kenneth French introduced an extension to the Capital Asset Pricing Model where two variables, besides the market factors, are presented. In the Fama French Thee Factor Model (1993), both size and book-to-market ratio, together with the market factor, is used in order to explain the cross-section of average returns on assets. Fama and French (1992) finds that β, used alone or in combination with other variables, gives limited information in regards to average returns, while size, leverage, E/P and book-to-market equity does however carry explanatory power. Especially the two variables size (ME) and book-to-market equity (BE/ME) are useful in explaining the cross-section of average stock returns. The size factor used by Fama and French (1993) is measured as "Small (market cap) minus Big" (SMB) and relies on the findings that small firms tend to outperform larger firms. The excess return from a portfolio consisting of firms with small market capitalization (S) is taken over the excess return of a portfolio with large capitalization firms (B). The book-to-market factor, "High (book-to-market) minus Low" (HML), is constructed in a similar fashion, taking excess return from a portfolio consistent of firms with a high book-to-market ration (H) over the excess return of a portfolio with low book-to-market ratio firms (L). This is due to the fact that firms with high book-to-market ratio, known as value stocks, have a tendency to outperform firms with low book-to-market ratio.

$$r = r_f + \beta_m(r_m - r_f) + \beta_s(SMB) + \beta_v(HML)$$

Where r denotes the expected rate of return, r_f is the risk-free rate and r_m represents the return of the market portfolio. β_m is analogous to the classical β but not equal to it, since we now have an additional two factors at play, *SMB* and *HML*. β_s and β_v denotes the sensitivity to these factors respectively.

2.5.2 Carhart Four Factor Model

Carhart (1997) builds on the Fama and French Three Factor Model (1993) by adding an additional factor to the model – momentum. While investigating the persistence in mutual fund performance, and building of the research on the Momentum factor by Jagadeesh and Titman (1993), Carhart (1997) found that stocks with a high return in the past tend to perform well in the next period as well. Thus, the momentum factor, "Up (performance) Minus Down" (UMD) was added and constructed by taking the excess return of past winners (U) over past losers (D). The Carhart Four Factor Model is depicted below:

$$r = r_f + \beta_{\square}(r_m - r_f) + \beta_s(SMB) + \beta_v(HML) + \beta_{mom}(UMD)$$

Where r denotes the expected rate of return, r_f is the risk-free rate and r_m represents the return of the market portfolio, β_m represents the sensitivity to the market, β_s, β_v and β_{mom} denotes the sensitivity to the size-, value- and momentum-factor respectively.

2.5.3 Fama and French Five Factor Model

Fama and French (2014) also built on their original Thee Factor Model by including an additional two factors – Profitability and an Investment factor. The argument for adding these two variables can be derived from the Dividend Discount Model (Fama, French 2006), which supplies further evidence that profitability and investment add to the description of average returns provided by the book-to-market ratio. The book-to-market ratio can be calculated by the following formula:

$$\frac{M_t}{B_t} = \frac{E_t\left[\sum_{\tau=1}^{\infty}. \frac{Y_{t+\tau} - dB_{t+\tau}}{(1+r)^{\tau}}\right]}{B_t}$$

Where M_t is the market value of the firm at time t, B_t is the firm's book equity, $Y_{t+\tau}$ is the total equity earnings for period $t + \tau$, $dB_{t+\tau}$ is the change in book value (i.e firm investment), r is the internal rate of return of dividends (a proxy for expected return). It follows that, all else held equal, differences in expected profitability $Y_{t+\tau}$ should, in the cross section, be related to the rate of return, r. Keeping the market-, book-value and the expected change in book value constant, the variation in expected earnings should be related to the variation in the rate of return. High expected profitability predicts a high rate of return, just as high valuation, $\frac{M_t}{B_t}$, and high rates of expected investment should predict a lower expected return.

Novy-Marx (2012) and Aharoni, Gundy and Zeng (2013) identify relationships between expected profitability and average return, and investment and average return separately, why an augmented version of the Fama and French Three Factor Model (1993) is included with the two factors. In the Fama and French Five Factor Model (2014), the profitability factor is measured by the difference in returns of portfolios with "Robust" (R) and "Weak" (W) profitability - "Robust minus Weak" (RMW). The investment factor is measured in a similar fashion between portfolios of low (C) and high (A) investment stocks – "Conservative minus Aggressive" (CMA).

$$r = r_f + \beta_m(r_m - r_f) + \beta_s(SMB) + \beta_v(HML) + \beta_p(RMW) + \beta_i(CMA)$$

Where r denotes the expected rate of return, r_f is the risk-free rate and r_m represents the return of the market portfolio, β_m represents the sensitivity to the market, β_s, β_v, β_p and β_i denotes the sensitivity to the size- value-, profitability-, and investment-factor respectively.

2.6 Conditional Beta

The Sharpe-Lintner-Black model (SLB) is based on the assumption of a positive risk-return tradeoff, and asserts that the expected return for any asset can be derived by a positive function of three variables: Beta, the risk-free rate and the expected market return. This implies that the only cause for systematic differences in returns between assets depend on the asset's responsiveness to market movements. Although early empirical tests, such as Fama and MacBeth (1973), supported the validity of the SLB model, the usefulness of Beta as the sole measure of risk for a security has been challenged by at least three arguments. The first argument suggests that Beta is not the most efficient measure of systematic risk, but rather systematic responsiveness to macroeconomic variables should be measured (Chen et al, 1986). The second argument relies on empirical evidence that security returns are affected by unsystematic risk (Lakonishok and Shapiro, 1986). The third argument states that there is empirical evidence which indicates the absence of a systematic relationship between Beta and security returns (Fama and French 1992). Therefore, the question of Beta's efficiency and completeness arises, and whether or not Beta does in fact measure risk and if there is a risk-return tradeoff.

By using realized market returns as a proxy for expected market returns and assuming an inverse relationship between realized returns and Beta, when the realized market returns fall below the risk-free rate, Pettengill, Sundaram and Mathur (1995) are able to find a significant and systematic relationship between Beta and returns. Their evidence of a positive risk-return tradeoff, when Beta is used as a measure of risk, supports Betas usefulness as a measure of risk, although it might not be direct support of the Sharpe-Lintner-Black model.

3 Method

In the following section, the methodology of the thesis is described and justified. The gathering of data and the process of analyzing it is presented in detail, in combination with critical perspectives and delimitations throughout the process.

3.1 Determining Geography

The American equity market is chosen due to several reasons. First, the amount of data is extensive, both in terms of completeness in a historical context, but also in terms of accounting standards. This mitigates some data selection issues. The universe of investable stocks is also larger than for instance the European equity market. Furthermore, the American stock market is likely one of the more well-functioning markets in a global context, in terms of efficiency.

3.2 Collecting Data

The Bloomberg terminal was used to collect the data. The data has been gathered from stocks incorporated in the Wilshire 5 000 index, the broadest equity index available in the US market. Due to data issues, historical performance and reliable accounting variables were only retrievable from January 1993. However, throughout the time period investigated, there have been several financial and economic crises. Also, during this time period, a lot of progress has been done in terms of factor investing, which may affect the results and the pervasiveness of the Quality anomaly as well as other, today, well documented factors. Altogether, we believe that the time period captures these aspects and thus provides an interesting and sufficient time window for the purpose of the thesis. Furthermore, all data is collected monthly, providing a total of 288 observations during the period examined, for each asset and each portfolio. Financial stocks are not excluded from the data, due to data issues. We observe the universe one time each year, and follow the same process as Fama and French (1992). The average number of stocks used each year in the quintiles is 350. With monthly data, spanning over a period of 24 years and across 12 metrics, this results in a total of over 2,4 million data observations. The stocks are weighted equally and as such, each stock has a weight of approximately 0,3 %.

3.3 Constructing a Quality Screen

Quality can be defined in various ways but is typically associated with profitable companies with low leverage and stable earnings. It is rooted in fundamental analysis and thus makes use of an assortment of financial data extracted from financial reports of companies. Although the definition of Quality varies in complexity, stretching from simple, one-dimensional metrics such as ROE, to multi-metric definitions encompassing a multitude of accounting ratios, the most frequently used characteristics of the quality definition can be group in three main categories: Profitability, Safety and Quality of earnings.

Profitability is defined as a company's ability to generate earnings as compared to its expenses, why profitable firms are often referred to as quality firms. Profitability can be measured by different accounting ratios such as gross profit over assets (GP/Assets), operating cash flows over assets (CF/Assets) and various net profit-based measures, for instance return on equity (ROE), return on assets (ROA) and return on invested capital (ROIC). The different ratios provide different insights in regards to the financial state of the company, where gross, net and operating margins suggest how well the company is at managing its expenses, while ratios such as ROE and ROA give insights to the company's ability in deploying its capital in order to generate returns. Since these ratios can, to some extent, be affected by accounting choices, the best suitable metric to represent profitability is a subject of controversy. Novy-Marx (2013) argue that the cleanest measure of profitability can be found in gross profit, since this metric is relatively unaffected by accounting estimates for accruals and non-cash expenses. Index providers however, argue that net profit-based metrics are better suited to represent profitability, since net profit measures the profit which accrues to common shareholders rather than stakeholders. These are metric such as ROE and ROA (Norges Bank, 2015). Seasonality is another factor which can affect the profitability metrics, rendering certain metrics unsuitable to be compare across different industries.

High-quality companies are also often defined as safe and stable. A company exhibiting excessive leverage carries greater risk of financial distress, since it may be jeopardizing its ability to service its debt. Therefore, safety is often regarded as having a strong balance sheet – low leverage, high current ratios and high interest coverage ratios.

A high-quality company is often regarded as one which generates a stable and persistent stream of earnings, since this might indicate the presence of a competitive advantage, good management and a strong market position. Furthermore, earnings stability can be measured by both the volatility of earnings, or profitability metrics such as ROE and ROA, and by its growth. Earnings variability tends to vary by industry and by company age, where younger companies exhibit more volatile earnings than more established, older companies.

In the following table, the Quality definition by different authors/practitioners is presented:

Author/ Practioner	Quality definition
Novy-Marx (2013)	Gross profits / assets
Fama and French (2014)	Operating income before depreciation and amortization minus interest expense scaled by assets
Greenblatt (2010)	Return on invested capital (ROIC)
Sloan (1996)	Difference between cash and accounting earnings scaled by assets (earnings quality)
Piotroski (2000)	(1) Return on assets, (2) Operating income, (3) Cash flow, (4) Quality of earnings, (5) Net income, (6) Leverage, (7) Liquidity equity issuance (8) Gross margins, (9) Asset turnover
Asness, Frazzini and Pedersen (2014)	Z-scores based on: • profitability: gross profits over assets, return on equity, return on assets, cash flow over assets, gross margin, low accruals • growth: 5-year prior growth of profitability • safety: low beta, low idiosyncratic volatility, low leverage, low bankruptcy risk, low ROE volatility • payout: equity and debt issuance and total net payout over profits
GMO white paper (2004)	• ROE • Leverage • Profit volatility
Graham (1973)	Adequate size of enterprise, sufficiently strong financial position, earnings stability, dividend record, earnings growth, moderate P/E and P/B ratios
FTSE Quality indices	• Return on assets • Accruals • Operating cash flow to debt”

Due to the controversy surrounding the definition of quality, a total of 12 metrics from the three main categories were chosen to represent the characteristics of Quality. The aim is to perform statistical test in order to determine the best proxy for Quality, although this method does imply a certain amount of selection bias. To somewhat account for the selection bias, the metrics of choice are frequently used individually in previous research, as can be seen from the table above. Furthermore, to mitigate the selection bias effect, a higher level of significance is imposed in the regressions. ROIC, GP/Assets, CF/Assets, Operating margin, ROE and ROA were chosen to represent the profitability characteristic of Quality, Leverage, Debt/Equity and Net debt for the Safety category, and EPS Stability, Dividend 5-year growth and Equity Variability as a measure of

Earnings.

Following the methodology of Asness and Frazzini (2013), portfolios for each metric is formed by a long position in the quintile of highest ranking stocks for each metric, and short the quintile of stocks with the lowest ranking, resulting in 12 net portfolios which are rebalanced yearly.

In order to construct an aggregated Quality portfolio, each of the fundamental metrics are first regressed on the CAPM, Fama French 3 Factor Model, Carhart's 4 Factor Model and Fama French 5 Factor Model. The statistical result from these regressions, combined with relevant literature regarding the choice of fundamental metrics to represent the Quality characteristics, will form the foundation in constructing the aggregated Quality portfolio consisting of three metrics – one from each category.

Once the three metrics have been decided upon, each variable is converted into ranks and standardized to obtain a z-score through a methodology that follows that of Asness and Frazzini (2013). Thus putting each measure on equal footing and making it possible to combine them. The z-score is computed as follows:

$$r_i = rank(x_i)$$

Where x is the variable of interest and r the vector of ranks.

$$z(x) = z_x = (r - \mu_r)/\sigma_r$$

Where μ_r and σ_r are the cross sectional mean and standard deviation or r.

$$Quality = z(z_{Metric\ 1} + z_{Metric\ 2} + z_{Metric\ 3})$$

With the Quality factor defined, an aggregated Quality portfolio is formed by a zero-investment portfolio, taking a long position in the 20% (10%) highest ranking stocks and a short position in the 20% (10%) lowest ranking stocks, following the methodology of Fama and French (1993) and Asness and Frazzini (2013). The portfolio is then rebalanced yearly. Furthermore, the aggregated Quality portfolio is regressed in CAPM, Fama French 3 Factor Model, Carhart's 4 Factor Model and Fama French 5 Factor Model, in order to investigate the portfolios factor loadings.

3.4 Decomposing the Quality Screen

Following the portfolio approach of Fama & Macbeth (1973), Fama and French (1992, 1993 and 1996) and Asness and Frazzini (2013), the Quality Factor is decomposed by conditional sorts, first sorting on size and then on quality. Size is sorted on Small and Big, while Quality is sorted on Low, Medium and High, thus forming 2x3 portfolios.

In order to further investigate the characteristics of Quality, differences in the long Quality portfolios and short Quality portfolios are studied. Firstly, the riskiness of High- and Low-Quality stocks is assessed by volatility and Beta. Volatility is calculated by a rolling 200day window, while Beta is measured with a window of 60 months and the following formula:

$$\beta_{im} = \frac{Cov\ (r_i, r_m)}{Var\ (r_m)}$$

The average market capitalization in each portfolio is also compared in order to illustrate size-differences between high- and low-quality stocks. Lastly, the price-to-book ratio is measured, as illustrated in **2.5.3**, in order to test for price-differences between the two, which can be derived to the characteristics of Quality.

3.5 Fama Macbeth and the Cross Sectional Regression

Risk factors are frequently used to explain asset returns in asset pricing models, and one of the preferred models for this endeavor is the Fama and Macbeth method. This approach involves a two-pass estimation methodology. The first part consists of estimating market Betas using the linear regression model developed in the CAPM framework. Step two involves using these Betas together with other variables that are considered important to explain the variation in returns (Fama and MacBeth, 1973). Even if this methodology was developed more than 45 years ago, numerous studies have relied on it when investigating factors affecting equity return. Skoulakis (2008) suggests that this framework is the preferred methodology to determine factor risk premium and cross section of returns.

By following the logic from Fama Macbeth's cross- sectional approach, we are able to include additional risk factors in the model and test to what extent these and the Betas describe the stock returns. In this sense, the framework fits the overall purpose of this essay well. However, in empirical tests of the CAPM, Black, Jensen, and Scholes (1972), Fama and MacBeth (1973), Fama and French

(1992) and Frazzini and Pedersen (2013), come to a similar conclusion; the market Beta is smaller than what is predicted by the CAPM. Furthermore, Davis, Fama, and French (2000) find a similar result for the multivariate Beta in the Fama-French Three Factor Model. Their findings suggest that the predictions of models that include a standard market factor are too high for assets with market Betas greater than 1.0 and too low for assets with Betas less than 1.0. In this sense, the argument for using a cross sectional model is strong, however, there are also some drawbacks since the estimation of Beta impose some difficulties, more on this later.

The general cross-sectional model, which is used in this essay, is:

$$\boldsymbol{R}_t = \gamma_{0t}\iota + \gamma_{1t}\boldsymbol{\beta}_{\boldsymbol{m}} + \mathbf{X}_{\boldsymbol{t}}\boldsymbol{\Gamma}_{\boldsymbol{2t}} + \boldsymbol{\mathcal{E}}_{\boldsymbol{t}}$$

Where $\boldsymbol{R}_t$ is the return for each of our sub portfolios for each month, a column vector with dimensions 288x1, expressed as returns in excess of the risk-free rate. All prices are calculated as the log price change, they are adjusted for dividends (the return is expressed as total return) and the prices are adjusted for new stock issuance, buybacks and splits. As a proxy for the risk free rate, the three month treasury rate is used, transformed into its monthly equivalent.

$\boldsymbol{\Gamma}_{2t}$ is the (k x 1) vector of the coefficients (γ_{2t} , … , $\gamma_{2k+1,t}$) of the k:th additional explanatory variable, $\mathbf{X}_t$. Since the market Betas are assumed not to be known, the first step in the Fama Macbeth regression consists of estimating these by separate time series regression for each portfolio, using the following model:

$$R_{it} = \alpha_{it} + \beta_i R_{mt} + \mathcal{E}_{it}$$

Where *i* denotes the portfolio, and *i* = 1, …, 6. The variance of the returns at time *t* may differ across the portfolios, and the returns might be correlated over time. This implies that the disturbance terms of the monthly cross-sectional model may be both heteroskedastic and correlated. Therefore, the OLS estimator of the parameters of the cross-sectional regression may be inefficient. We therefore use the GLS approach to estimate the single index model, outlined above. The following method is used to estimate the coefficients:

$$\hat{\boldsymbol{\Gamma}}_{\boldsymbol{t}(\boldsymbol{GLS})} = (\mathbf{H}'\,\boldsymbol{\Psi}_{\boldsymbol{t}}^{-1}\mathbf{H}_{\boldsymbol{t}})^{-1}\,\mathbf{H}_{\boldsymbol{t}}'\,\boldsymbol{\Psi}_{\boldsymbol{t}}^{-1}\mathbf{R}_{\boldsymbol{t}}$$

Where, $\mathbf{H}_{\boldsymbol{t}}$ is (N x K+2) matrix containing all the explanatory variables of the model. The properties of $\boldsymbol{\Psi}$ makes it possible to derive the best linear unbiased estimator for the Betas, thus obtaining error terms that are homoscedastic and exhibit no autocorrelation. $\boldsymbol{\Psi}$ has the following characteristics:

$$\boldsymbol{\Psi}_{\boldsymbol{t}}^{-\mathbf{1}} = \mathbf{P}'\mathbf{P}$$

Where $\mathbf{P}$ is a square, non singular matrix and $\boldsymbol{\Psi}$ is positive definite. Also,

$$\mathbf{P}\boldsymbol{\Psi}\mathbf{P}' = \mathbf{P}\mathbf{P}^{-\mathbf{1}}(\mathbf{P}')^{-\mathbf{1}}\,\mathbf{P}' = \mathbf{I}$$

Consequently, the following applies:

$$\mathbf{E}\,[\mathbf{P}_{\boldsymbol{\varepsilon}}|\mathbf{X}] = \mathbf{0}$$

And,

$$\mathbf{Var}\,[\mathbf{P}_{\boldsymbol{\varepsilon}}|\mathbf{X}] = \sigma^{2}\mathbf{P}\boldsymbol{\Psi}\mathbf{P}' = \sigma^{2}\mathbf{I}$$

The Gauss Markow assumptions are thus fulfilled. In this sense, we use the heteroscedasticity and autocorrelation consistent (HAC) standard errors, also referred to as Newey–West standard errors, for all the regressions. The White covariance matrix assumes that the residuals of the estimated equation are serially uncorrelated, a fact not observed in our data sets. It is a well-known fact that financial data tend to exhibit excess kurtosis and volatility clustering, so it is not reasonable to anticipate that the residuals are serially uncorrelated. Therefore, the HAC consistent covariances are a better fit. Furthermore, we test for autocorrelation using the Durbin–Watson test (Durbin and Watson, 1950). A value close to 2 indicates that the first-order autocorrelation coefficient is close to zero. If the value however is much smaller than 2, it is an indication of first order autocorrelation.

As mentioned above, there is another issue with financial data in time series analysis. Due to volatility clustering effects, the data is not likely to be stationary. Engle (1982) proposed the framework of autoregressive conditional heteroscedasticity (ARCH), which incorporates this mechanism. In the ARCH framework, the variance of the error term depends on the squared error terms from previous periods. In this essay, we use the augmented Dickey–Fuller (ADF) test to see whether the data is stationary or not, but we do not impose any model on volatility, such as ARCH or GARCH. The statistic generated from the ADF test is a negative number, and a low value indicates that the hypothesis, the presence of a unit root, is rejected.

Another aspect of GLS is that it might be sensitive to outliers in the data set, since it assigns equal weights to all the observations. A popular approach in finance is to winsorize the data set, in order to exclude the outliers. Since this method removes data (for instance, it is common the remove the top and bottom percentile). We follow this approach to avoid overfitting. Furthermore, all the stocks are assigned equal weights. The average stock weighs 0.3 % in our net quality portfolio, based on quintiles. This implies that even if a few stocks has some outlying values, this is not likely to affect the overall result.

Important to note is that including the Betas generated from equitation 1 into the cross-sectional regression causes errors-in-variables problem. Thus, the Cross Sectional Regression Model will be likely to underestimate the Beta and overestimate the other coefficients. The overestimation of the other coefficients depends on the level of correlation between the variables (Kim, 1995). Fama and MacBeth (1973) addresses this issue by using portfolios instead of individual stocks. However, it is important to note that this procedure not entirely solves the errors-in-the-variables problem (Ho, Strange and Piesse, 2006). Applying a portfolio approach might on the other hand cause a loss of information, an issue discussed by for instance Asgharian and Hansson (2000). Fama and French (1992) in contrast to Fama and MacBeth (1973) use portfolios in order to estimate the Betas and subsequently assign the Beta values to the individual stocks. The analysis is then carried out on the individual stocks (Fama and French, 1992). This approach is also used in this essay, in order to mitigate the errors-in-variables problem.

The discussion above constitutes the prerequisites for estimating the Cross Sectional Model, outlined below:

$$\boldsymbol{R}_t = \gamma_{0t}\iota + \gamma_{1t}\boldsymbol{\beta_m} + \gamma_{2t}\boldsymbol{\beta_{SMB}} + \gamma_{3t}\boldsymbol{\beta_{HML}} + \gamma_{4t}\boldsymbol{\beta_{MOM}} + \gamma_{5t}\boldsymbol{\beta_{Quality}} + \boldsymbol{\varepsilon_t}$$

This model might be subject to autocorrelation, heteroscedasticity and multicollinearity. However, according to Verbeek (2012), there is nothing wrong with including variables in a model that are correlated, however this needs to be controlled for to ensure a good fit. We use the Variance Inflation Factor (VIF) to detect multicollinearity.

The VIF is given by the following formula:

$$VIF(b_k) = \frac{1}{1 - R_k^2}$$

The VIF indicates the factor by which the variance of b_k variables is inflated, compared with the hypothetical situation when there is no correlation between the dependent variable and any of the other explanatory variables. There is no consensus on how high the VIF can be in order to constitute a problem, but if the value is higher than 2.50 (corresponding to a R^2 of 0.6), there is clearly a correlation issue amongst the factors. We also compare the correlation between the different factors, i.e construct both a correlation and a covariance matrix of the underlying data sets for Market, SMB, HML, MOM and Quality to investigate the robustness of the regression.

Once the model is specified, the coefficients of γ_{it} are estimated by using t-statistic values. The t-statistic values are defined as follows:

$$t_j = \frac{\hat{\gamma}_j}{\hat{\sigma}(\widehat{\gamma_j})}$$

Where

$$\hat{\gamma}_j = \frac{1}{T}\sum_{t=1}^{T} \hat{\gamma}_{jt}$$

And,

$$\hat{\sigma}^2(\hat{\gamma}_j) = \frac{1}{T}\frac{1}{T-1}\sum_{t=1}^{T}(\hat{\gamma}_{jt} - \hat{\gamma}_{\square})^2$$

Where t_j follows a t-distribution with (T-1) degrees of freedom. As discussed in section 3.6 the conventional level of significance of 2.0 or higher, is considered to be too low due to the process employed in selecting the portfolio.

3.6 Critical Perspectives

Numerous researches have raised the possibility that the quality anomaly as well as other anomalies may be a result of measurement errors, methodological biases and data mining. As discussed by for instance Lo and MacKinlay (1990) and Black (1993), researchers tend to test their hypothesis and conduct their analysis on the same investment universe, with the goal to uncover anomalies. Since researchers may only publish the most statistically significant findings, the process is prone to selection bias. Following this logic, it is not a surprise that an interesting pattern or an anomaly emerges from time to time, simply by chance. Another aspect, that is particularly relevant for the quality premium, is the fact that it is not well defined and leaves the researches with a lot of choices and potential exclusions. In light of this, there have been a lot of discussion regarding to what extent the findings are reliable or not, as it is subject to data snooping and selection bias.

Harvey, Liu and Zhu (2015) discuss this issue, and propose that the t-stat level of 2.0 needs to be adjusted upwards, to 3.5 or more, in order to control for the selection bias and data snooping aspects. This would imply that many of the papers published by researchers on the quality anomaly would not be able to pass the threshold of significance.

Another methodological aspect, in addition to the multiple testing bias, is overfitting. Novy-Marx (2015) states that when combining multiple metrics, each prone to predict high risk adjusted returns, conventional two-sided significance tests are no longer reliable. Intuitively, suppose a researcher tests 25 randomly selected metrics, and then concludes that ten of them predict higher returns, due to chance. If these are aggregated and back tested, the performance is likely to be very high.

Another important aspect is firm size. Davis (1994) who investigated the American Stock Market, excluded all the small firms, since small firms make up a great portion of the market, and thus may skew the findings and making them more difficult to generalize. We do not exclude any firms based on size, but it is important to note that this might skew the data. Since smaller firms are likely to be affected by idiosyncratic aspects to a greater extent than large firms, the conclusions might not be as robust. By dividing our sample in different size categories, it is intended to control for this.

A popular perspective to raise when discussing factor investing is survivorship bias. Most of the studies in this field rely on data from the COMPUSTAT or CRSP database. As Kim (1997) and Kothari, Shanken and Sloan (1995) describes, there is an element of survivorship bias regardless of what source of information is used. Large and profitable firms are more likely to be entered and maintained into the databases. However, this bias was likely more predominant some decades ago, since the process of updating data frequently and on a large set of stocks is more convenient today. This an important implication as to why we use the Wilshire 5 000 Index, rather than S&P 500 as the underlying universe.

Davis (1994) outlines another important aspect, affecting the selection process. Generally, firm policies and accounting standards are not coherent over time. To name an example, stock buybacks have to a large extent acted as a substitution of dividends since the late 90's in the US, which for instance might impose some selection issues when dividend models are used (Ogden, Jen and O`Connor, 2003).

One final aspect to outline is the case of non-synchronous trading. As for example Morelli (2007) and Ho, Strange and Piesse (2006) point out, the estimated Beta will not be efficiently estimated, thus resulting in an overall spurious regression. Since monthly returns have been used, this effect is to some extent mitigated. Furthermore, exclusions of stocks with insufficient data, such as low liquidity or missing data due to accounting variables, have been made.

4 Empirical Findings

The fourth chapter of the essay contains the results from data analysis as presented in the methodology. The empirical findings, in combination with the theories presented in chapter three, will serve as the foundation for the discussion and analysis of the thesis.

4.1 Building a Quality Screen

When creating the 12 zero investment portfolios, based on the fundamental metrics as described in section 3.3 the results, in terms of significance, varied depending on which factor model the zero investment portfolios were regressed on. The alpha for each of the 12 metrics and underlying models are presented in appendix 1.

From these results, the metrics with the most significant alpha values, and most consistent data, within the three different dimensions of Quality were chosen to act as proxies. From profitability, the CF/Assets metric was used. The reason for this is due to data issues; for instance, was GP/Assets was inconsistent and for some years, and a large part of the universe lacked information for this metric. It was therefore not considered to be representative. Additionally, cash flow metrics (accruals) are frequently mentioned in the literature as a suitable proxy for the profitability premium.

From the safety dimension, the leverage metric demonstrated highest significance and was therefore selected. EPS stability was chosen as a proxy for the earnings quality premium, since dividend, for the same data reason as GP/Assets, was excluded. Furthermore, the spread within the dividend metric was very high, thus concluded to be unrepresentative for the equity quality premium.

Below, the factor loadings (Carhart's Four Factor Model) for the different metrics are presented, for each of the dimensions of Quality. Regardless of which dimension the metric belongs to, factor loadings differ quite significantly, i.e the dimensions do not tend to be coherent in terms of factor loadings.

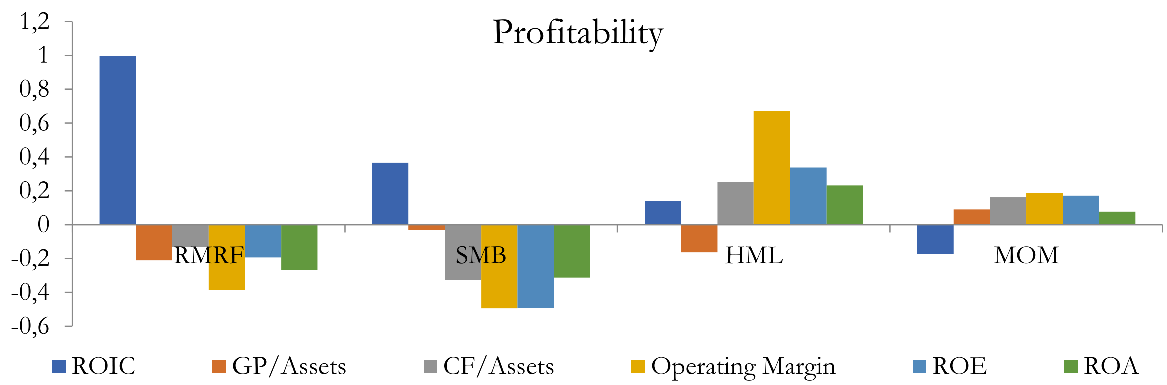


Figure 1

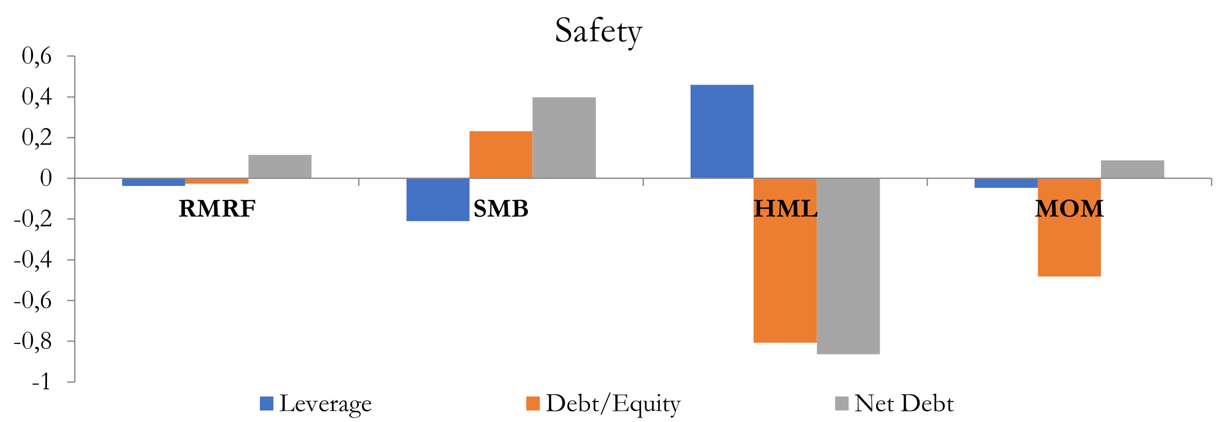


Figure 2

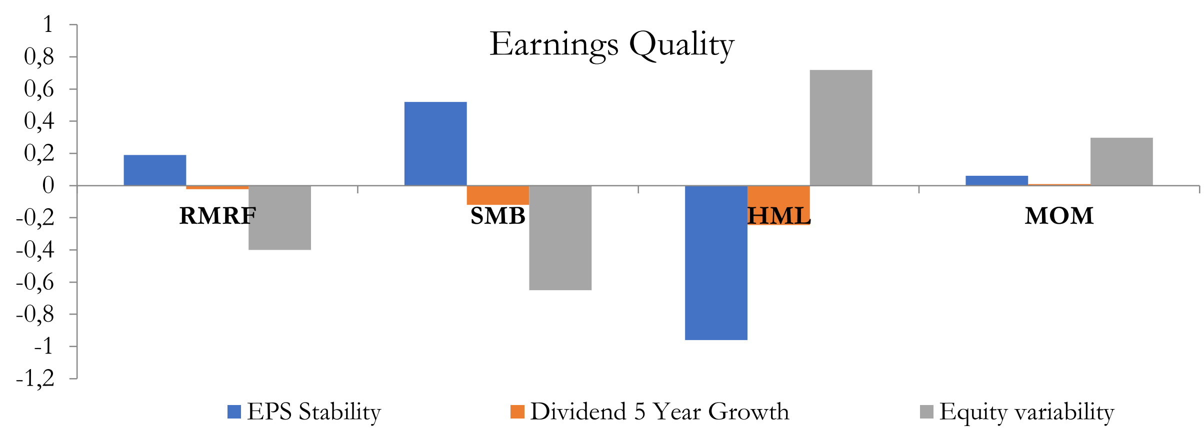


Figure 3

The performance of the three dimensions of Quality are depicted in Appendix 2.

As a result of defining the Quality screen, a ranking of our stock universe based on a combined z-score of the three metrics was enabled. Each of the stocks in the universe was assigned an individual rank for each year, and from this the zero investment portfolios were formed. The portfolios consisted of long positions in the 20 % (10 %) highest ranking stocks, and short the 20 % (10 %) lowest ranking stocks. This resulted in a net Quality portfolio. The net portfolios, presented as both deciles and quintiles, are depicted below. The average number of stocks in the quintiles is about 350 each year, thus considered to be well diversified. The portfolios were rebalanced yearly.

Figure 4

The graph below shows the cumulative returns from the Quality portfolio, in excess of the Wilshire 5000 Index. The majority of the excess returns can be linked to adverse equity market conditions.

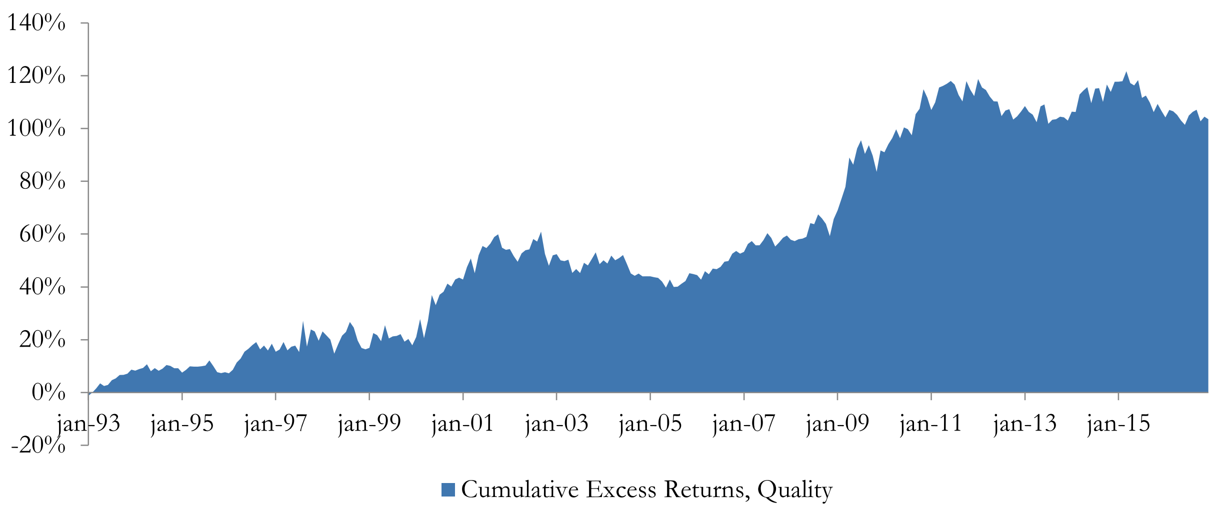


Figure 5

Below, the zero investment portfolio (consisting of a long position in the 10 % highest quality stocks and short the 10 % lowest) is depicted against the performance of Wilshire 5 000. The zero investment portfolio performs rather poorly, in comparison to the Wilshire 5 000, during normal and strong market conditions, but earns high excess returns in periods of market distress.

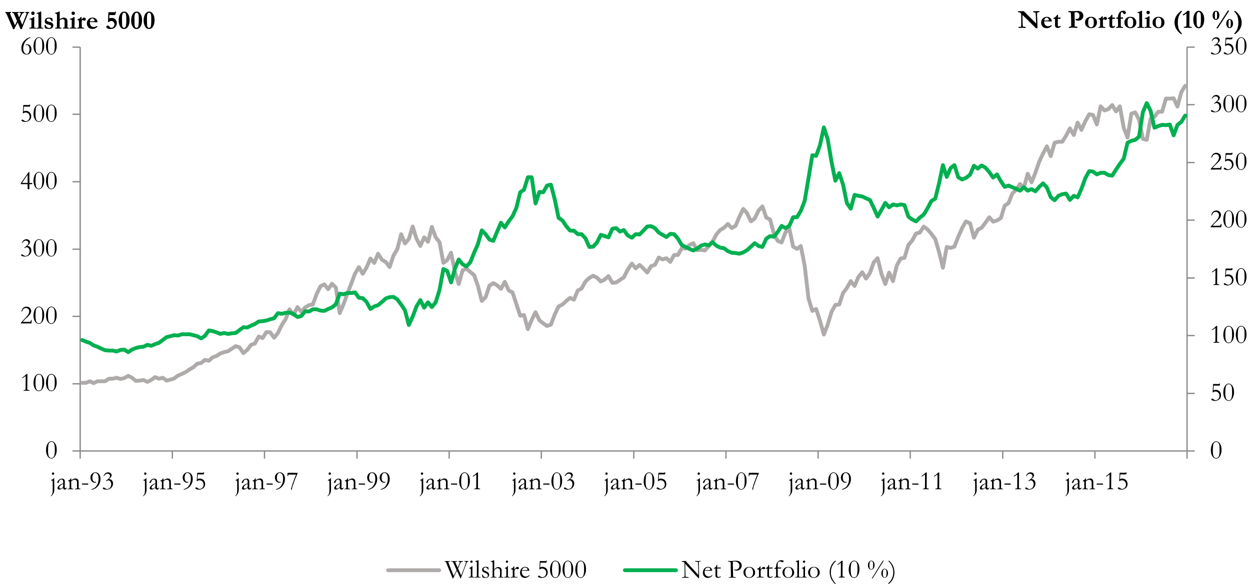


Figure 6

The risk and return characteristics of the long Quality portfolio, the SHB portfolio, the S&P500 and Wilshire 5 000 are presented in the table below. When compared to the indices, the Quality portfolios exhibit higher annualized returns but lower volatility, earning a significantly higher Sharpe-ratio. Worth noticing is that the SHB portfolio has an average of 165 stocks each year, about half of the Quality portfolio. Thus it is not as diversified, and the performance might thus be subject to chance, rather than systematic sources of risk.

	Annualized returns	Annualized volatility	Sharpe	Max	Min
Long Quality Portfolio	10,15%	13,37%	0,76	16,41%	-18,55%
SHB Quality Portfolio	13,19%	13,36%	1,01	16,69%	-17,87%
S&P500	7,23%	13,36%	0,54	11,38%	-17,71%
Wilshire 5000	7,30%	14,82%	0,49	11,39%	-17,74%

Table 1

The zero-investment portfolio was then regressed on CAPM, FF3, Carhart's Four Factor Model and FF5. From the table below, we can conclude that the Quality factor exhibits positive, significant alphas across all models. Furthermore, it loads negatively on Beta, SMB and HML, but positively on MOM, RMW and CMA. The results of the regressions are depicted below.

	alpha	Beta	SMB	HML	MOM	RMW	CMA
CAPM	0,5161%***	-0,284134***					
FF3	0,549%***	-0,268983***	-0,131828***	-0,070671			
Carhart	0,4311%***	-0,211661***	-0,154784***	-0,019977	0,151111***		
FF5	0,4422%***	-0,212143***	-0,095365*	-0,179917***		0,139688*	0,178546*

Table 2

The result of corresponding regressions for the SHB portfolio is presented below. The SHB loses the significant alpha when regressed on the FF5, which might be due to the high factor loading on RMW, i.e the risk return relationship is more or less explained (alpha not significant) by the factor loadings.

	Alpha	Beta	SMB	HML	MOM	RMW	CMA
CAPM	0,00383***	0,808652***					
FF3	0,002608**	0,829988***	0,077547**	0,327369***			
Carhart	0,002938**	0,813987***	0,083955**	0,313219***	.-0,042181*		
FF5	0,000299	0,952815***	0,250762***	0,190301***		0,487793***	0,064346

Table 3

Below, the variance-co-variance matrix for the factors used in our regressions is presented. The correlation is quite low, and the VIF is always below 2.5.

	Net Quality	RMrf	SMB	HML	MOM	RMW	CMA
Net Quality	0,080%						
RMrf	-0,052%	0,184%					
SMB	-0,021%	0,032%	0,109%				
HML	0,003%	-0,019%	-0,031%	0,096%			
MOM	0,049%	-0,058%	0,015%	-0,029%	0,252%		
RMW	0,025%	-0,059%	-0,054%	0,037%	0,011%	0,082%	
CMA	0,010%	-0,031%	-0,009%	0,042%	0,005%	0,017%	0,044%

Table 4

Furthermore, Quality exhibits higher persistence than other factors. In the table below, the percentage of positive outcomes of a rolling 12 months window of net return is presented for the different factors.

SMB	HML	MOM	RMW	CMA	Quality
54,9%	56,7%	70,8%	67,9%	61,4%	73,3%

Table 5

4.2 Conditional Beta Analysis

Following the research from Pettengill, Sundaram and Mathur (1995), a conditional Beta test was performed on the Quality portfolio, with the purpose of understanding how Quality performs during different market conditions. The results were not statistically significant, but indicate that a Quality strategy earns higher excess returns during bad market conditions than during good. Worth mentioning though, is that Quality generates excess returns in both up, as well as down market conditions. The results from the conditional Beta test are presented below.

	Aggregated Quality Portfolio			
Variable	Coefficient	Std, Error	t-Statistic	Prob,
Alpha	0,008865**	0,00317	2,797098	0,049
Beta, Up	0,010999	0,012699	0,866153	0,4353
Beta, Down	-0,01218	0,010049	-1,212066	0,2922

Table 6

4.3 Decomposing the Quality Factor

In the spirit of Fama and French, we formed sub-portfolios from the findings of our Quality screen by sorting on size, and then by Quality. Thus, we ended up with 6 sub portfolios, ranging from High to Low Quality, for two sub-universes - small cap stocks and large cap stocks. The returns from the 6 portfolios are presented below. As can be seen, higher quality outperforms lower quality, regardless of size. There is also a tendency for small cap stocks to perform better than large cap stocks, with the Small-Cap-Low-Quality portfolio being the exception.

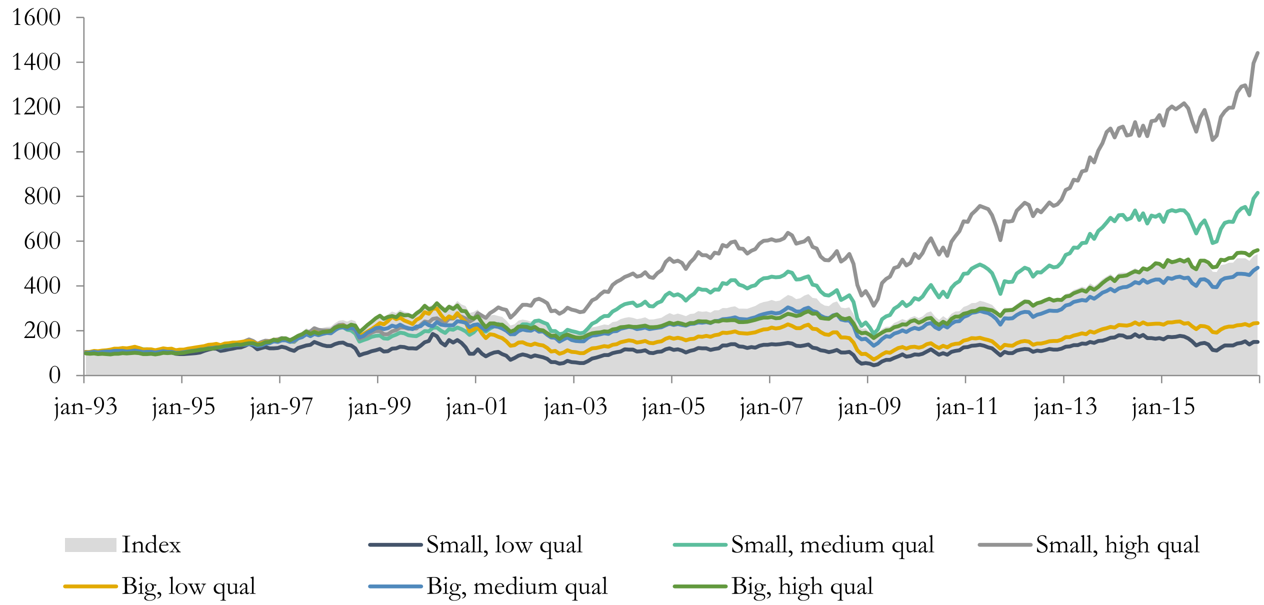


Figure 7

4.4 Fama Macbeth (Cross Sectional Regression) and Systematic Risk Premiums

The regression models from Fama Macbeth (1973) were used to regress the 6 sub-portfolios. According to the Sharpe-Lintner CAPM, γ_0 should be equal to 0, and $\gamma_1$should be positive, indicating a positive market risk premium. In this sense SMB, HML and Quality carries some explanatory power in the cross sectional variation of stock returns, however, the results are not high enough in terms of t-values. As previously discussed, when several layers of selection is used in deriving the portfolio, the t-value should be closer to 3.5 in order to infer significance. This implies that we cannot reject that these factors are important in explaining the variation. Neither can we reject that alpha differs significantly from zero.

		Beta	SMB	HML	MOM	Quality
	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5
Mean	0,00887	-0,0017	0,0006	0,00393	-0,0156	0,0265
Median	0,00164	0,0006	0,00103	0,00222	-0,0052	0,02409
Skewness	0,172	-0,0663	-0,0565	0,28966	-0,4043	0,06811
Kurtosis	3,32007	3,54846	6,43597	6,98009	4,46127	2,52242
Jarque-Bera	2,6493	3,82069	141,824	194,121	33,4689	2,95969
Sum	2,55551	-0,4856	0,17209	1,13257	-4,4889	7,63236
Sum Sq. Dev.	2,67935	3,47153	0,90598	0,62628	25,9154	9,53526
t-stat	1,55844	-0,2602	0,18063	1,42881	-0,8802	2,46737
Probability	0,2659	0,14803	0	0	0	0,22767

Table 7

The corresponding coefficients for the portfolios from the cross sectional regressions are presented below. There is a clear relation that higher Quality portfolios are associated with a lower Beta, lower SMB, lower HML, and higher MOM. All the coefficients for the various sources of risk are depicted, as calculated from the Fama Macbeth cross sectional regression. In short, High Quality stocks are often larger companies with lower Beta values and high P/B ratios.

	Portfolio	Beta	SMB	HML	MOM
Small Cap Stocks	Low Quality	1,225735	1,023901	-0,00167	-0,21272
	Medium Quality	0,958258	0,731187	0,362973	-0,10568
	High Quality	0,880476	0,665005	0,257789	-0,10064
Large Cap Stocks	Low Quality	1,206858	0,180431	0,122814	-0,0244
	Medium Quality	1,003436	-0,08184	0,175092	-0,00868
	High Quality	0,912115	-0,13542	-0,18904	0,017724

Table 8

4.5 Comparing Factors

From the earlier empirical findings, we have seen that Quality is generating returns in excess of the market, and produces significant alpha. Below, we illustrate the risk adjusted returns, according to Markowitz Mean Variance Optimization approach. In order to compare the return of factors, we collected data from other, well documented factors in the equity markets. Due to data issues, the time frame spans from January 1999 – January 2017. In addition to investigating other factors, we also test how mean variant efficient Quality is in relation to the Wilshire 5 000 Index.

In the optimization setting, maximum weights were set to 20 %. The optimized portfolio generated a combination of factors, rather than using the market index.

Factor	**High Dividend**	**Quality**	**Low Volatility**	**Russel 2000**	**Value**	**Momentum**	**Growth**	**Small**	**Wilshire 5 000**
Ann. Return	10,29%	5,82%	9,14%	8,15%	5,82%	3,52%	4,85%	10,22%	5,45%
Ann. Std	13,60%	13,51%	11,19%	19,68%	15,56%	16,57%	15,15%	18,46%	14,69%
Sharpe, rf=0	0,76	0,43	0,82	0,41	0,37	0,21	0,32	0,55	0,37
Weights	20,0%	20,0%	20,0%	0,0%	0,0%	20,0%	20,0%	0,0%	0,0%

Table 9

Below, the Optimized portfolio is presented and compared to the Wilshire 5000 index. The Mean Variance Optimized Portfolio performs better in terms of risk-adjusted returns and has less Value at Risk, from a historical distributional perspective.

	Annualized returns	Annualized Std	Sharpe	Max drawdown	VaR (99,5 %, Basic Historical Simulation)
Portfolio	6,72%	12 %	0,51	-33,04%	10,6%
Wilshire 5 000	5,45%	14,69 %	0,37	-43,97%	11,0%

Table 10

As illustrated below, the factors chosen mostly engage positions outside the efficient set. All returns are presented in excess of the risk free rate.

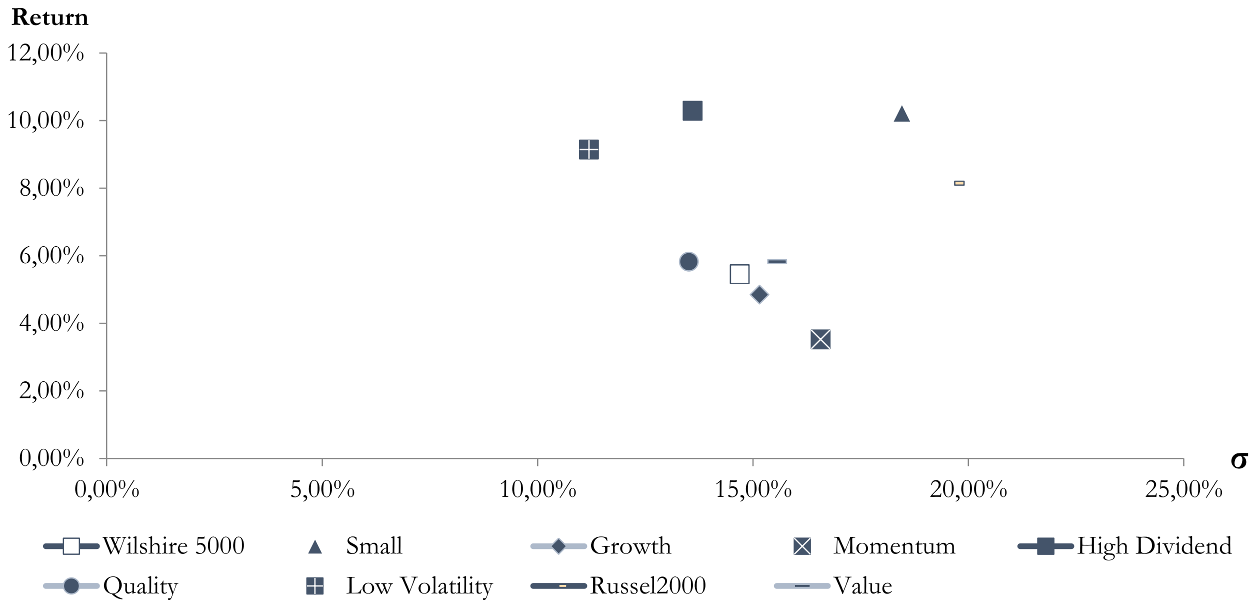


Figure 8

4.6 Descriptive Statistics

4.6.1 Performance Statistics by Sub-Periods

Further empirical analysis has been conducted during various market conditions, to test the Quality factor during shorter time periods. The events below are defined as follows. The Russian/Asian crisis is from January 1997 - January 2000, the Dotcom crisis spans from January 2000 - January 2003, the Great Financial Crisis is measured from January 2007 - January 2010 and the Aftermath of the Great Financial Crisis, also including the fiscal crises in Europe, is from January 2010 - January 2013. The results are depicted below.

	Alpha	Beta	SMB	HML
Russian/Asian Crisis	0.011498**	0.564042	0.159239	-0.010890
Dot-com Bubble	0.005705	0.664003	0.151088	0.044388
Great Financial Crisis	0.007230***	1.009342***	0.766948***	.-0.138088***
Aftermath	0.016181**	1.053517	0.687059**	-0.321837

Table 11

The results can be illustrated further, and below we present the aggregated Quality portfolio in terms of annualized returns, volatility and Sharpe-ratio, as well as the market portfolio. Throughout the six year periods, Quality performs in line with the market portfolio, but risk adjusted returns are higher during the defined events, i.e in various market distress conditions.

Aggregated Quality Portfolio

Event/Year	Annualized return	Annualized Volatility	Sharpe	Min Monthly	Max Monthly
Russian/Asian Crisis	17,74%	13,33%	1,49	-13,44%	7,43%
Dot-com Bubble	-11,41%	15,34%	-0,82	-9,33%	7,65%
Great Financial Crisis	6,17%	19,19%	0,30	-18,55%	16,41%
Aftermath	15,91%	19,44%	0,82	-9,59%	15,01%
1993 - 1998	14,39%	10,58%	1,57	-13,44%	7,43%
1999 - 2004	2,83%	13,13%	0,36	-9,33%	8,47%
2005 - 2011	10,40%	16,24%	0,64	-18,55%	16,41%
2012 - 2017	12,98%	13,54%	1,23	-8,74%	15,01%

Table 12

Market Portfolio (Wilshire 5000)

Event/Year	Annualized return	Annualized Volatility	Sharpe	Min Monthly	Max Monthly
Russian/Asian Crisis	22,86%	17,49%	1,37	-16,08%	7,72%
Dot-com Bubble	-16,37%	20,00%	-0,82	-10,72%	7,94%
Great Financial Crisis	-1,24%	17,84%	-0,04	-17,23%	10,19%
Aftermath	12,67%	15,78%	0,88	-7,89%	11,35%
1993 - 1998	16,37%	11,91%	1,74	-16,08%	7,33%
1999 - 2004	2,32%	15,69%	0,50	-10,72%	8,22%
2005 - 2011	4,91%	14,74%	0,48	-17,23%	10,19%
2012 - 2017	13,52%	11,67%	1,40	-7,59%	11,35%

Table 13

4.6.2 Characteristics of Quality

In the following two figures, Figure 9 and Figure 10, a similar relationship can be observed. Figure 9 and 10 show monthly excess returns of the Quality portfolio (10% and 20%) and the market excess returns. The aggregated Quality portfolio exhibits a mild positive convexity, indicating that it benefits from crisis rather than portraying a crash risk. The sample runs from January 1993 to December 2016.

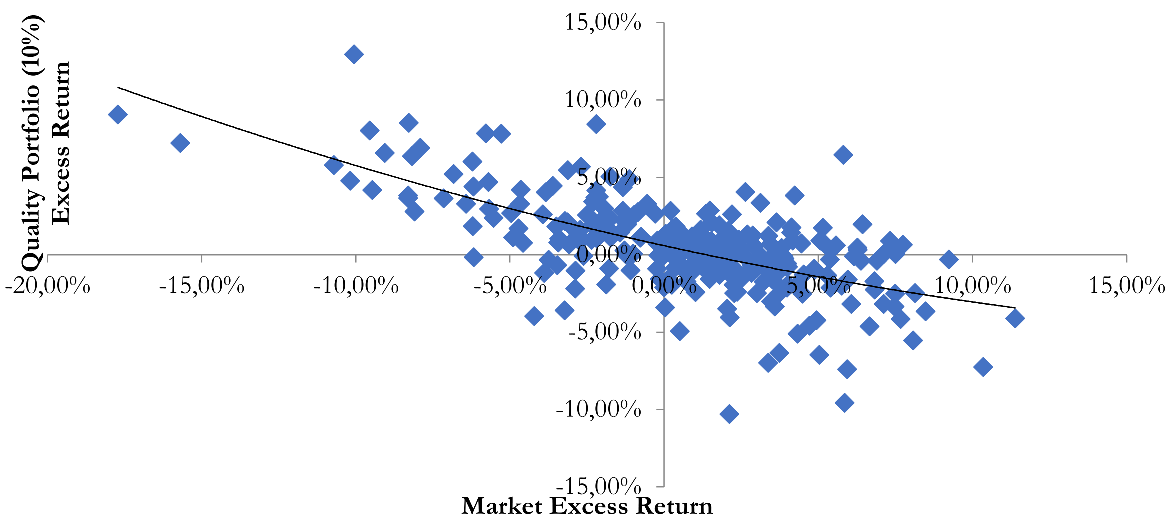


Figure 9

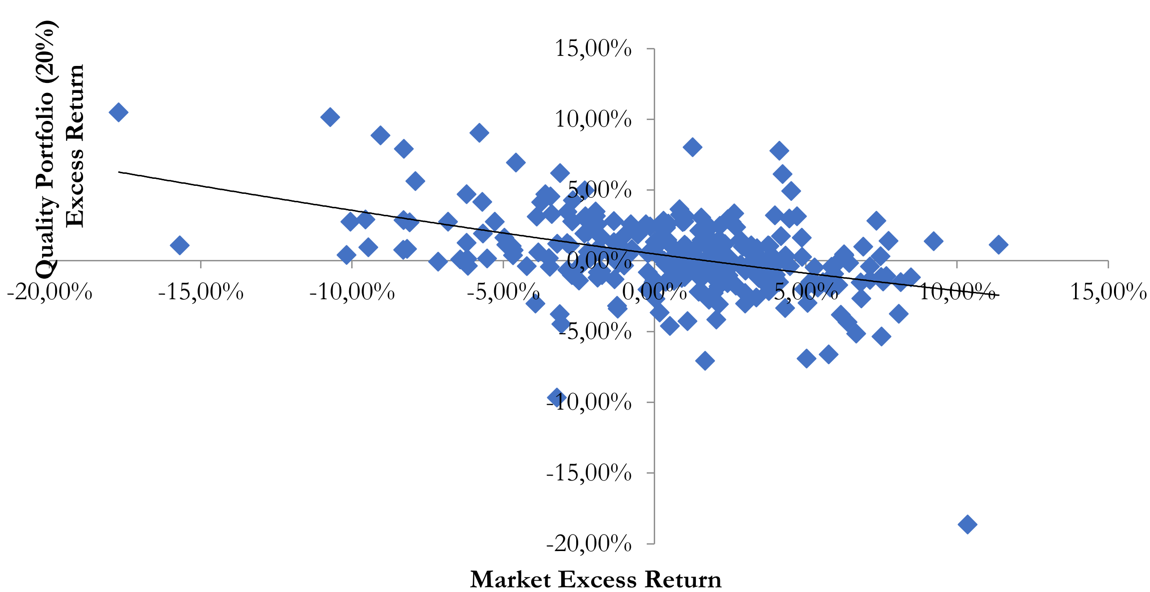


Figure 10

When studying the characteristics of Quality, by observing the differences in the Long- and Short-Quality portfolios, certain features are brought forth. Riskiness, measured by both volatility and Beta, tends to be higher for Low Quality stocks in both cases. The results are illustrated in the figure below.

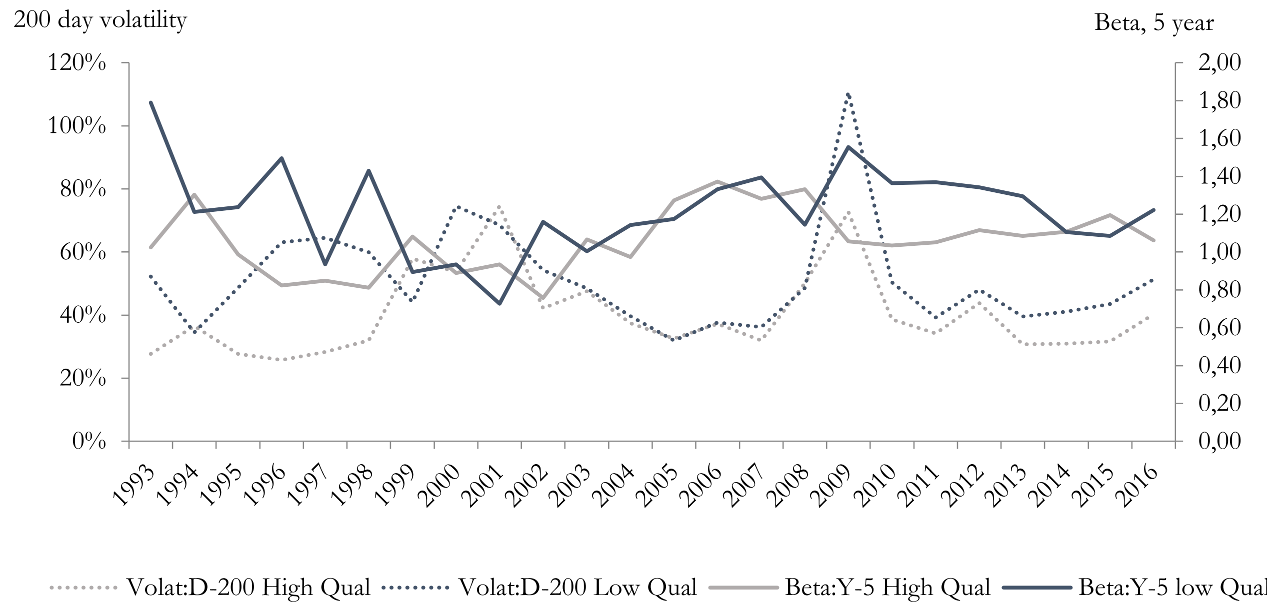


Figure 11

When comparing the size-difference between Low- and High Quality stocks, measured by market capitalization, the empirical results portray a positive relationship between quality and size. The results are shown in the figure below.

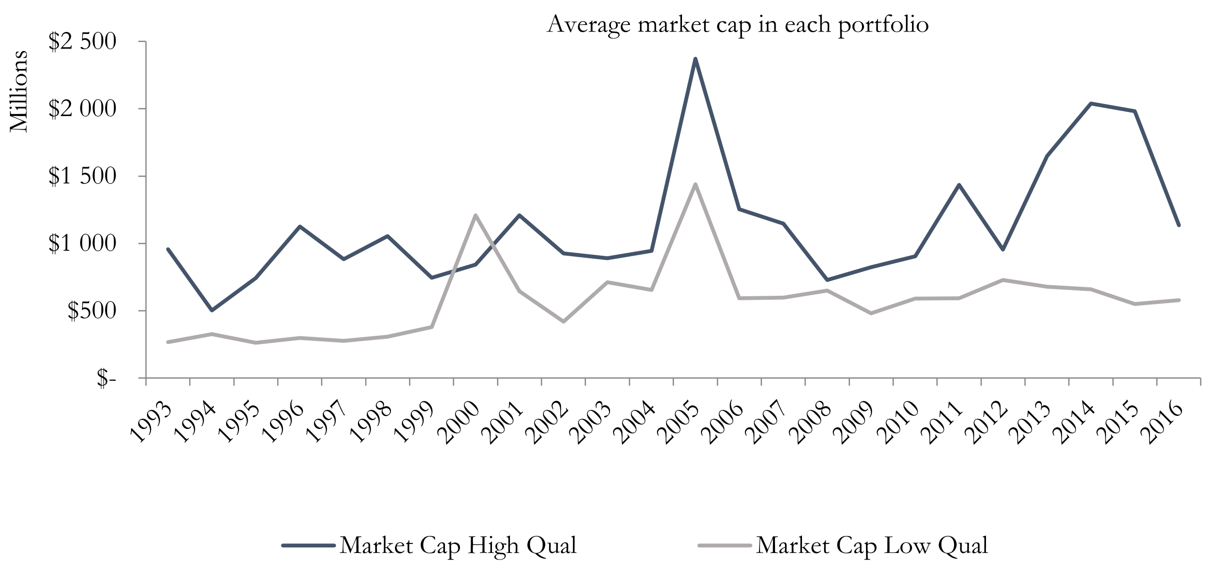


Figure 12

Lastly, when investigating price-differences between Low- and High Quality stocks, measured by price-to-book ratio, High Quality stocks tend to be more expensive than it Low Quality counterpart. This is illustrated in the figure below.

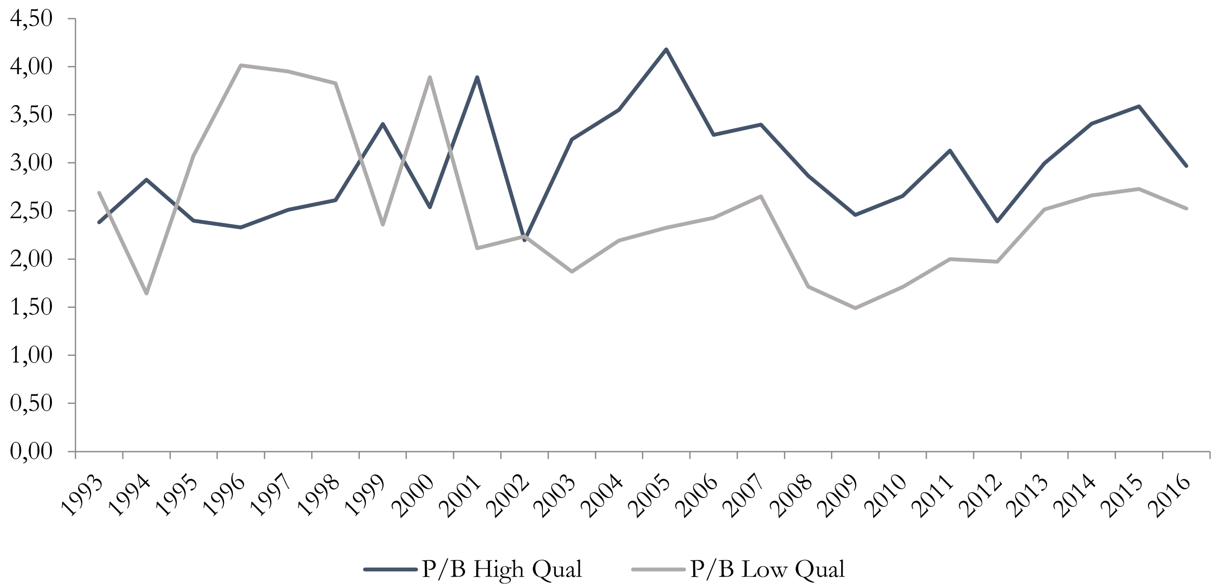


Figure 13

5 Analysis

The following chapter contains the analysis and discussion surrounding the empirical findings in relation to the theoretical framework presented, relevant to the main topic of the thesis.

When the three dimensions of our proxies for Quality were set up, an interesting feature came to light. The first remark can be made regarding the variance and magnitude of significant alphas, for each metric in the different regressions. For instance, the net portfolio based on ROIC exhibited positive, significant alpha at the 10 % level in the Carhart regression, but not in the other regressions. Similar patterns were observed in Dividend, Debt/Equity and Net Debt. The persistence of the metrics, and their ability to generate alpha, therefore seem to mostly be a matter of chance and be dependent on the sample at hand as well as the time frame. Also, in the CAPM setting only 4 out of 12 metrics exhibited significance, only 3 out of 12 metrics were significant in FF3, in the Carhart model, 5 out of 12 obtained significance, and finally, 5 out of 12 metrics were significant in the FF5 model (see Appendix 1). Furthermore, the metrics tend to load on different factors, and since the selection of metrics imply different factor loadings, the merits of selection are likely to produce different results.

The 12 metrics chosen have been frequently used in previous research and they are all, to a varied extent, validated in terms of proxies for Quality. However, it is evident that each of the metrics, used alone, is subject to selection bias. An interesting finding in the literature is that independently of metrics selected to proxy for Quality, there are indications that the combination of metrics should capture a "Quality effect". However, this line of thinking is contradicted by the different factor loadings of the metrics. The factor loadings of the metrics in the Profitability screen exhibit the most coherent factor loadings, except from ROE. Concerning the other dimensions, Safety and Earnings quality, the metrics vary and each of the dimensions exhibit inconsistent patterns (see Figure 1, 2 & 3). This implies that the metrics are not as coherently defined as in the Profitability screen, indicating that factor loadings seem to be linked to the selection of metrics. A more extensive research effort regarding how factor loadings vary for different combinations of metrics would therefore be of interest. The differences in terms of factor loadings that can be seen from our overall Quality portfolio compared to the SHB portfolio, also indicates that a similar selection process of metrics may result in different factor loadings (see Table 2 & 3).

An interesting note regarding this aspect is that the SHB portfolio loads on other factors whilst exhibiting Quality characteristics. The correlation between the two amounts to nearly 80 %, the major difference being the higher amplitude found in the SHB portfolio. We interpret this as an indication of selection bias in the screen. The SHB screen loads highly on Beta, which is contradictory with our Quality screen. Despite the differences in factor loadings, the characteristics are similar. Thus, the lack of a coherent Quality screen continues to be a problem when used as proxy for Quality.

The aggregation of the three dimensions into a zero investment portfolio indicates that Quality earns a lot of the excess returns during bear market conditions. As can be seen in Figure 4, the net portfolio generates consequent excess returns during crises. I.e, the spread of High versus Low Quality stocks seems to increase during times of distress. This can also be seen in section 4.6, Descriptive statistics, that during market distress, Quality performs better than comparable indices.

When the Net Portfolio is regressed on traditional factor models, Quality generates significant monthly alpha in between 0.431 - 0.549 %. Furthermore, the Quality portfolio loads significantly negative on market Beta, and tendencies are observed on negative factor loadings on SMB and HML, although the level of significance varies. It is interesting to note that regardless of factor model, the Net Quality Portfolio seems to earn its excess return from filtering out large cap companies with stable business models (as indicated by the negative Beta) and higher valuations (as indicated by negative loading on HML).

This result is further emphasized in the Conditional Beta Analysis presented in Table 6. Even if the regression does not impose significant result on neither down nor up market conditions, the tendency lean towards an increase in excess return during down market conditions. This can also be linked to the characteristics of the Low and High Quality portfolios, respectively. As can be seen in section 4.6, Figure 11, the risk, measured as either average volatility of all assets in the portfolio or average Beta, is lower in the High Quality portfolio. Also, High Quality stocks tend to be more expensive (measured as P/B) and larger in terms of market capitalization, as shown in Figure 12 and 13. The dynamics of the High versus Low Quality portfolios indicates that there are tendencies toward flight to Quality. For instance, during the Dot-com bubble, the two portfolios were inversely correlated in terms of average P/B ratios, implying that the highly valuated, pre-crisis, tech stocks where not captured by Quality. Not investing in these stocks led to stable returns relative to the overall market in the subsequent year. This valuation spread (high tech stocks vs High Quality stocks) led to a large spread in the High versus Low Quality portfolios from September 1999 until May 2000, thus contributing to earning high returns for the zero investment portfolio. Furthermore, there is a spike in volatility and Beta for both portfolios during the great financial crises. Also, the spread in terms of volatility and Beta increases dramatically. For instance, the amplitude of the volatility is more than 50 % higher for the Low Quality portfolio throughout 2008 - 2009. The fact that High Quality exhibits lower market risk seems to be an important source for explaining the risk adjusted returns.

When dividing the stock universe into two size categories and conduct a sort based on Quality, it is evident that High Quality performs better than Low Quality, irrespective of size. The difference between the highest Quality portfolio within the large cap universe, compared to the lowest Quality for large cap stocks, is 0,78 % monthly. The corresponding figure for the small cap portfolios amounts to 4,14 % monthly. This indicates that our proxy for Quality is significantly higher for small cap stocks. Previous research has outlined that there exists a small cap premium in the equity market due to un-diversifiable sources of risk. Furthermore, small cap stocks do not tend to be as efficiently priced as stocks of larger companies. Thus, our findings are in line with the previous research; the magnitude of the premium in the Quality portfolio is higher for small cap stocks, due to either market inefficiencies or errors in expectations among investors. However, since the results are coherent based on our sort on Quality, we conclude that Quality cannot be rejected to carry a systematic source of risk.

Another interesting aspect of the 6 sub-portfolios is, as presented in Table 8, that regardless of size, the factor loading on Beta diminishes consequently as Quality increases. Also, lower Quality commands negative loading on the Momentum factor. According to the CAPM, low Beta portfolios should not exhibit return in excess of the market, as our Quality portfolios do. From this we can therefore argue for the presence of a premium that is not explained by traditional factor models.

This, combined with a t-statistic higher than 2 for Quality in the CSR, implies that we cannot reject that Quality constitute a systematic risk premium. However, adding Quality to the CSR does not improve the explanatory power sufficiently, to explain all the variation of the stocks, which weakens the argument. Also, as Harvey, Liu and Zhu point out, the value of our t-statistic is not a high enough level when the process of designing the portfolio is subject to several layers of subjective selection.

In a Markowitz optimization setting, the Quality portfolio is “outside” the efficient set. This is also the case for the high dividend and low volatility strategies. This should not be possible according to the Markowitz framework for extended time periods. As can be seen in Table 9, the most efficient portfolio is a combination of factors, and the market portfolio is not assigned any weight. Since the factor portfolios have performed well historically, the results are expected and further strengthen the argument that factor investing can harvest systematic alpha. However, the Markowitz framework has been known to produce unreliable results in terms of future performance and pervasiveness.

Therefore, we conclude that our proxies for Quality aim to capture returns in excess of the market. The main questions that remain are why, and if the Quality factor will persist as a source of systematic alpha. From the cross sectional regression results, we conclude that we cannot reject that Quality is a systematic source of risk, i.e has explanatory power in the cross sectional variation of stocks. As previously discussed, there are tendencies that Low Quality underperforms High Quality, irrespective of size, and the persistence of the net portfolio is higher than the persistence of other factors.

In order to constitute a systematic source of risk, the Quality factor should imply exposures to un-diversifiable factors such as macro-economic factors, value, size or momentum: premiums that command addition return. There is no evident result that points in this direction. Instead, the Quality portfolio earns its return in excess of the market during longer time periods and more turbulent times. An interesting point to be made in this regard is that investors tend to exhibit irrational exuberance and over confidence. During the Dotcom bubble for instance, High Quality was “underpriced” relative to Low Quality. The subsequent result was a mean reversion; Low Quality stocks fell dramatically, whilst High Quality stocks performed well on a relative basis. This seems to be the virtue of High Quality: irrational investor behavior, such as chasing winners, is avoided. The correlation between the market and the Quality portfolio is quite high during “normal” market conditions, and the returns are slightly lower during normal market conditions (which is also indicated by the lower Beta-values).

Furthermore, when the Quality anomaly is analyzed during shorter time horizons, the anomaly is less evident. We believe that this might be consistent with previous research made on investor preferences. Most investors have an investment horizon of 3 -5 years, and the Quality anomaly becomes evident first during longer time periods or during crises. Thus, there seem to be a holding premium for Quality for taking on the horizon risk. Investing in large capitalization, stable stocks with low financial risk, does in other words not seem to attract investors with a shorter horizon to the same extent as riskier investments (i.e companies with lower P/B ratios). Even if High Quality stocks are associated with higher valuations, the pricing of Low Quality stocks are more volatile. Also, the average volatility and Beta-values of the individual stocks in the Low Quality portfolio are higher. Thus, we conclude that the Quality anomaly rather is a result of the miss-pricing of stocks that do not exhibit High Quality traits. It appears to be in times of poor performance of these kinds of stocks that High Quality stocks earn most of their returns in excess of the market. This is indicated by the mean reversion pattern, shown in for instance Figure 6. Therefore, the results indicate that the premium is more likely to be due to systematic errors than a systematic source of risk.

6 Conclusions

The four main conclusions from the analysis above, aimed at responding to the central questions and the purpose of the thesis, are presented in the section below.

6.1 Lack of coherent definitions impose selection bias

The first conclusion to be drawn is that there are several elements of selection bias when constructing a Quality screen. This applies to several aspects of the Quality anomaly. First, defining what metrics to use as a proxy for Quality involves a subjective element. As shown in Figures 1, 2 and 3, the factor loadings across the three dimensions of Quality differs a lot, implying that even on a category level of fundamental metrics, inconsistency can be found. Interesting to note however, is that Profitability (Figure 1) shows the least deviation in terms of factor loadings, a finding that aligns well with previous research such as Novy-Marx. This also reflects upon the persistence, where Profitability exhibits higher persistence than the other dimensions.

Furthermore, the selection bias is also evident when the Quality portfolio is compared to the SHB portfolio. Despite a similar process of defining Quality, the factor loadings are quite different, as shown in Table 2 and 3. However, both portfolios exhibit significant positive alphas, indicating that the usage of a Quality screen captures some premium, and even though the factor loadings differ, the correlation between the two amounts to roughly 80 %.

The main conclusion is thus that the difficulty in defining Quality coherently makes it too prone to several selection biases, which in turn is likely to affect the overall performance of the Quality factor. Still, the return characteristics of both portfolios are left unexplained by traditional factor models, indicating a presence of a Quality premium.

All conclusions from this point on will be based on the Quality factor as outlined by this essay.

6.2 High performance in down markets – flight to Quality

The second conclusion is that stocks that exhibit Quality characteristics earn most of the returns in excess of the market during adverse market conditions. This can for instance be inferred from

Figures 4, 5 and 6. The peaks in excess returns of the Quality portfolio tend to coincide with the various financial crises, and the cumulative excess returns remain relatively flat in the periods in between. This is further illustrated in Table 12 and 13, where the Quality portfolio outperforms the market portfolio in most of the crisis. Similarly, this is indicated from the Conditional Beta Analysis, showing that the benefit of owning Quality stocks is higher when equity markets are in distress.

Another feature amongst Quality stocks is that they are less risky. As illustrated in Figure 12, High Quality stocks exhibit lower volatility and Beta than their Low Quality counterparts, thus rendering them less risky. The Quality portfolios' positively convex traits, as shown in Figure 9 and 10, indicates a flight to Quality which benefits from crisis rather than implying a crash risk. When comparing the Quality portfolio to the Wilshire 5000 index, it is also evident that the Quality portfolio is less risky since the Sharpe-ratio is higher. This is further strengthened by the fact that the Quality portfolio engages a position outside the efficient set, as can be seen in Figure 8.

6.3 Presence of a Quality Premium

The third major conclusion to be drawn is that there seems to be a systematic premium associated with Quality. The portfolio's performance and consequent alpha cannot be explained by traditional factor models, and we find a t-statistic higher than 2 in the CSR, displayed in Table 6. However, adding Quality to the CSR does not improve the explanatory power sufficiently, to explain all the variation of the stocks. Also, as Harvey, Liu and Zhu point out, the value of our t-statistic is not a high enough level when the process of designing the portfolio is subject to several layers of subjective selection. Thus, we conclude that Quality cannot be rejected to carry a systematic source of risk.

The conclusion above can be extended to the 6 sub-portfolios, sorted on size and Quality. As is shown in Table 7, the factor loading on Beta diminishes consequently as Quality increases, regardless of size.

The fact that High Quality exhibits lower market risk seems to be an important source for explaining the high risk adjusted returns. The high performance relative to the benchmark in benign market conditions accounts for the majority of excess returns. This argument, combined with the fact that traditional factor models cannot explain the alphas, indicates the existence of a Quality premium.

6.4 Systematic errors rather than systematic risk

The final conclusion to be drawn is that the premium derived from the Quality factor is more likely to be due to systematic errors rather than systematic risk. As discussed above, Quality earns the majority of its excess return during market distress. A conclusion to be drawn in this sense is linked to behavioral aspects. Investing in High Quality stocks leads to a more rational investor behavior, such as avoiding chasing winners. The correlation between the market and the Quality portfolio is quite high during "normal" market conditions, and the returns are slightly lower during normal market conditions.

We thus conclude that the Quality anomaly rather is a result of the miss-pricing of stocks that do not exhibit High Quality traits. This is indicated by the mean reversion pattern, shown in for instance Figure 6. The phenomena of flight to Quality, shown in Figures 9, 10, further strengthen this conclusion. Therefore, the results indicate that the Quality premium is more likely to be due to systematic errors, foremost in terms of Low Quality stocks, than a systematic source of risk.

In order to constitute a systematic source of risk, the Quality factor should imply exposures to un-diversifiable factors such as macro-economic factors, value, size or momentum: premiums that command addition return. There is no evident result that points in this direction.

Another aspect strengthening this conclusion is that most investors have an investment horizon of 3 -5 years, and the Quality anomaly becomes evident first during longer time periods or during crises. Thus, there seem to be a holding premium for Quality associated with taking on horizon risk. Investing in large capitalization, stable stocks with low financial risk, does in other words not seem to attract investors with a shorter horizon to the same extent as riskier investments.

7 Future research

During the process of dissecting the Quality anomaly in order to answer the main questions of the thesis, certain elements and indication have been brought forth which are subject to further research. These findings are presented in the chapter below.

When constructing the Quality factor used in this thesis, it became evident that the factor loadings varied between the metrics, even in regards to factors of the same category. Subsequently, the factor loadings of the Quality factor based on a combination of metrics might therefore differ, conditional on which metrics are chosen to proxy for the factor. Further research in regards to the factor loadings of the metrics, as well as the persistence of the metrics, would therefore be beneficial for defining a Quality factor, and suitable metrics to proxy for it. The differences in terms of factor loadings that can be seen from our overall Quality portfolio compared to the SHB portfolio, also indicates that a similar selection process of metrics may result in different factor loadings (see Table 2 & 3).

Another interesting topic to investigate further, is whether different Quality screens can be proxied for by other factors, such as low volatility or dividend strategies. The similarities in terms of risk and return of these strategies, outlined briefly in this essay, indicates that there might be some common sources that generate the excess returns. As such, it would be interesting to see if there are any common elements in terms of risk premiums, which carry explanatory power.

Furthermore, Liquidity is a commonly used risk factor in explaining returns. It would therefore be of interest to adjust the Quality Screen, and for instance form sub portfolios sorted on Liquidity. It would also be interesting to conduct a CSR based on this kind of screen, to further dissect the Quality Anomaly in equity returns.

8 References

Ang, A., W. N. Goetzmann and S. M. Schaefer (2009), "Evaluation of Active Management of the Norwegian Government Pension Fund – Global"

Arnott, R., Berkin, A., & Ye, J. (2000). "How Well Have Taxable Investors Been Served in the 1980s and 1990s?" Journal of Portfolio Management, 26; 4, 84 - 93.

Bank of International Settlements (2017). Available online: https://www.bis.org/

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

Barras, L., Scaillet, O. och Wermers, R. (2010). "False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas", The Journal of Finance 65, 179 – 216.

Berk, J. and van Binsbergen, J. (2012), "Measuring Managerial Skill in the mutual fund industry", National Bureau of Economic Research. Working Paper No 18184.

BlackRock (2017). Available on: https://www.blackrock.com/

Bogle, J. (2009). Common Sense on Mutual Funds. Wiley.

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

Chan, Louis K.C., Yasushi Hamao, and Josef Lakonishok, 1991, Fundamentals and stock returns in Japan, Journal of Finance 46, 1739–1789.

Chen, Nai-fu, Richard Roll, and Stephen Ross (1986), "Economic forces and the stock market". Journal of Business 59: 383–403.

Cohen, Randolph B., Paul A. Gompers, and Tuomo Vuolteenaho, 2002, Who underreacts to cash-flow news? Evidence from trading between individuals and institutions, Journal of Financial Economics 66, 409–462.

Connor, Gregory (1995), "The Three Types of Factor Models: A Comparison of Their Explanatory Power," Financial Analysts Journal, May/June 1995, Vol. 51, No. 3: 42-46.

CS Investing (2013). Available on: http://csinvesting.org

Daniel, Kent, and Sheridan Titman, 2006, Market reactions to tangible and intangible information, Journal of Finance 61, 1605–1643.

Davis, James L (2001). "Explaining the Stock Returns: A Literature Survey" Dimensional Fund Advisor's Library (https://www.ifa.com/pdfs/explainingstockreturns.pdf)

Egloff, D., M. Leippold, and L. Wu (2010): "The term structure of variance swap rates and optimal variance swap investments," Journal of Financial and Quantitative Analysis, 45, 1279–1310.

ETFGI (2017). Available on: http://etfgi.com/

Eugene F. Brigham and Myron J. Gordon (1970), "Leverage, Dividend Policy and the Cost of Capital" The Journal of Finance Vol. 25, No. 4 (Sep., 1970), 904-908

Fairfield, Patricia M., Scott Whisenant, and Terry Lombardi Yohn, 2003, Accrued earnings and growth: Implications for future profitability and market mispricing, The Accounting Review 78, 353–371.

Fama, Eugene F. 1991. "Efficient capital markets: II." The Journal of Finance 46.5: 1575-1617.

Fama, Eugene F., and French, Kenneth R., 2008, Dissecting Anomalies, Journal of Finance, 63, 1653 – 1677.

Fama, Eugene F., and Kenneth R. French (2010), "Luck versus Skill in the Cross-section of Mutual Fund Returns", Journal of Finance Vol. 65, No. 5

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, Journal of Finance 47, 427–465.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, Journal of Financial Economics 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 1995, Size and book-to-market factors in earnings and returns, Journal of Finance 50, 131–156.

Fama, Eugene F., et al. 1969. "The adjustment of stock prices to new information." International economic review 10.1: 1-21.

Fama, Eugene F., Lawrence Fisher, Michael C. Jensen, and Richard J. Roll. 1969. "The adjustment of stock prices to new information." International economic review 10.1: 1-21.

Fama, Eugene. F. Foundations of Finance. New York: Basic Books, 1976, p. 319.

Frazzini, Andrea, David Kabiller, and Lasse Heje Pedersen (2012), "Buffett's Alpha," Working Paper, AQR Capital Management, New York University.

Geczy, Christopher and Mikhail Samonv (2013), "212 Years of Price Momentum (The World's Longest Backtest: 1801-2012)," Working Paper, University of Pennsylvania (The Wharton School), OctoQuant.com.

Gruber, Martin J. (1996). "Another Puzzle: The Growth in Actively Managed Mutual Funds". Journal of Finance 51, 783 – 810.

Haugen, Robert A., and Nardin L. Baker, 1996, Commonality in the determinants of expected stock returns, Journal of Financial Economics 41, 401–439.

Hotelling, H., 1933. Analysis of a Complex of Statistical Variables Into Principal Components, Journal of Educational Psychology, volume 24, pages 417-441 and 498-520.

Hyde, Charles E (2015). "The Piotroski F Score in the Australian Market: Performance & Fundamental Drivers" Available at SSRN: https://ssrn.com/abstract=2577181

Ikenberry, David, Josef Lakonishok, and Theo Vermelean, 1995, Market underreaction to open market share repurchases, Journal of Financial Economics 39, 181–208.

Ishares, 2017. Available at: http://www.ishares.com/

Jegadeesh, Narasimhan, 1990, Evidence of predictable behavior of security returns, Journal of Finance 45, 881–898.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, Journal of Finance 48, 65–91.

Johnson, A. R. & Wichern, W. D. (2009), "Applied Multivariate Statistical Analysis." 6th edition. Pearson Prentice Hall NJ

Jolliffe, I.T., 2002. Principal Component Analysis, second edition, New York: Springer-Verlag New York, Inc.

Jones, R.C. and R. Wermers (2011), "Active Management in Mostly Efficient Markets," Financial Analysts Journal 67(6), 29-45.

Kinnel, R. (2010). "How Expense Ratios and Star Ratings Predict Success", Morningstar.

Kosowski R., Timmermann A. och Wermers R. (2006) "Can mutual fund 'stars' really pick stocks? New evidence from a bootstrap analysis". Journal of finance 61, 2551 – 2595.

Lintner, John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, Review of Economics and Statistics 47, 13–37.

Litterman, R., and J. Scheinkman (1991): "Common factors affecting bond returns," Journal of Fixed Income, June, 54–61.

Loughran, Tim, and Jay R. Ritter, 1995, The new issues puzzle, Journal of Finance 50, 23–51. Pontiff, Jeffrey, and Artemiza Woodgate, 2008, Share issuance and cross-sectional returns, Journal of Finance 63, 921–945.

Malkiel (1995), "Returns from Investing in Equity Mutual Funds 1971 to 1991," Journal of Finance 50, 549-572

Malkiel, B. (2012). “A Random Walk Down Wall Street”. W. W. Norton & Company.
Miller, Guy (2006), “Needles, Haystacks, and Hidden Factors,” Journal of Portfolio Management, Vol. 32, No. 2: pp. 25-32.

Mok, William, Jennifer Bender, and P. Brett Hammond (2013), “Can Alpha be Captured by Risk Premia?” Journal of Portfolio Management, 40; 2, 18 – 29.

Mossin, J. (1966), “Equilibrium in a Capital Asset Market", Econometrica, 34, 768–783.

MSCI, 2013, “Foundations of Factor Investing”, Available at: http://www.MSCI.com

Norges Bank, 2015, “The Quality Factor”

Pearson, Karl, 1901. On lines and planes of closest fit to systems of points in space, Philosophical Magazine, Series 6, vol. 2, no. 11, pp. 559-572.

Pontiff, Jeffrey, and Artemiza Woodgate, 2008, Share issuance and cross-sectional returns, Journal of Finance 63, 921–945.

Rosenberg, Barr, and Vinay Marathe (1976), “Common Factors in Security Returns: Microeconomic Determinants and Macroeconomic Correlates,” University of California Institute of Business and Economic Research, Research Program in Finance, Working paper No. 44.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, 1985, Persuasive evidence of market inefficiency, Journal of Portfolio Management 11, 9–17.

Ross, Stephen (1976), "The arbitrage theory of capital asset pricing,” Journal of Economic Theory 13(3), 341–360.

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, Journal of Finance 19, 425–442.

Sloan, Richard G., 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings? The Accounting Review 71, 289–315.

Stock, J. H., and M. W. Watson (1999): “Forecasting Inflation,” Journal of Monetary Economics, 44, 293–335.

Titman, Sheridan, K.C. John Wei, and Feixue Xie, 2004, Capital investments and stock returns, Journal of Financial and Quantitative Analysis 39, 677–700.

Titman, Sheridan, K.C. John Wei, and Feixue Xie, 2004, Capital investments and stock returns, Journal of Financial and Quantitative Analysis 39, 677–700.

Treynor, J. L. (1961), “Market Value, Time, and Risk.” Unpublished manuscript. Rough draft dated 8/8/61, #95-209.

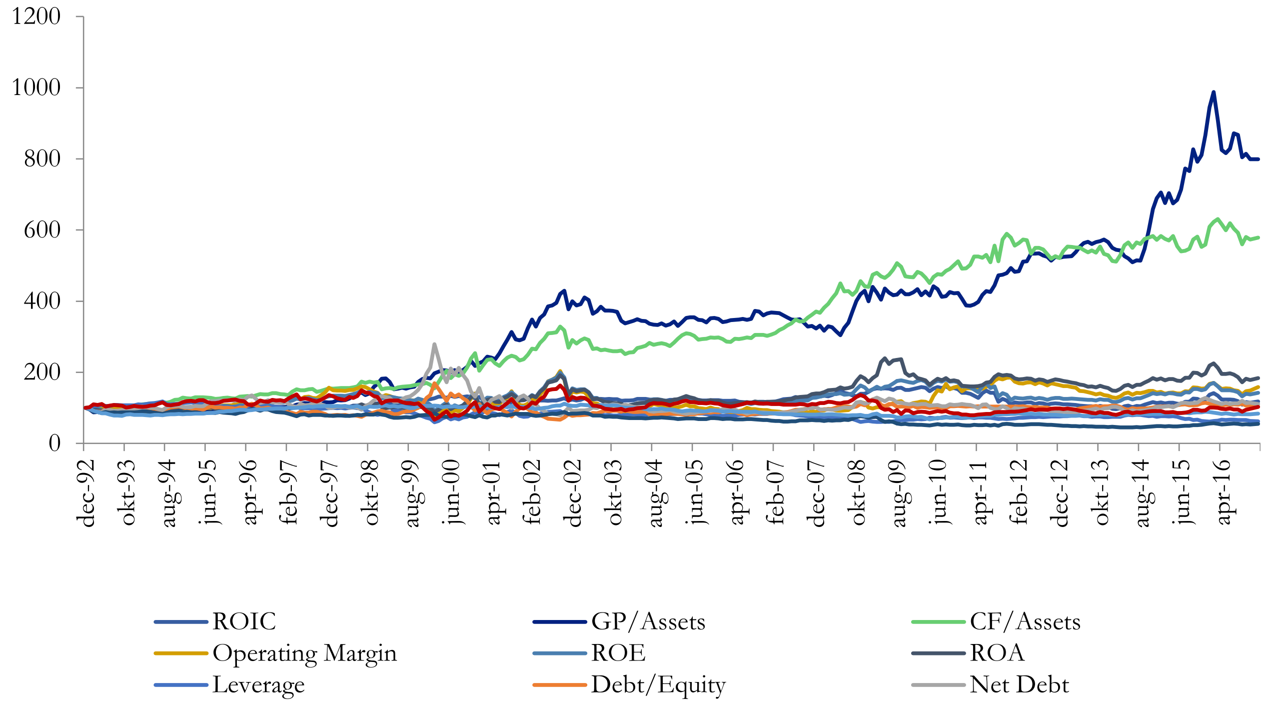
Verbeek, Marno (2012). "A Guide to Modern Econometrics", 4th Edition, John Wiley & Sons Inc
Wermers, R. (2003), "Is Money Really 'Smart'? New Evidence on the Relation between Mutual Fund Flows, Manager Behavior, and Performance Persistence." Unpublished paper, University of Maryland (May).

Appendix

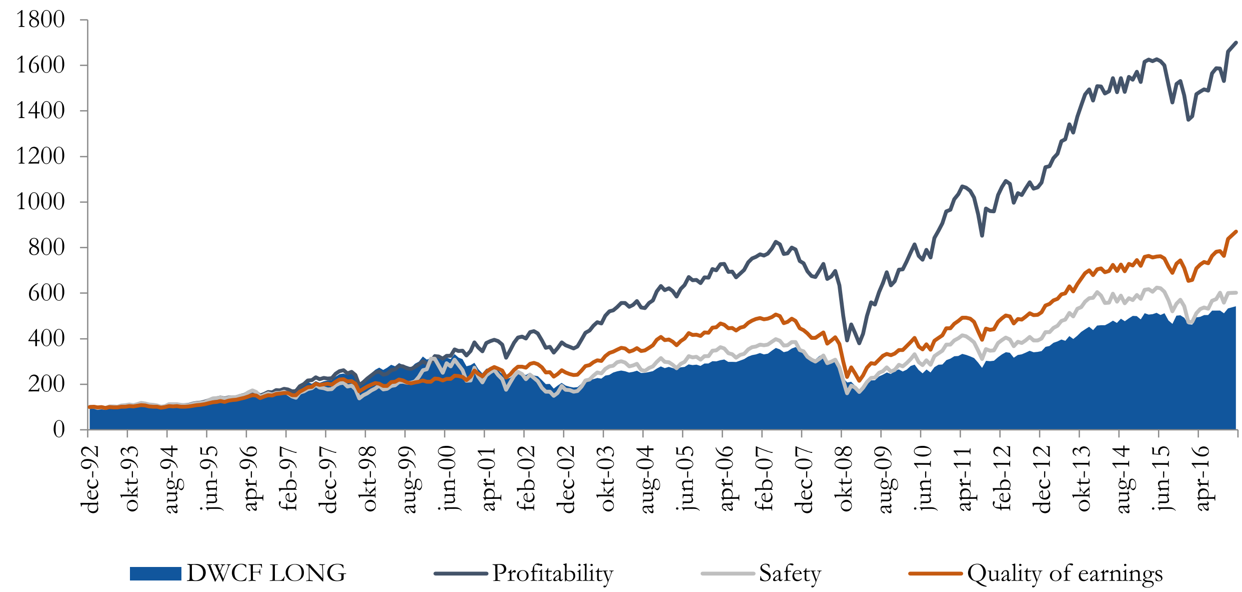
Alfa	ROIC	GP/Assets	CF/Assets	Operating Margin	ROE	ROA
CAPM	0.001655	0.007089***	0.006121***	0.004608	0.002240	0.003139
FF3	0.000753	0.007797***	0.005577***	0.002688	0.001494	0.002566
Carhart	0.002095*	0.007094***	0.004311**	0.001217	0.000154	0.001972
FF5	0,000082	0.006265***	0,002245	0.000343	-0.001511	0.000452

Alfa	Leverage	Debt / Equity	Net Debt	EPS Stability	Dividend	Equity Var
CAPM	-0.002909	-0.001777	-0.001566	0.002706*	0.002347**	0.002631
FF3	0.004497***	0.001207	0.001665	0.001647	0.001386	0.000878
Carhart	0.004130***	0.000514	0.001185	0.002533*	0.001464	0.000752
FF5	0.006125***	0.003205*	0.004595**	0.002221	0.001892*	0.002777

Appendix 1



Appendix 2



Appendix 3