CBS COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

Master's Thesis

MSc in Advanced Economics and Finance (cand.oecon)

Copenhagen Business School

May 15, 2020

Quality Investing and Industry Heterogeneity

Authors:

Johannes Lundh (123108)

Sergi Estorach Cardona (124374)

Supervisor:

Mads Stenbo Nielsen

Number of characters (including spaces): 246 836

Number of pages: 116

Abstract

This thesis aims to investigate if there exist industry differences regarding how well various definitions of the style-factor “quality” can explain stock returns. It also investigates if it is possible to improve portfolio performance by including industry heterogeneity in a quality-investing strategy.

Common definitions of firm quality are used as proxies for the quality factor, namely; gross profitability, operating profitability, return on equity, return on asset, return on invested capital, investments, and debt to equity ratio. Their power to explain stock returns is tested while controlling for size, value, and momentum. Using cross-sectional tests as well as Fama-MacBeth regressions, we find some evidence of industry heterogeneity. Technology and industrials are the industries where most metrics are significant. Energy and consumer staples are the ones with fewest significant metrics. Industry heterogeneity is also found regarding the quality metrics relative coefficient sizes.

Based on this, we test whether our definitions of quality can generate a return premium and if this premium can be improved by taking the observed industry differences into account. This is tested using portfolio sorting strategies; sorting based on firm quality and controlling for other factor loadings using Carhart’s four-factor model. We develop quality sorting strategies that take multiple firm quality metrics into account: one benchmark sorting strategy, where all industries are treated equally, and two strategies that take scaling amounts of industry heterogeneity into account.

The benchmark portfolios generate significant alphas in eight out of the eleven industries. However, these alphas are only significant for the high-quality portfolio in three of these portfolios. Furthermore, the sorting strategy is improved in two out of these eight industries when considering industry heterogeneities. However, when incorporating industry heterogeneities, we find signs of improvement as the magnitude of the alphas is enhanced for most industries, regardless of their significance.

Keywords: Quality investing, asset pricing, industry heterogeneity, quality metrics, explanatory power, cross-sectional analysis, portfolio alphas, sorting strategies.

Acknowledgments

We would like to thank our supervisor Mads Stenbo Nielsen for all the helpful feedback and engaging discussions that have contributed greatly to our thesis. We would also like to thank the CBS library staff for their guidance with regard to our data collecting process.

Table of contents

1 Introduction

Factor investing, or style investing, is an investing style where investors chose what assets to invest in based on certain, predetermined, characteristics of that asset. Based on asset pricing research, the investor can identify asset characteristics that, historically, have shown to be associated with superior performance. Superior performance can either be in terms of return, risk, risk-adjusted returns, or even transaction costs. A style investor will tilt his/her portfolio towards assets with the characteristics that the investor believe is associated with better future performance, and by doing so, hopefully, boost portfolio performance.

The academic evidence supporting the existence of return driving factors is not a new phenomenon. Sharpe (1964), Lintner (1965), Treynor (1961) and Mossin (1966) showed early on that stock returns were correlated to their co-movement with the market portfolio, a theory that later became the Capital Asset Pricing Model. Other return driving factors that showed to be associated with superior stock performance has since then been well identified, such as value, size, momentum and quality. Factors that has become known as style factors. Along with more thorough research within the field and the fact that style factors have proven themselves worthy over time, style investing has become a common methodology used by practitioners to achieve superior results (Warren & Quance, 2019).

While there seems to be a certain consensus amongst academic researchers and practitioners on how to define and quantify style factors such as value (book to market ratio), size (market value) and momentum (trailing 12 months past returns), the same does not hold for the quality factor. Hsu et al. (2018) identify a clear discrepancy between how practitioners, such as institutional investors and index providers measure quality in their investing strategies. As with practitioners, academics also show a clear discrepancy when it comes to defining and quantifying quality. There is not only a vast amount of ways used to measure quality, but the differences between the quality metrics have also shown to be, in some cases, very large. In fact, the quality metrics can vary between being a measure of, for example; profitability, capital structure and asset growth. So, as it is right now, quality as a style factor does not lack recognition with regards to its ability to capture superior performance. However, it does lack consensus with regards to how high-quality should be defined. From an investor's point of view, knowing what quality metric poses the best ability when it comes to explaining stock returns, is of great importance. One consensus that does exist with regards to the quality factor is that “quality” is something investors should be willing to pay more for. Similarly to purchasing most other assets, such as a house, most would agree that a house of high quality should be more expensive. However, how to measure quality in a house is not that clear cut, and the same holds for stocks. The characteristic that best defines equity quality is not obvious. Is it profitability, low leverage, low asset growth, or a well thought-off combination of these measures?

Regardless of the quality metric discrepancy, academia has managed to present robust results pointing towards the existence of a quality factor. Empirics across time, and across markets have all shown clearly that high-quality firms do outperform the market. Vyas and van Baren (2019) widened the style factor research when they showed that some quality metrics, along with other conventional factors, are priced differently across industries. Thus, they captured something that other papers had overlooked. Cheng et al. (2019) and Baca et al. (2019) further concluded that a rather large part of equity return variability was a result of industry differences and that industry-specific effects are gradually becoming a larger driver of equity variability. Based on these two observations, one could argue that it is plausible that various quality metrics perform differently depending on the industry on which they are applied. If that is the case, it could be of much importance to quality investors to know what metric, or metrics, is best suited for what industry. With factor investing growing in popularity, and industry effects becoming a larger driver of equity variations, we believe that a better understanding of industry heterogeneities' effect on quality-investing is necessary.

This thesis therefore aims to unravel some of the mystique around the quality metric. Both qualitative and quantitative methods will be used to analyze the explanatory power of commonly used quality metrics across different industries in the US equity market and if firm quality can generate superior portfolio returns. Furthermore, we will analyze if it is possible to improve a quality investing strategy by incorporating industry heterogeneities.

1.1 Research question

By studying the field of factor investing and equity returns, we came across two interesting findings. First, the lack of consensus on how to measure firm quality and secondly, the growing importance of industry-specific effects on equities are questions that we believe need further investigation. We consider it being of importance for investors to understand if the way quality is defined yields varying results depending on the industry on which these are applied. We also believe that, if various ways of measuring firm quality explain stock returns differently depending on the industry, that it would be useful to investigate if the portfolio sorting methodology could be improved by taking these differences into account. With that said, the questions that we want to answer with this thesis is the following;

Is there a difference in how well various quality metrics explain stock returns in different industries, and is it possible to improve one's quality sorting strategy by taking these differences into account?

This research question can be framed as two separate hypotheses;

Hypothesis 1: *There exist industry differences with regards to how well various quality metrics explain stock returns.*

Hypothesis 2*: Taking industry differences into account when developing a quality sorting strategy will improve portfolio performance.*

By answering this research question, we believe that this thesis will add knowledge to the existing literature on quality investing as well as bring useful insights to practitioners.

1.2 Thesis delimitation

The empirical research of this thesis is limited to analyzing various quality metrics across the US stock market. Hence, the findings of this thesis might not be applicable in other geographical areas or for other asset classes, such as bonds. The main reasons for the geographical delimitation is; differences in market conditions (such as varying investor types, liquidity differences and currency effects) and the difference in accounting standards between countries. Non-homogeneous accounting standards could result in varying definitions of the same quality metric, which would make a fair comparison difficult. One difference between the US accounting standards, GAAP, and the international standard, IFRS, is related to the treatment and reporting of assets. For example, fixed assets are valued at historical cost under GAAP, while IFRS allows valuation to be based on market value. These differences affect the definition of assets, which in turn affects profitability and leverage related ratios. The US equity market is also the largest and most developed equity market in the world, holding roughly 44% of the world's equity value in terms of market capitalization (The world bank, 2020), which allows us to gather a broad enough dataset. We also focus on equities since this is the main asset on which quality factor strategies are applied. Besides this, to ensure good and homogenously reported data, we focus on equities traded on the two main US stock exchanges, the New York Stock Exchange and NASDAQ.

We will also focus on the 11 larger industries in our industry segmentation, using a well-known and acknowledged industry segmentation benchmark, the Industry Classification Benchmark. This means that, if a similar analysis were to be done using more a narrow industry segmentation, results could be different from what this thesis will produce. However, we believe that our 11-industry segmentation approach manages to cover most of the industry variation without segmenting the industries too narrowly and thus negatively affecting sample sizes.

The thesis is also delimited to analyze the quality metrics that have shown potential in previous research. Hence, if different quality metrics are analyzed, the outcome might differ from what is presented in this thesis. Also, this thesis neglects the impact of transaction costs on portfolio returns, which in practice will affect portfolio returns. However, since we aim to explore performance differences between various quality metrics, not between different style factors, we argue that the difference in transaction costs between quality metrics is small enough not to impact the concluding findings of this thesis. Findings have also shown that transaction costs related to the quality factor are relatively low compared to other conventional style factors. This is more thoroughly discussed in *section 3.6*.

1.3 Research design

This study uses a descripto-explanatory design since it utilizes a combination of descriptive and explanatory techniques. A descriptive approach is used since we want to observe if there exists "… a difference in how well various quality metrics explain stock returns in different industries" and if it is "possible to improve one's quality sorting strategy by taking these differences into account?". We then apply an explanatory framework to those findings in order to analyses the relationship between potential industry heterogeneities and quality metrics. The thesis is built upon quantitative analysis. The main analysis is empirical in nature since it aims to find and analyze statistical relationships between firm quality and stock returns in data. A modest amount of content analysis is used to identify patterns in previous research.

1.4 Thesis structure

This thesis is structured such that previous research related to our topic will be presented first to give the reader an overview of how well the topics have been covered so far. Then, necessary background information related to factor investing will be covered as well as theories related to the topic and the techniques used in this thesis. This includes covering econometric and statistical concepts used during the empirical research. We will then cover the data and the data related work preceding the empirical analysis, and the methodologies applied in this thesis as well as the main findings and results from our work. Finally, we will end this thesis by thoroughly discussing the main conclusions that can be drawn from the findings of this paper.

A more detailed structure of the project has the following outline:

Chapter 1: This first chapter introduces the reader to the topic and the reason behind our research. It also covers the research question and the thesis delimitations.

Chapter 2: This part will cover the previous research that has been conducted with regards to quality investing or within the scope of quality investing. We will also cover research that has studied the industry-specific effects on stock returns. We end this chapter by discussing how we hope to contribute to the existing research.

Chapter 3: This chapter provides a thorough discussion on factor investing, including quality investing and other popular investing strategies, including other topics related to quality investing or factor investing.

Chapter 4: This section will cover the theories on which this thesis is built upon.

Chapter 5: This chapter deals with the econometric and statistical concepts utilized in the empirical research.

Chapter 6: Here, we discuss the data collecting methodology used when gathering data for our empirical research. We also discuss the techniques used to clean and treat outliers in the data.

Chapter 7: This chapter will cover the methodologies and structure of our empirical research. This chapter is designed to help the reader understand the procedure we use to reach our results.

Chapter 8: Here, we will present, interpret, and dissect the results of our findings.

Chapter 9: The final section summarizes our findings and discusses the conclusions drawn as well as presents possible topics for future research.

2 Literature review

A modest amount of research has been conducted with regards to the quality factor and its ability to forecast and explain stock returns. Academia agrees on the quality factor's ability to forecast and explain stock returns; however, no consensus has been established regarding the construction of the quality factor. It is common in the literature to see the definition of quality as a profitability proxy, with operating profitability, gross profitability as the definitions, closely followed by earnings factors such as return on equity, return on assets and return on invested capital, but other quality metrics have also shown power. Also, little known literature focuses on comparing the explanatory power of different quality definitions, and there is limited research that focuses on whether there is a performance variation for different quality definitions throughout various industries. In the following section, we will conduct a review of some of the most important literature related to this project. The content and findings of these articles will also be used at a later stage when conducting an initial screening of potentially promising quality definitions.

2.1 Previous research on the quality factor and quality metrics

2.1.1 Hou et al. (2012)

Built on a rich theoretical literature on investment-based asset pricing, Hou, Xue & Zhang's paper *Digesting Anomalies: An Investment Approach*, provides a four-factor model consisting of the market factor, a size factor, an investment factor, and a return on equity factor. Using a two-period investment-based asset pricing model, Hou et al. (2012) present a theoretical relationship between expected future return on equity, investments to assets, and expected stock return. They show that, given current return on equity levels, stock return decreases with investments to assets, and given current investments to assets, stock returns increase in future expected return on equity. Further, they used current return to equity as a proxy for future expected return on equity. Combining these two return predicting measures with the size factor of Fama-French (1993), they present their q-factor model. They argue that their so-called "q-factor" model is less prone to data mining than Fama-French's three-factor model due to it being more related to theory and less based on ad hoc empirics. They further test this model across a broad and long sample of US equities, and they find that their q-factor model manages to outperform

Fama-French and Carhart's three and four-factor models when it comes to explaining excess stock returns.

Even though this paper does not label return on equity or growth in investments as quality metrics, this paper provides a theoretical foundation as well as empirical evidence for these metrics' ability to predict stock returns. Both return on equity and growth in investments have later become commonly used proxies for firm quality.

2.1.2 Novy-Marx (2013)

The Other Side of Value: The Gross Profitability Premium (Novy-Marx, 2013) focuses on the profitability part of the quality factor, namely gross profitability. Similar to Hou et al. (2012), Novy-Marx uses a theoretical framework, the Dividend Discount Model, to show that profitable firms should outperform unprofitable firms and thus deliver superior stock returns. Novy-Marx focuses on analyzing the predictive power of gross profitability since he argues that this is the purest measure of firm profitability. Besides this, he also tests other profitability measures, free cash flow, and income, both scaled by book equity. The empirical analysis is done using Fama-MacBeth regressions and a portfolio sorting strategy. Gross profitability, measured as gross profits (revenue minus cost of goods sold) scaled by total assets, showed to have significant explanatory power on returns in the cross-section, even when controlling for firm size, book to market ratio, and momentum. Free cash flow and income, scaled by book equity, both showed to have less but significant predictive power in the Fama-MacBeth regressions. In the portfolio setting, by allocating firms to five different portfolios based on their gross profitability, highly profitable firms earn a significantly higher average return compared to unprofitable firms. The most profitable firms earned a 0.32% extra return above the least profitable firms on a monthly basis. A long-short portfolio of low and high profitability stocks achieved a 0.52% monthly alpha above Fama and French's three-factor model. Novy-Marx explains the predictive power in gross profitability as a way of acquiring productive capacity cheaply by financing the purchase of productive assets through the sale of unproductive assets.

2.1.3 Novy-Marx (2014)

As an extension of Novy-Marx (2013), *Understanding Defensive Equity* (Novy-Marx, 2014) further digs into the profitability part of the quality factor. Like Novy-Marx (2013), Fama & French (2014), and Ball et al. (2014), this paper analyses the explanatory power in profitability, but in this paper, defined as operating profits (Sales-COGS-SG&A) scaled by total assets. Novy-Marx finds that operating profitability, as well as gross profitability, have robust explanatory power with regards to forecasting expected returns. Also, its shown that factor investing strategies that are labeled as "defensive" by overweighting "safe" and "defensive" stock (often defined by stock's volatility or beta) tend to be tilted towards high profitability stocks. Besides that, value and volatility seem to be negatively correlated with profitability, which makes profitability an arguable risk adjusting strategy for value and volatility loaded portfolios. Even though this paper mainly focuses on explaining the low volatility factor, it does show the predictive power in operating profitability; a metric commonly used to proxy firm quality.

2.1.4 Fama & French (2015)

Based on Novy-Marx's (2013) and Titman, Wei, and Xie's (2004) arguments, that the previous Fama & French three-factor model failed to capture much of the return variations, Fama and French decided to improve their old model. Based on Novy-Marx's (2013) findings, they decide to add operating profitability and investments as a proxy for earnings growth in their paper *A Five-Factor Asset Pricing Model*. Like Novy-Marx (2013), they use the Dividend Discount Model to argue that profitability should be associated with greater stock returns. They also argue for a negative relationship between stock returns and investments, defined as annual asset growth scaled by total assets. Fama & French (2015) forms portfolios of June year t, where profitability is measured in the fiscal year ending $t-1$ as revenues minus cost of goods sold, interest expense, and selling, general and administrative expense, all divided by book equity at the end of fiscal year $t-1$. Even though their five-factor model (including market, size and value factors) does not fully explain the expected stock returns, the model manages to capture between 71% to 94% of the cross-section variance of expected returns. Hence, this model is an improvement of their previous three-factor model (including market, size and value factors). They further find signs that, when controlling for profitability and investments, the value factor becomes redundant.

2.1.5 Ball et al. (2015)

Building on previous work that relates superior stock returns to various profitability factors, Ball, Gerakos, Linnaimaa, and Nikolaev's 2015 paper, *Deflating Profitability*, shows that the quality measure operating profitability outperforms other profitability factors such as gross profitability and net income. Similar to earlier work, Ball et al. (2015) use both cross-sectional analysis and portfolio sorting in their empirical work. They also show that gross profitability outperforms net income since net income is usually deflated by the book value of equity, while gross profitability uses book value of assets to deflate gross profits. They conclude that operating profitability predicts returns as far as ten years ahead. Hence, this paper further solidifies that profitability metrics, commonly used as quality proxies, have significant explanatory power when it comes to stock returns. However, they further show that operating profitability outperforms gross profitability in their sample.

2.1.6 Kyosev et al. (2016)

Does Earnings Growth Drive the Quality Premiuim by Kyosev, Hanauer, Huij, and Lansdorp (2016) sets out to analyze the power of some of the more commonly used accounting-based factors, "...also referred to as quality variables". Thus, it focuses on further evaluating the predictability power in some of the more common definitions of quality. The analysis is conducted using three steps. First, the authors decide on which factors to analyze based on their power in previous research. Secondly, Fama-MacBeth regressions are used to test the explanatory power of the metrics. As a third and final step, the performance of portfolios sorted on their quality metrics is analyzed. Based on the first step, the authors decide to focus on ROE, ROE growth, margins, earnings variability, gross profitability, accruals, and total asset growth. In the second and third steps, they find that quality metrics that are associated with future growth, namely, gross profitability, accruals, and investments all outperform the total market portfolio. However, quality metrics that are not related to future growth; ROE, ROE growth, leverage, and earnings variability, fails to generate significant outperformance. Also, they find that a combined quality factor (gross profitability, accruals, and investments) performs better than the stand-alone factors, with a 6% annual alpha above Carhart's four factors. These results differ slightly from some other previous research. For example, Hou et al. (2012) find that ROE does have significant predictive power and Asness et al. (2018) find leverage and earnings variability to have significant predictive power.

Further, the paper tries to explain why the quality factor has predictability power regarding future stock performance. They find that "The potential predictive power of quality measures for stock returns can be fully attributed to their predictive power for future earnings growth" and is thus a result of mispricing rather than quality being associated with greater risk.

2.1.7 Bouchaud et al. (2016)

Similarly to Kyosev et al. (2016), *The Excess Returns of "Quality" Stocks: A Behavioral Anomaly* by Bouchaud, Ciliberti, Landier, Simon, and Thesman analyses the causes of the quality factor, which they argue to be one of the strongest anomalies in the equity market. They test whether the higher return of high-quality stocks is a result of their riskier behavior or if it is a result of an underestimation of quality stock's true value and they find evidence in favor of the latter. Quality is here defined as; return on assets (EBIT/Total assets), return on equity (Net Income/Common equity), and cash flow to assets (net operating cash flows/total assets). One argument they use to explain their findings is that investors are over-optimistic and that this bias is a result of investors being too earnings per share focused and overlooking cash flow and profitability factors, such as return on equity, return on assets, and cash flow to assets.

2.1.8 Hsu et al. (2017)

Hsu, Kalesnik, and Kose's paper, *What is Quality*, is one of the first papers that deep dives into the wide variety of ways used to construct the quality factor. Like the argument made by Kyosev et al. (2016), the authors argue that the quality factor lacks a commonly accepted definition. They analyze the performance of the most common definitions used by practitioners, such as index providers. They identify a total of 35 quality definitions on the market. These are then divided into seven groups based on their types; Growth in profitability, Profitability, Accounting quality, Payout/Dilution, Investment, Capital structure, and Earnings stability. They test factor robustness by 1) estimating the factor premium using data from various non-US regions, and 2) perturbing the definitions for constructing the factor portfolio. Based on this procedure, they show that profitability, accounting quality, payout/dilution, and investment tend to be associated with premia; further, profitability and investment-related characteristics tend to capture most of the quality-related premia. Capital structure, earnings stability, and growth in profitability show little evidence of premia.

2.1.9 Asness et al. (2018)

Asness, Frazzini, and Pedersen evaluate the quality factor's ability to forecast and explain abnormal returns in their paper *Quality Minus Junk* (2018). Much in line with other prior work, they base their analysis on a theoretical model. Based on the Gordon Growth Model, they derive a dynamic model of firm quality with time-varying profits, growth and risk. Based on this, they show that stock, in theory, should command a higher stock price if profitability or growth is higher or if the stock is safer. Thus. they defined high-quality stocks as having;

- High profitability, measured as gross profitability, return on equity, return on assets, cash flow over assets, gross margin, and low accruals.
- High growth, namely, five-year growth in residual per-share profitability (ex, accruals)
- Safe, namely, low beta, low leverage, low bankruptcy risk, and low return on equity volatility.

Profitability, growth, and safety, as well as a combined measure of these, all show a clear association with higher stock prices in the cross-section. This holds over a broad and wide sample, across industries and market cycles. They find a monthly average profitability premium of 40 basis points in the US. Quality-sorted portfolios have an increase in excess return that scales monotonically as they go from "junk" to "quality" and significantly larger risk-adjusted return. They find a monthly alpha of 53 basis points in the US that is not explained by any of the conventional factors, market, value, size, momentum, and investment. Quality-stocks also seem to be associated with a lower risk in terms of beta and tend to perform well during periods of market distress. Apart from this, they present evidence that points towards analyst's bias being a favorable explanation for quality mispricing. Even though this paper presents strong evidence of the existence of a quality anomaly, it does not test whether the various quality specifications vary in its explanatory power across industries.

2.1.10 Frazzini et al. (2018)

Frazzini, Kabiller, and Pedersen's paper *Buffet's Alpha* (2018) aims to explain the performance of Berkshire Hattaway using, amongst others, the quality factor. With an average annual excess return of 18,6% compared to the 7,5% of the market and a Sharpe ratio of 0,79 compared to 0,49 of the market, Berkshire Hattaway has managed to outperform the market since 1976. Frazzini et al. (2018) show that the conventional factors, market, size, and momentum, fails to explain Berkshire's alpha. However, when controlling for value and quality, most of the alpha was explained. They further show that one general feature of Buffet's portfolio is that he invests in high-quality stocks with high profitability that are stable, growing, and with high payout ratios. This, yet again, shows that quality, in terms of Quality-minus-Junk (Asness et al., 2018), does have explanatory power when it comes to predicting future stock returns.

2.1.11 Vyas & van Baren (2019)

Vyas and van Baren's 2019 paper *A Novel Template for Understanding Priced Factors* aims to analyze industry differences and their effect on factor investing strategies. The authors argue that there is substantial heterogeneity in stock characteristics for stocks in the same industry. They argue that this needs to be controlled for when investing based on factors, such as value and quality. If an investment decision is made based on predetermined factors, then the investor wants to invest based on the factors, assuming all else equal. However, if sector differences are not taken into account, we cannot assume all else equal, and the result can be that a factor investing strategy becomes an industry bet. Vyas and van Baren (2019) construct a methodology to measure whether a factor is priced within or across sectors. Unlike the value factor, the quality factor is not as clearly priced within industries, and different factor definitions vary in the potential for earning within- and across-industry premiums. ROA and ROE seem to be priced across-industry, while asset growth and growth in profitability is priced within industries. Gross profitability, equity issuance, and ROIC are found to be priced both within and across industries. Vyas and van Baren (2019) argue that the within-industry premiums are a result of the quality factors being subject to heterogeneity across industries.

2.2 Previous research on industry variances in equity returns

2.2.1 Baca et al. (2019)

Baca, Garbe, and Weiss's 2019 paper *The Rise of Sector Effects in Major Equity Markets* quantifies the industry-specific effects on equity returns and compares those to the country-specific effects. By studying ten major industries across seven large equity markets, over 20 years starting in the '90s, they manage to quantify how much of the equity return variance stems from industry versus country-specific effects. They find that the gap between pure country returns and pure industry returns has reduced significantly since the early '90s. In the later years, the importance of industry effects has been roughly equal to country effects, where industry effects stood for roughly 10% of return variance. In comparison, country effects stood for roughly 12%. This shows that industry-specific characteristics contribute more to equity return variation than previously believed. They argue that the increased globalization of economies and more integrated capital markets are the main reasons why the country-specific effects are becoming less important while industry-specific effects are growing in importance. This is a prime example of why it is essential to include an industry perspective when analyzing factor investing strategies.

2.2.2 Cheng et al. (2019)

Similarly to Baca et al. (2019), *Sector Effects in Developed vs. Emerging Markets* by Cheng, Bennett, and Zheng (2019) analyze the impact of pure country-specific effects and pure industry-specific effects on the mean absolute deviation of equity returns across emerging and developed markets. They find that the pure industry effects on equity return variance, depending on measurement techniques, have either surpassed or is closing in on the pure country-specific effect. They argue that this is not due to a fall in the country-specific effects, rather a result of the industry-specific becoming more important due to the rapid growth of the telecommunications and technology industries. Based on this, the authors argue that investors cannot afford to ignore the industry effect when deploying capital in developed countries. This is another example of why it could be necessary for investors to consider an industry perspective when analyzing factor investing strategies.

2.3 Summary of previous research

The research supporting quality as an alpha-generating style factor is unanimous. However, the ways that quality is defined and quantified are varying, with measures ranging from profitability, to capital structure, to return metrics, to earnings variability, and more. Thus, research has failed to establish a single set of superior quality metrics.

Also, research has managed to show that quality metrics can be priced within industries, which tells us that accounting for the industry might be useful when choosing how to quantify quality. Besides that, research has also shown that a growing part of equity returns variability stems from industry differences and that investors should not ignore industry differences when investing.

In similarity to Kyosev et al. (2016) and Hsu et al. (2017), we conduct our initial factor screening approach based on previous research. In other words, we will decide on what quality metrics to focus on by analyzing which quality metrics have shown to be promising on earlier research.

Like Novy-Marx (2013) and Kyosev et al. (2016), we will also use regression methodologies to analyze the ability of our quality metrics to explain stock returns in a multifactor setting.

To investigate if eventual industry heterogeneities can influence portfolio return premiums, we will use a portfolio sorting methodology, a commonly used technique in asset pricing research. A few papers that have used this technique to analyze the quality metric are Hou et al. (2012), Fama and French (2014), and Asness et al. (2018).

However, our thesis will include an industry perspective, a perspective that has not been thoroughly analyzed in previous quality investing research. The idea to focus on different industries is an approach that, to some extent, is based on the findings of Vyas and van Baren (2019), Baca et al. (2019), and Cheng et al. (2019).

2.4 Contribution to previous research

As can be seen by reviewing previous research regarding the quality factor anomaly is that there seems to be a consensus about the existence of the anomaly. Different methods have been used to evaluate the pricing power of quality. Tests have been done over various markets and through different periods in time. Nevertheless, the conclusion that can be drawn from academia is that high-quality stocks outperformance their low-quality peers.

Most research has focused on analyzing the quality factors persistence over time and across markets. Little research has been covering the quality anomaly across industries or sectors. Also, most research has focused on proving the existence of a quality anomaly given a specific definition of quality. This thesis aims to contribute to research by shedding some light on whether the chosen quality metric matters across industry sectors. We also argue that our discoveries could contribute to investor's ability to implement quality into their investment strategies optimally.

3 Background on factor investing and the quality factor

3.1 Factor investing: What is it, where does it come from, and how is it used?

Factor investing, or style investing, has grown to become one of the cornerstones of modern portfolio strategy, with 70% of institutional investors using factor investing as a part of their investing strategy (Warren & Quance, 2019). According to Fidelity (Nielson et al., 2016), by 2016, factor investing styles had generated net capital inflows of nearly $250 billion over a five-year period, and The Economist (2018) estimated the amount of money invested in factor strategies in 2018 to $1 trillion.

The idea that stock returns are driven by underlying exposure to various factors was, to some extent, initiated by the Capital Asset Pricing Theory. CAPM argued that stock performance was relying on the asset's exposure to market-wide risk. This idea was later on extended to allow for multifactor exposure by Ross in his Arbitrage Pricing Theory. Based on the findings of Ross, and other empirical findings, more factors emerged that proved to have predictive power with regards to stock returns, one of which is quality. The more prominent and recognized factors will be further discussed in this thesis.

Then, what is factor investing, and why is it popular? BlackRock (2020) explains factor investing as "an investment approach that involves targeting specific drivers of return across asset classes," and Invesco explains it as "… an investment strategy in which securities are chosen based on certain characteristics with the goal of achieving a given investment outcome or to improve long-term risk and return."(Warren & Quance, 2019). Another way to explain a factor is that it is a 'characteristics of a security that is shared with other securities', or, as senior portfolio manager Nick Zylkowski explains it, one can compare factors to how films are divided into genres (Russell Investments, 2018). Hence, the aim of factor investing strategies is to identify firm or stock characteristics that are associated with future good performance. By constructing portfolios that invest in stock with these characteristics, investors can capture future expected excess returns and achieve superior portfolio performance. A point worth mentioning is that, even though factor investing often is associated with equity investing, these strategies are becoming more common for other asset classes as well, such as fixed income (BNP Paribas Asset Management, 2020).

Unlike traditional stock-picking strategies, factor investing strategies are implemented in a more systematic and rule-based manner. Stock picking uses a narrower approach where investors leverage their unique skills or information to identify good investments. Factor investing is based on a set of rules regarding asset characteristics and targets a broader set of potential investments. Factor investing also relies more heavily on the factor or the set of factor characteristics that the investor sets up and their ability to find investments. Stock-picking tends to involve more case-to-case analysis by the investor. These differences have also led factor-based strategies to not only be cheaper than actively managed stock picking strategies but also to be more transparent (Russell Investments, 2018).

Factors are typically divided into two groups, macro factors and style factors. Macro factors relate to the effect of macro related fluctuations on asset prices, such as economic growth, inflation, and interest rate fluctuations. Macro factors can work as predictive and risk management tools in terms of understanding how security prices respond to market-wide economic fluctuations. However, this thesis will not cover macro factors.

The second category is the so-called style factors. These are factors that are based on firm-specific characteristics. Style factors have also become what one tends to refer to when mentioning factor investing these days. The most known and used factors today are size, value, momentum, and quality[1]. These factors themselves are not directly measurable; all of them need to be proxied. We will go through these four conventional factors in more detail below.

3.1.1 The value factor

Value is one of the first factors that was widely accepted as an extension of the conventional CAPM. Value is a part of Fama and French's famous three-factor model, alongside the size and the market factor. The idea behind investing in value is that investing in assets that are priced at a discount will, in the long run, generate abnormal returns. Research has shown that cheaper stock, or value stock, do generate higher returns than expensive stock, or growth stock. Value stocks are stocks that are traded at "a low multiple" or, in other words, a stock that is traded at a value below its intrinsic value. Growth stocks are traded at higher multiples, often due to its future potential (Penman & Reggiani, 2018). The definition of value does vary somewhat amongst investors and academia. Book to market ratio is the most conventional definition, but dividend yield and price to earnings ratio are also used (Warren & Quance, 2019). Fama

[1] The market factor, beta, is also a factor that relates to expected stock returns. However, the market factor is usually seen as a benchmark that investors try to beat and is therefore not considered a style factor.

and French (1993) show that value, along with size, has significant explanatory power on returns above the CAPM market factor. One argument for why value investing works is that value firms tend to be capital heavy, which allows them to recover quickly during an economic recovery. However, value has shown to underperform during periods of economic slowdown. (Framsted, 2019).

3.1.2 The size factor

The size factor aims to capture the overperformance made by small firms compared to that of medium and large firms. The ability for smaller firms to outperform their larger counterparts were identified by Rolf W. Banz in his 1981 paper *The relationship between return and market value of common stocks*. Banz (1981) shows that smaller firms, measured by market capitalization, had, on average, a higher risk-adjusted return compared to larger firms over a 40-year period. He also saw that most of the difference in return came from very small firms, while above-average sized firms showed little difference to large firms. Banz did raise the important point that size might not perform due to size per se, but due to its correlation with other factors, such as risk. Along with value, size proved to have significant explanatory power over the market factor in Fama and French's three-factor model (Fama & French, 1993).

3.1.3 The momentum factor

Momentum builds on the idea that previous outperformers will continue outperforming in the future. Momentum has little to no theoretical foundation and is yet to be explained. However, behavioral finance theories such as herd mentality and anchor bias seem to be the best theories so far. Momentum builds on Jegadeesh and Titman's (1993) findings that stock performance seemed to prevail over multiple periods. A stock that had performed well in previous periods was prone to perform well in the near future, and stock that had performed poorly in previous periods was prone to continue to perform badly in the near future. Momentum was incorporated into Carhart's four-factor model where it showed to have significant explanatory power despite controlling for market, size, and value effects. Past12-month return is the most conventionally used proxy for momentum; however, 6- and 3-month return are used as well. (Warren & Quance, 2019).

3.1.4 The quality factor

The quality factor is based on the fundamental idea that companies with high quality should outperform companies with lower quality. The quality factor is often based on a series of accounting-based factors, such as various profitability measures, leverage ratios, and asset growth ratios. These accounting measures are often seen as important measures of a firm's quality, and thus the name "quality factors". The asset management firm, Schroders, argues that high-quality firms are able to generate a return premium since they can keep their competitive advantage longer compared to low-quality firms (Schroder, 2017).

In similarity to the previously mentioned factors, value, size, momentum, the quality factor stems from the Arbitrage Pricing Theory. In the same sense that market, value, size, and momentum are factors in a k-factor model, quality is as well. In the same sense that value, size, and momentum are factors designed to explain pricing anomalies or to capture risk premium from non-diversifiable risk not captured in previous factor models, quality is too. However, regardless of the similarities between the quality factor and other conventional factors, there are vast differences.

The main things that distinguish the quality factor from the four conventional asset pricing factors are; the wide variety of ways quality can be defined, and, the lack of a common quality definition. However, one could argue that this problem exists for other factors as well. Value can be defined in more ways than just book to market ratio, for example, earnings to price and dividends to price, and, similarly, one could argue that momentum can be defined as more than 12-months price performance. Nevertheless, the factor defining problem for value and momentum is dwarfed compared to that of quality. Hsu et al. (2018) visualizes the quality definition problem quite well as they identify 35 metrics used by practitioners to define quality. Another thing that differentiates quality as an investment strategy from other factor strategies, such as size and value, is that the quality investing strategies offered by practitioners are, in most cases, consisting of a composition of many quality metrics. So, more than one quality metric is used when evaluating a firm's level of quality (Hsu et al. 2018).

It is not only practitioners that have a hard time agreeing on the definition of quality, academia as well. As can be seen in *chapter 2*, "Previous research", the definition of a quality-capturing factor is not clear cut. As stated by Hsu et al. (2018), one of the more common ways of measuring quality is by looking at profitability. Novy-Marx (2013, 2016), Fama and French (2014) and Ball et al. (2015) define profitability as operating profitability. Hou et al. (2012) overlooks operating profitability in their q-factor model and focuses on investments (asset growth) and return on equity instead. Asness et al. (2018) extends their quality definition and, on top of profitability, include growth in profitability and safety as criteria in their

QMJ-factor. As can be seen in the previous research-section, more than the definitions mentioned here are used as proxies for quality.

3.2 Theoretical soundness of the quality factor

Then, what is it with firm quality that makes it capable of predicting future returns? Most research has pointed towards the quality factor's ability to predict future economic profitability. Kyosev et al. (2016) argue that predictive power in quality comes from its power in predicting true economic profitability. More specifically, they show that predictive power in quality metrics can be attributed to their ability to forecast earnings growth. This is an argument that Sloan (1996), Novy-Marx (2013), and Akbas et al. (2017) also make. Novy-Marx (2013) and Fama and French (2015) both use a theoretical framework, namely the Dividend Discount Model (further explained in *section 4.2*), to argue for why their metrics of quality should translate to superior returns. Fama and French (2015) show that there is a positive relationship between a firm's profitability, defined as return on equity (ROE), and its value and a negative relationship between a firm's investments and value. Asness et al. (2018) show that profitability, as well as firm growth and firm safety, should be related to greater firm value using the Gordon Growth Model (further explained in *chapter 4.3*). Hou et al. (2012) also use a theoretical approach to argue for their use of ROE and investments (asset growth) as return predicting factors using the relationship between ROE and discount rates and investments and the net present value of assets. Hence, the quality factor has, to a large extent, a theoretical foundation to stand on, unlike factors such as momentum.

An optional theory could be that the relationship between high quality and future stock performance is simply a result of the quality premium being correlated with higher distress risk. Although it might seem counterintuitive that high-quality stocks would be riskier than their low-quality counterparts, it could be possible for the quality premium to be more common amongst high-risk firms. One argument for this theory is that high-quality firms (measured in profitability) are frequently taking on more profitable, but also riskier, projects compared to low-quality companies. High-quality firms are therefore risker to own from an investor perspective since they are operating in a more profitable but also riskier segment of the industry (Bouchaud et al., 2016). However, Kyosev et al. (2016) found that "The potential predictive power of quality measures for stock returns can be fully attributed to their predictive power for future earnings growth" which speak against the theory that the quality premium is a risk premium. Bouchaud et al. (2016) also analyze the causes of the quality factor. More precisely, whether the higher return of high-quality stocks is a result of their riskier behavior, or if it is a result of an underestimation of quality stocks true value. Like other scholars, they found evidence of the latter. They argue that the quality premium is a result of investors being over-optimistic and that this bias is a result of investors overlooking factors such as profitability in favor of earnings per share ratios.

3.3 Practical use of the quality factor

As the case with other conventional asset pricing factors, such as size, value, and momentum, the quality factor has become a commonly used ingredient in portfolio solutions offered by institutional investors. Hsu et al. (2018) argue that an increasing amount of index providers are now offering quality-based investment strategies or multifactor investment solutions, where quality is one of the strategies used to screen for possible investment solutions. They identify six index providers using quality as part of their investment strategy. Besides index providers, quality strategies are also a reoccurring trait in ETFs (exchange-traded funds) and hedge funds. Norges Bank (2015) also argues that quality has become increasingly common as a systematic investment style by practitioners in recent years. Also, Invesco Global Factor Investing Study found that over 70% of institutional investors used factor investing strategies in 2018, including quality strategies, and over 60% were planning to increase their use (Quance, 2018).

Some of the more well-known index providers that provide quality-based indices are MSCI, FTSE, S&P, and Deutsche Bank. Hsu et al. (2018) identify that the variation in the specification of quality is large between these index providers. With their findings of a total of 35 different quality-based metrics used by practitioners, it is quite clear that there is close to no consensus regarding how to measure quality. Many other authors, academics, as well as practitioners, also recognized this phenomenon. This, of course, raises questions. How can it be that academia and institutional investors fail to agree on an optimal set of quality measures? It also complicates the evaluation process of different index providers and institutional investors since it is hard to know if good performance is due to good portfolio management or a direct result of their specific definition of quality. Even though these questions are outside the scope of this thesis, they show why the quality factor as a phenomenon deserves further investigation.

Some of the larger flag-bearers of the quality factor as investment strategy are Exchange Traded Funds (ETFs). ETFs are funds that, similarly to stocks, are traded on exchanges, and by 2016, factor investing ETFs had more than $800 billion in assets under management (Li et al. 2019). Some of the larger quality investing ETFs are; iShares Edge MSCI USA Quality Factor, JPMorgan US Quality Factor, WisdomTree US Quality Sharehld Yld, and Invesco S&P 500 Quality. These funds all vary concerning how they define high-quality stocks. Reoccurring quality metrics for these funds are ROE, ROA, earnings growth, leverage, and profitability (Shriber, 2019).

3.4 Other style factors

The three above mentioned factors, along with quality, are the more recognized and commonly used factors. Besides these, low volatility is also considered by some to be a widely used factor. However, it has not been integrated into a well-known and accepted multifactor model such as the three- or the four-factor models of Fama-French and Carhart or the q-factor model of Hou, Xue, and Zhang.

However, there are plenty of factors out there besides the more conventional ones—some with backup in research, some without. As pointed out by Feng et al. (2019), finance research has provided academia with over 100 potential style factors that, to some extent, affect returns in the cross-section. However, they argue that a significant share of these factors are redundant when tested in a multifactor setting, such as the three- or the four-factor models of Fama-French or Carhart. In Harvey, Liu, and Zhu's 2016 paper *...and the Cross-Section of Expected Returns*, they identify 316 factors published in top journals and working papers. This further proves the vast universe of potential style factors. However, these factors will not be covered in this thesis.

3.5 Multifactor strategies

Multifactor strategies are exactly what they sound like; strategies that combine the power of various factors into one joint factor strategy. The goal with multifactor strategies is to capture, not only one factor's ability to forecast performance but a few. The more obvious reason for why multifactor strategies are popular is because it allows for investors to capture multiple pricing anomalies simultaneously. Furthermore, there are possible synergy effects between factors in a multifactor setting that does not exist when investing using a set of single-factor strategies. One example is the fact that different factors tend to perform well during various periods in the economic cycle. For example, all conventional factors have performed well over a longer horizon, but all of them have also faced periods of lousy performance. Factors also tend to have a relatively low correlation, or in some cases, even negative correlation. Multifactor strategies can exploit this by moving in and out of factors that, at that time, should perform well. Thus, a combination of different factors allows for portfolio diversification, similar to when investors combine different asset classes to diversify and reduce portfolio risk. Therefore, multifactor strategies do not only allow for the possibility to achieve better returns but also to lower the risk and thus achieve better risk-adjusted performance. When comparing a multifactor strategy that is based on quality, value, momentum, and low volatility, to their respective single factor strategies, the multifactor strategy fared better than all single factor strategies on a risk-adjusted basis (Innes, 2018).

Multifactor strategies can be constructed in numerous ways and adjusted based on individual investor preferences. This includes things such as; deciding on which factors to include, how much each factor should contribute, and the appropriate rebalance frequency. The more famous multifactor models are Fama and French's three and five-factor models, Carhart's four-factor model, and Hou, Xue & Zhang's q-factor model.

Just like factor investing has grown in popularity over the years, multifactor strategies have as well. Multifactor strategies have grown from having $3.8 billion in assets as of 2009 to $70 billion in assets in 2018. In 2018, 197 ETFs were offering multifactor exposure (Ang & Framsted, 2020).

3.6 Transaction costs

One of the delimitations of this thesis (discussed in *section 1.2*) is the absence of transaction costs in our analysis. A transaction cost is a cost that occurs when an investor sells or buys an asset. Logically, this cost reduces the actual return generated by an investor, and in some cases, it can be the actual difference between an investment strategy being successful or not.

When it comes to transaction costs, two factors are affecting; asset turnover volume and the actual costs associated with selling or buying an asset, such as commissions, taxes, and ticker charges. Turnover volume can be defined as the number of transactions that need to be done to put an investment strategy into practice. Since turnover volume is the main driver behind transaction costs, investment strategies that require frequent and extensive portfolio rebalancing generate higher costs. Turnover volume is also affected by how often investors decide to rebalance their portfolio. For long-short portfolios with monthly rebalancing, momentum has the highest turnover rates, while quality and size have the lowest turnover volumes amongst the conventional factors. For long-short quality strategies, there is a fairly equal division of turnover between the legs, with the short leg having a slightly larger turnover ratio (Rabener, 2018a). Rabener (2018b) also finds that the turnover ratio of quality strategies are relatively unaffected by the rebalancing frequency due to the infrequent disclosure of new fundamental financial company data.

However, Li et al. (2019) argue that, for larger portfolios, there exists an indirect transaction cost as well. This cost originates from the fact that when larger portfolios rebalance their asset allocations, a large purchase of an asset might drive up the current asset price, resulting in lower returns.

3.7 Industry differences

The existence of country-specific effects on equity behavior has been a fact for a while. This has affected investors' decision making when it comes to deployment of capital, diversification methodologies, and investment strategies. Academia on quality factors, as well as on other return predicting factors, tend to claim robustness in their tests by controlling for country-specific effects; however, they tend to neglect industry-specific differences.

Research has established quite robust results showing the relative importance of the industry effect in equity markets. Baca et al. (2019) show that roughly 10% of equity return variability can be attributed to industry differences and that the industry effect is steadily becoming as important as the country effect. Chen et al. (2019) find similar results and further argues that investors should not neglect the industry effect when deploying capital or when setting up investing strategies. Phylaktis and Xia (2007) also found that the industry effect had had a major upswing since 1999 and that North America was one of the regions in which this upswing was more prominent. They also showed that the increasing importance of industry effects seemed not to be a temporary phenomenon.

Since the industry-specific effect is growing, and the recommendations for investors is to take this into account when developing investment strategies, one could argue that industry differences in factor investing strategies could be of interest for investors. Since there exists a vast amount of ways to define the quality factor, it could be useful for quality-investors to understand if the optimal way of constructing the quality factor varies across industries. For example, if an investor wants to diversify away some of the industry-specific risks, he/she needs to allocate capital across industries. If this investor is a quality investor, that is, he/she uses quality as a way to determine what stock to invest in, it is useful for him/her to know if different quality metrics have varying predictive power in different industries. If industry differences with regards to the quality factor exist, it might be the case that the investor could develop superior stock selecting strategies by taking industry heterogeneity into account. Similarly, Vyas and van Baren (2019) argue that factor investors should take industry heterogeneity into account if not, the result can be that a factor investing strategy merely becomes an industry bet.

4 Theory

This section will cover, in a more detailed manner, the main theories and concepts that this thesis is built upon. This includes fundamental asset pricing theories and various corporate finance concepts. Following this chapter, *section 7* will demonstrate how these theories are put into practice in order to investigate our hypotheses empirically.

4.1 Return calculations

Deriving the returns based on price fluctuations is one of the more obvious steps when it comes to analyzing equity performance. However, how to derive equity returns most optimally is not always as clear. The simple approach is to solely focus on returns in terms of appreciations or deprecations in stock prices. This can be mathematically expressed as;

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

R_t represents the return for period t, P_t and P_{t-1} are stock prices at time t and $t\text{-}1$, respectively.

This captures most of the stock returns for stocks that pay little or no dividend or that conducts few or no stock buybacks. However, when a company redistributes a large amount of its wealth in terms of dividend or stock buybacks back to shareholders, the simple return formula misses a large part of the actual return that investors get. To capture investor returns stemming from redistributions, a formula that includes dividends and buybacks is needed. To bridge the gap between returns from price fluctuations and actual returns, including redistributions, we use DataStream's Total Return Index. This index shows "...a theoretical growth in value of a share held over a specified period, assuming that dividends are reinvested to purchase additional units of an equity or unit trust at the closing price applicable on the ex-dividend date." (Eikon's DataStream, 2020). In other words, this index aims to capture the total return to investors from both price fluctuations and dividend distributions. The Total Return Index (RI), is mathematically defined as;

$$RI_t = RI_{t-1} * \frac{P_t}{P_{t-1}}$$

if t is a date when dividend is paid out then;

$$RI_t = RI_{t-1} * \frac{P_t + D_t}{P_{t-1}}$$

Where RI is the total return index, P is the stock price, and D is the actual gross dividend paid. t and $t-1$ are time periods.

However, since the total return index does not take stock buybacks into account, it is not a perfect measure of true return. However, it captures more of the actual return compared to merely focusing on price fluctuations. Another consideration worth discussing is that the total return index assumes that dividends are reinvested into the stock by the investors. This is a vague assumption since some investors prefer to hold divided as cash and possibly reinvest it at a later stage. The total return index also does not take tax on dividends or re-investment costs into account, which makes it less realistic. However, despite its drawbacks, we believe that the total return index is a more accurate measure of actual returns compared to only focusing on stock price fluctuations.

Since the total return index is an index, it is not in itself a measure of returns. The change in the total return index is used to calculate returns.

$$R_t = \frac{RI_t - RI_{t-1}}{RI_{t-1}}$$

Where R_t represents the total return to investors, including changes in stock price and payments of dividends.

4.2 The Dividend Discount Model

The well-known equity valuation model, the Dividend Discount Model, is used by Novy-Marx (2013) and Fama and French (2014) to rationalize the theoretical soundness of some of the most common quality metrics, namely the profitability and investment metrics. Since these metrics are analyzed in this thesis, the Dividend Discount Model is equally important for this thesis. The Dividend Discount Model builds on the theory that the value of a stock to an investor is based on the expected future cash flow of that stock. Future cash flow includes both the future expected dividend and the future expected price at which the stock can be sold. However, an assumption made in the Dividend Discount Model is that the price at which the stock can be sold in the future, itself depends on the expected dividend forecasted at that time. Thus, in the case of a predetermined holding period, the Dividend Discount Model can be written as;

$$V_0 = \frac{E(D_1)}{1+k} + \frac{E(D_2)}{1+k^2} + \cdots + \frac{E(D_h) + E(P_h)}{1+k^h}$$

where V_0 is the stock price at time 0, D_1 is the expected dividend at time 1, and k is (approximately) the long-term average expected stock return, or more precisely, the internal rate of return on expected dividends. P_h is the present value, at time h, of all expected future dividends (Bodie et al., 2014).

Whereas if the holding period runs in perpetuity, the Dividend Discount Model can be written as;

$$V_0 = \frac{E(D_1)}{1+k} + \frac{E(D_2)}{1+k^2} + \cdots$$

The Dividend Discount Model says that if at time t, two stocks have the same expected dividends but varying prices, the expected long-term average return should be higher for the stock with a lower price (Fama & French, 2015). Further, Fama and French (2015) shows that the Dividend Discount Model can be rewritten as;

$$\frac{M_t}{B_t} = \frac{\sum_{t=1}^{\infty} E(Y_{t+\tau} - d\beta_{t+\tau})/(1+r)^t}{B_t}$$

where $Y_{t+\tau}$ is the total equity earnings for period $t+\tau$ and $d\beta_{t+\tau}$ is the change in total book equity. M_t is the current market value of the stock.

Through this formula, some interesting relationships can be derived. As mentioned in *section 3.2*, Novy-Marx, as well as Fama and French, use these to rationalize their choice of investments and profitability as additional factors. First, assuming all else equal, a lower current stock price indicates a higher expected return. Secondly, higher expected earnings implied a higher expected return. Put merely, profitable firms should, in theory, outperform unprofitable firms. Finally, higher expected growth in investments implies a lower expected return. These three relationships tell us that M_t, current stock market price, responds to forecast of earnings and investments (Fama & French, 2015).

4.3 The Gordon Growth Model

A version of the Dividend Discount Model is the Gordon Growth Model. The formula has come to be one of the most common tools in fundamental valuation models, such as the discounted cash flow model. The Gordon Growth Model is one of the more straight forward ways of determining a firm or a stock's value and is useful in cases where the investor expects the dividend to have stable growth in perpetuity. The Gordon Growth Model is written as the following;

$$P_t = \frac{D_t(1+g)}{r-g}$$

where P_t is the stock price at time t, D_t is dividend at time t, g is the expected dividend growth rate, and r is the cost of capital or discount rate (Pages, 1999).

Just as Novy-Marx (2013) and Fama and French (2014) used the Dividend Discount Model to rationalize their decision to use profitability and investments as factors in their factor models, Asness et al. (2018) use the Gordon Growth Model to rationalize their choice of factors to use as proxies for firm quality. They show that the Gordon Growth Model can be re-written as follows;

$$\frac{P}{B} = \frac{dividend}{required\ return - growth} = \frac{\frac{profit}{B} \times \frac{dividend}{profit}}{required\ return - growth} = \frac{profitability \times payout\ ratio}{required\ return - growth}$$

This relates prices (scaled by assets) to a few interesting variables, variables that Asness et al. (2018) argues are proxies for quality. They also argue that since prices and returns are linked, the price of quality can be used to predict the future return to quality factors. The extended Gordon Growth formula, or more precisely, the right-hand-side variables, profitability, required return (safety) and growth, forms the basis for their choice of quality metrics.

All else equal, more profitable companies should be valued higher. Further, a growing profit is also a characteristic that should lead to higher firm value. Finally, a stock with a lower internal rate of return, or as Asness et al. (2018) also defines it, a safer stock, should command a higher price.

4.4 Capital Asset Pricing Model (CAPM)

Capital asset pricing model, from now on referred to as CAPM, is a fundamental asset pricing model first introduced, independently, by Sharpe (1964), Lintner (1965b), Treynor (1961), and Mossin (1966). CAPM has been proved useful for a wide variety of tasks, where one of them is to assess portfolio performance. CAPM relies on a set of assumptions that, at best, are approximately true. These assumptions can be summarized to;

- Investors are rational, mean-variance optimizers. Hence, investors care only about expected asset returns and the standard deviation of these returns.
- Investors have one common single-period planning horizon.
- Investors have homogeneous expectations. Hence, they optimize their behavior based on the same information.
- Assets are publicly held and traded on public exchanges, short positions are allowed, and investors can borrow or lend at a common risk-free rate.
- All information is publicly available.
- There are no taxes.
- There are no transaction costs associated with selling or buying assets.

(Bodie et al., 2014)

CAPM relates the systematic market risk of an asset, also known as beta (β), to the expected return of the asset (Finch et al., 2011). CAPM predicts that an asset's risk premium (the additional marginal return demanded per unit of risk) results from a single risk factor, the return of the market portfolio of all risky assets. The theoretical CAPM model is given by the intuitive but highly useful formula;

$$E(R_i) = R_f + \beta_i\left[E(R_m) - R_f\right]$$

Where $E(R_i)$ is the expected return of asset *i*, R_f is the risk-free interest rate, and $E(R_m)$ is the expected return of the market portfolio, often approximated using a well-diversified portfolio such as S&P 500 (Hull, 2018). In practice, investors use ex-post values to estimate the model, then CAPM is written as;

$$R_{it}^{ex} = \alpha_i + \beta_i R_{mt}^{ex} + e_{it}$$

Where R_{mt}^{ex} is the ex-post excess values of the market portfolio, R_{it}^{ex} is the excess return at time t for asset i.. α_i is the intercept, which represents the unexplained part of asset $i's$ return and e_i is defined as an orthogonal, mean zero residual, also known as idiosyncratic and diversifiable risk (Pennacchi, 2018).

Since the idiosyncratic risk can be diversified away, investors are not compensated for taking on this risk. Compensation comes solely from the systematic risk, β (Mullins, 1982). The systematic risk of the given asset i, beta, is the non-diversifiable risk of an asset, and this is the risk that, according to the CAPM-model, is priced. Beta is defined as;

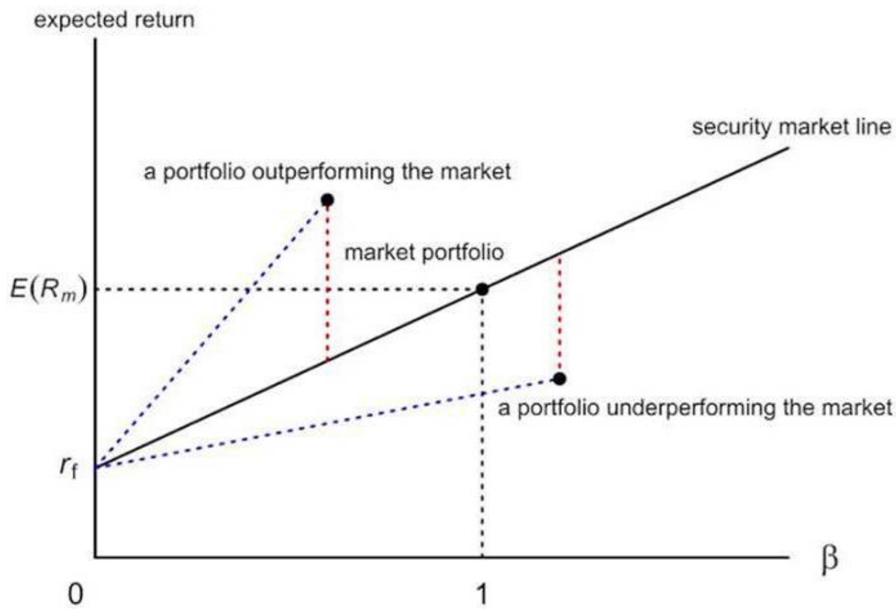
$$\beta_i = \frac{Cov(R_m, R_i)}{Var(R_m)}$$

(Tofallis, 2008)

Beta can also be estimated as the slope by regressing the excess return of the asset on the excess return of the market portfolio. Beta simply represents the co-movement of the given asset i, and the market portfolio. Theoretically, the beta of the market portfolio should be one. Assets with a beta larger than one can expect its return, above the risk-free rate, to, on average, exceed that of the market portfolio. Vice versa holds for assets with a beta below one.

This risk-return relation can be more clearly illustrated by the Security Market Line (SML). Following the reasoning behind CAPM, the expected return of an asset is a function of β_i, and the market risk premium, $R_m - R_f$. In the CAPM universe, all assets, if they priced correctly, should line up on a straight line, where the expected return is increasing monotonically with β_i.

Figure 1: Security Market Line.



Source: What is the security market line? Tickeron.com, Marsh 2020

Assets above the security market line would be considered undervalued according to the logic of CAPM. Investors would then increase their share in that asset – driving its price back up to its fair value and the security market line. Thus, assets below the security market line would be considered overvalued according to the logic of CAPM. Investors would then reduce their share in that asset – driving its price back down to its fair value and the security market line (Mullins, 1982)

Regardless of CAPMs popularity, the criticism does not run short. One of the more recognized criticisms is that of Roll (1977). Roll (1977) points out the rather obvious, that the market portfolio, on which the CAPM is based, is unobservable. The market portfolio should contain all assets in the economy, including such unobservable assets as, for example, privately owned jewelry. Authors also argued that the market factor was not enough to capture the risk-return relationship, which has lead to the introduction of new, alternative, equity price driving factors.

4.5 Arbitrage Pricing Theory (APT)

The arbitrage pricing theory (APT) was established in 1976 by Stephen Ross and is the leading theory behind a possible multi-factor model. Like CAPM, APT relates expected asset returns to asset pricing factors. However, one main difference between the two is related to how they define systematic risk. Ross (1976) argued that the return of a security was driven by two components, expected return and a surprise component. For individual stocks, the surprise component was depending on one systematic part that affects all stock but not equally much, and a stock-specific part that solely affected one given stock.

However, CAPM derives a single factor model, where the sole price-driving factor is the assets co-movement with the market portfolio, beta. APT allows for several factors that capture market-wide risk to affect asset prices. APT argues that the stochastic process that generates asset returns are expressed as a linear function of a set of K risk factors;

$$R_i = E(R_i) + \beta_{i1} F_1 + \cdots + \beta_{ik} F_k + e_i$$

(Cuthbertson, 1996).

Worth to mention is that APT does not suggest any particular factors in itself; it only facilitates a theoretical model that opens up for multi-factor asset pricing models. This has led to many possible asset pricing multifactor models.

APT is also based on a few assumptions;

- A factor model exists that can describe asset prices.
- Markets are efficient – eventual arbitrage opportunities are trade away effectively.
- Enough assets are available in order to diversify away idiosyncratic risk.

(Bodie et al., 2014)

We will describe these assumptions and their implications in further detail below.

4.5.1 The existence of a k-factor model

Bodie, Kane, and Marcus (2014) explains this as the idea that there exists a model, with an arbitrary amount of factors, k, such that the value of any asset can be explained. They argue that there are two sources of variability in stock returns, macro related risk, also known as systematic risk, and firm-specific variability, also known as idiosyncratic risk. The multifactor APT model can be described as;

$$R_i = E(R_i) + \beta_{i1}F_1 + \cdots + \beta_{ik}F_k + e_i$$

Where $E(R_i)$ is the expected excess return of stock i, e_i is the idiosyncratic risk component, F_1 to F_k are the factors, β_{i1} to β_{ik} are the stock's exposure to that specific factor, called factor-beta or coefficient. Both the systematic factors and the firm-specific ones have zero expected return. If the systematic factor has a value of 0 in any period, then the value of the asset is its previous expected value, $E(R_i)$, plus any shocks to idiosyncratic risk, e_i.

With a k-factor model that captures the exposure to each source of risk, stock prices can be explained. Similarly to the CAPM, the k-factor asset pricing model can be related to the Security Market Line. Each securities exposure to a risk factor is based on that factor's exposure to that risk times the risk premium of that particular factor (Bodie et al., 2014).

4.5.2 No arbitrage condition - Law of one price

One of the more powerful theorems in theoretical asset pricing, and an important proposition in the APT, is the law of one price (LoOP). LoOP states that if two portfolios have the same payoff in every state of nature, they must have the same price. Stated differently, LoOP says that investors cannot make instantaneous profits by repackaging portfolios. LoOP describes a market in equilibrium, and if the theorem is violated, investors will exploit this such that the opportunity disappears, and the market returns to equilibrium. If this does not hold, an arbitrage opportunity exists since it allows for investors to sell the expensive portfolio and buy the cheap portfolio and by doing that, earn a risk-free return (Cochrane, 2005).

The no-arbitrage condition states that one may not get a portfolio for free that might pay off positively. Thus, an arbitrage opportunity is one where an investor can earn a riskless profit without making a net investment. One example of an arbitrage opportunity could be if stock A costs $10 on NYSE but only $9 on NASDAQ. One could then sell the stock on NYSE and buy it on NASDAQ, earning a riskless profit of $1. This arbitrage opportunity is a result of the same asset having two different prices, thus a violation of the LoOP. The fact that the arbitrage opportunity arises as the LoOP is broken is also what makes the LoOP being withheld in equilibrium. When LoOP is broken, arbitrageurs will bid up the price of the stock on NASDAQ (where it was cheap) and push the price down on NYSE (where it was expensive) until the arbitrage opportunity is eliminated, thus bringing the market back to a state where the LoOP holds, also known as market equilibrium.

Together with the concept of diversification, the no-arbitrage condition ensures that the risk premium of a riskless and well-diversified portfolio is zero. Thus, the expected return on any well-diversified portfolio P is merely a product of its level of risk and the risk premium;

4.5.3 Diversification

The idea of diversification is a powerful and critical tool for most asset pricing and portfolio optimization theories, including the APT. Simply put, it can be shown that in a portfolio containing a large number of assets, the asset-specific risk can be reduced to neglectable as the number of assets grows large enough.

Bodie et al. (2014) shows that for an n-stock portfolio;

$$R_p = E(R_p) + \beta_p F + e_p$$

Where e_p is the portfolios non-systematic risk or asset-specific risk. β_p is the weighted average of β_i and risk-premiums of n securities, $\beta_p = \sum w_i \, \beta_i$. Similarly, e_p is the weighted average of e_i of the individual n stocks, $e_p = \sum w_i \, e_i$. They show that the portfolio variance can be separated into one systematic and one asset-specific part;

$$\sigma_p^2 = \beta_p^2 \sigma_F^2 + \sigma^2(e_p)$$

Where σ_F^2 is the variance of factor F, and $\sigma^2(e_p)$ is the non-systematic risk. This assumes that the asset-specific risk e_i is uncorrelated across assets. They further show that the portfolio variance is given by;

$$\sigma^2(e_p) = \frac{1}{n} \bar{\sigma}^2(e_i)$$

Thus, the portfolios asset-specific risk approaches zero as the number of assets in the portfolio, n, grows large. They further show that this holds for any portfolio where the weight of each asset shrinks as the number of included assets grow. However, one shortcoming with the ATP relates to the idea of well-diversified portfolios. The concept is theoretically sound, but it might be hard for an investor to manage a portfolio with a large enough number of stocks for it to be considered well-diversified.

4.6 Fama-French's three-factor model

The arbitrage pricing theory opened up for the possibility to construct k-factor pricing models with the aim to more optimally price different assets, such as stocks. Parallel to this, criticism towards the single-factor pricing model, CAPM, had been rising. Reinganum (1981) argued, "The accumulation of empirical evidence inconsistent with the simple one-period capital asset pricing model of Sharpe (1964), Lintner (1965) and Black (1974) indicates that alternative models of capital market equilibrium deserve investigation". Previous research found that market value and book to market equity had explanatory power with regards to stock returns. With these two observations, Eugene Fama and Kenneth French developed their famous three-factor model that, still to this day, is one of the more popular multifactor models. Fama and French (1993) develops a three-factor model by adding these two factors to the CAPM to capture the value and size effects;

$$R_{it} - R_{ft} = \alpha_i + \beta_{iM}\left(R_{Mt} - R_{ft}\right) + \beta_{iSMB}SMB_t + \beta_{iHML}HML_t + e_{it}$$

Where $R_{it} - R_{ft}$ is the return of asset i over the risk-free rate, f. α_i is the return of asset i that is not explained by the factors. β_{iM}, β_{iSMB}, and β_{iHML} are factor coefficients explaining asset i's sensitivity to the market-, SMB- and HML factors. $R_{Mt} - R_{ft}$ is the market factor, illustrating the return of the market portfolio, M, over the risk-free rate, f. SMB_t is the size premium, illustraiting the return of the Small-minus-Big portfolio. HML_t is the value premium, illustrating the return of the High-minus-Low portfolio. e_{it} is the orthogonal, mean zero residual, known as idiosyncratic and diversifiable risk

The two additional factors, SMB and HML, are chosen based on the empirical findings that firm's size, measured as market value, and firm's value, measured as their book to market ratios, predicts stock returns. Fama and French argue that firms with high book to market ratios (value stocks) and small stocks are able to generate higher returns compared to their counterparts, growth stocks, and large stocks. These factors could then, arguably, capture systematic risk not captured by the conventional market factor (Bodie et al., 2014). Fama and French (1993) developed a portfolio construction methodology to generate factor portfolios, namely the Small-minus-Big (SMB) and High-minus-Low (HML) portfolios. By ranking the stocks in their sample based on their market value (size) and their book-to-market ratios (value), they constructed six portfolios consisting of the top 30%, mid 40%, and bottom 30% value and size stocks. Each portfolio is then value weighted based on the individual stock's market value.

When applying the SMB and HML portfolios as additional right-hand side variables in the CAPM formula, Fama and French (1993) find that the three-factor model does a good job explaining the returns of these portfolios. Goyal (2012) further shows that the three-factor model is an improvement over CAPM when applied to 25 value and size sorted portfolios (Bodie et al., 2014). Fama and French's three-factor model have become one of the more commonly used multifactor models, often used for portfolio performance evaluation.

4.7 Carhart's four-factor model

To further add on to the Arbitrage Pricing Theory and the three-factor model, Carhart (1997) introduces the four-factor model. The model, based on the findings of Jegadeesh and Titman (1993), introduces the Momentum factor. Jegadeesh and Titman (1993) discovered that stock performance was sticky in the sense that performance, good or bad, tended to last over several periods. They found that, a long-short strategy that bought stocks that had performed well in the past and sold stocks that had performed poorly in the past, generated significant positive returns over a 3- to 12-month period. On top of this, they found that the performance of this strategy was not a result of exposure to the conventional three factors of Fama and French. By adding the momentum factor, constructed in a similar manner to that of the SMB and HML factors, Carhart's four-factor model has become a common tool when analyzing the performance of stock portfolios (Bodie et al., 2014).

Carhart's four-factor model;

$$R_{it} - R_{ft} = \alpha_i + \beta_{iM}(R_{Mt} - R_{ft}) + \beta_{iSMB}SMB_t + \beta_{iHML}HML_t + \beta_{iMOM}MOM_t + e_{it}$$

Where $R_{it} - R_{ft}$ is the return of asset i over the risk-free rate, f. α_i is the return of asset i that is not explained by the factors. β_{iM}, β_{iSMB}, β_{iHML}, and β_{iMOM} are factor coefficients explaining asset i's sensitivity to the market-, SMB- and HML factors. $R_{Mt} - R_{ft}$ is the market factor, illustrating the return of the market portfolio, M, over the risk-free rate, f. SMB_t is the size premium, illustraiting the return of the Small-minus-Big portfolio. HML_t is the value premium, illustrating the return of the High-minus-Low portfolio. MOM_t is the momenum premium, illustrating the return of the Up-minus-Down portfolio. e_{it} is the orthogonal, mean zero residual, known as idiosyncratic and diversifiable risk

The MOM-factor is constructed by ranking the universe of stocks based on their prior 12-month average performance. The bottom 30% performing stocks equal-weighted average performance is subtracted from the top 30% performing stocks equal-weighted average. Carhart applies this methodology to a broad sample of US mutual funds and finds that, by buying previous 12-months top-performers and selling previous 12-months bottom performers, yields an 8% return per year, and a significant additional annual excess return (Carhart, 1997). Further, Carhart found that much of the alpha generated by mutual funds was a result of funds loading on the momentum factor (Bodie et al., 2014). Just like for the value factor, alternative proxies for momentum exists, such as 6 and 3-months previous return.

Traditional financial theory struggles to explain the momentum factor. Some argue that it is merely a result of data mining. However, the more prevailing theory stems from behavioral finance. Jegadeesh and Titman (2001) test the hypothesis that momentum profits "are a result of delayed overreactions that are eventually reversed". They manage to show evidence supporting this theory; however, they also state that their findings should be tempered with caution. To this day, there is no robust theory that manages to explain the momentum anomaly.

4.8 Fama-MacBeth regressions

Since this thesis aims to investigate if the optimal quality factor definition varies across industries, the first step is to analyze whether the potential quality metrics have explanatory power with regards to stock returns. A possible first step is to analyze the explanatory power in the cross-section. However, with panel data, running from $t = 1,2,...,T$ and across asset $i = 1,2,...,N$, it becomes less trivial. By running cross-sectional regression for each time period, we can exanimate explanatory power, but we lose power since the number of possible observations are reduced significantly. To overcome this, another option is to stack the cross-sectional (i) and the time (t) together, called Pooled OLS. However, with firm-specific data over time, this technique is highly likely to result in cross-sectionally correlated errors (heteroskedasticity), which leads to faulty standard errors. This intervenes with inference testing. With heteroskedasticity, we cannot test the significance of the estimated coefficients. Finally, one can use the Fama-MacBeth technique (Cochrane, 2005).

When attempting to analyze the risk premium associated with the market factor, Fama and MacBeth (1973), develops the Fama-MacBeth regression technique. What makes the Fama-MacBeth technique stand out is that it estimates standard errors that are corrected for cross-sectional correlation in a panel data setting, something that simple cross-sectional or times-series regression do not adjust for. The methodology is therefore a complement for running cross-sectional regressions and for producing standard errors and t-statistics. The technique works in both single and multi-factor settings. We will use it in a multifactor setting, as describes in the Methodology chapter, *section 7.4.3*.

Conceptually, the Fama-MacBeth procedure is conducted by running separate cross-sectional regressions for each time period, and by doing so, obtain a time-series of the regressor coefficient of interest. This methodology allows us to test if a factor, or a set of factors, are priced and by how much. As mentioned before, the model corrects for cross-sectional correlation. However, it does not correct for potential autocorrelation in the time-series regression, which might exist when regressing using corporate finance data, as we do (Cochrane, 2005). This can be accounted for by using autocorrelation adjusted standard errors, such as Newey-west standard errors.

The procedure consists of two steps, where the first step is to, in the cross-sectional setting, estimate the coefficients of the desired independent factors on our dependent factor at each time period, *t*. The regression is expressed as follows;

$$\beta_{it} = \frac{Cov(R_{mt}, R_{it})}{Var(R_{mt})}$$

The second step is conducted by running T $(t = 1, 2, ..., T)$ number of cross-sectional regressions where the returns of asset *i* are the dependent variables and the regression coefficients estimated in step 1 work as the independent variables. In other words, the following regression is run for every time period;

$$\bar{R}_i^p = \lambda_0 + \lambda_1\beta_i + \lambda_2\beta_i^2 + \lambda_3\sigma_{\varepsilon i}^2 + \eta_i$$

Where $\bar{R}_i^p$ is the portfolio returns for each time period. β_i is the regression coefficients estimated in step 1. $\sigma_{\varepsilon i}^2$ is the variance of the error terms estimated in the first regression. η_i are the second-step error terms.

As a result, the regression coefficients from step 2 will form a time series, running from $t = 1$ to $t = T$. The effect is then aggregated by taking the average of the regression coefficients over the time series. This procedure allows for inference tests to be performed on these averages to evaluate the existence of significant explanatory power in the independent variables (Cuthbertson & Nitzsche, 2004).

Fama and MacBeth (1973) initially developed this regression procedure to price the market risk premium. However, papers such as Fama and French (2008), Novy-Marx (2013) and Kyosev (2020) show that this procedure can be used to price other factors as well, including firm-specific factors such as quality metrics. They also show that the procedure can be used in a multifactor setting where control variables are included.

4.9 Quality metrics

This part will cover, in a detailed manner, the definitions of the quality metrics that are investigated in this thesis. The motivation behind why these quality metrics are chosen will be further discussed in *section 8.1*.

4.9.1 Return on Equity (ROE)

Return on equity (ROE) is one of the most commonly used measures of profitability. ROE is also one of the most common metrics used as a proxy for firm quality, according to Hsu et al. (2018) and Norges Bank (2015). ROE is also commonly used in the research previously covered in this thesis.

ROE measures earnings per dollar employed in terms of equity. Something that differentiates ROE to return on assets and return on invested capital, is that ROE disregards return that belongs to debt holders, and solely focus on returns to equity holders (Bodie et al., 2014). We use Worldscopes definition of ROE;

$$Return\ on\ equity = \frac{Net\ Income - Preferred\ Dividend\ Requirement}{Average\ of\ last\ and\ current\ years\ common\ equity}$$

(Worldscope Database, 2015)

Since ROE is a measure of a firm’s profitability, one can motivate ROE’s price predicting ability using the Gordon Growth Model, as described in *section 3.2* and *4.3*.

4.9.2 Return on Assets (ROA)

Similar to return on equity, return on assets (ROA) is commonly used as a proxy for profitability in the quality factor. This is the case in both academia and practice.

ROA describes the earnings per unit of capital. However, unlike ROE, ROA takes both debt and equity into account. Since ROA takes all assets into account, it measures how well management is managing the assets entrusted to them, not only the capital entrusted to them by shareholders.

Unlike ROE, ROA is not fully standardized, which means that a few various definitions are commonly used.

$$Return\ on\ assets = \frac{EBIT}{Total\ assets}$$

$$Return\ on\ assets = \frac{EBIT \times (1 - Tax\ rate)}{Total\ assets}$$

$$Return\ on\ assets = \frac{Net\ Income}{Total\ assets}$$

(Bodie et al., 2014)

However, we decide to use ROA as defined by Worldscope. This is also the definition used by Asness et al. (2018). The tax rate applied is 35%.

$$Return\ on\ assets = \frac{Net\ Income + (Interest\ expense - Interest\ Capitalized) \times (1 - Tax\ rate)}{Average\ of\ last\ years\ and\ current\ years\ total\ assets}$$

(Worldscope Database, 2015)

Where total assets are defined as the sum of total current assets, long term receivables, investment in unconsolidated subsidiaries, other investments, net property plant and equipment, and other assets. Worldscope's definition of total assets varies somewhat for financial companies.

4.9.3 Return on Invested Capital (ROIC)

Return on invested capital is, similarly to ROE and ROA, commonly used as a profitability proxy. ROIC tells an analyst how efficient the firm was in investing capital in profitable investments. It measures the after-tax profits generated by the business and relates that to the capital raised from equity and debt holders. ROIC does not take capital held as cash into consideration, and thus, it relates generated profits to the capital that already has been deployed. ROIC is seen as the most useful measure of the performance of a company's operations (Berk & DeMarzo, 2017).

We use Worldscope's definition of Return on Invested Capital;

$$Return\ on\ Invested\ Capital = \frac{Net\ Income + (Interest\ expense\ on\ debt - interest\ capitalized) \times (1 - Tax\ rate)}{Average\ of\ last\ years\ and\ current\ years\ Invested\ capital}$$

(Worldscope Database, 2015)

Where invested capital is defined as the sum of total capital, short-term debt and long-term debt.

In order to analyze the value creation in the company, ROIC needs to be compared to the weighted average cost of capital (WACC). If ROIC is larger than the firms WACC, then value is created within the firm.

4.9.4 Gross Profitability (GP)

Gross profitability is one of the most commonly used quality measures and is, as we argue for in *section 8.1*, a well-established quality metric. Gross profitability focuses on the top line of the income statement and is thus less prone to accounting manipulation compared to metrics based on bottom-line items. However, it does not capture all factors that affect profitability.

Novy-Marx (2013), Asness et al. (2018) and Kyosev et al. (2016) defines gross profitability as;

$$Gross\ profitability = \frac{Revenue - Cost\ of\ goods\ sold}{Total\ assets}$$

We use Worldscope's definition of revenues/net sales and total assets.

4.9.5 Operating Profitability (OP)

Operating profit, deflated by the book value of assets, is another common quality proxy. Operating profit is an item further below gross profits on the income statement. It aims to explain the profit-generating ability of a firm's core business. Ball et al. (2015) analyses operating profitability in their paper *Deflating Profitability* where they defined operating profitability as;

$$Operating\ profitability = \frac{Revenue - Cost\ of\ goods\ sold - SG\&A}{Book\ value\ of\ total\ assets}$$

Yet again, there is some discrepancy with regard to its definition. However, since Ball et al. (2015) finds that their definition of operating profitability is optimal for explaining returns, we decide to use their definition. More specifically, we use Worldscope's definition of revenues, or as they call it, net sales or revenues, which they define as gross sales and other operating revenue less discounts, returns, and allowances. We use Worldscope's definition of total assets as our measure of book value of total assets.

4.9.6 Investments (INV)

Investments is another quality metric analyzed in this thesis. The computation of the investment metric varies somewhat between previous papers, but we decide to use the definition of Hou et al. (2012) since they argue that this is the most comprehensive measure of investments.

$$Investments = \frac{Total\ assets_t - Total\ assets_{t-1}}{Total\ assets_{t-1}}$$

Thus, we measure investments as annual growth in total assets. As before, we use Worldscope's definition of total assets.

4.9.7 Leverage ratio (DTE)

Along with profitability, safety, or "safe" stocks, is one of the most common characteristics used in academics and practice when defining the quality of a company. One of the arguments for the safety measure is that safer companies, ex through lower leverage, suffer a lower risk of ending up in financial distress (Norges Bank, 2015).

Like profitability, one could use the Gordon Growth Model to argue for why safety should drive firm value. Asness et al. (2018) make the argument that all else equal, an investor should pay a higher price for a stock that has a lower required return, or in other words, a safer stock. Safety then enters the model in the denominator through the required return.

As with the other metrics of quality, there is no standardized measure of safety. Prior research and practitioners use both fundamental metrics, such as leverage ratio and current ratio, and return based metrics such as beta and volatility. One could also argue for earnings stability to enter as a safety measure. Stable earnings, such as stable profitability and less volatile earnings per share or dividends per share, could very well be associated with a safer stock.

We decide to measure safety as the leverage ratio, namely, debt to equity ratio (used by Asness et al. (2018)). This leverage metric measures the ratio of debt on a company's balance sheet in relation to its share of common equity.

$$Debt\ to\ equity = \frac{Total\ debt}{Common\ equity}$$

Where total debt is measured using Worldscope's definition of total debt; $long\ term\ debt + short\ term\ debt + current\ portion\ of\ long\ term\ debt.$

The leverage ratio is a useful way to evaluate the capital structure of a company. A high leverage ratio can be associated with a higher risk of future default due to higher debt payments. It could also result in more volatile earnings. This ratio might be high for industries that require a high level of capital expenditure to manage its operations. (Hayes, 2019)

4.10 Performance measures

The Sharpe ratio and Jensen's alpha are the two main measures of portfolio performance used to evaluate quality-sorted portfolios' performance in this thesis.

4.10.1 Sharpe Ratio

In light of the assumptions of risk-averse and rational investors, investors require higher compensation for higher units of risk. Thus, when measuring an investing strategy or portfolio performance, one should take risk into account. A strategy that earns better returns at lower risk should rationally be considered superior. The Sharpe ratio, initially named reward-to-volatility ratio, was introduced by William F. Sharpe in his 1966 paper "Mutual Fund Performance".

Sharpe (1966) argued that, with the assumptions mentioned before, as well as the assumption that investors share the same predictions concerning future security performance, all efficient portfolios would fall on the straight line formed by:

$$E_i = R_f + b\sigma^2$$

Where E_i is the expected return of asset i and R_f is the risk-free return. b is the risk premium and $\sigma^2$is the standard variation in the returns of asset i.

Sharpe (1966) further states that an investor, based on the above-stated assumptions, can reach any efficient portfolio on the line following;

$$E = R_f + \frac{(E_i - R_f)\sigma}{\sigma_i}$$

thus, achieve any desired level of risk-return ratio. Moreover, based on this rationale, the optimal portfolio is the one that achieves the highest $\frac{(E_i - R_f)}{\sigma_i}$. However, since the rationale used here is based on expected values of future returns, it cannot be used in practice. Therefore, an ex-post methodology is used instead where actual excess returns in relation to actual historical standard deviation of excess return is measured.

$$R_i = \frac{(R_i - R_f)}{\sigma_i}$$

This ratio relates the excess return earned on the portfolio to its measured actual, risk, measured as standard deviation. Hence, this ex-post model has proven to be a useful tool for portfolio managers to relate their achieved excess return to their inhered level of risk (Sharpe, 1966).

Another argument for why focus lies on excess returns rather than absolute returns goes as follows; with the assumptions of risk-averse and rational investors, investors will demand higher expected returns per unit of risk. Without reimbursement to higher risk, risk-averse investors would simply invest in the risk-free rate. Therefore, investors are interested in the excess return they can achieve on their investments by investing in risky assets instead of the risk-free rate (commonly notes as the US T-bill rate). With the assumption that investors can borrow and lend at the risk-free rate, investors price the assets in such a way that the return of an investment reflects its increased risk compared to that of the risk-free rate. Therefore, the most intuitive way of measuring performance is through the excess return, not absolute return (Bodie et al., 2014).

Even though the Sharpe-ratio is one of the most used measures of performance, it is not flawless. Two of the more apparent flaws are related to cross-asset correlation within the portfolio and the upside volatility. The first issue has to do with the fact that the Sharpe-ratio does not take cross-asset correlations into account, which, if it is high, poses a higher risk on the portfolio. The second problem has to do with how risk is measured. The measure of risk, standard deviation, classifies any volatility, upside or downside, as units of risk. Even if the value of the portfolio's assets is rising, which for investors with long positions translates to higher returns, this is still seen as risky. Therefore, the risk could be seen as high even though most of the volatility comes from increases in asset prices, not retracements (Oxford Capital Strategies, 2020).

4.10.2 Jensen's Alpha

Another performance measure used in this thesis is Jensen's Alpha, which is a commonly used measure of risk-adjusted return for portfolios and funds. Michael C. Jensen first introduced the measure in his 1968 paper *The Performance of Mutual Funds in the Period 1945-1964*. Jensen constructs a measure that evaluates the performance of portfolios by relating their excess return to their level of risk. Jensen's alpha differs from the Sharpe ration in their definition of risk. Jensen used the same risk measure as in the CAPM model, the portfolios co-movement with the market portfolio, commonly referred to as market beta.

Jensen's alpha is measured by estimating the intercept in the following regressing:

$$R_{it} - R_{ft} = \alpha_i + \beta\left(R_{mt} - R_{ft}\right) + \varepsilon_{it}$$

(Cuthbertson & Nitzsche, 2004)

Where R_{it} is the return on the portfolio of interest in period t, R_{ft} is the risk-free interest rate in period t, β is portfolio i's market risk, R_{mt} is the return on the market portfolio in period t, ε_{it} is the error term and α is Jensen's alpha. If α were to be 0, the function would be identical to the CAPM. A positive α, therefore, represents a portfolio that has managed to deliver an excess return above what was expected with regards to its risk exposure and thus above what was given by CAPM or the Security Market Line. Vice versa holds for a negative alpha. Portfolio i's and the market portfolio's expected excess return is the theoretically correct values to use. However, since these are unobservable, historical values are used instead.

Jensen’s alpha may also be used in a multifactor setting such as Fama and French's three-factor model or Carhart’s four-factor model. Jensen's alpha will then measure the portfolios over/underperformance after controlling for the other variables. For example, a portfolio that manages to achieve a positive alpha in a setting using Carhart’s four factors has managed to capture excess return that cannot be explained by its market beta nor its loading on size, value, or momentum.

In this thesis, Jensen’s alpha will be used to analyze whether portfolios sorted on various definitions of firm quality have been able to achieve an excess return that cannot be explained by its market beta or its loading on size, value, or momentum. By doing so, we can exclude that our portfolio performances are a direct result of our portfolios being tilted towards stocks that have preferable size, value and momentum characteristics.

4.11 Industry Classification Benchmark

In order to analyze the power of various quality definitions across industries and how this can impact investors factor investing strategies, an industry identification tool is needed. With a few potential contesters available, such as the North American Industry Classification System (NAICS), the Standard Industrial Classification (SIC), and The Global Industry Classification Standard (GICS), we decide to use the Industry Classification Benchmark (ICB). ICB is a well-known and recognized sector definition-tool, developed by FTSE Russell, a British provider of stock market indices, in cooperation with Dow Jones, an American publishing and financial information firm. ICB is a globally utilized standard for the categorization and comparison of companies by industry and sector. It is designed to be a “...detailed and comprehensive structure for sector and industry analysis, facilitating the comparison of companies across four levels of classification...”. ICB is designed in close collaboration with asset owners, portfolio managers, and investment analysts and is designed to meet needs across the whole investment ecosystem, from exchanges to investment managers to analysts and economists. ICB has a governance framework with the purpose of ensuring relevance and reliability. The ICB’s ground rules, structure, and guidelines are established based on research and market trends and is designed in cooperation with independent market practitioners. ICB is widely adopted by institutional clients and has a wide range of possible applications, including working as an assisting tool in strategy execution for asset managers and as a categorizing tool for exchanges such as NASDAQ and NYSE.

The ICB classification system is based primarily on revenues, but earnings and market participation is also taken into consideration. The ICB is a segmentation of the current market into 11 industries, each industry containing a subset of supersectors and sectors.

- Technology
  - Software and computer services, Technology hardware and equipment.
- Telecommunications
  - Telecommunications equipment, Telecommunications service providers.
- Health Care
  - Health care providers, Medical equipment and services, Pharmaceuticals and biotechnology.
- Financials
  - Banks, Mortgage real estate investment trusts, Investment banking and brokerage services, Finance and credit services, etc.
- Real Estate
  - Real estate investment and services development, Real estate investment trusts.
- Consumer Discretionary
  - Automobiles and parts, Consumer services, Household goods and home construction, Retailers, Media, etc.
- Consumer Staples
  - Beverages, Food producers, Tobacco, Personal care, Drug and grocery stores.
- Industrials
  - Construction and materials, Aerospace and defense, Industrial engineering, Electronic and electrical equipment, General industrials, Industrial support services.
- Basic Materials
  - Industrial materials, Industrial metals and mining, Precious metals and mining, Chemicals.
- Energy

    - Oil, gas and coal, Alternative energy.
- Utilities
    - Electricity, Gas, water and multi-utilities, Waste and disposal services

(FTSE Russel, 2020)

These broader industries are then further specified into 20 supersectors, 45 sectors, and 173 subsectors. However, in order to keep a holistic approach and to prevent too narrow segmentation specifications, we decided to limit this research to the 11 industry classifications. This broader segmentation also allows for larger industry sample sizes that further improves statistical power. Nevertheless, with this delimitation in mind, some within-industry variation might exist, especially for some of the broader industries, such as the Consumer Discretionary industry.

We decide to include financials in our research even though previous research on the existence of a quality factor premium in many cases has excluded this industry. Previous authors have excluded financials with the argument that this industry is subject to unique accounting standards and has financial metrics that widely differ from those of non-financial companies. Similar arguments can be made with regards to the utilities and real estate industries since utility companies, to a larger extent than others, tend to be partly government owned and real estate firms are diverse and hard to compare with other industries. However, since our thesis aims to analyze quality metrics within-industry power, we will mainly relate firms to other firms of the same industry classification. We will therefore not face this problem.

5 Econometrics

Since this thesis aims to explore the relationship between stock's fundamental values, classified as quality metrics, and the performance of the stock in question, we will, to a big extent, focus on empirical research. In order to analyze the explanatory power of quality factor metrics on stock returns, excess returns, and prices, we will, as explained in *chapter 7*, apply various statistical and econometric methods to historical stock and company-specific data. The data used in this thesis will mainly be of cross-sectional and panel data type. This section will go over the statistical and econometrical concepts needed to understand the methodologies and concepts used in our empirical research.

5.1 Cross-sectional data

Cross-sectional data refers to a sample of data points drawn from a population universe at a given point in time, where the dataset subjects run from $n = 1, 2, ..., N$. Within-sample datapoint variation comes from differences between sample subjects, not from time-related differences. This differentiates cross-sectional data from time-series data, where one subject is observed over multiple time periods (Wooldridge, 2013). For the case of this thesis, cross-sectional data is utilized when analyzing quality metrics' effect on returns or prices at one given period, across a set of firms.

5.2 Panel data

Panel data refers to a dataset that contains both cross-sectional and time-series dimensions. A set of panel data follows a given number of subjects over a period of time, thus collecting data for one or many factors, for a set of subjects, over time. Panel data dimensions are cross-sectional from $n = 1, 2, ..., N$ and time-serial from $t = 1, 2, ..., T$. For the case of this thesis, panel data is utilized when analyzing quality metrics' effect on stock returns or prices, across a set of firms, over several time-periods. Panel data can also be pooled into a broader cross-sectional set of data by ignoring the time dimension. This allows for a larger dataset, which could enrichen the power of statistical tests (Wooldridge, 2013).

5.3 Ordinary Least Squares

When it comes to regression analysis, we will base our analysis on the ordinary least squares (OLS) methodology. OLS aims to, linearly, describe the relationship between a given dependent variable, often referred to as the Y-variable or the dependent variable, and a single, or a set of, independent variables, also often referred to as X-variables, regressors or independent variables. OLS is used in a wide variety of fields and is not limited to financial research.

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Illustrated above is the simple linear regression model for a cross-sectional setting. x_i is the independent variable that we aim to use in order to explain y_i, our dependent variable. u_i denotes the so-called error term. The error term captures relationships between y_i and other unobserved independent variables, not captured by x_i. The more interesting parameters are β_0 and β_1. β_0 illustrates the intercept of the linear regression and is known as alpha in most financial risk-return models, such as CAPM. β_1 is the slope parameter. Assuming all else equal, β_1 explains the effect of changes in the independent variable on the dependent variable. Thus, the variable coefficient, β_1, measures the marginal effect of changes in the independent variable on the dependent variable. OLS estimates the coefficients by fitting the regression line such as to minimizes the sum of the squared residuals. Residuals are the difference between the true values and the fitted value estimated by the regression. The fitted value of the regression is given by;

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

While residuals are given by;

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 + \hat{\beta}_1 x_i$$

By minimizing the sum of the squared residuals, a regression line can be fitted that, to its best effort, estimates the relationship between our dependent and independent variables.

The OLS concept can be extended to models where multiple explanatory variables are used to explain the dependent variable, so-called multiple regression models. Multiple regression models are commonly seen throughout finance and asset pricing and are directly linked to the arbitrage pricing theory's k-factor model and other factor pricing models such as Fama-French's three-factor model and Carhart's four-factor model. The multiple regression model is expressed as a direct extension of the simple regression model;

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i \qquad k = 1, 2 \ldots, n$$

(Wooldridge, 2013)

5.3.1 Ordinary Least Squares assumptions

OLS regressions that work properly are called BLUE (best linear unbiased estimator) and refers to regressions that are unbiased and effective, where effective refers to having the smallest variance possible. For an OLS regression to be BLUE, certain assumptions need to hold, the so-called Gauss-Markov assumptions. If the assumptions hold, the OLS regression is the optimal one. However, if one or more assumptions fail, then the regression coefficients might be biased or inefficient, which can result in inference being disrupted or coefficients having the wrong sign. The assumptions for BLUE are similar to the least squares' assumptions, but a bit stronger.

The assumptions are:

1) The linear regression model is linear in parameters. This allows the regression model to be written as seen above.

2) Random Sampling. The sample gathered should be picked randomly from the population.

3) No Perfect Collinearity. This states that the independent variables of the regression are not allowed to be perfectly correlated. Or, in other words, one independent variable may not be a perfect linear combination of another independent variable. However, even though some correlation between independent variables is allowed, the less, the better.

4) Zero conditional mean. This states that the mean of the error term, u, should have an expected value of zero, regardless of the values of the independent variables. In other words, there should be no relation between the value of the x-variables, and the expected mean of the error should be zero. Mathematically, this is written as;

$$E(u|x_1, x_2, \ldots, x_k) = 0$$

Under these four assumptions, the OLS regression is unbiased. For the OLS regression to be efficient as well, a fifth assumption is needed.

5) Homoskedasticity. This assumption states that the error term, u, has the same variance regardless of the values of the independent variables. In other words, in a multiple regression setup, the variance in u, should not depend on the values of the x-variables. Mathematically, this is written as;

$$Var(u|x_1, x_2, \dots, x_k) = \sigma^2$$

(Wooldridge, 2013).

5.3.2 Homoskedasticity

Homoskedasticity refers to uncorrelated error terms across sample subjects, $Var(u|x_1, x_2, \dots, x_k) = \sigma^2$, and it allows the OLS to correctly estimate the standard errors, which, in turn, gives us correct t-statistics and p-values. Correct t-statistics and p-values are needed in order to evaluate the significance level of regression coefficients accurately. Thus, with heteroskedastic errors, it not possible to evaluate if a coefficient, for example, a given quality metric, has a significant impact on returns. Homoskedasticity is therefore important.

There exist a few different tests that can be used to detect possible heteroskedasticity in a linear regression model. However, the test uses in this thesis is a test that was developed in 1972 by Trevor Breusch and Adrian Pagan, the so-called Breusch-Pagan test. Without digging too deep into the theory behind the test, it uses the estimated squared residuals and the estimated R-squared from the linear regression model to test the null hypothesis of homoskedasticity. First, the test regresses the estimated squared residuals, $E(u^p)$, as the dependent variable, and the regressors from the linear regression model as independent variables. The Lagrange-Multiplier test statistic is then derived as $LM = nR^2$ (of the second regression), where the LM-statistic is asymptotically Chi-distributed with $p - 1$ degrees of freedom. To perform the Breusch-Pagan test in R, we use the "bptest" function.

To correct for heteroskedastic standard errors, heteroskedasticity-consistent standard errors, also known as robust standard errors, could be used, such as Newey-West standard errors. These correct the standard errors and the inference testing on the regression coefficients (Woolridge, 2013).

5.3.3 Autocorrelation

Autocorrelation, or serial correlation, is another important property that comes in play when regressing using OLS. Similar to heteroskedasticity, autocorrelation refers to the correlation between error terms. However, autocorrelation refers to the correlation between error terms across time, not across sample subjects. Thus, autocorrelation comes to play in time-series and panel data. The presence of autocorrelation leads to more substantial variance in the regression coefficients, thus increases the risk of the estimated coefficients ending up far from the true value. Also, autocorrelation causes the estimated standard errors to be smaller than their true values. This leads to the estimated t-statistics being overestimated. This can lead to coefficients being reported as significant even though they are not. Autocorrelation may also lead to goodness-of-fit measures being incorrectly estimated. For panel data as well as time-series data, where the dependent variables are equity returns, the risk of autocorrelation is likely.

To test for possible autocorrelation, we use the LM-test, also known as the Breusch-Godfrey test, named after Trevor Breusch and Leslie Godfrey. The test is similar to the Breusch-Pagan test for heteroskedasticity; however, it tests the null hypothesis of no autocorrelation of any order of lag up to p. First, the residuals from the initial regression are obtained. Then, the test regresses the estimated squared residuals, $E(u^p)$, as the dependent variable and the regressors from the linear regression model and the lagged estimated squared residuals as independent variables. One lag is commonly used in this step. The LM test-statistic is then derived by multiplying the estimated R-squared from the second regression with the number of observations used in the regression, $LM = nR^2$. The LM-statistic is asymptotically chi-distributed.

As with the case of heteroskedasticity, autocorrelation can be adjusted for with Newey-west adjusted standard error (Wooldridge, 2013).

5.3.4 Heteroskedasticity and autocorrelation-consistent standard errors

To account for possible heteroskedasticity and/or autocorrelation, one can use robust standard errors. When heteroskedasticity and/or autocorrelation is present, the estimated standard errors in a regression is likely to be underestimated. Underestimated standard errors result in overestimated t-statistics. Since we use t-statistics to estimate the statistical significance of a regression coefficient, an overestimation of the test statistics can lead to a regression coefficient appear significant even though it is not. Since a big part of our analysis is depending on analyzing regression coefficients statistical significance, we believe it to be important to account for possible heteroskedasticity and autocorrelation.

One commonly used way of accounting for heteroskedasticity and autocorrelation is by using heteroskedasticity and autocorrelation-consistent standard errors. Newey and West (1987) proposed a variance estimation that accounted for heteroskedasticity and autocorrelation. Newey-west adjusted standard errors have thus become a common tool used to get more conservative significance testing, and this will be the tool used when estimating standard errors and deriving t-statistics in this thesis.

5.3.5 Multicollinearity

One important distinction that needs to be made with regards to multicollinearity is the difference between perfect and imperfect multicollinearity. As mentioned in assumption nr 3, perfect multi-collinearity is not supported by OLS. If two or more independent variables are a perfect linear combination of another independent variable, one independent variable needs to be dropped; otherwise, OLS will not work. However, the situation is not as clear cut in the case of imperfect multicollinearity. Imperfect multicollinearity is acceptable to some degree, but too large correlation between independent variables is associated with inflation of the variance in regression parameters, which jeopardizes the accuracy of inference testing. Multicollinearity could therefore make it hard to test the significance of the regression coefficients (Wooldridge, 2013).

When testing the explanatory power of various quality metrics, it is necessary to look at possible high multicollinearity between the quality metric of interest and the other control variables. One way of finding possible multicollinearity between regressors is to analyze the cross-correlation between the independent variables of the regression using a correlation matrix. This allows us to find possible high correlation between the regressors. For multicollinearity to be a problem, cross-correlation needs to be large. Dohoo et al. (1996) argue that a cross-correlation of 0.9 or larger is a clear sign of multicollinearity.

In cases when the correlation between independent variables is high enough to suspect a problematic level of multicollinearity, but not high enough to be certain, we turn to the variance inflation factor, VIF. VIF measures the impact on the estimated variance of coefficient j, that is a result of independent variable j not being uncorrelated to other independent variables in the regression. The VIF for regressor j is given by;

$$VIF_j = \frac{1}{1 - R_j^2}$$

This is the term that is determined by the correlation between X_j and other independent variables when estimating the variance of a coefficient;

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j\left(1 - R_j^2\right)}$$

Thus, the VIF represents how much higher $Var(\hat{\beta}_j)$ is as a result of correlation between X_j and other regressors. A large VIF is commonly interpreted as a sign of possible high multicollinearity. The rule of thumb is that a VIF of around 10 or higher is a sign of problematic levels of multicollinearity (Wooldridge, 2013). We use the "vif" command found in the package "car" to estimate the VIF in R.

5.4 Goodness-of-fit

Goodness-of-fit helps explain how well the overall model, consisting of a set of independent variables, explains the dependent variable. A regression model with a high goodness-of-fit does a good job explaining the patterns in the dependent variable. However, a model with a high goodness-of-fit might still contain independent variables that do not have significant explanatory power over the dependent variable.

The most common metric of goodness-of-fit is the coefficient of determination, also known as R-squared. R-squared is defined as the ratio of the explained variation in the dependent variable in relation to the total variation. In other words, it tells us how much of the sample variation in the dependent variable that can be explained by the model's independent variables. R-squared is mathematically expressed as;

$$R^2 \equiv \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

Where SSE is the explained sum of squares (sample variation in $\hat{y}_i$), SSR is the residual sum of squares (sample variation in $\hat{u}_i$), and SST is the total sum of squares (total sample variation in y_i). R-squared ranges from 0 to 1, with 1 being the case when the model perfectly explains the sample variation in y. However, a low R-squared does not necessarily mean that the model is useless. The relationship between the dependent and independent variables might still be well explained by the regression. Parallel to this, a model with a high R-squared might still contain independent variables that have no significant explanatory effect on the dependent variable. Adjusted R-squared is an extension of R-squared, where the model's goodness-of-fit is punished for containing unnecessary independent variables.

$$\overline{R^2} = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

Where n is the number of observations and p is the number of regressors.

(Wooldridge, 2013).

5.5 Outliers

In statistical work, outliers are defined as "unusual" observations. For a given sample, an outlier can be seen as observations that majorly differ from the rest of the sample. The impact of outliers is larger, the smaller the sample size is, since each observation comprises a more considerable weight of the total sample. Since OLS-regressions aim to minimize squared residuals, outliers can have a sizeable skewing effect on the OLS-estimations. Not only do outliers affect the regressions goodness-of-fit and the regressor coefficients, it also affects the robustness of statistical inferences due to a higher standard deviation, which translates to smaller test statistics (Woolridge, 2013).

The most straightforward way of identifying possible outliers is to study the distribution of the data, ex through boxplots, where possible outliers are clearly visible.

The is no clear way of how to treat outliers. One common method is to simply drop identified outliers from the dataset. However, this, of course, leads to a reduced sample size, which in turn reduces statistical precision. Another commonly used methodology is winsorizing. This allows for adjustments of the outliers to reduce their skewing effect on the dataset, without reducing the sample size. Winsorizing is done by adjusting the values of outliers to a level that is more in-line with the rest of the sample. Values that exceed a predetermined percentile, for example, the 1- and 99 percentiles, are assigned the values of the 1- and 99 percentiles. By doing this, the effect of extreme outliers is reduced, but the sample size is not trimmed, which reduced the sample bias compared to simply excluding these observations. In order to reduce data bias, percentile levels should be chosen in such a way that affects the data as little as possible but large enough to capture clear outliers. We will use winsorizing to address outliers in this thesis.

6 Data

In this chapter, we will discuss the procedure used to screen, collect, and clean the data needed for the empirical part of this thesis.

6.1 Stock screening process

It is clear from the thesis research question and methodology that we need company-specific data. Both stock related data, such as returns, but also company-specific data used to calculate the quality metrics. Since we are analyzing cross-sectional explanatory power as well as time-series performance, a panel dataset is needed.

When screening for stocks, we want to have a broad sample, namely, a sample with a broad cross-section of stocks. With a broader cross-sectional sample, we can include more relevant company-specific information and get more robust results. A broad cross-sectional sample is also necessary since we are doing an industry-specific analysis. With a narrow sample, the industry divided samples can become small for some industries and for some metrics, which, of course, would reduce the statistical power of our results.

Besides a broad cross-section, we also want a long sample. A long sample refers to a long historical time series data set. The upsides with a long dataset for our analysis are a few. First of all, just like a broad cross-sectional sample increases the sample size, a long dataset does as well. Secondly, a time series data set that covers a longer time span, also captures more of the macroeconomic and financial fluctuations, such as trends, crashes, and booms. That allows us to analyze the impact of these larger fluctuations on our models and theories, which not only makes the statistical conclusions more robust, it also makes the analysis more interesting. However, most important, since the portfolios sorting analysis is based on time-series regressions, a longer sample means a larger sample size to regress upon.

However, a long and broad dataset is also associated with a lot of challenges and downsides. One of the more obvious downsides is the case of missing data. The likelihood of missing data points increases the longer back in history the data sample runs. Missing data is not always a defining problem; however, when the amount of missing data becomes too vast, it poses a problem. Another challenge of missing data is related to the broad sample, and that is that smaller and less liquid shares tend to have less complete and coherent data. This, just like in the case of the long sample, imposes a problem if the amount of missing data becomes too vast. Other potential issues with a broad and long sample is that the financial markets today might behave differently than prior decades due to structural changes and that micro-cap stocks tend to behave differently from non-micro-cap stocks. Micro-cap investments tend to have higher costs related to information gathering and liquidity risk (Uppaluri, 2019).

Another question that needs to be considered is the decision on what market to focus on. The market in focus might lead to different conclusions with regards to the explanatory power of the quality factor in various industries. However, previous research, such as Asness et al. (2018), have shown that the results are similar across developed markets. As discussed in *section 1.2*, country-specific differences are still a factor due to differences in market conditions, such as varying investor types, liquidity differences and currency effects and the difference in accounting standards between countries. We therefore believe that the scope of the research would be far more extensive, without leading to any significant differences in results, if country-specific effects would be considered. Therefore, we have decided to conduct the research solely based on the US market.

The US market is the market of choice since it is the most developed equity market in the world and it is the one that offers the broadest and longest sample of stock and company-specific information. We focus on stocks traded on NYSE and Nasdaq since these are the largest stock exchanges in the US (and the world) with liquid and transparent data. Since 2009, the NYSE delisting requirement requires stocks that are listed to hold a market cap of not lower than $15 million over 30 sequential days. Before this, the minimum was capped at $25 million (Spicer, 2009). For Nasdaq, the same number is around $5 million. These requirements reduce the risk of micro-cap or nano-cap stocks having a too large weight in the sample, which is particularly important since we equally weight, rather than value weight, our portfolios. This will be further discussed in the *methodology chapter*.

We also restrict the screening criteria to only include stocks that have been active throughout the full sample period. However, by excluding stocks that have defaulted during the sample period, the sample is prone to suffer from an upward bias since the poorest performing stocks are excluded, also known as survivorship bias. We do not believe that this potential bias will have an effect on our ability to assess whether quality metrics' ability to predict stock returns are varying across industries. However, it might have an upward bias on the alphas and the Sharpe ratios estimated through this thesis.

We decide to focus or research on the time period of 2000 up until 2019. A sample size containing 20 years of data is enough to conduct robust empirical testing. The sample also contains periods of market-wide financial distress, namely the dotcom crash in the early 2000s and the global financial crisis, as well as periods of considerable stock market growth.

When it comes to the frequency of the data, we decided to use monthly data. This is the more commonly used in other similar papers. Even though most data related to a firm's financial statements is updated on a quarterly or annual basis, we believe that annual stock price data misses a lot of relevant stock price information.

Stock screening and data gathering are done in Eikon's Datastream database. Eikon's Datastream offers the world's most comprehensive database for cross-asset financial data, with data from the 1950s. Datastream is used to gather company-specific data such as financial reporting data and stock data. Financial reporting data refers to the metrics needed to compute quality-factors, stock return data mainly refers to stock's total return index values, used to calculate returns. In order to conduct industry-specific research, we also gather the company-specific ICB sector number. Based on the screening criteria presented above, our initial dataset consisted of 1793 companies. However, after cleaning the data by removing firms with a vast amount of missing values and outliers, the sample shrunk to between 1077 and 1492 depending on the specific quality metric in question. With the bulk of the data gathered based on the screening criteria, the data was exported to either excel or R, where most of the initial data work was conducted.

6.2 Risk-free rate

A big part of our empirical research requires a risk-free interest rate. The risk-free interest rate is an important component in most asset pricing models, such as CAPM, Fama-French three-factor model, and Carhart's four-factor model, as well as when calculating Sharpe ratios.

The risk-free interest rate represents the rate of return one could expect on capital without being exposed to any level of risk. Since the Federal Reserve is considered a risk-free creditor, US government bonds are commonly used as a proxy for the risk-free rate.

The arguments for the ideal risk-free interest rate proxy differs amongst practitioners and academia. Fama and French (1993) and (2014), as well as Kyosev et al. (2016), uses 1-month US Treasury bill rates as a proxy for the risk-free interest rate in their models. 1-month US treasury bill rates also coincide with our monthly data frequency. We, therefore, believe that the 1-month US Treasury bill rate is a suitable proxy for the risk-free interest rate in our research. We gather the risk-free interest rate data from Kenneth R. French's data library.

6.3 Data cleaning procedure

As the case with most data-heavy and empirically focused projects, data structuring and data cleaning is needed, and this is no different in our case. Most of the work related to data cleaning is focused on handling missing values. There is no optimal method regarding how to treat missing values, and there are a few things to consider. First, how many missing values may one company have before it needs to be excluded from the sample? With too strict rules, the dataset would be reduced considerably, which reduces its explanatory power. But, by including a large portion of the data points as missing values and adjusting these, the data loses some of its reliability. Also, how should missing values, that are not excluded in its entirety, be adjusted? The methods used to solve these practical hurdles could have a significant impact on the results, and thus deserves a thorough discussion.

To not include stocks with a too-large portion of missing values, we decided to identify and exclude companies with more than 14 missing values over the data period. To adjust the remaining missing values, we decided to use a rather simple but effective method. For the remaining data points identified as missing, we replaced it with the previous value. If the previous value is missing, we replaced it with the following period’s value. If no data on previous values exist, we use the nearest future observation. This method reduces the risk of creating too large artificial fluctuations in the data. This could have been the case if an industry average was used. Another possible method could have been to use the mean of the firm’s prior values, e.g. 1-year prior values. However, if the factor in consideration were expanding or decreasing continuously over a longer period, then a mean of the 1-year prior values could lead to a big and unrealistic jump in the value. Our methodology avoids this problem.

Since cross-sectional work requires equal amounts of independent variables and dependent variables, we also needed to align return data with the quality-factor metric data. When conducting multiple regression analyses that include control variables, such as market value, book to market ratio, and momentum, we also need to align return and the quality metric of interest with these measures. This, in turn, leads to a reduced dataset if missing values on these measures exist for a firm. We also exclude firms with negative book to market ratios when regressing using logarithmic book to market ratios as control variable. For the portfolio sorting analysis, we extend this procedure to only include companies that are present in each metric sample. By doing so, we create portfolios that are fully comparable to the expense of smaller sample sizes (this is further discussed in *section 8.4.1*)

Finally, in order to conduct industry-specific analysis, we segment the data based on their industry affiliation. This was done using the ICB industry identification number.

6.4 Treatment of outliers

Prior to conducting the cross-sectional, Fama-MacBeth, and the portfolio sorting analysis, we scan our data for possible outliers using boxplots. This is done for every sample used in the regression analysis and for the z-scores in the portfolios sorting analysis. For the first regression analysis and the Fama-MacBeth analysis, where we pool the data based on industries and thus neglect the time dimensions, we winsorize our eleven industry samples individually. For the annualized regression analysis, every industry-year sample is analyzed and winsorized individually. For the last step of our analysis, the portfolio sorting, we winsorize the individual quality metric z-scores before combining them into a composite quality a-score. By adjusting every sample individually, instead of the whole dataset at once, we get datasets that are better adjusted for the particular analysis at hand.

7 Methodology

The methodology of our research will allow us to identify the more powerful quality metrics based on previous research, then analyze the explanatory power of these quality metrics across a broad sample and in 11 industry divided samples. The analysis of the explanatory power will be done in three different multiple regression settings. Based on that, we can identify if there exist clear differences in various quality metrics return-explaining power in different industries. Besides that, we will also be able to find which quality metrics have explanatory power in what industry. As the last step, we will evaluate if portfolios, sorted on quality, are able to generate a premium return above the market, size, value, and momentum factors. In doing this, we will also compare if portfolios that are sorted based on the cross-sectional findings can outperform portfolios that are sorted using a benchmark strategy that treats each industry equally. If that is the case, we can argue that an investor should consider industry differences when choosing what quality metric(s) to use as a proxy for firm quality.

7.1 Descriptive statistics

We start out the analysis by reporting descriptive statistics of our data. To get a better understanding of the performance of our stock concerning their quality scores, we start by looking at the performance of portfolios of stock sorted on the quality metrics individually. In other words, for every month t, we sort all the stock in our sample based on the quality metric of interest, from high-quality to low-quality. We gather the top 30% stock into one high-quality portfolio and the bottom 30% into one low-quality portfolio. This is a commonly used portfolio sorting methodology, similarly used in papers such as Asness et al. (2018), and Novy-Marx (2013). All portfolios, including the portfolios in the portfolio analysis (*section 7.4*), are equally weighted. This prevents our results from being largely driven by the performance of large-cap companies. However, one drawback of equally weighting, compared to value weighting, is that smaller firms can get a relatively large effect on our portfolio performances. Average annual Sharpe ratios of the high and low portfolios are then calculated using their $t + 1$ returns and standard deviations. We focus on Sharpe ratios rather than returns since we want to rake risk into account; see *chapter 4.11.1* for more details. This allows us to compare if the portfolios that are sorted on high-quality outperforms portfolios sorted on low-quality in risk-adjusted terms. Something that is important to keep in mind is that this method does not allow us to single out the cause of the outperformance. If a top 30% portfolio outperforms that of a bottom 30% portfolio, we will not be able to pinpoint the reason for the outperformance. However, this is still a good initial guideline.

We also analyze sample distributions and conduct adjustments of outliers if necessary. Furthermore, we check for signs of strong multicollinearity and test for possible autocorrelation and/or heteroskedasticity in our regressions when necessary.

7.2 Initial quality metric screening

From studying previous research, it is clear that the number of ways "quality" is defined is quite large. For example, Hsu at al. (2019) identified 35 different quality factor metrics used by practitioners. To analyze 35 factors, or even 10 factors, over 11 industries would be too much to cover in a thesis like this.

Thus, as the first procedure to narrow down the number of quality metrics to include, we use a similar procedure to that of Kyosev et al. (2016) and Hsu et al. (2019). That is, we base our initial screening on previous research. We include those factors that have shown to have explanatory power across industries in other research. This procedure allows us to narrow down our sample of quality metrics considerably and it reduces the risk of including redundant factors. We also argue that it is very likely that factors that have explanatory power in the broader sample have it for some, or many, of the industry-specific samples.

One clear drawback with this methodology is that by not analyzing all possible quality metrics that exist, there could be metrics that have not yet been covered by research that would have industry-specific explanatory power in our tests. However, we consider this to be relatively unlikely, and we argue that this delimitation is necessary for the scope of the thesis not to become too large.

7.3 Cross-sectional and Fama-MacBeth analysis

Based on the set of quality metrics left over from the initial factor screening, we move over to our cross-sectional analysis. This section will analyze the explanatory power of a given quality metric, x, on stock returns, y. This is done cross-sectionally for the whole sample as well as for the individual industries.

By doing this, we can find and analyze potential differences between quality metrics and their ability to explain returns in different industries. As our hypothesis follows, we believe that there exists some degree of industry variation when it comes to finding significant quality metrics. For example, there might exist quality metrics that have significant explanatory in some industries, but that becomes redundant in others. The cross-sectional section of our analysis will consist of three tests that all will play a role in evaluating whether a quality metric has power explaining stock return in a given industry. The more robust a quality metric has shown to be in these tests, the more likely it is for it to be considered relevant for that industry. For changes in regressors to have the time to be reflected in returns, all tests are using one-period lagged regressors, as Novy-Marx (2013), Kyosev (2019), and others.

The samples used for the cross-sectional variables are matched such that we only include companies that have information on the quality metric of interest. This is also the case for our three control variables, market value, book to market ratio, and past 12-month returns. By doing this, we maximize the possible sample size available for every given regression. We also exclude companies that have a negative book-to-market ratio. We acknowledge that these regressions could suffer from potential missing variable bias as well as possible heteroskedasticity, autocorrelation, or multicollinearity. We look for multicollinearity using correlation matrices and the variance inflation factor. We test for heteroskedasticity and autocorrelation when necessary. For cases where autocorrelation or heteroskedasticity is found, Newey-west robust standard errors are used. This is discussed in more detail in the *Econometrics* chapter. We discuss the three cross-sectional tests in more detail below.

7.3.1 Pooled cross-sectional analysis

The first test that we conduct in order to analyze the cross-sectional explanatory power of our different quality metrics is a fairly simple regression analysis based on a pooled dataset. The data is initially in a panel data setting, with both cross-sectional and time series dimensions, N and T. However, by pooling the datasets, we can run cross-sectional testing. By pooling the dataset, the number of observations used to estimate the regression becomes larger. However, time-varying effects are neglected. Cross-sectional testing is done using Ordinary Least Squares (see *chapter 5.3*), where we try to estimate the marginal effect of our independent variable (the quality metric) on the dependent variable, stock returns. T-statistics are estimated to test whether the explanatory power of our regressors are significantly different from zero. The regression model looks like the following;

$$R_i = \alpha_i + \beta_{iQuality} Q_i + \beta_{iValue} \log\left(\frac{B}{M}\right)_i + \beta_{iSize} \log(MV)_i + \beta_{iMom}(t - 12\ month\ return)_i + e_i$$

Where R_i is the return of stock i, α_i is the intercept, Q_i is the quality metric (independent variable) of interest, and $\beta_{iQuality}$ is the regression coefficient of that quality metric. The regression coefficients are the variables of interest in these regressions. We aim to analyze the sign, magnitude, and significance of these. If a given regression coefficient is significant, we can conclude that according to this test and in this sample, the regressor in question has a marginal effect on stock returns. For example, if we regress using gross profitability as our regressor and we find that it has a significant and positive regression coefficient, it tells us that, for this specific test and sample, higher gross profitability is associated with higher stock return. To reduce the risk of missing variable bias, we control for the size factor (market value), the value factor (book to market ratio), and momentum (prior 12-month return). To proxy momentum, we use dummy variables that take a value of 1 if the prior 12-month return is positive, otherwise 0. By controlling for the factors value, size, and momentum, we also reduce the risk of finding significant quality metrics that are significant merely because the model is explaining stock returns poorly. However, there exists a vast amount of variable one could control for, but including more control variables increases the risk of other problems, such as multicollinearity and spurious relationships.

Stock returns behave fairly stationary and are usually assumed to follow a normal distribution (*Figure 2*: Average monthly portfolio returns, broad sample, 2001-2019 shows signs of stationary returns). However, market values are accumulating and clearly trending. Thus, in order to de-trend the market value and make it follow a more normalized distribution, we use logarithmic values. Regarding the book to market ratios, since negative observations are dropped, the book to market ratio data is heavily skewed. In order to make the book to market distribution more normally distributed, we log these as well. Since the cross-sectional analysis is done on a company level, not on a portfolio level, we do not control for the market factor.

This analysis is done for all quality metrics of interest. We test the quality metrics in the broader sample, or in other words, a sample that does not distinguish between stocks' industry affiliation. However, the more important analysis is done when we control for industry affiliation. This is done by conducting the analysis on the different industry segmented samples. This allows us to analyze potential industry differences when it comes to the cross-sectional significance of various quality metrics.

7.3.2 Annualized cross-sectional analysis

The first test pooled the entire dataset without taking years into account. This allowed for more observations in each regression, thus improving statistical power. Nevertheless, by analyzing each year individually, we can find varying time differences in our sample. The case could be that a quality metric has shown to be significant in the first test, but when treating each year as an individual sample, we find that the regressor behaves differently depending on the year. This allows us to find possible time-varying trends in the power of the regressor as well as analyzing whether possible significance is consistent over time. In order to do this, the sample is divided into 19 individual year-samples. Moreover, based on those samples, we run the same procedure as in the previous test.

$$R_i = \alpha_i + \beta_{iQuality} Q_i + \beta_{iValue} \log\left(\frac{B}{M}\right)_i + \beta_{iSize} \log(MV)_i + \beta_{iMom}(t - 12\, month\ return)_i + e_i$$

Where R_i is the return of stock i, α_i is the intercept, Q_i is the quality metric (independent variable) of interest, and $\beta_{iQuality}$ is the regression coefficient of that quality metric. One flaw with this procedure is that the samples used for each regression are significantly smaller compared to that of the pooled OLS, since each sample now only contains observations from 1 year. This could affect the regressions' ability to produce significant results, especially in those industry samples that contain fewer companies.

7.3.3 Fama-MacBeth regressions

To improve the robustness, we complement the cross-sectional analysis with Fama-MacBeth regressions. This procedure was first introduced by Fama and MacBeth in their 1993 test of the CAPM but have since then become a commonly used tool for evaluating the explanatory power in various asset pricing factors. Due to some of the econometric advantages with the Fama-MacBeth technique, it allows us to get more robust results and reduces some of the flaws in the two other cross-sectional approaches. The details behind the Fama-MacBeth regression technique and its econometric advantages are further discussed in *section 4.8.*

We use a similar approach as Novy-Marx (2013) and Kyosev et al. (2019), where we control for size, value, and momentum. This further adds to the robustness of the tests since it checks if the quality metric is significant even above the other factors, size, value, and momentum. This will also allow us to analyze the power in the other factor relative to our quality metrics. As argued for in *section 7.4.1*, we use logarithmic market value as a proxy for size, logarithmic book to market ratio as a proxy for value, and last 12-month return as a proxy for momentum. Newey-west adjusted standard errors will be used to adjust for potential autocorrelation.

7.4 Portfolio analysis

With a portfolio analysis procedure, we aim to explore the possible outperformance of the quality metrics that showed promising results for a given industry in the cross-sectional analysis. We want to analyze if, within an industry, a portfolio sorted on the quality metrics that have been hand-picked based on cross-sectional performance manages to outperform a portfolio sorted on all seven quality metrics. This will, first of all, tell us if portfolios sorted on quality can generate alpha above the conventional factors, size, value, and momentum and if this varies across industries. Secondly, it will also tell us if sorting on only the quality metrics that have shown power in that given industry, one can outperform a similar portfolio that is sorted taking all seven quality proxies into account. By doing this, we will be able to test if a given benchmark quality sorting strategy can be improved by taking industry differences into account. The portfolio analysis will therefore be used to test our second hypothesis "*Taking industry differences into account when developing a quality sorting strategy can improve portfolio performance*".

We follow the procedure of previous research, such as Fama and French (1993), Asness et al. (2018), Novy-Marx (2013), and many more. The initial step of this procedure is to create portfolios based on the quality-metric of the companies, and the final step is to check whether these portfolios are able to generate excess returns that outperform what can be expected with regards to their risk profile and factor loadings.

We rank the companies in our sector portfolios at each period t, from low quality to high quality. For each period, the 33% highest quality companies, the middle 33% quality companies, and the bottom 33% quality companies are identified and segmented. We rebalance the portfolios in June every year, and we base the sorting on $t-1$ to $t-12$ months' average quality. Therefore, we allow the whole previous year's quality to influence the portfolios sorting procedure. Worth mentioning is that we use one month lagged quality metrics when sorting the stock. This means that the portfolio will go from June in year

t to June in year $t+1$, but the rebalancing is based on the quality metrics observed in the period of May year t to May year $t-1$. There are two main reasons why we decide to rebalance our portfolios annually. The first reason is the frequency of our quality variables, that in many cases are changing on an annual basis since they are based on accounting data. If a large amount of the quality metrics on which the portfolios are sorted is changing on an annual basis, there is no value added by rebalancing the portfolios monthly. The second reason is related to turnover ratios and transaction costs. From an investor's perspective, to rebalance a portfolio costs money, which reduces the net returns. A portfolio that is rebalanced more frequently will act on market information more frequently and thus stay more updated, but it will also have larger transaction costs compared to a portfolio that is rebalanced less frequently. With that in mind, it does not make sense to rebalance the portfolio more frequently than annually for an investor that is investing based on firm characteristics that, in many cases, are changing annually. Rabener (2018b) argues that due to the infrequent release of quality-related information, rebalancing frequency should not have a significant impact on quality investing, something he also found empirical findings of. Also, as argued for in *section 7.2,* we decide to equally weight our portfolios.

In order to rank stocks based on more than one single quality metric at a time, we need to standardize the quality metrics in order for them to become comparable. In order to do this, we will conduct a z-score ranking of the stocks based on each quality metrics ranking. Thus, we will be able to convert the separate rankings into one composite quality metric. By ranking the stock based on the composite z-score, we will take all quality metrics of interest into account when ranking the stocks. The z-score is created by ranking the variable of interest, x, as $r_i = rank(x_i)$ then subtracting the cross-sectional mean of r, u_r, and divide by the standard deviation of r, σ_r. This gives us;

$$Z(x) = Z_x = \frac{r - u_r}{\sigma_r}$$

This gives us a standardized score for quality metric x. This is done on all quality metrics of interest. With a standardized score for each quality metric, we can average the individual quality metrics' z-scores to get a composite z-score that takes all quality metrics of interest into account,

$$Z = \frac{z(ROE, ROA, ROIC, operating\ profitability, gross\ profitability, debt\ to\ equity\ ratio, investments)}{7}$$

To clarify, z-score are used for two reasons. First, it standardizes the various metrics which allows us to rank stocks based on multiple quality charcteristics simultaneously. Secondly, it allows us to identify and determine potential large deviations for a given metric, which can be adjusted accordingly using winsorizing. We subtract the z-scores for debt to equity and investments since these metrics are expected to have a negative impact on stock returns.

Z is the composite quality score that takes all various quality metrics into account. This is the one used to create the benchmark portfolios; that is to say, we apply this same procedure to all the industries. However, since we aim to test if we can improve the sorting methodology by incorporating industry differences, we also create a industry specific z-score, name it Z_2. This Z_2 varies among industries, and it has the following representation, $Z_2 = \frac{\sum z(Q_i)}{n}$, where Q_i represents the quality metric that is significant for that industry given the initial cross-sectional tests and n is the number of quality metric chosen for that industry. By comparing the performance of portfolios sorted on the benchmark z-score versus portfolios that are sorted on an industry-specific z-score, we will be able to analyze if it is possible to generate a return premium above the benchmark by taking industry differences into account. To further incorporate cross-sectional industry differences, we create a second industry-specific z-score where we weight each quality metric based on the relative size of its coefficient from the pooled OLS regression. By doing this, we put a heavier emphasis on the quality metric that has a larger marginal effect on returns. To make the quality metric coefficients more comparable, we multiply them with the average of its respective variable.

Similarly to the previous empirical parts of this thesis, we need to take outliers into account. By winsorizing the top 99% and the bottom 1% of each quality metric before calculating their final z-scores, we adjust for potential outliers. Thus, the composite z-score is adjusted in a way that accounts for extreme quality metrics, and by doing that, we make sure that the low and high-quality portfolios are not heavily skewed by quality outliers.

By sorting our industry-segmented portfolios based on their quality metrics, we can analyze the quality sorted portfolios ability to generate alpha, where alpha is the return generated above the conventional market, size, value and momentum factors, also known as the four factors in Carhart's factor model (see *chapter 4.7* for further details). If our quality sorted sector portfolios are able to generate alpha, it is a sign that our quality measure can explain some of the portfolio returns that the conventional factors are not able to explain. In order to test this, we need to incorporate control portfolios for the size, value, and momentum factors. We use two approaches with regards to the control factors, one where we construct them on our own and one where we use Kenneth R. French's pre-constructed factors. The factor portfolios are constructed as follows.

$$Size = SMB = \frac{1}{3}(SL + SN + SH - BL - BN - BH)$$

$$Value = HML = \frac{1}{2}(SH - SL + BH - BL)$$

$$Momentum = MOM = \frac{1}{2}(BHR - SHR + BLR - SLR)$$

The samples used for the portfolio sorting analysis are matched such that we only include companies that have information on all seven quality metrics of interest as well as for our three control variables, market value, book to market ratio, and past 12-month returns. By doing this, we make sure that we compare the performance of portfolios sorted on the same sample of stocks. We also exclude companies that have a negative book-to-market ratio.

8 Analysis

This chapter will present, discuss, and analyze all the major findings of this thesis. We will first present portfolio Sharpe ratios to get a better understanding of the performance of our industry samples. We will then discuss the quality metrics that we have decided to analyze based on previous academic findings. After that, we will move on to discuss the descriptive statistics of the gathered data. We will then discuss and interpret the cross-sectional and Fama-MacBeth findings to investigate if there exists a difference between industries regarding the explanatory power of our different quality metrics. Based on those findings, we will also decide what quality metrics we believe have explanatory power in what industry. As a final step, we will test if our quality metrics can generate a return premium and test if those results can be improved by adapting each quality sorting strategy for every industry.

8.1 Initial quality metric screening

With a relatively rigorous sample of research that has analyzed the quality factor or metrics that could be considered possible proxies for firm quality, we managed to come up with seven quality metrics that we thought had shown promising results. And these are the metrics we procede to analyse. The quality metrics discussed here are further explained in the theory *section 4.9* of this thesis.

The first metrics that we identify as reoccurring in previous research are factors associated with profitability. Gross profitability, namely revenues less cost of goods sold, deflated by assets, is shown to have a theoretical connection to firm value, as well as empirical power in previous research. Novy-Marx (2013, 2014), Asness et al. (2018), Kyosev et al. (2016) and Hsu et al. (2019) all conclude that gross profitability has explanatory power in the cross-section of stock returns, as well as superior returns in a portfolio sorting setting.

Operating profitability, namely revenues less cost of goods sold and selling, general and administrative expenses, deflated by assets, is the second metric that has shown to be promising in previous research and thus is chosen for further analysis in this thesis. Novy-Marx (2014), Fama and French (2014), and Hsu et al. (2019) all show the predictive power and superior returns of operating profitability. Ball et al. (2015) also show that operating profitability outperforms both gross profitability and net income when it comes to predicting differences in average returns.

Return on equity is argued for by Asness et al. (2018), Hsu et al. (2019), Vyas, and van Baren (2019), Bouchaud et al. (2016) and Hou et al. (2012). Return on equity's twin, Return on assets, is not left behind. Asness et al. (2018), Hsu et al. (2019), Vyas and van Baren (2019), and Bouchaud et al. (2016) all speak in favor of return on assets as a justifiable quality metric. These papers all argue for the possibility of return on equity and return on assets as return predictors and possible measures of quality and are thus chosen for further analysis in this thesis. Return on equity and return on asset's cousin, Return on invested capital, is not as thoroughly covered as the previous two. However, backed by Hsu et al. (2019) and Vyas and van Baren (2019), we decide to include this metric as well.

The next quality metric we decide to analyze further is investments. Investments has been used in a few of the papers covered in this thesis, including Hou et al. (2012), Fama and French (2014), Kyosev et al. (2016) and Hsu et al. (2017). Investments, defined by Hou et al. (2012) as annual total asset growth divided by 1-year lagged total assets, has a theoretical inverse relationship to stock returns, as shown in *chapter 4.2.* All papers found investments to have significant explanatory power on stock returns as well as an outperformance above the conventional factors when analyzed in a portfolio sorting setting. Since investments have shown promising empirical results as well as having a theoretical relevance, we believe that this is a quality metric that is interesting to analyze more thoroughly.

Leverage is another balance sheet ratio that has been reoccurring in a few previous papers and that we believe has shown results that are promising enough for it to be considered interesting for further analysis. Papers such as Asness et al. (2018), Hsu et al. (2019), and Kyosev et al. (2016) have all used leverage ratios as metrics to define firm quality. However, as with a few previously covered metrics, there exists a discrepancy amongst previous papers on how to define leverage, but we decide to use the debt to equity ratio as our metric for leverage (see *chapter 4.9.7* for details).

Based on these findings, the quality definitions that we believe have shown promising results in previous research and thus the ones further investigated in this thesis are:

Gross profitability (**GP**), Operating profitability (**OP**), Return on equity (**ROE**), Return on assets (**ROA**), Return on invested capital (**ROIC**), Debt to equity (**DTE**) and Investments (**INV**).

8.2 Sharpe ratio performance

To get an initial overview of the performance of our sample, we conduct a simple portfolio sorting comparison of portfolio Sharpe ratios, as described in *section 7.1*. This is done for the broad sample and the segmented industry samples. This procedure allows us to get an overview of the performance of our stocks, sorted on different quality metrics, on a risk-adjusted basis. We use the Sharpe ratio to take risk into account and capture potential differences in portfolio standard deviation. But, without further analysis, we cannot single out if potential superior performance is due to higher firm quality or not. However, we still believe that it can work as a useful initial guideline by giving us a better understanding of our dataset.

Table 1:Sharpe ratio outperformance

	GP	OP	ROA	ROE	ROIC	DTE	INV
Technology	17	17	18	16	17	10	0
Telecommunications	14	14	15	14	14	9	1
Health Care	14	16	16	15	15	11	1
Financials	17	13	15	17	16	12	2
Real Estate	15	13	14	15	14	11	6
Consumer Discretionary	13	16	19	18	16	11	0
Consumer Staples	11	16	15	14	15	11	4
Industrials	16	18	19	18	19	10	0
Basic Materials	11	16	16	14	17	12	2
Energy	16	16	18	19	18	10	5
Utilities	11	12	16	17	16	9	6
Broad Sample	18	16	17	18	18	12	1

Number of years from 2001 to 2019 in which the 30% top sorted stocks' Sharpe ratio outperformed the 30% bottom sorted stocks' Sharpe ratio

When sorting based on ROE, the segmented industry samples performed well. The energy industry had the most number of outperforming periods, 19. When sorting on ROA, the consumer discretionary and industrials industries performed best by outperforming in all 19 years. Worst was it for the real estate industry that saw their high- portfolio outperform in 14 out of the 19 years. ROIC saw similar performance across the industries. However, ROIC outperformed in all years for the industrials industry. Stocks picked based on these three metrics all showed a better risk-adjusted performance in the broad sample as well as across the eleven industries, thus aligning with our expectations and what previous researchers found. These findings are in-line with those of Bouchaud et al. (2016); they found that a long-short quality portfolio, ranked based on ROE, ROA, and cash-flow to assets outperformed the sample with roughly 0.7.

Both operating and gross profitability showed promising results, however, slightly worse than that of ROE, ROA, and ROIC. There were industries, such as utilities and basic materials that only outperformed in 11 years. The rest of the industries had around 16 to 17 of the years outperforming. Stocks picked based on these two metrics all showed a better risk-adjusted performance in the broad sample as well as across the eleven industries, thus aligning with our expectations and what previous researchers found.

For debt to equity and investments, the results were unexpected. In the broad sample, low-investment portfolios only managed to outperform high-investment portfolios in 1 out of 19 years. The same number was 12 for debt to equity. These are both non-promising results for investments and debt to equity. Utilities and the real estate industry were the industries that showed the best results for the investment metric, managing to outperform in 6 of the 19 years. This is not only a sign of low-investment firms failing to outperform, but it also speaks towards high-investment firms being superior, which speaks against the findings of previous scholars. Based on the findings of Hou et al. (2012), Fama and French (2014), and Kyosev et al. (2019), there should be an inverse relationship between returns and high investment ratios. Asness et al. (2018) and Kyosev et al. (2019) argued for a similar inverse relationship between leverage ratios and returns; however, this relationship was also not evident in our findings. Regardless of the ambiguous results for debt to equity, portfolios sorted on low debt to equity managed to outperform high debt to equity ratio portfolios in roughly half of the periods.

To summarize this initial Sharpe ratio analysis, we found that our five profitability factors behaved according to our expectations as well as according to previous findings. Portfolios consisting of high-quality stocks, measured as high ROE, ROA, ROIC, gross profitability, or operating profitability, had a higher risk-adjusted return compared to portfolios with low-quality stocks. Debt to equity failed to show the inverse relationship to returns as was expected, and investments seemed to have a positive relation to returns, opposite of what was expected. With this as an initial guideline, and having our discussion in *section 4.11* in mind, we have reason to believe that at least some of our seven quality metrics are good proxies for firm quality in our 11 industries. In order to more clearly pinpoint the relationship between quality characters and stock returns we will in the following sections use various methods to more thoroughly explain the marginal effect of firm quality on stock returns by also controlling for other potential return driving firm characteristics, such as company size, book to market ratio and momentum.

8.3 Cross-sectional and Fama-MacBeth analysis

The second step in our analysis aims to explore the explanatory power of our seven chosen quality metrics. We aim to find and analyze potential differences between quality metrics and their ability to explain returns in different industries. We will first present some descriptive and summarizing statistics of the dataset used for these analyses.

8.3.1 Descriptive statistics

The sample used in this analysis consists of 1537 stocks after the cleaning procedure is done, as described in *section 6.3* (Data cleaning procedure). The sample is also industry segmented in order to distinguish between the quality metrics behavior in the different industries. The summary statistics of each industry are presented below, including the number of stocks in each industry as well as the total market capitalization of the different industries in relation to the whole sample.

Table 2: Summary statistics

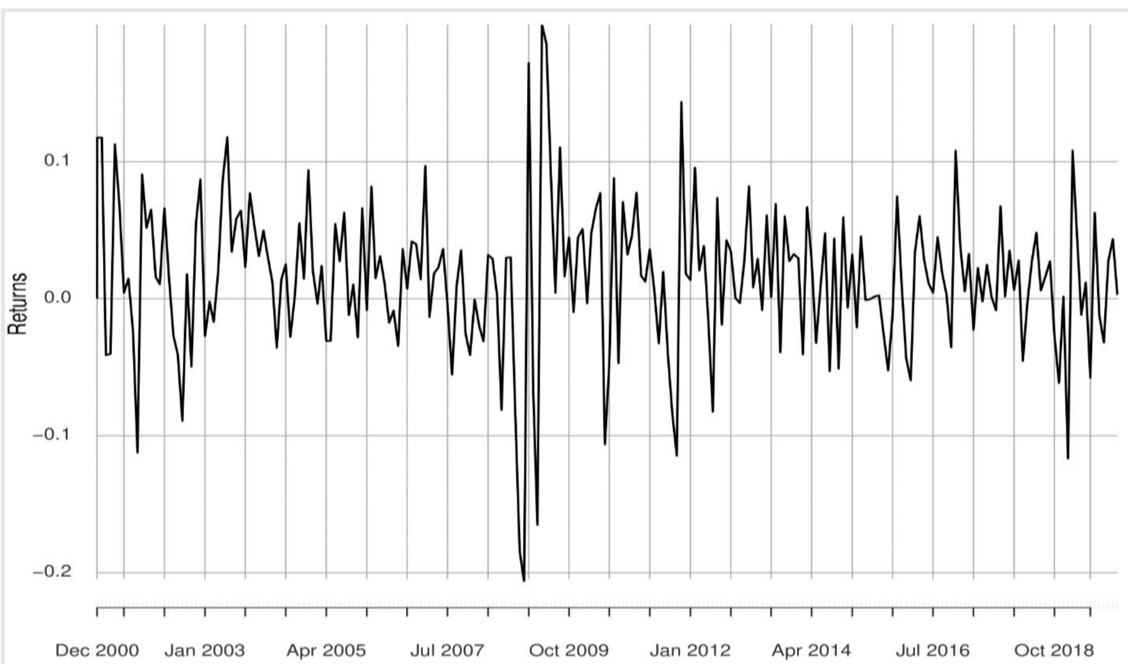
	1.	**2.**	**3.**
Technology	161 (10%)	11% (20%)	6.1 (23.1)
Telecommunications	40 (3%)	9% (5%)	20.4 (27.8)
Health Care	132 (9%)	22% (15%)	15.2 (26.0)
Financials	335 (22%)	12% (15%)	3.2 (8.9)
Real Estate	90 (6%)	3% (4%)	2.7 (9.2)
Consumer Discretionary	223 (15%)	10% (10%)	4.1 (10.7)
Consumer Staples	76 (5%)	11% (7%)	11.6 (16.4)
Industrials	275 (18%)	13% (12%)	4.1 (10.6)
Basic Materials	65 (4%)	1% (2%)	1.3 (6.2)
Energy	76 (5%)	3% (5%)	3.5 (11.5)
Utilities	64 (4%)	5% (5%)	7.4 (18.5)
Broad Sample	1537		

1. Number of firms (as % of the whole sample)
2. Size of the industry as a % of the whole sample, average of 2001-2019 (2019)
3. Average market value, 2001-2019 (2019), billion $

Worth to notice is that these numbers will differ slightly when the data is aligned with the respective quality variable since firms that do not have data on a given quality metric will be neglected in that industry-metric sample.

The smallest and the largest industries in terms of the number of companies are telecommunications and financials, with 40 and 335 companies respectively. If we look at the whole sample period, 2001-2019, we find that the industry that is largest as % of the whole sample measured as market value is the health care industry with 22% of the whole sample's market value. The smallest industry is the basic materials industry, where the same figure is 1%. In 2019, the technology industry becomes the largest, measuring 20% of the whole sample's market value, and the basic materials industry grows to 2% but remains the smallest industry. When looking at the average market value per company, we find that the telecommunication firms are the largest, measuring an average market value of $27.8 billion in 2019. The basic materials industry is the one that has, on average, the smallest companies, with an average market value of $6.2 billion in 2019. When looking at the average monthly portfolio returns, we find that the returns behave according to expectations. The returns are fluctuating around a mean of zero and behaves rather stationary. They also have large swings during periods of market decline, such as around the period of the financial crisis.

Figure 2: Average monthly portfolio returns, broad sample, 2001-2019



8.3.2 Data distribution and outliers

As described in *chapter 6.4*, we analyze the distribution of our various quality metrics across industries and as a broad sample in order to detect outliers. Boxplots of the various datasets made it clear that all quality metrics had outliers that could have major effects on our cross-sectional analysis as well as our portfolio sorting analysis. We found that winsorizing at a 1- and 99 percentile level was enough to capture extreme outliers across our samples, as can be seen is this plot, captures most of the extreme outliers.

Figure 3: ROE (Broad sample) boxplots: Non-winsorized vs winsorized at 1 and 99% level

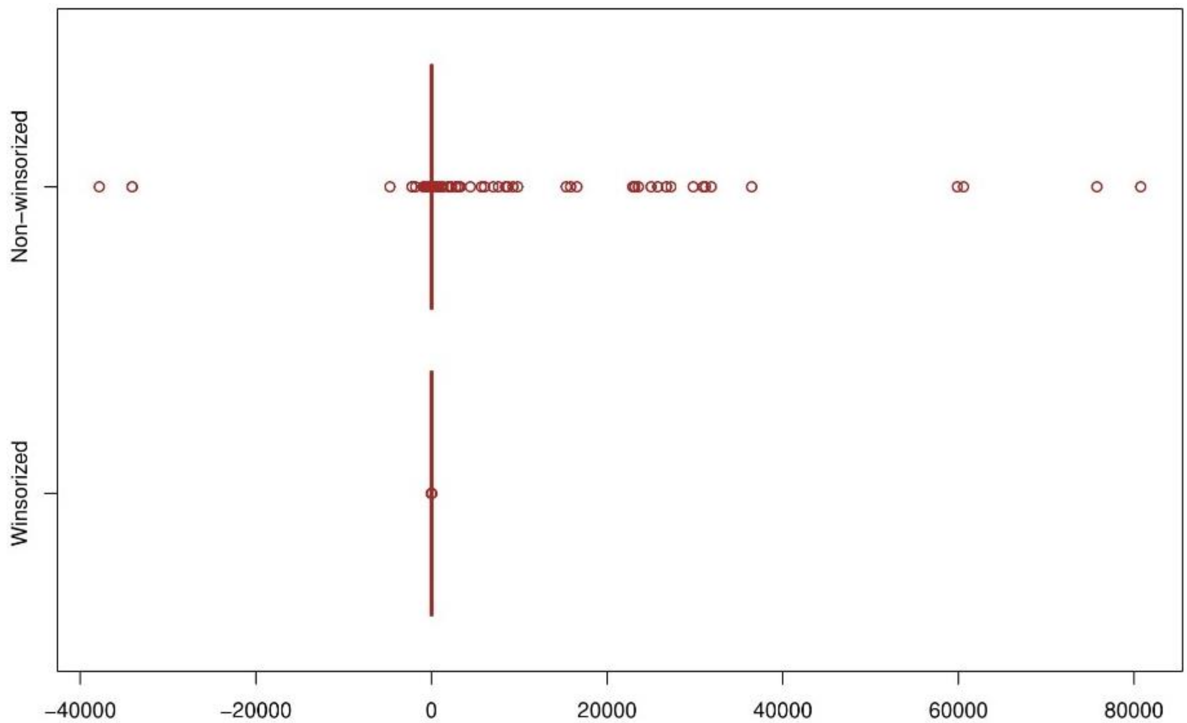
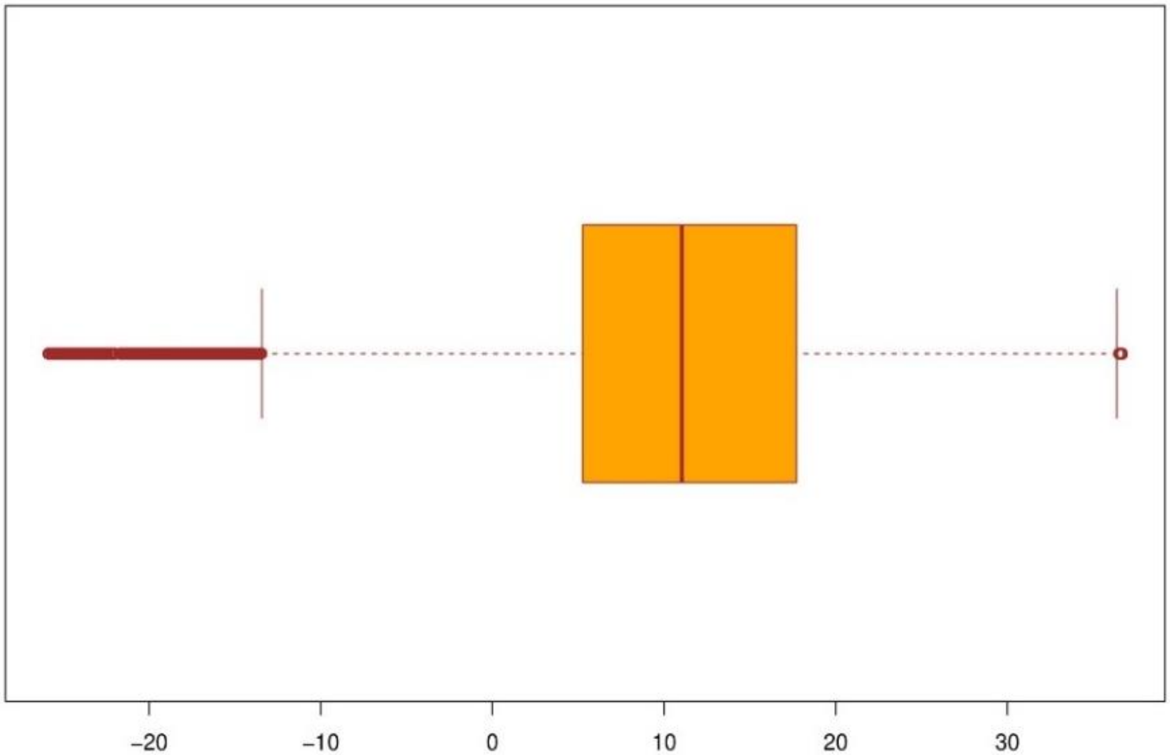


Figure 4:ROE (Broad sample) boxplot: Winsorized at 1 and 99% level



As can be seen in the two above boxplots for ROE in the broad sample, the raw dataset had extreme outliers that, if not adjusted, would skew our results. When winsorizing the data at the 1- and 99 percentiles, extreme outliers were adjusted to acceptable levels. The same procedure was conducted for all quality metrics and all industries, and we found that a 1 and 99 percentile winsorize was optimal for all samples used throughout this thesis.

8.3.3 Cross-correlation matrices and VIF

We created cross-correlation matrices for all seven quality metrics in order to analyze the risk of multicollinearity between our quality metrics and the three control variables, market value, book-to-market ratio, and momentum. This is relevant for the cross-sectional analysis as well as the Fama and MacBeth regressions.

Table 3: Cross-correlation matrix of the quality metrics and the control variables

	MV	**BTM**	**MOM**
DTE	0.00	-0.07	-0.01
ROA	0.01	-0.24	0.01
ROIC	-0.01	-0.28	0.01
ROE	0.00	0.06	0.00
INV	0.02	0.00	0.07
GP	-0.15	-0.26	0.01
OP	0.02	0.00	0.07

As can be seen in the table above, none of the seven quality metrics are correlated with the control variables to a level that would be of concern. The most substantial correlation is between ROIC and the book-to-market ratio, -0,27, which is not large enough for multicollinearity to be a threat.

When we look at cross-correlations between our seven quality metrics, we find that some quality metrics have low levels of joint correlation. In contrast, some have levels of correlation that could potentially lead to high multicollinearity.

Table 4: Cross-correlation matrix of the quality metrics

	DTE	**ROA**	**ROIC**	**ROE**	**INV**	**GP**	**OP**
DTE	1.00	-0.18	-0.17	0.08	0.08	-0.28	-0.15
ROA	-0.18	1.00	0.96	0.54	0.50	0.39	0.71
ROIC	-0.17	0.96	1.00	0.59	0.54	0.40	0.70
ROE	0.08	0.54	0.59	1.00	0.92	0.22	0.42
INV	0.08	0.50	0.54	0.92	1.00	0.21	0.40
GP	-0.28	0.39	0.40	0.22	0.21	1.00	0.59
OP	-0.15	0.71	0.70	0.42	0.40	0.59	1.00

As can be seen in this table, factors such as ROA and ROIC, as well as ROE and investments have correlations above 0.9, which points towards problematic levels of multicollinearity. Debt to equity is the metric that is least correlated with the other factors. Multiple quality measures have correlation pairs that make multicollinearity a possibility but where further analysis is necessary. These are correlation-pairs such as gross profitability and ROA with a correlation of 0.4 or ROIC and ROE with a correlation of 0.58. To further analyze these, we turn to the VIF. As mentioned in *chapter 5.3.5*, VIF above 10 is considered to constitute a level of multicollinearity that is problematic. As expected from the correlation matrix, we found that 4 out of 7 quality metrics had a VIF above 10. To avoid potential problems with multicollinearity, we decide to analyze the return-explaining power of each quality metric individually, and by doing so, reduce the risk of estimating the regression coefficients wrongfully.

8.3.4 Pooled cross-sectional analysis

The first procedure is done by running a pooled OLS across our panel data sample, both in the broad sample and in samples segmented by industry. Regressors are lagged one month, and all regressions are tested for possible heteroskedasticity and autocorrelation using the Breusch-Pagan and Breusch-Godfrey tests. We find signs of heteroskedasticity in almost all regressions and autocorrelations in most. Thus, we decide to adjust for this using heteroskedasticity and autocorrelation consistent standard errors (Newey-west) for all regressions. This allows us to get a more conservative inference testing without risking having inflated t-statistics. As covered in the Methodology section of this thesis, we run multiple regressions models where we control for size, value, and momentumThis reduces the risk of omitted variable bias, and it reduces the risk of the quality metric coefficient being overestimated. The summarized results can be found in the table below.

Table 5: Summary of the quality metric coefficients of the pooled cross-sectional analysis

	GP	OP	ROA	ROE	ROIC	DTE	INV
Technology	0.56***	1.26***	1.11***	0.43***	0.64***	0.02	0.33***
Telecommunications	0.46***	0.57**	0.74***	0.16*	0.44***	-0.05***	0.18***
Health Care	0.3***	0.24***	0.53***	0.15***	0.25***	-0.023	0.18***
Financials	0.08	0.46***	1.82***	0.08***	1.66***	-0.003*	0.28***
Real Estate	0.56**	1.07***	0.79***	0.07***	1.02***	0.02***	0.06**
Consumer Discretionary	0.16***	1.16***	1.72***	0.03***	1.44***	-0.003	0.37***
Consumer Staples	0.1***	1.06***	1.28***	0.48***	1.24***	0.006	0.28***
Industrials	0.35***	1.48***	1.74***	0.35***	1.23***	0.01	0.31***
Basic Materials	0.92***	2.61***	2.01***	0.33***	0.16***	0.03	0.22***
Energy	0.32***	0.53***	0.01***	0.15***	0.0001***	-0.024	0.25***
Utilities	-0.05	0.76***	1.74***	0.4***	1.21***	-0.01	0.08
Broad Sample	-0.06***	0.75***	1.08***	0.22***	0.75***	-0.01***	0.23***

Coefficients are scaled by 10 $\quad$ *p<0.1; **p<0.05; ***p<0.01

We found that all seven quality metrics showed significance in the broad sample. However, gross profitability had a negative coefficient, and investments had a positive coefficient, which speaks against financial theory and previous empirical findings. We found similar results when regressing our quality metrics within our 11 industries separately. Starting with gross profitability, 9 industries, including the broad sample, have a significantly positive gross profitability coefficient, align with previous academic findings and theory. The average magnitude for a 10% increase in gross profitability for the 9 industries in which gross profitability had significance was a 0,4% increase in return. However, the coefficient magnitudes were quite varying across industries, ranging from 0.1% to 0.9%.

For operating profitability, the results were even more clear-cut. We found operating profitability to have positive and significant explanatory power across all 11 industries. The average magnitude for a 10% increase in operating profitability was roughly a 1% increase in return, which is more than twice that of gross profitability. This is in-line with the findings made by Ball et al. (2015), where they found operating profitability to be superior to gross profitability when it came to predicting stock returns. One reason for this could be that operating profitability is a profitability measure that takes more of the actual expenses (Sales, Administrative and General) into account and thus is a more accurate measure of firm profitability. As for gross profitability, the marginal effect of operating profitability on stock returns was fairly varying across industries, ranging from 0.2% for the health care industry to 2.6% for the basic materials industry.

If we move on to ROE, ROA, and ROIC, we found that all three quality measures performed well across all industries. Both ROA and ROIC had positive and significant explanatory power on stock returns in all industries. The same was found for ROE except for in the telecommunications industry. ROA and ROIC were also found to have a larger regression coefficient compared to ROE. A 10% increase in ROA or ROIC would result in a 1.2% and 0.8% increase in average monthly return across industries. The same figure was much lower for ROE, at around 0.2%. However, the marginal effect for ROA and ROIC was more varying in magnitude across industries compared to that of ROE.

Debt to equity was found to have a negative marginal effect in 5 of the 11 industries. However, it was only in the telecommunications industry that the effect was significant. In the financial industry, debt to equity had a negatively significant explanatory power, but only at a 10% significance level. One surprising finding is that for the real estate industry, the debt to equity ratio had a significant positive effect on returns, which speaks against the arguments made by Asness et al. (2018), that leverage has an inverse relationship to returns. For the industries in which debt to equity showed power, the marginal effect was small compared to previous quality metrics. For the telecommunications industry, a 10% reduction in leverage only yielded a 0.05% increase in returns, and for the financial industry, it only yielded a 0.0035% increase in returns. For the investments measure, we found significant results in all industries except for the utility industry. However, the regression coefficient was positive, not negative, as expected based on findings made by Hou et al. (2012), Fama and French (2014), Kyosev et al. (2016) and Hsu et al. (2017). With a positive marginal effect on returns of 0.2% per 10% increase in asset growth, these findings are puzzling.

With regards to our control variables (appendix *Table 12: Summary of the pooled cross-sectional regressions per industry and quality metric*), the value proxy (book to market ratio) had a positive and significant effect on returns in almost all samples. The size factor proxy (market value), and the momentum proxy (12-month prior returns), had less power across our samples, which is puzzling. One reason for this could be that our quality metrics partly captures the size and the momentum effect.

To summarize, the findings of the initial cross-sectional analysis showed that the quality metrics had a rather strong explanatory power, more than initially suspected. Investments showed significance but solely positive significance, which speaks against theory and previous academic findings. The telecommunications industry was the industry that had the most promising metric results. The utility industry was the industry with fewest quality metrics with power; operating profitability, ROA, ROE, and ROIC. The quality metrics that showed the most promising results across industries were operating profitability, ROA, ROE, and ROIC. We also found that the significance of the quality metrics did not vary much across industries, however the magnitude of the coefficients did.

8.3.5 Annualized cross-sectional analysis

With the intention to increase the robustness of our cross-sectional analysis, we complement the first cross-sectional test by running regressions that capture yearly variations. Along with removing some fixed effects between periods and removing possible autocorrelation within observations, these regressions allows us to analyze if and how explanatory power varies over time. For this test, we turn to annual data. In order to do so, we run lagged annual regressors on annual stock returns. The result is 1 observation per company per year, a total of 19 observations per company (2001 to 2019), instead of the

228 that we were using in the previous analysis. As previously, we control for the other conventional style factors.

We test for heteroskedasticity in each regression, and we find varying results. Some yearly industry samples showed signs of heteroskedasticity, and some did not. We therefore decided to use heteroskedastic robust standard errors (Newey-west) for all regressions. This will make our inferences more conservative but more robust. Since there are no time dimensions, testing for autocorrelation is not necessary.

The assessment of this analysis is done through a percentage of accuracy. A quality metric that has an accuracy level of 100% would be one that is significant for all 19 years, and the sign of the coefficient for all those years aligns with previous literature. This means that we look for a positive coefficient for all metrics except for investments and debt to equity. The level of significance we use to label a coefficient as significant is 5%. The reason why we use a percentage of accuracy is that, unlike for the two other regression settings (pooled cross-sectional and Fama-MacBeth regressions), this analysis does not provide one outcome per industry-variable pair, but 19 different ones, one per year-industry-variable pair. So, if the two other analyses have 11 x 7 outcomes, this one has 11 x 7 x 19. Thus, the accuracy ratios allow us to transform the results from this analysis to ones that are comparable to the pooled cross-sectional and Fama-MacBeth regressions.

The yearly accuracy-test, presented in table 6 below, tells us that 2003 and 2015 are the years with highest accuracy ratios (40 and 36% each). Worst performers are 2014 and 2004 with accuracy levels of 9 and 10% each. Most of the other years move around 15 and 25% of accuracy, with a mean of 23%. These results hints that the power of quality varies across time. Turning to the industry accuracy results, presented in table 7 below, the technology and telecommunications industries are the ones having the highest average accuracy, of 33 and 26%, respectively. Out of these two, the technology industry is the one that exhibits the highest harmonized median of around 32%. On a similar basis as technologies, we found that utilities exhibit an average accuracy of 25%. This suggests that, on an annual basis, these industries are the ones most affected by quality metrics. Finally, turning to the individual quality metrics, the accuracy tables show that ROA and ROIC are the quality metric that has the most prominent annual impact, with an average accuracy of 33% across industries. The rest of the quality metrics had accuracy levels of between 17 and 11%, but investments that exhibit an accuracy of 2%, the lowest amongst all the quality metrics.

Table 6: Yearly accuracy ratios

2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
34%	22%	40%	10%	31%	19%	18%	25%	34%	13%
2011	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
21%	25%	14%	9%	36%	26%	13%	25%	30%	

However, since the goal with the cross-sectional analysis is to assess how different measures of firm quality behaves in various industries, the most relevant part of this annual cross-sectional test is to analyze the accuracy ratios for industries. To do this, we calculate the average annual accuracy ratio per industry and quality metric. By doing this, we get a picture of how often a given quality metric had significant explanatory power in a given industry. What we found was that the accuracy ratios were surprisingly low, with the highest being 58% and the lowest 0%.

Table 7: Accuracy ratios per industry and quality metric

	GP	OP	ROA	ROE	ROIC	DTE	INV	Average	Average*
Technology	21%	32%	58%	21%	53%	16%	0%	29%	33%
Telecommunications	11%	21%	47%	16%	47%	16%	0%	23%	26%
Health Care	5%	21%	21%	16%	32%	21%	0%	17%	19%
Financials	21%	0%	26%	21%	47%	16%	0%	19%	22%
Real Estate	11%	11%	21%	11%	21%	16%	5%	14%	15%
Consumer Discretionary	0%	21%	53%	5%	32%	11%	0%	17%	20%
Consumer Staples	5%	5%	21%	5%	21%	0%	5%	9%	10%
Industrials	11%	32%	42%	16%	42%	11%	0%	22%	25%
Basic Material	5%	21%	26%	11%	16%	16%	5%	14%	16%
Energy	16%	5%	5%	26%	11%	11%	0%	11%	12%
Utilities	11%	16%	42%	26%	42%	11%	5%	22%	25%
Average	11%	17%	33%	16%	33%	13%	2%		

*Excluding investments

Starting with analyzing ROA, we discover that despite being the variable with the highest average accuracy ratio in the broad sample, along with ROIC, it presents a different behavior per industry. The metric performs relatively well for the technology industry, where ROA shows an accuracy level of 58%, which means that in 11 out of the 19 observed years, the metric is positive and significant. The telecommunications, consumer discretionary, industrials, and utility industries also show relatively high accuracy ratios for this metric, scoring between 47% and 42%. When looking at the rest of the industries, we find accuracy ratios of about 20% for the health care, financial, consumer staples, and basic material industries and an accuracy ratio as low as 5% for the energy sector. For ROIC, the technology industry is yet again the industry that has the highest accuracy ratio, 53%. ROIC performs relatively well for the telecommunications, financials, industrials, and utility industries, with accuracy ratios of 47%, 47%, 42%, and 42%, respectively. ROIC performed worst for the energy industry with an accuracy ratio of 11%. ROE did not manage to deliver results in-line with those of ROIC and ROA. For ROE, the energy and utility industries are the best industry samples with accuracy ratios of 26%, respectively. Both the consumer staples and consumer discretionary industries had a ratio of low 5%. The rest of the industries had accuracy ratios of 10-15%. Based on this, we can see that while ROA and ROIC performed quite similarly, ROE tends to perform worse. We also see that there exists a fair amount of variation between industries.

Gross profitability was the lowest scoring quality metric after investments with an accuracy ratio of average 11% in the broad sample. As for ROA and ROIC, the industry in which gross profitability scored the highest was the technology industry and the financial industry, both with a ratio of 21%. Only one year showed significance in the health care, consumer staples, and basic material industries, and no years had significance for the consumer discretionary industry. As shown by Ball et al. (2015), operating profitability did outperform gross profitability in this test as well, with a mere 17% accuracy ratio across all industries compared to the 11% for gross profitability. Again, along with the industrials industry, the technology industry was that industry in which we saw our quality metric perform best, with an accuracy ratio of 32%. No year saw operating profitability being positively significant in the financial industry, and only one saw it in the energy industry.

Moving over to the debt to equity ratio, we found an average accuracy ratio of 13% across industries, which yet again is surprisingly low. The accuracy ratios for the debt to equity ratio was somewhat less varying across industries. The debt to equity ratio had the highest accuracy ratio in the health care industry (21%) and the lowest for consumer staples (0%), with the rest of the industries around 11-16%. Aligning with the results found in the pooled cross-sectional analysis; investments performed poorest amongst our seven quality metrics. 7 industries saw no year in which investments had a negatively significant explanatory power on stock returns. For the remaining 4, only 1 out of 19 years showed negative significance. As we saw in the previous test, investments did have a surprisingly significant positive explanatory power on stock returns. This is surprising since it contradicts financial theory and previous empirical findings, such as those found by Hou et al. (2012), Fama and French (2014), Kyosev et al. (2016) and Hsu et al. (2017).

To summarize, the annualized cross-sectional analysis found surprisingly few occasions in which our quality metrics showed clear power. We also found signs of industry heterogeneity as the accuracy ratio varied a lot between industries. However, one should bear in mind that the sample sizes for these regressions are drastically smaller than the samples used for the previous analysis. With smaller samples, the power of the regressions are reduced, and the chance of finding significance is lower.

8.3.6 Fama-MacBeth regressions

The next step in our procedure to analyze our seven chosen quality metrics across industries is conducted using Fama-MacBeth regressions. For a more rigorous explanation of the details and the theory behind this procedure, see *section 4.8*. Like the previous cross-sectional analysis, we regress using lagged regressors, and the reported t-statistics are Newey-west adjusted to account for potential autocorrelation. We also include proxies for three of the other conventional factors to make sure that our quality metric has explanatory power in a multifactor setting. Ones againm,we control for the conventional factors, similarly conducted by Novy-Marx (2013) and Kyosev et al. (2019).

Table 8: Summary of the quality metric coefficients from the Fama-MacBeth regressions

	GP	OP	ROA	ROE	ROIC	DTE	INV
Technology	0.45***	1.15***	0.94***	0.35***	0.54***	0.01	3.03***
Telcommunications	0.44***	0.68**	0.8***	0.16	0.46***	-0.04*	0.24***
Health Care	0.26***	0.19**	0.85***	0.14***	0.27***	-0.003	1.71***
Financials	-0.45	1.54	1.3***	0.05***	1.6***	-0.002	0.19***
Real Estate	0.19	0.64**	0.79***	0.05**	0.8***	0.004	0.05
Consumer Discretionary	1.26***	0.84***	1.39***	0.03***	1.12***	-0.008	0.36***
Consumer Staples	0.08**	0.9***	1.16***	0.52***	1.21***	-0.004	2.54***
Industrials	0.28***	1.21***	1.48***	0.23***	1.01***	0.003	0.24***
Basic Materials	0.69***	3.25***	1.69***	0.3***	0.17***	0.019	0.3***
Energy	0.62***	0.61***	0.02***	0.13***	0.001	-0.02	1.4***
Utilities	-0.18	0.69***	1.78***	0.6***	1.33***	-0.02	0.1
Broad Sample	-0.05***	0.62***	0.91***	0.16***	0.63***	-0.01***	0.19***

Coefficients are scaled by 10 $\quad$ ${}^{*}p<0.1$; ${}^{**}p<0.05$; ${}^{***}p<0.01$

Starting by regressing using gross profitability as our quality proxy, we find some level of industry variation. We find, to our surprise, that gross profitability has significantly negative power in the broad sample, which is a finding that is not in-line with our previous findings nor the findings of Novy-Marx (2013, 2014), Asness et al. (2018), Kyosev et al. (2016) and Hsu et al. (2019). However, when regressing using our industry samples, we find positive and significant power in 8 out of 11 industries. Across these 8 industries, the average effect of a 10% increase in gross profitability was a 0.7% increase in returns. The highest coefficient magnitude is found in the consumer discretionary industry, 1.26%, and the lowest significant marginal effect is found in the consumer staples industry, where this figure was 0.08%. With gross profitability being insignificant in 3 industries, and the regression coefficient being mostly varying for the industries in which it is significant, gross profitability seems to be promising for some industries. Looking at the control variables (see appendix *Table 13*: *Summary of the Fama-MacBeth regressions*), we find that market value is negatively significant in the broad sample, which is in line with expectations. However, when looking at the industries, market value tends to be positively significant more often than negatively significant. The book to market ratio behaves more according to expectations by being positively significant in most industries. Momentum is positively explaining returns in most industries, as expected, but with weak significance.

For operating profitability, we find positive and significant power in the broad sample as well as in all individual industries, except for the financial industry. The average increase in stock returns as a result of a 10% increase in operating profitability was 0.9% across the 10 industries in which significance was found. That operating profitability has a slightly larger effect on returns compared to gross profitability is in line with the findings of Ball et al. (2015). Operating profitability is coherent across industries when it comes to statistical significance, but the marginal effect on returns has a large variation across industries. This becomes evident when looking at the basic materials industry, where the marginal effect was 3.25%, which is 3.5 times larger than the average and almost 16 times larger than that of the health care industry where the marginal effect was 0.19%. For the control variables, we find that market value is negatively significant in the broad sample as well as in the technology industry. However, it lacks significant in the other industries. Book to market also behaves as expected by being positively significant in all samples except 2. As when regressing on gross profitability, momentum is mostly positive, however, it is only significant in two industries.

Moving over to the ROA, we find that it has a positive and significant explanatory power across all industries, similar to what we found in the pooled OLS analysis. And in the broad sample, the regression coefficient was higher than for previously covered metrics. The average magnitude of the regression across industries is 1.1%, which is higher than for gross and operating profitability. For ROE, we found positive significance for all industries, including the broad sample, except for the telecommunications industry. However, the average regression coefficient for the industries in which ROE was significant, is a mere 0.2%, which is lower than for the quality metrics mentioned so far. Industries such as financials and real estate had coefficients as low as 0.05%. Hence, a 10% increase in ROE is estimated to increase next month's stock returns by merely 0.05% in these industries. The findings for ROIC are also strong. ROIC has positive and significant power in all industries except in the energy industry. A 10% increase in ROIC is estimated to, on average, be associated with 0.8% higher stock returns. ROIC shows signs of industry heterogeneity as the regression coefficients are varying a lot in magnitude, from 0.17% on the low end to 1.33% on the high end. The findings for ROA, ROE, and ROIC are very similar to those found in the pooled cross-sectional analysis; all three metrics behave somewhat similar across industries in terms of significance, but they vary in terms of coefficient magnitude. That these three metrics can explain stock returns in the cross-section is in line with previous academic findings. Looking at the control variables, we find that market value is negatively significant in the broad sample as well as in roughly half of the industries when regressing with either ROA, ROE, or ROIC. By being positively significant in all industries, including the broad sample, the book to market ratio behaves according to expectations. Momentum is yet again puzzling by not showing strong positive contribution; however, for those samples in which momentum is significant, it is also positive.

When regressing using debt to equity as our quality proxy, the results were far less promising. Debt to equity did show negative and significant explanatory power in the broad sample, findings that coincides with previous academic findings. However, when regressing using the segmented industry samples, we found that only one industry showed any case of significance, telecommunications. The coefficients for the broad sample and the telecommunications industry were reasonably small compared to those of other metrics, only at –0.01% and -0.04%, respectively. These findings are similar to that of the pooled cross-sectional analysis, where a significant marginal effect of –0.05% was found in the telecommunications industry. When regressing on debt to equity, the control variables behave similarly as when regressing or the other quality metrics, with book to market being the only control variable that entirely behaves according to expectations. When regressing using investments as a proxy for quality, we find equally puzzling results as in the pooled cross-sectional analysis. Investments is found to be significant in 9 industries and in the broad sample. However, the coefficient is significantly positive, not negative as expected based on financial theory and previous academic findings. When regressing on investments, the only difference with regards to the control variables, is that the size proxy tends to be positively significant, thus not in line with our expectations.

In summation, the Fama-MacBeth multiple regression analysis showed promising results for many of the quality proxies. We found some degree of industry variation between our seven chosen quality proxies. All quality proxies, except for investments, showed promising results in the broad sample. For the industry samples, all variables showed promising results expect for investments and debt to equity. Besides debt to equity and investments, gross profitability was the quality proxy that showed the least promising results by being insignificant in 3 industries.

8.3.7 Summary of the cross-sectional analysis

The cross-sectional and Fama-MacBeth analysis was conducted to analyze the explanatory power of our seven quality metrics across industries. Based on the results of this analysis, we have a better understanding of how different quality metrics' explanatory power varies across industries and which quality metric has power in what industry. The findings were, to some extent, in line with our expectations as well as previous research. What became evident was that there is a reasonably big variation between the quality measures explanatory power in the different industries. The industries behaved reasonably similar with regards to the statistical significance of the quality metrics. However, it was a relatively large variation concerning the coefficient size. This speaks in favor of our hypothesis that different quality metrics explain stock returns differently depending on the industry. We will summarize the results of each test for every industry in order for the reader to get a better overview of what factor we believe works in what industry. The findings are summarized in table 9 below. Based on the performance in the cross-section of our quality metrics across our industries, we will decide on what metrics we believe have shown promising results enough to be included in a portfolio setting.

Due to the smaller sample sizes used in the annualized pooled cross-sectional regressions, we decided to emphasis this test less when deciding on what quality metrics to include in the portfolio sorting analysis. With that said, if we find that a quality metric shows promising results in both the pooled cross-sectional analysis and the Fama-MacBeth regressions, we conclude that this quality metric should be included in the portfolio analysis for that given industry. However, if the results in the annualized pooled cross-sectional analysis are found to be considerably bad, we take this into account as well.

One immediate finding is that neither debt to equity nor investments managed to produce promising results in any industry. Debt to equity was significant in both the pooled cross-sectional analysis and the Fama-MacBeth regressions for the telecommunications industry, but the sign of the coefficients was conflicting. The findings for the investment metric was rather ambiguous. Investments did produce significant results on both the pooled cross-sectional analysis and the Fama-MacBeth regressions in many industries. However, the coefficient was positive, which contradicts previous academic findings as well as financial theory. We therefore decide not to include neither of these two metrics when creating the industry-specific quality sorted portfolios. Table 9 below summarized the three previously conducted tests.

Table 9: Summary of the quality metric coefficients of the cross-sectional and Fama-MacBeth analyses

	Test	GP	OP	ROA	ROE	ROIC	DTE	INV
	1.	0.056***	0.13***	0.11***	0.04***	0.06***	0.002*	0.03***
Technology	*2.*	21%	32%	58%	21%	53%	16%	0%
	3.	0.045***	0.115***	0.094***	0.035***	0.054***	0.001	0.303***
	1.	0.046***	0.057**	0.07***	0.016*	0.04***	-0.005***	0.018***
Telecommunications	*2.*	11%	21%	47%	16%	47%	16%	0%
	3.	0.04***	0.07**	0.08***	0.02	0.05***	0	0.02***
	1.	0.03***	0.02***	0.05***	0.015***	0.025***	-0.002	0.018***
Health Care	*2.*	5%	21%	21%	16%	32%	21%	0%
	3.	0.026***	0.019**	0.085***	0.014***	0.027***	0	0.171***
	1.	0.008	0.05***	0.18***	0.008***	0.17***	0	0.028***
Financials	*2.*	21%	0%	26%	21%	47%	16%	0%
	3.	-0.045	0.154	0.13***	0.005***	0.16***	0	0.019***
	1.	0.057**	0.11***	0.08***	0.008***	0.1***	0.002***	0.007**
Real Estate	*2.*	11%	11%	21%	11%	21%	16%	5%
	3.	0.019	0.064**	0.078***	0.005**	0.08***	0	0.005
	1.	0.017***	0.12***	0.17***	0.004***	0.144***	0	0.037***
Consumer Discretionary	*2.*	0%	21%	53%	5%	32%	11%	0%
	3.	0.126***	0.084***	0.139***	0.003***	0.11***	-0.001	0.036***
	1.	0.01***	0.11***	0.13***	0.05***	0.12***	0.001	0.028***
Consumer Staples	*2.*	5%	5%	21%	5%	21%	0%	5%
	3.	0.008**	0.09***	0.116***	0.052***	0.12***	0	0.254***
	1.	0.036***	0.15***	0.17***	0.035***	0.12***	0.001	0.031***
Industrials	*2.*	11%	32%	42%	16%	42%	11%	0***
	3.	0.028***	0.121***	0.148***	0.023***	0.101***	0	0.024***
	1.	0.09***	0.26***	0.2***	0.03***	0.016***	0.003	0.022***
Basic Materials	*2.*	5%	21%	26%	11%	16%	16%	5%
	3.	0.069***	0.325***	0.168***	0.03***	0.017***	0.002	0.03***
	1.	0.03***	0.05***	0.002***	0.015***	0***	-0.003	0.025***
Energy	*2.*	16%	5%	5%	26%	11%	11%	0%
	3.	0.062***	0.061***	0.002***	0.013***	0	-0.002	0.14***
	1.	-0.006	0.07***	0.04***	0.04***	0.12***	-0.001	0.009
Utilities	*2.*	11%	16%	42%	26%	42%	11%	5%
	3.	-0.018	0.069***	0.178***	0.06***	0.133***	-0.002	0.01

Coefficients are scaled by 10 — *p<0.1; **p<0.05; ***p<0.01
Test 1. Pooled cross-sectional analysis
Test 2. Annualized cross-sectional analysis
Test 3. Fama-MacBeth analysis

As can be seen in the table above, there is somewhat of an industry variation for the other five quality metrics. ROIC is the quality metric that makes it into most industries. ROIC produced positive and significant results in both the pooled cross-sectional analysis and the Fama-MacBeth regressions in all industries except for the energy industry, in which only test 1 was significant. However, in the energy industry, ROIC produced weak results in the annualized pooled cross-sectional analysis as well, with an accuracy ratio of only 10% and is therefore not seen as powerful enough for this industry. ROE showed promising results in 10 out of the 11 industries. ROE was also neglected in the consumer staples and the consumer discretionary industries since the annualized pooled cross-sectional analysis performed very bad, with an accuracy ratio of low 5%. ROA showed promising results in all industries except for the energy industry, where the accuracy ratio was too low. Operating profitability manages to deliver promising results in 8 out of the 11 industries, but the findings are considered inconclusive for the financial, consumer staples, and energy industries. Besides debt to equity and investments, gross profitability is the metric that shows inconclusive results in most industries. This finding is in-line with the one made by Vyas and van Baren (2019) that gross profitability is the quality metrics with most industry heterogeneity. Gross profitability is considered to have strong cross-sectional power in the technology, telecommunications, industrials, and energy industries.

We end up with between two to five quality metrics per industry. ROIC and ROA are the quality metrics that are represented in most industries, and the consumer staples and the energy industries are the industries with only two quality metrics. While for ROE, the same figure is only three. This shows to tell that even for measures that are quite similar, there exists industry variation with regards to their return predictive power. Debt to equity, and to some extent, also investments, did show promising results in some individual tests. However, none of the two quality metrics showed to be robust enough to be included in the portfolio analysis.

Table 10: Summary of the quality metrics that will be included in the industry-specific portfolio sortings

	GP	**OP**	**ROA**	**ROE**	**ROIC**	**DTE**	**INV**
Technology	Yes	Yes	Yes	Yes	Yes	No	No
Telecommunications	Yes	Yes	Yes	No	Yes	No	No
Health Care	No	Yes	Yes	Yes	Yes	No	No
Financials	No	No	Yes	Yes	Yes	No	No
Real Estate	No	Yes	Yes	Yes	Yes	No	No
Consumer Discretionary	No	Yes	Yes	No	Yes	No	No
Consumer Staples	No	No	Yes	No	Yes	No	No
Industrials	Yes	Yes	Yes	Yes	Yes	No	No
Basic Materials	No	Yes	Yes	Yes	Yes	No	No
Energy	Yes	No	No	Yes	No	No	No
Utilities	No	Yes	Yes	Yes	Yes	No	No

8.4 Portfolio sorting analysis

8.4.1 Descriptive statistics

For the portfolio setting analysis, we decide to align the data such that only companies that have observations for all the quality metrics and the control variables are included. We decide to do this because we want the samples used to create the quality portfolios to consist of the same companies and be fully comparable. However, the obvious negative impact of this decision is that the samples used when creating the portfolios become considerably smaller. An optional method where the samples are only aligned based on the quality metrics used in that sample would lead to larger sample size for most portfolios. However, it would be hard to compare the performance of those portfolios to the benchmark portfolio since the companies included in the two portfolios would be different. As in the previous cases, we also remove companies with negative book to market ratios, an approach used by Fama and French (1993), Novy-Marx (2013), and others.

After implementing this data alignement procedure, we end up with a total sample of 858 stocks, which is further segmented based on industries. These industry samples are varying in the amount of included stocks from 192 for the largest industry (industrials) to 7 for the smallest industry (financials). The financial industry sample was reduced significantly due to a large amount of negative book to market ratios.

8.4.2 Portfolios Analysis

In this new and smaller sample, running from June 2001 to May 2019, we create portfolios that start in June 2002 and end in May 2019. There are two reasons why we sort the portfolios using one-year lagged company characteristics. The first reason is; investors that invest at year t will have to rely on last year's financial statements when sorting the portfolios, but they will generate the return of year t. The second reason has to do with the fact that most of the quality metrics used in this project are presented in the financial statements of the companies. Usually, the end of the fiscal year for a company does not coincide with the end of the natural year. By lagging the quality metrics, we account for the fact that companies' fiscal year ends at different dates.

To construct the annual quality sorted portfolios we follow the procedure described in *section 7.4*, where four quality sorted portfolios are constructed (high, medium, low, and high-low). The portfolios are equally weighted and rebalanced annually. To make sure that potential outperformance of the high-quality sorted portfolio and under-performance of the low-quality portfolio is not a result of the portfolios having heavy loadings on the other four conventional factors, we regress our portfolio's excess returns against the market, value, size and momentum factors. As mentioned in *chapter 7*.5, we use factors that are constructed based on our sample of stocks as well as pre-constructed factors from Kenneth R. French's website. After analyzing both, we found that the factors constructed by French behaved more in line with our expectations when regressed on our quality-sorted portfolios, most likely due to them being constructed using a larger dataset. We therefore decide to conduct the following analysis using Kenneth R. French's pre-constructed factors.

As stated in *section 7.5*, we first perform a benchmark quality sorting, where each of the seven quality metrics has an equally large impact in each industry, regardless of the findings in the cross-sectional and Fama-MacBeth analysis. These portfolios, labeled as the benchmark portfolios, aim to represent the performance of a quality strategy that does not take industry differences into account. By comparing the performance of the benchmark portfolios to the performance of portfolios that are sorted on the quality metrics that were found to be significant in the previous analysis (*Table 10*: Summary of the quality metrics that will be included in the industry-specific portfolio sortings), we can analyze if taking industry differences into account generates additional returns. To further incorporate cross-sectional industry differences, we develop another quality sorting strategy where we weight each quality metric based on the relative size of its coefficient obtained in the pooled cross-sectional analysis, as described in *section 7.5*. By doing this, we do not only put emphasis on a quality metric's significance in a given industry, but we also incorporate its relative coefficient magnitude. However, something that is worth keeping in mind is that this procedure does present a time-consistency problem, or put differently, a forward-looking bias. The cross-sectional coefficients that are used as weights when deriving the composite quality z-score are estimated using the whole sample period, 2001-2019. Since portfolios are rebalanced annually, we rebalance the portfolios using coefficients that are estimated based on partly future data, data that is unavailable for the investor at that time. An investor would, of course, use historical data to estimate the coefficients used when weighting his composite z-scores; however, this could be problematic with a limited sample size, as in our case. However, with more historical data assume that the effects of the quality metrics on the industries would be fairly similar to the ones obtained in our analysis.

Regarding cross-correlations between quality metrics, *Table 4*: Cross-correlation matrix of the quality metrics shows that metrics such as ROIC and ROA as well as ROE and investments are rather high. If the correlation between two metrics is very high, it might be redundant to include both metrics as sorting criteria. However, we do not believe that the sorting strategy becomes worse if correlations between metrics are high.

The findings from the portfolio analysis, as well as the quality metric weights, are summarized in table 11 below.

Table 11: Summary alphas from regression of quality sorted portfolios using three different sorting strategies

		High	Middle	Low	High-Low	GP	OP	ROA	ROE	ROIC	DTE	INV
	Benchmark	0.52**	0.26	0.52	-0.01	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%
Technology	*Non-weighted*	0.56***	0.25	0.49	0.07	20%	20%	20%	20%	20%	-	-
n=120	*Weighted*	0.61***	0.39	0.31	0.3	45.47%	31.89%	9.2%	6.65%	6.79%	-	-
	Benchmark	0.52*	0.34	-0.02	0.5	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%
Telecommunications	*Non-weighted*	0.5	0.35	-0.01	0.51	25%	25%	25%	-	25%	-	-
n=24	*Weighted*	0.56	0.46	-0.19	0.75	71.98%	22.26%	3.85%	-	1.9%	-	-
	Benchmark	0.6***	0.78***	0.72**	-0.12	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%
Health Care	*Non-weighted*	0.62***	0.61***	0.82**	-0.2	20%	-	20%	20%	20%	-	-
n=92	*Weighted*	0.64***	0.51**	0.89***	-0.25	37.21%	-	28.93%	16.23%	17.62%	-	-
	Benchmark	0.48	1.2**	0.35	0.13	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%
Financials	*Non-weighted*	0.36	1.21***	-0.02	0.38	-	-	33.33%	33.33%	33.33%	-	-
n=7	*Weighted*	0.54	1.05***	0.03	0.51	-	-	42.3%	4.13%	53.56%	-	-
	Benchmark	0.48	0.51	0.45	0.03	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%
Real Estate	*Non-weighted*	0.54	0.43	0.43	0.11	-	20%	20%	20%	20%	-	-
n=61	*Weighted*	0.48	0.53	0.39	0.09	-	51.67%	19.49%	2.16%	26.68%	-	-
	Benchmark	0.47**	0.35	0.36	0.11	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%
Consumer Discretionary	*Non-weighted*	0.49**	0.29	0.4	0.08	-	33.33%	33.33%	-	33.33%	-	-
n=161	*Weighted*	0.48**	0.33	0.37	0.11	-	43.16%	25.8%	-	31.04%	-	-
	Benchmark	1.23***	0.66***	0.5**	0.7***	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%
Consumer Staples	*Non-weighted*	1.22***	0.57***	0.6**	0.62***	-	-	50%	-	50%	-	-
n=55	*Weighted*	1.23***	0.57***	0.59**	0.63***	-	-	43.28%	-	56.76%	-	-
	Benchmark	0.52***	0.47**	0.39*	0.13	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%
Industrials	*Non-weighted*	0.53***	0.45**	0.43**	0.11	20%	20%	20%	20%	20%	-	-
n=193	*Weighted*	0.54***	0.43**	0.44**	0.1	21.8%	34.77%	17.69%	7.17%	18.56%	-	-
	Benchmark	0.9***	0.1	0.5	0.4	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%
Basic Materials	*Non-weighted*	0.6*	0.35	0.4	0.2	-	25%	25%	25%	25%	-	-
n=52	*Weighted*	0.64**	0.32	0.49	0.16	-	54.82%	19.52%	4.19%	21.47%	-	-
	Benchmark	0.46	0.44	0.38	0.1	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%
Energy	*Non-weighted*	0.59	0.42	0.21	0.38	50%	-	-	50%	-	-	-
n=52	*Weighted*	0.56	0.39	0.36	0.2	99.8%	-	-	0.2%	-	-	-
	Benchmark	0.56***	0.37	0.51**	0.05	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%	14.29%
Utilities	*Non-weighted*	0.48**	0.55**	0.42	0.06	-	25%	25%	25%	25%	-	-
n=39	*Weighted*	0.59***	0.58**	0.28	0.31	-	38.27%	7.85%	17.86%	36.02%	-	-

*p<0.1; **p<0.05; ***p<0.01

Alphas represent monthly equally weighted average excess-returns to portfolios sorted on quality during the period of 2002-2019. The "%" indicates the weight of the quality metric in each portfolio's z-score

One sign of the quality strategy being relevant in a given industry is if alphas are significant and decreasing. In a scenario where a quality investing strategy was able to generate a return premium, we could expect a positive and significant alpha for the high-quality portfolio that would decrease in size for the medium and low-quality sorted portfolios. Besides analyzing the behavior of our portfolio-alphas, we also analyze expected returns (the regressions' fitted values). We argue that the expected returns should follow the same pattern as the alphas if the sorting strategy works for that industry. We also believe that if the quality-sorting strategies that take industry differences into account improve the models' return-predictive power, the high-low portfolio alpha should grow.

The full results of our three different quality-sorted portfolio strategies can be found in the appendix. *Table 14, 15* and *16* include information on alphas, factor loadings, adjusted R-squared, and expected returns for each industry for the benchmark portfolios, the equally weighted portfolios as well as the coefficient weighted portfolios. We will now cover the results for each industry individually and discuss whether quality can generate a return premium and if the performance is improved when the industry differences found in the previous analysis is considered.

In the technology industry, consisting of 120 stocks, the high-quality benchmark portfolio managed to generate a significant alpha of 0.52%. This tells us that high-quality stocks, where all seven quality metrics are equally weighted, manages to generate a return premium. However, the middle and the low-quality portfolios fail to generate significant alphas. The expected return was around 1.4% for the high-quality portfolio, but surprisingly, the low-quality portfolio presented larger expected returns, 1.6%. If we move over to the industry-specific portfolios, where the significant quality metric(s) are weighted equally (gross profitability, operating profitability, ROA, ROE, and ROIC), we find an alpha of 0.56% for the high-quality portfolio that is significant at 1% level. The middle and the low-quality portfolios fail to deliver significant alphas yet again. However, the low-quality portfolio has a somewhat smaller coefficient, which can be seen as a sign in favor of this model. The expected return grows for the high and shrinks for the low-quality portfolio compared to that of the benchmark portfolio. For the weighted industry-specific sorting, we find that the alphas for the high-quality portfolio grows again, to 0.61%, and stays significant at a 1% level. The middle portfolio grows as well, and the low-quality portfolio shrinks even further. However, both of these portfolios remain insignificant. The same change that alphas experience can be seen for the expected return. So, quality seems to be able to generate a positive return premium for high-quality stocks. However, the findings are not significant for the middle and low portfolios, and we can tehrefore not reject that these alphas are zero. However, we do see that as we take more industry differences into account, our quality sorting yields stronger results.

In the telecommunications industry, consisting of 24 stocks, we find a decreasing pattern in the benchmark sorted portfolios, where the strongest alpha (0.52%) is in the high-quality portfolio and the lowest one (-0.02%) in the low-quality portfolio. However, only the alpha for the high portfolio is significant(on a 10% level). Besides these, the long-short portfolio alpha is 0.5% but not significant. Also, the expected return moves in a decreasing pattern, speaking in favor of quality as a premium generating sorting strategy. When we move over to the industry-specific portfolios, where the significant quality metrics are weighted equally (gross profitability, operating profitability, ROA, and ROIC), we find no significant alphas, but the decreasing pattern is more evident. For the weighted industry-specific sorting, we yet again have no significant alphas. However, the coefficients grow in magnitude for the high-quality portfolio and shrink in magnitude for the low-quality portfolio, suggesting that an improvement in the model compared to the benchmark. A similar improvement is found for the expected returns. The quality-sorted portfolios fail to show any clear significant alphas in the telecommunications industry; however, this could be a result of the small sample size. We find that by using the more industry-specific sorting styles, the coefficient magnitudes are larger for high quality and smaller for low quality, hinting that the model was improved compared to the benchmark model.

In the health care industry, with 92 stocks, we find highly significant alphas for all three portfolios in the benchmark sorting strategy. Nevertheless, to our surprise, we find that the low-quality portfolio, manages to outperform the high-quality portfolio's alpha for all three sorting strategies. This which speaks against our intuition since this suggests that low-quality, in the health care industry, is related to superior stock return. This inverse pattern is seen in expected returns as well. When only sorting on the quality metrics that were significant in the cross-section (operating profitability, ROA, ROE, and ROIC), we find that the alphas rise slightly in magnitude. However, the inverse pattern is still evident. When including the coefficient weights from the cross-section in the sorting strategy, we find no signs of improvement with regards to the inverse relationship of the alphas. However, as with the previous case, the alphas become slightly larger. Thus, for the health care industry, portfolios sorted based on quality do not seem to be able to earn a return premium and this is not resolved when industry heterogeneity is taken into account.

With a sample of only 7 stocks, it is hard to draw any conclusions from the financial industry. Nonetheless, we find that the high-quality portfolio outperforms the low-quality portfolio in all three portfolio sorting strategies, but none of these alphas are significant. Surprisingly, for all three portfolio sorting strategies, it is the middle portfolio that performs best, with a significant alpha of 1.2% for the benchmark portfolio, 1.21% for the equally weighted industry-specific portfolio, and 1.05% for the coefficient weighted industry-specific portfolio. This could most likely be explained by one or two medium-quality companies performing uncharacteristically well during 2002-2019. Furthermore, the high-low portfolio also generates an insignificant premium of 0.10%. However, with the small sample size in mind, firm quality does not seem to play an important role for the financial industry. The only sign of improvement when we include the findings from the previous analysis into our portfolio sorting strategy is that the high-minus low portfolio alphas grew to 0.38% and 0.50%.

For the real estate industry, with a sample of 61 stock, we find vague signs in favor of quality. In the case of the benchmark portfolios, none of the high, middle, or low-quality portfolios manages to produce significant alphas. However, the alphas are more substantial for the high-quality portfolio than for the low-quality portfolio, but the expected return moves in the opposite direction. For the industry-specific quality sorting strategy, where each significant quality metric (ROA, ROE, ROIC, and operating profitability) are weighted equally, we find similar results. No significant alphas, but a larger alpha for the high-quality portfolio compared to the low-quality portfolio. The high-quality alpha is larger compared to that of the benchmark portfolio, and the low-quality alpha is lower than the benchmark portfolio one, which is a sign of slight improvement. The expected returns also grow, but they still move in an inverted direction. For the coefficient weighted case, we find that the alpha of the low-quality portfolio shrinks further to 0.38%, however, the alpha of the high-quality portfolio falls back to the level of the benchmark portfolio, of around 0.48%. Thus, the findings in favor of quality are weak in the real estate industry, but we find some improvement when taking industry heterogeneity into account.

For the consumer discretionary industry, with a sample of 161 stock, we find a significant alpha of 0.47% for the high-quality portfolio using the benchmark sorting strategy. We also see that the alpha is lower for the low-quality stocks and insignificant. When we sort using only the quality metrics that were found to be significant in the previous analysis (operating profitability, ROA, and ROIC), we find that the alpha for the high-quality portfolio grows in magnitude. However, so does the alpha for the low-quality portfolio, though it is still insignificant. Thus, this sorting strategy seems to be a slight improvement to the benchmark strategy, at least for the high-quality sorted portfolio. When including the relative cross-sectional coefficient weights in the sorting strategy, we find that the low-quality alpha is reduced to a level similar to that of the benchmark portfolio, 0.37%, and the alpha of the high-quality portfolio is slightly improved compared to the benchmark case. However, it is lower than the alpha of the equally weighted industry-specific sorting. All in all, we find quality to have significant power for all sorting strategies for high-quality stocks, but the results are puzzling for middle and low-quality portfolios. We also find small to no improvement in alphas when we compare the benchmark sorting strategy to the two industry-specific strategies.

In the consumer staples industry, with a sample of 55 stocks, we find that all the 4 different types of portfolios present positive and significant alphas, regardless of the sorting strategy. In the case of the benchmark strategy, we find a significant alpha of 1.23% for the high-quality portfolio, 0.66% for the middle portfolio, and 0.5% for the low-quality portfolio, all of which are highly significant. The decreasing pattern and the significant high-quality alpha speak in favor of quality being able to generate return premiums in this industry. However, as the case with the previous industries, the low-quality alphas are not negative. The high-low portfolio generates a significant premium of 0.70%. We find the same decreasing pattern for the expected returns. However, when we move over to the industry-specific sorting strategies, we find that the high-quality portfolios get a slightly lower alpha (1.22%), and the low-quality alphas increase to 0.59%. This tells us that when we only use the quality metrics that were significant in the cross-section as the sorting criteria, the model is worsened compared to the benchmark sorting. This is extra surprising for the consumer staples industry since only 2 out of the 7 quality metrics (ROA and ROIC) are included in the industry-specific sorting strategies, making it the industry where the difference between the industry-specific sorting strategies and the benchmark strategy is largest. Quality as a premium generating factor seems to work in the consumer staples industry; however, adapting the sorting strategy to the cross-sectional findings does not seem to lead to an improvement.

The largest industry, industrials, with a sample of 193 stocks, presents significant alphas for all three portfolios using the benchmark sorting strategy. The high-quality portfolio generates the largest and most significant alpha, 0.52%, and the low-quality portfolio generating the smallest and least significant alpha, 0.39%. However, expected returns do not follow this decreasing pattern. Moving over to the industry-specific sorting strategies, where only gross profitability, operating profitability, ROE, ROA, and ROIC are included as sorting criteria, we find that all alphas increase in magnitude, even those of the low-quality portfolio. The coefficient weighted strategy outperforms the equally weighted strategy for the high-quality portfolio but not for the low-quality portfolio; similar findings are found for expected returns. Our quality metrics thus seem to be able to generate return premiums for high-quality stocks, and their premium grows in magnitude when we take industry heterogeneity into account. However, the same cannot be said for low-quality stocks.

The basic materials industry, with a sample of 52 stocks, presents a significant alpha of 0.9% in the high-quality portfolio and an insignificant alpha of only 0.5% in the low-quality portfolio, results pointing towards quality having a positive impact on portfolio returns. When only including operating profitability, ROA, ROE, and ROIC as quality sorting criteria, the alpha shrunk in magnitude for the high-quality portfolio to 0.60%; however, the low-portfolio decreases to (0.4%). Moreover, when incorporating the cross-sectional coefficients for the significant quality metrics, we see a slight improvement for the high-quality portfolio but not enough to generate more abnormal returns than the benchmark sorting portfolio, while the low-quality portfolio stays insignificant. Thus, quality seems to have the ability to generate abnormal returns in the basic materials industry. However, when taking industry heterogeneity into account, we do not find any improvement.

The energy industry has a sample consisting of 52 stocks, and it is the only industry where no significant alpha is found across all sorting strategies and portfolios. However, we do find that the alpha is larger for the high-quality portfolios compared to the low-quality portfolios for all three sorting strategies. The same pattern is found when looking at the expected returns. This is somewhat of an indication that quality, to some extent, is associated with a return premium. When applying the equally weighted industry-specific sorting, where only gross profitability and ROA are included as quality sorting criteria, we find that the alphas grow in magnitude for the high-quality portfolio and shrinks in magnitude for the low-quality portfolio. This indicates that when industry heterogeneity is taken into account, the portfolio performance might be improved. Weak signs of improvement is also found when we incorporate the relative coefficient magnitudes into our sorting strategy. However, it does not lead to an improvement compared to the case where we equally weight gross profitability and ROA. Thus, even though quality fails to generate any significant alphas in the energy industry, we weak signs of improvement when taking industry heterogeneity into account.

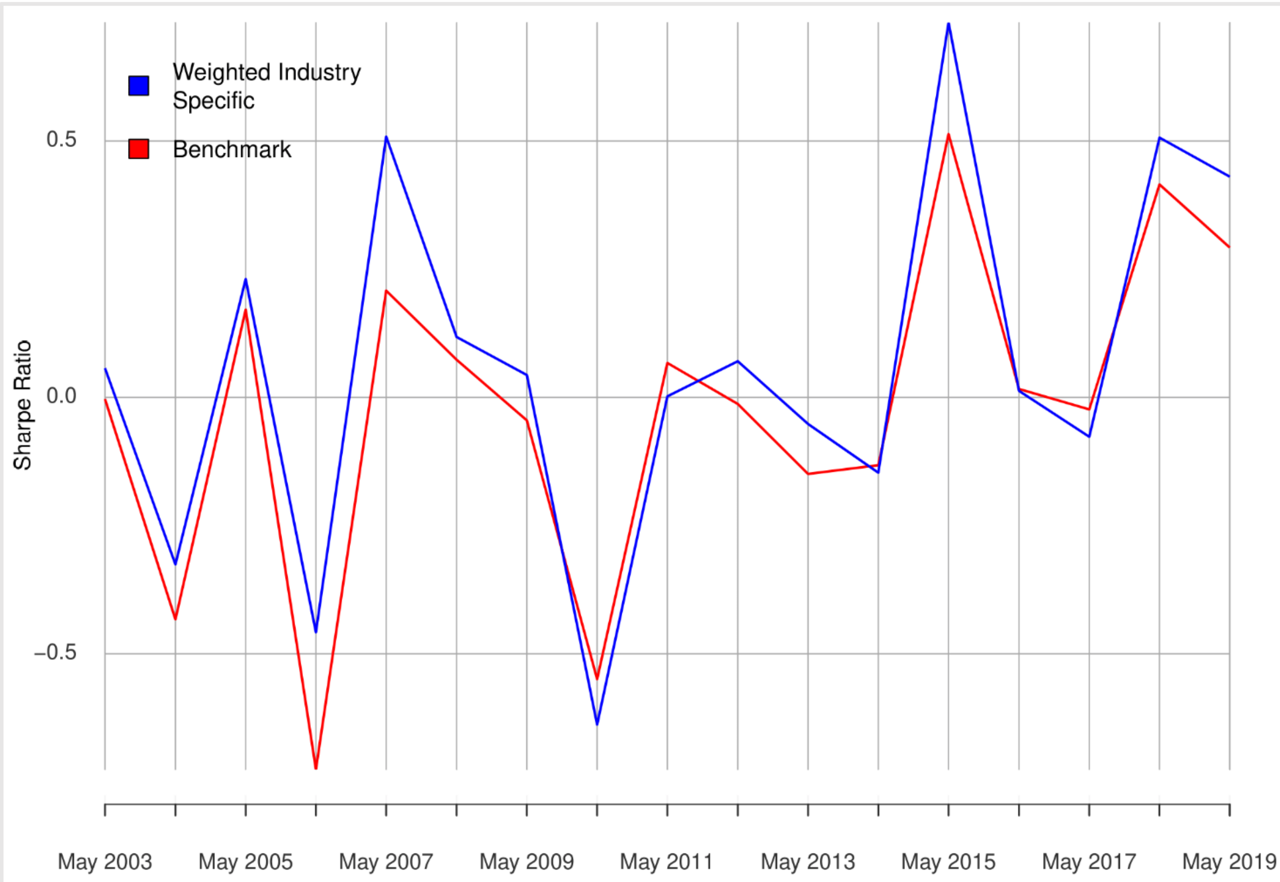
For the utilities industry, with a sample comprising 39 stocks, we find significant alphas for the high and low-quality portfolios in the benchmark sorting case; however, they are similar in magnitude, 0.56%, and 0.51% respectively. When applying the industry-specific sorting, sorting only on operating profitability, ROA, ROE, and ROIC, we see that the high-quality alpha decreases in magnitude. However, the low-quality alpha becomes insignificant, which tells us that quality as a sorting strategy improves somewhat when incorporating industry heterogeneities. This improves further when we incorporate the relative coefficient magnitudes as additions sorting criteria. The alpha of the high-quality portfolio grows both in magnitude (0.59%) and in significance, and the low-quality alpha drops by half in magnitude and stays insignificant. Thus, when the benchmark strategy is used, the decreasing pattern in alphas is not as evident, but when taking industry heterogeneity into account, the pattern becomes clear, and the alpha of the high-quality portfolio grows.

Furthemore, both Asness et al. (2018) and Novy-Marx (2013) find that higher-quality portfolios tend to have heavy negative loadings on the value factor, suggesting that high-quality firms tend to be growth firms. Asness et al. (2018) argue that this is expected since high-quality stocks should be more expensive, while the HML-factor is long cheap stocks. Novy-Marx (2013) argues that value strategies can be improved by controlling for profitability and vice versa. From observing the correlation between our quality metrics and the book to market ratio, we find that the correlation is negative for four out of seven cases, and 0 in two cases. We could therefore expect to find similar loadings in our portfolio analysis. However, by looking at the HML loadings in *tables 14*, *15*, and *16*, we do not find this clear relationship across all industries. We only find clear negative HML loading in the technology industry. Despite this, we find positive HML loadings in industries such as utilities and real estate, while for most of the industries, the evidence is weak. These findings are therefore not in line with those of Asness et al. (2018) or Novy-Marx (2013). Our findings tell us that one should consider the industry loadings of one's portfolio and not assume the HML-loadings to be negative for all industries.

We find that the loadings on the SMB-factor are more significant for high-quality portfolios than for low-quality portfolios. This is a consistent finding with only minor deviations in all three sorting strategies and across all industries. This relationship between quality and size is in line with our expectations as we could expect larger companies to have higher quality. One could expect larger companies to generate superior margins due to synergies such as, for example, economics of scale. This would most likely have a positive impact on quality metrics that are related to profitability. However, this is not evident in the correlation matrix, where all quality metrics are basically uncorrelated with market value, except for gross profitability where the correlation is –0.15. Despite this, the factor loadings suggest that a strategy that longs high-quality stocks would be a strategy that also longs large stock. This makes quality a possible hedge with regards to the size factor.

To complement the performance analysis using Jensen's alpha, we calculate annual Sharpe ratios of the high-low portfolios for the benchmark strategy and the coefficient weighted industry-specific strategy. This is done for every industry, and the graph for the technology industry is presented on the following page. The graphs covering the other industries can be found in the appendix, *section 11.3*. The Sharpe ratios behave somewhat similar across industries. For some of the industries, the benchmark Sharpe ratio outperforms that of the coefficient weighted industry-specific portfolio; however, the case is the opposite for most industries, speaking in favor of our second hypothesis. However, for almost all industries, there is rarely a portfolio that consistently outperforms the other over the 18-year period. The industries in which the industry-specific and coefficient weighted sorting strategy outperforms the benchmark strategy are technologies, financials, telecommunications, consumer discretionary, industrials, and utilities. While for the consumer staples, real estate, and basic material industries, the case was the opposite. These findings are in-line with the alphas above.

Figure 5: Annual Sharpe ratios of the High-Low Technology industry quality sorted portfolios



Henceforth, we find that quality, as we have defined it, does have a premium generating ability for most industries; however, not for all of them (8 out of 11 industries). We also find that when we take industry differences into account by incorporating the findings from the cross-sectional analysis, the premiums found grew in magnitude, but this new sorting rarely improves the significance level. The high-quality portfolio's alphas declined, on average, by 0,02 when going from the benchmark to the industry specific sorting but they grew with 0.12 when including the coefficient magnitudes. For the low-quality portfolios, the alphas declined by, on average, 0.045 when going from the benchmark to the industry specific sorting and they declined by 0.064 when including the coefficient magnitudes. The alphas of the high-low portfolios therefore grew by 0,03 when going from the benchmark to the industry specific sorting and by 0.081 when including the coefficient magnitudes. So, by constructing a sorting strategy that takes industry heterogeneities into account, an investor could improve his/her portfolio performance somewhat. Thus, we do find some evidence in support of our second hypothesis. However, the improvement that we found was not evident in all industries. We also believe that the insignificance problem could be addressed with larger sample sizes. However, there is no reason to believe that this would affect the relationship between the portfolios. Furthermore, we believe that the improvement over the benchmark sorting strategy by taking industry heterogeneity into account would be larger if we did not solely include companies that had observations for all quality metrics. One could argue that an investor's attention is spread on the amount of available information. Therefore, one or a few specific quality metrics become less important for an investor when all seven metrics are available. If we also included companies that only had information available for the quality metrics that were actually used when sorting in that industry, the outcome might have been different. However, then the companies included in the benchmark sample would differ for the ones included in the industry-specific sortings. This would make it less comparable.

We can also see that for some industries, where the alphas are insignificant, we also find a relatively low R-squared. This tells us that, first of all, the four factors of Carhart does not explain the portfolio returns well. Secondly, the quality factor, as we define it, fails to capture some of the unexplained parts. For these industries, it could be the case that stock returns are driven by other measures of quality or other style factors.

One puzzling finding is that the portfolios sorted on low-quality firms, do in more cases than not, generate a positive return premium instead of a negative. This leads to the high-low portfolio premium, in many cases, being lower than that of the high-quality portfolio. We have reason to believe that the alphas of the low-quality portfolios might be inflated due to survivorship bias. If we included companies that defaulted during the period, they would most likely have ended up in the low-quality portfolios since low profitability and high leverage are common characteristics of poorly performing companies. This would have a reducing impact on the excess return and alphas of the low-quality portfolios. Another possible explanation for why the alphas of the low-quality portfolios are larger than expected could be because the portfolios are equally weighted, not value-weighted. We could expect smaller companies to have lower quality scores. For example, smaller companies tend to have lower margins compared to larger companies, since they lack the synergies and economics of scale that larger enterprises have. Thus, we suspect that the low-quality portfolios are more commonly consisting of smaller stocks that gets a larger impact since we use equal weighting. Since smaller stocks, on average, has higher returns, we could expect the heavy loading on the small stock in the low-quality portfolio to boost returns and alphas (Banz, 1981). We can see signs of this in the more substantial loadings on the SMB-factor in the low-quality portfolios. However, since we are analyzing potential performance enhancements between the sorting styles, we are more concerned of the effect on alphas between sorting style than of the differences of alphas between the high-quality portfolios and the low-quality portfolios. We therefore argue that the decision of equal versus value weighted portfolios do not have a substantial impact on our conclusions.

9 Conclusion and further research

We initiated this paper by showing how quality investing is growing in popularity while, at the same time, exposing how the quality factor lacks a consensual definition amongst practitioners and academics. Parallel to this, we also emphasized how industry-specific effects are growing in importance for equity markets. Based on these observations, we argue that it is beneficial for investors to better understand how these observations interact. More precisely, we argue that it would be beneficial for quality investors that diversifies their portfolios in terms of industry exposure to have a better understanding if the way quality is defined will explain returns differently depending on the industry. If this is the case, it is necessary to investigate if a quality investing strategy can be improved by implementing these industry differences.

In order to bring clarity to these questions, we formulated our research question: *"Is there a difference in how well various quality metrics explain stock returns in different industries, and is it possible to improve one's quality sorting strategy by taking these differences into account?".*

To answer the first part of this question, we state the first hypothesis *"There exist industry differences with regards to how well various quality metrics explain stock returns.".* Due to the vast amount of ways to define firm quality, we decide to narrow down the existing quality metrics and only analyze those that are backed by financial theory, and that has shown promising results in previous research. To answer the first hypothesis, we conducted a thorough analysis of seven quality metrics across industries segmented according to the Industry Classification Benchmark.

Based on a dataset consisting of 1537 US equities, comprised between the years 2001 and 2019, we uncover empirical findings that suggest that the way one measures quality does matter depending on the industry. Based on three different regression analyses, we find that five out of seven quality metrics (gross profitability, operating profitability, ROA, ROE, and ROIC) explains stock returns in-line with theory, previous research, and our expectations. However, for debt to equity and investments, the results are puzzling. Most importantly, we observe that there exists an industry heterogeneity with regards to the quality metric's significance, but even more so in terms of their coefficient magnitudes. This industry heterogeneity can be best exemplified by comparing the findings of ROIC and gross profitability. When firm quality is measured as ROIC, it manages to significantly explain stock returns in 10 out of the 11 industries. Whereas, when quality is measured as gross profitability, this effect can only be observed in four industries. While if we focus on the coefficient magnitude of gross profitability in the pooled cross-sectional analysis, we see that it varies between 0.1 and 0.92 across different industries. Also, in industries where a number of quality metrics have a significant effect on stock returns, we find that the relative effect of each metric varies considerably. Based on these findings, we argue that there exist industry differences with regards to how well various quality metrics explain stock returns, in-line with hypothesis 1. These findings take us to the second part of our research question*, "... is it possible to improve one's quality sorting strategy by taking these differences into account?"*.

In order to investigate whether it is useful for investors to incorporate the previous findings into their investing strategy, we set up three different portfolio sorting strategies. One where all industries are treated equally, and two, where various degrees of industry heterogeneity are considered. Z-scores are used to create composite quality sortings, where we combine the effects of various quality metrics. These portfolios are then tested using Carhart's four-factor model, where we control for the market, size, value, and momentum effects. This allows us to analyze whether the quality sorting is able to generate a return premium and if those returns premiums are improved by incorporating industry heterogeneity. First of all, we find that quality is able to generate a significant premium in most industries, but not in all of them. We further find that for most of the industries in which quality seems to generate a return premium, that premium tends to improve when industry differences are taken into account. On average across industries, monthly average excess return premiums (alphas) of our high-low portfolios grew from 0.18 to 0.26 in magnitude, but we do not find much change in significance. However, improvements of alpha magnitudes are not found in five out of the eleven industries. This does not necessarily mean that quality is useless or that one should neglect industry differences in these industries. However, it does raise the question if our definitions of quality are sufficient to cover all industries. Nevertheless, we find patterns that suggest that an investor's quality investing strategy can be improved when more industry differences are taken into account, in line with hypothesis 2.

Another finding worth highlighting is that portfolios sorted on low-quality stocks have a significantly positive alpha more often than a significantly negative alpha. If high-quality stocks are related to return premiums, one would expect low-quality firms to underperform, a pattern found by other scholars researching the quality premium. However, we have reason to believe that this puzzling finding is related to potential survivorship bias in our sample and to our portfolios being equally weighted rather than value-weighted. We also believe that the premium generating ability of our portfolios would be more robust with a larger sample at hand. We also find that the there is no clear inverse relationship between the quality factor and the value factors, as could be expected by looking at our quality metrics correlation with the book to market ratio. However, we do find that high quality portfolios tend to be loading more heavily on larger firms, suggesting a negative correlation to the size-factor.

To summarize, previous scholars have done a good job analyzing the explanatory power of various quality metrics, across time and markets. However, our research does not only bridge the gap between quality metrics and industry heterogeneity, but it also takes it one step further by investigating whether it is possible to enhance portfolio performance by incorporating industry differences into the quality investing strategy. We therefore add knowledge, not only to academia, but also to the continuously growing field of factor investing. If Invesco's (Quance, 2019) and Baca et al.'s. (2019) predictions are accurate, then, not only will factor investing strategies keep growing in popularity, but industry-specific effects on equity returns will also continue to rise in importance, predictions that showcase the relevance of our research. Based on our findings, we believe that investors should take industry differences into account when deciding on how to optimally define firm quality and when developing their quality-sorting strategy.

Moving forward, we believe that it would be interesting for future researchers to expand our analysis to include more quality metrics than the seven covered in this thesis. One could also broaden the set of control variables in the cross-sectional and Fama-MacBeth regressions to possibly improve the estimation of the quality slope parameters. Another possible future consideration could be to expand this analysis to a global market, and thus add value to the discussion. Also, as noticed in the *annualized analysis*, there seem exist time differences with regards to quality metrics return explanatory power; it could be interesting to further analyze how industry cycles and trends are related to the explanatory power of various quality metrics. If such relationships exist, one could investigate if portfolio performance could be further improved by incorporating those findings into the quality sorting strategies. Furthermore, one limitation of this thesis is the neglection of transaction costs. By expanding the research to also cover transaction costs, future researchers could add robustness to our findings and further enhance the relevance of the topic.

10 Bibliography

Akbas, F., Jiang, C., & Koch, P. D. (2017). The trend in firm profitability and the cross-section of stock returns. *The Accounting Review*, *92*(5), 1-32. DOI: https://doi.org/10.2308/accr-51708

Ang, A., Framsted, H., (2020). Factor perspective – Multifactor Strategies. *Blackrock.com.* URL: https://www.blackrock.com/institutions/en-axj/insights/factor-perspectives/multi-factor-strategies

Asness, C. S., Frazzini, A., & Pedersen, L. H. (2019). Quality Minus Junk. *Review of Accounting Studies, 24*(1), 34–112. DOI: doi.org/10.1007/s11142-018-9470-2

Baca, S. P., Garbe, B. L., & Weiss, R. A., (2019). The Rise of Sector Effects in Major Equity Markets, Financial Analysts Journal, 56(5), 34-40. DOI: doi.org/10.2469/faj.v56.n5.2388

Ball, R., Gerakos, J., Linnainmaa, J. T., & Nikolaev, V. V., (2015). Deflating Profitability. *Journal of Financial Economics*, *117*(2), 225-248. DOI: 10.1016/j.jfineco.2015.02.004

Banz, R. W., (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, *9*(1), 3-18. DOI: doi.org/10.1016/0304-405X(81)90018-0

Berk, J., DeMarzo, P., (2017). *Corporate Finance* (4th). Edinburgh Gate, Harlow: Pearson Education Limited

BlackRock, (2020). *What is factor investing?.* URL: www.blackrock.com/us/individual/investment-ideas/what-is-factor-investing

BNP Paribas Asset Management, (2020). *Factor-Based Investing in Bond Markets*. URL: www.bnpparibas-am.lu/intermediary-fund-selector/factor-based-investing-in-bond-markets/

Bodie, Z., Kane, A., & Marcus, A. J., (2014). *Investments* (10th). Berkshire: McGraw-Hill Education

Bouchaud, J., Ciliberti, S., Landier, A., Simon, G., & Thesman, D., (2016). The Excess Return of “Quality” Stocks: A Behavioral *Anomaly*. *HEC Paris Research Paper, FIN-2016-1134*. URL: ssrn.com/abstract=2717447

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

Chen, J., Bennett, A., & Zheng, T., (2019). Sector Effects in Developed vs. Emerging Markets. *Financial Analysts Journal, 62*(6), 45-50. DOI: doi.org/10.2469/faj.v62.n6.4352

Cochrane, J. H., (2005). *Asset Pricing* (2nd). Princeton: Princeton University Press

Cuthbertson, K., (1996). *Quantitative Financial Economics – Stocks, Bonds and Foreign Exchange*. Chichester: John Wiley & Sons Ltd.

Cuthbertson, K., Nitzsche, D., (2004). *Quantitative Financial Economics – Stocks, Bonds and Foreign Exchange* (2nd). Chichester: John Wiley & Sons Ltd.

Dohoo, I.R., Ducrot, C., Fourichon, C., Donald, A., & Hurnik, D. (1996). An overview of techniques for dealing with large number of independent variables in epidemiologic studies. *Preventive Veterinary Studies*, *29*(3), 221-239. DOI: 10.1016/S0167-5877(96)01074-4

Eikon's Datastream, *Refinitive*. (2020). Database

Fama, E. F., & French, K.R., (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, 33*(1), 3-56. DOI: doi.org/10.1016/0304-405X(93)90023-5

Fama F. E., & French, K. R., (2015). A Five-factor Asset Pricing Model. *Journal of Financial Economics*, 116(1), 1-22. DOI: 10.1016/j.jfineco.2014.10.010

Fama, E. F., & MacBeth, J. D., (1973). Risk, Return, and Equilibrium: Empirical Tests. *The Journal of Political Economy, 81*(3), 607-636.

Feng, G., and Giglio, S., & Xiu, D., (2019). Taming the Factor Zoo: A Test of New Factors. *NBER Working Paper, 25481.* URL: https://www.nber.org/papers/w25481

Finch, H. J., Fraser, S. P., & Scheff, S. R., (2011). Teaching the CAPM in the Introductory Finance Course. *Journal of Economics and Finance Education, 10*(1), 77-82

Framsted, H., (2019). Is value investing dead?. *BlackRock Blog.* URL: www.blackrockblog.com/

Frazzini, A., Kabiller, D., & Pedersen, L. H., (2018). Buffet's Alpha. *Financial Analysts Journal*, *74* (4), 35-55. URL: ssrn.com/abstract=3197185

FTSE Russell, (2020). Industry Classification Benchmark (ICB), FTSERussell.com, URL: https://www.ftserussell.com/data/industry-classification-benchmark-icb

Goyal, A., (2012). Empirical Cross-sectional Asset Pricing: A Survey. *Financial Markets and Portfolio Management, 26*(1), 3-38: DOI: 10.1007/s11408-011-0177-7

Hall, M., (Aug 5th , 2019). What Are the Listing Requirements for the NASDAQ?. *Investopedia.com* URL: www.investopedia.com/ask/answers/nasdaq-listing-requirements/

Harvey, C. R, Liu, Yan., & Zhu, H., (2016). … and the Cross-Section of Expected Returns, *Review of Financial Studies*, *29*(1), 5-68. DOI: 10.3386/w20592

Hayes, A., (2019). Leverage Ratio Definition. *Investopedia.com. URL:* www.investopedia.com/terms/l/leverageratio.asp

Hsu, C. J., Kelsnik, V., & Engin, K. (2019). What Is Quality?. *Financial Analysts Journal, 67*(2), 81-95. URL: ssrn.com/abstract=2971185

Hull, J. C., (2018). *Options, Futures, and Other Derivatives* (10th). New York: Pearson Education

Innes, A., (2018). The Merits and Methods of Multi-Factor Investing, S&P Dow Jones Indices. URL:us.spindices.com/documents/research/research-the-merits-and-methods-of-multi-factor-investing.pdf

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

Jegadeesh, T., & Titman, S., (2001). Profitability of Momentum Strategies: An Evaluation of Alternative Explanations. *The Journal of Finance, 56*(2), 699-720. URL www.jstor.org/stable/222579

Kenneth R. French's Database. (2020). Data Library. URL: mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Kewei Hou, K., Chen Xue, C., & Lu Zhang, L., (2012). Digesting Anomalies: An Investment Approach. *The National Bureau of Economic Research,* Working Paper *18435.* URL: www.nber.org/papers/w18435

Kyosev, G., and Hanauer, M. X., Huij, J., & Lansdorp, S., (2020). Does Earnings Growth Drive the Quality Premium? . URL: ssrn.com/abstract=2794807

Li, F., Chow, T., Pickar, A., & Garg, Y., (2019) Transaction Costs of Factor-Investing Strategies. *Financial Analysts Journal, 75*(2), 62-78. DOI: 10.1080/0015198X.2019.1567190

Lintner, J., (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics, 47*(1), 13-37. DOI: 10.2307/1924119

Market capitalization of listed domestic companies (2020). Worldbank.org. URL: data.worldbank.org/indicator/CM.MKT.LCAP.CD?name_desc=true

Mossin, J., (1966). Equilibrium in a Capital Asset Market, *Econometrica, 34*(4), 768-783. DOI: 10.2307/1910098

Mullins, D. W., (Jan 1st, 1982). Does the Capital Asset Pricing Model Work?. *Harvard Business Review*. URL: https://hbr.org/1982/01/does-the-capital-asset-pricing-model-work

Nielson, D., Nielsen, F., & Barnes, B., (2016). An Overview of Factor Investing. URL: www.fidelity.com/

Norges Bank Investment Management, (2015). *The Quality Factor – Discussion Note*. URL: www.nbim.no/en/publications/

Novy-Marx, R. (2014). Understanding Defensive Equity. *The National Bureau of Economic Research*, Working paper *20591*. Retrieved from: www.nber.org/papers/w20591

Novy-Marx, R., (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics, 108*(1), 1-28. DOI: 10.1016/j.jfineco.2013.01.003

Novy-Marx, R., (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics, 108,* 1-28. DOI: doi.org/10.1016/j.jfineco.2013.01.003

Oxford Capital Strategies, (2020). Four Problems with the Sharpe Ratio. *Oxfordstrat.com*. URL: oxfordstrat.com/ideas/sharpe-ratio/

Pagés, H., (1999). A Note on the Gordon Growth Model with Non-Stationary Dividend Growth. *Bank for international settlements (working paper, 75).* URL: www.bis.org/publ/work75.pdf

Pennacchi, G., (2007). *Theory of Asset Pricing* (1st). New York: Pearson

Phylaktis, K., & Xia, L., (2007). The Changing Roles of Industry and Country Effects in the Global Equity Markets, *The European Journal of Finance, 12*(8), 627-648. DOI: doi.org/10.1080/13518470500460202

Quance, S., (2018). *Invesco Global Factor Investing Study 2018.* URL*: apinstitutional.invesco.com/dam/jcr:4526a7ee-2da9-4f0f-a873-d96d13857d26/Factor-Investing_20181105_Invesco-Global-Factor-Investing-Study-2018.pdf*

Rabener, N., (2018a). *Factor Portfolios: Turnover Analysis – What is my portfolio churn?.* URL: www.factorresearch.com/research-factor-portfolios-turnover-analysis

Rabener, N., (2018b). *Factor construction: Portfolio Rebalancing*. URL: https://www.factorresearch.com/research-factor-construction-portfolio-rebalancing

Reinganum, M. R., (1981). Empirical Tests of Multi-Factor Pricing Model - The Arbitrage Pricing Theory: Some Empirical Results. *The Journal of Finance, 36*(2). DOI: doi.org/10.1111/j.1540-6261.1981.tb00444.x

Roll, R., (1977), A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory. *Journal of Financial Economics, 4*(2), 129–176. DOI: 10.1016/0304-405X(77)90009-5

Ross, S. A., (1976). The arbitrage theory of capital asset pricing. *The Journal of Economic Theory*, 13(*2*), 341-360. DOI: 10.1016/0022-0531(76)90046-6

Russell Investments, (2018). *Multi-Factor Investing, Demystified: Part 1*. URL: seekingalpha.com/

Sharpe, W., (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. Journal of Finance, 19(3), 425-442. DOI: doi.org/10.1111/j.1540-6261.1964.tb02865.x

Sharpe, W., (1966). Mutual Fund Performance. *The Journal of Business, 39*(1), 119-138. URL: www.jstor.org/stable/2351741

Shriber, T., (2019). *5 Quality ETFs That Qualify for Your Portfolio*. URL: finance.yahoo.com/news/5-quality-etfs-qualify-portfolio-191302590.html

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

Spicer, J., (Jan 23rd, 2009). NYSE lowers market cap requirement for listings. *Reuters*. URL:uk.reuters.com/article/us-nyse/nyse-lowers-market-cap-requirement-for-listings-idUKTRE50M4G920090123

Stock, J. H., Watson, M. W., (2013). *Introduction to Econometrics - The statistical analysis of economic (and related) data* (3rd). URL: www.sas.upenn.edu/~fdiebold/Teaching104/Ch1-7_slides.pdf

The benefits of quality investing. (2017). *Schroder.* URL: schroders.com/getfunddocument/?oid=1.9.2655049

The Economist, (Feb 1, 2018). Maxing the factors - “Factor investing” gains popularity. *The Economist.* URL: www.economist.com

Tofallis, C., (2008) Investment Volatility: A Critique of Standard Beta Estimation and a Simple Way Forward. *European Journal of Operational Research, 187,* 1358-1367. URL: arXiv:1109.4422

Treynor, J., (1961). Market Value, Time, and Risk. DOI: dx.doi.org/10.2139/ssrn.2600356

Uppaluri, V. S., (2019). *Micro-Caps Under the Microscope - An up-close look at common characteristics of the market’s smallest names*. URL: morningstar.com/articles/957944/micro-caps-under-the-microscope

Warren, B., & Quance, S., (2019). *Foundational concepts for understanding factor investing.* URL: www.invesco.com/us/insights/?audienceType=investors

What is the security market line? (2020). Tickeron.com URL: tickeron.com/app/academy/articles/alias/what-security-market-line

Wooldridge, J. M., (2013). *Introductory Econometrics – A Modern Approach* (5th). Mason: South-Western, Cengage Learning

Worldscope Database – Data Definitions Guide (Issue 14.3), (2015). *Thomson Reuters.* URL: datastream.jp/wp/wp-content/uploads/2015/10/guide_Worldscope-Data-Definitions-Guide-Issue-14.3.pdf

Vyas, K., & van Baren, M., (2019). A Novel Template for Understanding Priced Factors. URL: ssrn.com/abstract=3423566

11 Appendix

11.1 Summary tables: Pooled cross-sectional analysis and Fama-MacBeth regressions

Table 12: Summary of the pooled cross-sectional regressions per industry and quality metric

		GP	OP	ROA	ROE	ROIC	DTE	INV
Broad Sample	*Quality metric*	-0.07***	0.75***	1.09***	0.23***	0.76***	-0.01***	0.23***
	MV	-0.02***	-0.01***	-0.01***	-0.01***	-0.01***	-0.001	-0.01***
	BTM	-0.17***	0.17***	0.16***	0.12***	0.15***	0.12***	-0.13***
	Momentum	-0.07***	0.02***	0.001	0.01***	-0.001	0.04***	-0.06***
Technology	*Quality metric*	0.57***	1.27***	1.11***	0.44***	0.64***	0.02	0.33***
	MV	0.03***	-0.02***	-0.02***	0.01	-0.01***	0.2**	0.05***
	BTM	0.28***	0.25***	0.2***	0.26***	0.15***	0.19***	0.27***
	Momentum	0.02	-0.02	-0.06***	-0.03	-0.06***	-0.01	0.17**
Telecommunications	*Quality metric*	0.46***	0.17**	0.74***	0.16*	0.45***	-0.05***	0.18***
	MV	-0.001	-0.07***	-0.02***	-0.02***	-0.03***	-0.001	0.02***
	BTM	0.27***	0.24***	0.13***	0.12***	0.18***	0.14***	0.17***
	Momentum	0.04	0.01	-0.02	0.01	0.03	0.02	-0.001
Health Care	*Quality metric*	0.31***	0.24***	0.54***	0.15***	0.25***	-0.02	0.19***
	MV	0.001	-0.01***	-0.01*	-0.02***	-0.02***	0.22**	0.02***
	BTM	0.16***	0.13***	0.16***	0.14***	0.14***	0.15***	0.18***
	Momentum	0.02	0.02	0.01	0.02	0.02	0.04*	-0.01
Financials	*Quality metric*	0.08	0.46***	1.83***	0.08***	1.66***	-0.001*	0.28***
	MV	0.001	-0.07	-0.01***	-0.001	-0.001*	0.001	0.02***
	BTM	0.14**	0.05	0.18***	0.13***	0.21***	0.13***	0.15***
	Momentum	-0.03	-0.03	0.04***	0.05***	0.04***	0.07***	0.04***
Real Estate	*Quality metric*	0.27**	1.07***	0.79***	0.08***	1.02***	0.02***	0.2**
	MV	-0.001	-0.01	-0.001	-0.02***	-0.01	-0.001	0.03***
	BTM	0.11***	0.14***	0.14***	0.13***	0.12***	0.14***	0.15***
	Momentum	0.01	-0.02	-0.001	0.001	0.001	0.02	0.01
Consumer Discretionary	*Quality metric*	0.17***	1.16***	1.73***	0.04***	1.45***	-0.001	0.37***
	MV	0.02***	0.01*	0.001	0.01	0.001	0.01*	0.06***
	BTM	0.21***	0.27***	0.25***	0.17***	0.27***	0.18***	0.26***
	Momentum	0.05***	0.04*	0.36**	0.06***	0.01	0.06***	0.02
Consumer Staples	*Quality metric*	0.1***	1.06***	1.28***	0.48***	1.25***	0.01	0.28***
	MV	0.01	0.02***	-0.01*	-0.01	-0.02***	-0.001	0.01***
	BTM	0.1***	0.23***	0.12***	0.13***	0.17***	0.09***	0.1***
	Momentum	0.001	0.04	-0.01	-0.01	-0.01	0.04*	-0.01
Industrials	*Quality metric*	0.36***	1.48***	1.75***	0.35***	1.23***	0.01	0.31***
	MV	0.03***	0.01	0.01***	0.001	-0.001	0.17**	0.06***
	BTM	0.23***	0.25***	0.25***	0.18***	0.24***	0.18***	0.27***
	Momentum	0.05***	0.04***	-0.01	0.01	0.01	0.06***	0.01
Basic Materials	*Quality metric*	0.92***	2.62***	2.01***	0.34***	0.16***	0.03	0.22***
	MV	0.04***	0.01	0.01	-0.01	-0.02***	-0.04***	0.02***
	BTM	0.32***	0.36***	0.3***	0.23***	0.19***	0.08***	0.15***
	Momentum	-0.02	0.03	-0.03	0.01	0.03	0.02	-0.01
Energy	*Quality metric*	0.33***	0.54***	0.02***	0.15***	0.001***	-0.02	0.25***
	MV	-0.03***	-0.03***	-0.03***	-0.03***	-0.03***	-0.02***	0.001
	BTM	0.23***	0.12***	0.05***	0.03***	0.05***	0.05***	0.02***
	Momentum	0.4**	0.09***	0.07***	0.23**	0.09***	0.08***	0.01
Utilities	*Quality metric*	-0.06	0.77***	1.74***	0.41***	1.22***	-0.01	0.09
	MV	-0.001	0.38**	-0.01	-0.001	0.46**	-0.001	0.15**
	BTM	0.15***	0.12***	0.18***	0.15***	0.15***	0.14***	0.14***
	Momentum	0.02	0.001	0.01	0.01	0.01	0.03*	0.02

Coefficients are scaled by 10 $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
MV and BTM are logarithmic values

Table 13: Summary of the Fama-MacBeth regressions

		GP	OP	ROA	ROE	ROIC	DTE	INV
Broad Sample	*Quality metric*	-0.048***	0.62***	0.907***	0.163***	0.634***	-0.008***	0.189***
	MV	-0.135***	-0.01**	-0.007**	-0.006**	-0.008***	-0.001	-0.008**
	BTM	-0.157***	0.12***	0.118***	0.08***	0.103***	0.078***	-0.119***
	Momentum	-0.050***	0.02	-0.004	0.008	-0.01	0.026*	-0.049***
Technology	*Quality metric*	0.452***	1.15***	0.943***	0.35***	0.537***	0.005	3.03***
	MV	0.021***	-0.02**	-0.017***	0.005	-0.013**	0.007	0.034***
	BTM	0.200**	0.17***	0.125***	0.174***	0.089***	0.121***	0.18***
	Momentum	-0.001	-0.04	-0.08**	-0.032	-0.06**	-0.019	-0.041
Telcommunications	*Quality metric*	0.440***	0.68**	0.803***	0.162	0.458***	-0.041*	0.24***
	MV	-0.00	-0.05	-0.016***	-0.015**	-0.021***	0	0.001
	BTM	0.159***	0.01	0.053**	0.01	0.068*	0.07**	0.095***
	Momentum	-0.080	-0.09	-0.064	-0.01	-0.04	0.018	-0.057
Health Care	*Quality metric*	0.255***	0.19**	0.851***	0.142***	0.27***	-0.003	1.71***
	MV	-0.000	-0.02**	-0.034**	-0.014***	-0.018***	-0.009	0.14***
	BTM	0.120***	0.07***	0.202***	0.091***	0.079***	0.09***	0.122***
	Momentum	0.030	0.05	-0.002	0.009	0.012	0.031	0.001
Financials	*Quality metric*	-0.450	1.54	1.304***	0.047***	1.598***	-0.002	0.19***
	MV	-0.250*	0.55	-0.004	0	0	0.002	0.011***
	BTM	-0.410	0.79	0.111***	0.077***	0.145***	0.089***	0.09***
	Momentum	0.210	1.28	0.019	0.032**	0.02	0.041***	0.01
Real Estate	*Quality metric*	0.188	0.64**	0.785***	0.051**	0.799***	0.004	0.048
	MV	0.00	-0.01	0	-0.007	-0.003	-0.001	0.018***
	BTM	0.030*	0.06**	0.059***	0.03*	0.44***	0.046**	0.05***
	Momentum	0.023	-0.03	-0.014	-0.032	-0.027	-0.015	0
Consumer Discretionary	*Quality metric*	1.260***	0.843***	1.392***	0.028***	1.105***	-0.008	0.36***
	MV	0.020***	0.006	0.002	0.007	0.001	0.006	0.07***
	BTM	0.131***	0.19***	0.182***	0.108***	0.188***	0.119***	2.72***
	Momentum	0.030*	0.038	0.03	0.063**	0.008	0.061***	0.017
Consumer Staples	*Quality metric*	0.079**	0.897***	1.162***	0.517***	1.205***	-0.004	2.54***
	MV	0.014**	0.013	-0.006	-0.007	-0.014**	-0.004	0.012**
	BTM	0.078***	0.171***	0.107***	0.108***	0.139***	0.05***	0.078***
	Momentum	-0.009	0.023	-0.011	-0.023	-0.016	0.016	-0.022
Industrials	*Quality metric*	0.28***	1.209***	1.481***	0.232***	1.013***	0.003	0.239***
	MV	0.015***	0	0.003	-0.004	-0.004	0.004	0.037***
	BTM	0.15***	0.164***	0.168***	0.101***	0.164***	0.101***	0.16***
	Momentum	0.04*	-0.009	-0.02	0.012	-0.015	0.022	0.002
Basic Materials	*Quality metric*	0.69***	3.245***	1.685***	0.297***	0.174***	0.019	0.3***
	MV	0.02**	-0.069	0.011	-0.012	-0.012	-0.024***	0.16**
	BTM	0.2***	0.894***	0.212***	0.135***	0.12***	0.037**	0.1***
	Momentum	-0.001	0.111***	-0.013	0.025	0.039	0.057	0.01
Energy	*Quality metric*	0.62***	0.608***	0.022***	0.128***	0.001	-0.015	1.4***
	MV	-0.02	-0.01	-0.013**	-0.012**	-0.012**	-0.009	0.001
	BTM	0.172***	0.092***	0.033***	0.016**	0.025*	0.028**	0.013*
	Momentum	0.015	0.061	0.033	-0.02	0.247	0.073*	0.01
Utilities	*Quality metric*	-0.18	0.686***	1.779***	0.601***	1.325***	-0.023	0.1
	MV	0.004	-0.132*	-0.008*	-0.003	-0.01**	-0.001	0.001
	BTM	0.083**	0.071*	0.109***	0.113***	0.121***	0.073**	0.08***
	Momentum	-0.01	0.01	-0.002	0	-0.015	0.011	0.02

Coefficients are scaled by 10 $\quad$ $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

MV and BTM are logarithmic values

11.2 Summary tables: Portfolio analysis

Table 14: Regressions of the benchmark portfolios on the market, size, value and momentum factors

		Quality α	Market β	SMB	HML	Momentum	Exp.Return	R^2
	High	0.523**	1.198***	0.462***	-0.51***	-0.177***	1.424	0.779
Technology	*Middle*	0.258	1.249***	0.749***	-0.398***	-0.167***	1.232	0.776
n=120	*Low*	0.525	1.434***	1.057***	-0.515***	-0.553***	1.661	0.705
	High-Low	-0.002	-0.236**	-0.595***	0.005	0.376***	-0.119	0.322
	High	0.521*	1.026***	0.686***	-0.05	-0.007	1.325	0.581
Telecommunications	*Middle*	0.336	1.138***	0.82***	-0.127	-0.097	1.236	0.618
n=24	*Low*	-0.017	1.167***	0.712***	-0.297	-0.426***	0.882	0.514
	High-Low	0.5	-0.141	-0.026	0.247	0.419***	0.2	0.085
	High	0.603***	0.782***	0.354***	-0.08	0.037	1.195	0.598
Health Care	*Middle*	0.777***	0.818***	0.582***	-0.137	-0.022	1.426	0.651
n=92	*Low*	0.722**	0.949***	0.974***	-0.442***	-0.155**	1.527	0.587
	High-Low	-0.1	-0.167**	-0.621***	0.361**	0.192***	-0.2	0.321
	High	0.481	1.038***	0.28*	-0.221	-0.202**	1.234	0.521
Financials	*Middle*	1.198**	0.928***	0.764***	0.379*	-0.063	1.927	0.389
n=7	*Low*	0.351	1.491***	0.499	0.378	0.085	1.43	0.32
	High-Low	0.1	-0.453*	-0.219	-0.599**	-0.287	-0.1	0.051
	High	0.479	0.829***	0.384***	0.36***	-0.096	1.083	0.523
Real Estate	*Middle*	0.512	0.886***	0.529***	0.557***	-0.223**	1.164	0.461
n=61	*Low*	0.452	0.998***	0.73***	0.423**	-0.397***	1.21	0.533
	High-Low	0.03	-0.169***	-0.347***	-0.063	0.3***	-0.1	0.318
	High	0.473**	0.887***	0.685***	0.076	-0.2***	1.169	0.724
Consumer Discretionary	*Middle*	0.349	0.985***	0.689***	0.404***	-0.2***	0.1	0.764
n=161	*Low*	0.358	1.149***	0.916***	0.549***	-0.4***	1.24	0.797
	High-Low	0.1	-0.262***	-0.231*	-0.473***	0.212***	-0.04	0.41
	High	1.234***	0.6***	0.502***	0.115	-0.007	1.713	0.46
Consumer Staples	*Middle*	0.659***	0.644***	0.128	0.229***	0.112**	1.113	0.493
n=55	*Low*	0.497**	0.709***	0.271**	0.218**	-0.091	1.009	0.538
	High-Low	0.7***	-0.109	0.231	-0.104	0.084	0.4	0.008
	High	0.523***	0.981***	0.641***	0.159*	-0.022	1.281	0.772
Industrials	*Middle*	0.474**	1.078***	0.781***	0.202**	-0.042	1.315	0.793
n=193	*Low*	0.394*	1.208***	0.942***	0.331***	-0.148***	1.339	0.849
	High-Low	0.1	-0.227***	-0.301***	-0.172***	0.126***	-0.03	0.396
	High	0.086***	9.821***	6.849***	2.075*	-0.348	0.016	0.633
Basic Materials	*Middle*	0.007	11.991***	7.277***	4.242***	-2.212***	0.01	0.733
n=52	*Low*	0.043	10.974***	6.355***	3.807***	-0.81	0.013	0.639
	High-Low	0.4	-0.115	0.049	-0.173	0.046*	0.2	0.035
	High	0.463	1.046***	0.563***	0.327*	0.006	1.247	0.444
Energy	*Middle*	0.443	1.112***	0.602***	0.256	-0.092	1.278	0.459
n=52	*Low*	0.377	1.108***	0.682***	0.445**	-0.154	1.21	0.488
	High-Low	0.1	-0.062	-0.12	-0.118	0.16***	0.02	0.027
	High	0.56***	0.643***	0.129	0.207**	0.169***	1.016	0.445
Utilities	*Middle*	0.371	0.578***	-0.038	0.101	0.143**	0.762	0.291
n=39	*Low*	0.514**	0.636***	0.056	0.19*	0.028	0.95	0.365
	High-Low	0.05	0.007	0.073**	0.017	0.141	0.03	0.034

*p<0.1; **p<0.05; ***p<0.01

* Dependent variable: Monthly excess stock returns (in percentage) on 3 portfolios sorted on basis of operating profitability, gross profitability., ROA, ROE, ROIC, investments and debt to equity

Table 15: Regressions of the equally weighted industry specific portfolios on the market, size, value and momentum factors

		Quality α	Market β	SMB	HML	Momentum	Exp.Return	R^2
	High	0.561***	1.196***	0.438***	-0.482***	-0.169***	1.458	0.781
Technology	*Middle*	0.251	1.287***	0.782***	-0.456***	-0.191***	1.258	0.767
n=120	*Low*	0.492	1.397***	1.048***	-0.485***	-0.537***	1.602	0.703
	High-Low	0.069	-0.201**	-0.61***	0.002	0.368***	-0.144	0.322
	High	0.499	1.12***	0.643***	-0.122	-0.011	1.363	0.587
Telecommunications	*Middle*	0.345	0.979***	0.86***	0.008	-0.036	1.139	0.609
n=24	*Low*	-0.004	1.233***	0.715***	-0.36*	-0.483***	0.941	0.544
	High-Low	0.503	-0.113	-0.072	0.238	0.471***	0.422	0.085
	High	0.618***	0.77***	0.308***	-0.156*	-0.002	1.196	0.621
Health Care	*Middle*	0.609***	0.842***	0.479***	-0.058	0.096*	1.261	0.64
n=92	*Low*	0.819**	0.938***	1.158***	-0.431***	-0.246***	1.639	0.614
	High-Low	-0.201	-0.167**	-0.85***	0.275**	0.244***	-0.442	0.321
	High	0.361	1.034***	0.291**	-0.278*	-0.223***	1.114	0.525
Financials	*Middle*	1.208***	0.912***	0.477***	0.45***	0.02	1.884	0.426
n=7	*Low*	-0.019	1.421***	0.704**	0.398	-0.021	1.039	0.314
	High-Low	0.38	-0.387*	-0.413	-0.676**	-0.202	0.076	0.051
	High	0.536	0.819***	0.437***	0.391***	-0.12	1.14	0.482
Real Estate	*Middle*	0.427	0.859***	0.454***	0.543***	-0.185*	1.051	0.459
n=61	*Low*	0.429	1.059***	0.695***	0.43**	-0.396***	1.223	0.55
	High-Low	0.107	-0.24***	-0.258***	-0.04	0.276***	-0.084	0.318
	High	0.488**	0.891***	0.689***	0.052	-0.17***	1.19	0.729
Consumer Discretionary	*Middle*	0.287	1.013***	0.704***	0.395***	-0.22***	1.058	0.759
n=161	*Low*	0.404	1.133***	0.896***	0.588***	-0.441***	1.27	0.792
	High-Low	0.084	-0.242***	-0.207*	-0.536***	0.271***	-0.08	0.41
	High	1.22***	0.633***	0.381***	0.142	0.022	1.703	0.461
Consumer Staples	*Middle*	0.57***	0.64***	0.2**	0.14	0.045	1.033	0.506
n=55	*Low*	0.596**	0.647***	0.38***	0.299***	-0.033	1.081	0.523
	High-Low	0.623***	-0.015	0.001	-0.157	0.055	0.622	0.008
	High	0.534***	1.006***	0.643***	0.151*	-0.02	1.308	0.769
Industrials	*Middle*	0.445**	1.095***	0.72***	0.205**	-0.035	1.289	0.793
n=193	*Low*	0.428**	1.169***	1.009***	0.338***	-0.168***	1.355	0.852
	High-Low	0.105	-0.163***	-0.367***	-0.186***	0.147***	-0.046	0.396
	High	0.601*	1.08***	0.57***	0.202	-0.036	1.413	0.587
Basic Materials	*Middle*	0.351	1.122***	0.75***	0.447***	-0.131**	1.204	0.776
n=52	*Low*	0.395	1.092***	0.72***	0.372***	-0.155*	1.226	0.631
	High-Low	0.206	-0.012	-0.15	-0.17	0.12*	0.188	0.035
	High	0.594	1.104***	0.586***	0.362*	0.038	1.421	0.449
Energy	*Middle*	0.42	1.101***	0.67***	0.223	-0.166	1.256	0.458
n=52	*Low*	0.214	1.103***	0.586***	0.358*	-0.136	1.034	0.486
	High-Low	0.38	0.001	0.001	0.004	0.173***	0.386	0.027
	High	0.478**	0.669***	0.153	0.177*	0.133**	0.956	0.445
Utilities	*Middle*	0.548**	0.517***	0.054	0.117	0.14**	0.909	0.282
n=39	*Low*	0.419	0.671***	-0.06	0.203*	0.067	0.862	0.356
	High-Low	0.06	-0.001	0.213**	-0.026	0.066	0.094	0.034

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Dependent variable: Monthly excess stock returns (in percentage) on 3 portfolios sorted according the chosen industry specific quality metrics. Each quality metric is weighted equally.

Table 16: Regressions of weighted quality specific metric portfolios on market, size, value and momentum

		Quality α	Market β	SMB	HML	Momentum	Exp.Return	R^2
	High	0.61***	1.193***	0.506***	-0.488***	-0.172***	1.514	0.775
Technology	*Middle*	0.385	1.283***	0.802***	-0.45***	-0.215***	1.391	0.747
n=120	*Low*	0.31	1.405***	0.96***	-0.485***	-0.511***	1.413	0.716
	High-Low	0.3	-0.212***	-0.454***	-0.004	0.339***	0.1	0.317
	High	0.564	1.155***	0.674***	-0.096	-0.149*	1.452	0.592
Telecommunications	*Middle*	0.464	0.904***	0.759***	0.007	-0.042	1.192	0.531
n=24	*Low*	-0.189	1.273***	0.785***	-0.385**	-0.339***	0.8	0.583
	High-Low	0.753	-0.117	-0.111	0.289	0.189*	0.652	0.014
	High	0.638***	0.779***	0.288***	-0.12	0.001	1.218	0.609
Health Care	*Middle*	0.509**	0.829***	0.501***	-0.104	0.083*	1.157	0.654
n=92	*Low*	0.89***	0.934***	1.142***	-0.412***	-0.243***	1.705	0.612
	High-Low	-0.252	-0.155**	-0.854***	0.292**	0.244***	-0.486	0.317
	High	0.539	1.03***	0.359**	-0.374**	-0.237***	1.303	0.527
Financials	*Middle*	1.054***	0.836***	0.459***	0.483***	-0.005	1.672	0.398
n=7	*Low*	0.031	1.453***	0.633**	0.455	0.13	1.103	0.314
	High-Low	0.508	-0.423**	-0.275	-0.828***	-0.367**	0.2	0.067
	High	0.478	0.822***	0.398***	0.377***	-0.109	1.078	0.514
Real Estate	*Middle*	0.534	0.827***	0.526***	0.492***	-0.213**	1.148	0.425
n=61	*Low*	0.387	1.071***	0.656***	0.48***	-0.389***	1.181	0.547
	High-Low	0.091	-0.249***	-0.258**	-0.103	0.28***	-0.102	0.291
	High	0.478**	0.9***	0.687***	0.064	-0.19***	1.185	0.736
Consumer Discretionary	*Middle*	0.33	1.001***	0.7***	0.389***	-0.198***	1.094	0.748
n=161	*Low*	0.372	1.135***	0.903***	0.581***	-0.444***	1.239	0.796
	High-Low	0.106	-0.234***	-0.216**	-0.517***	0.254***	-0.054	0.391
	High	1.222***	0.63***	0.379***	0.134	0.027	1.703	0.456
Consumer Staples	*Middle*	0.572***	0.645***	0.206**	0.147*	0.042	1.039	0.518
n=55	*Low*	0.592**	0.646***	0.375***	0.301***	-0.034	1.074	0.522
	High-Low	0.63***	-0.016	0.003	-0.167	0.06	0.628	0.012
	High	0.541***	0.998***	0.653***	0.153*	-0.021	1.311	0.768
Industrials	*Middle*	0.427**	1.107***	0.725***	0.204**	-0.043	1.28	0.797
n=193	*Low*	0.437**	1.168***	0.99***	0.344***	-0.158***	1.361	0.851
	High-Low	0.104	-0.171***	-0.337***	-0.191***	0.137***	-0.05	0.399
	High	0.644**	1.044***	0.574***	0.178	-0.049	1.433	0.595
Basic Materials	*Middle*	0.317	1.088***	0.743***	0.335***	-0.137**	1.151	0.757
n=52	*Low*	0.486	1.181***	0.659***	0.651***	-0.163**	1.356	0.658
	High-Low	0.159	-0.137*	-0.084	-0.473***	0.114	-0.026	0.151
	High	0.557	1.055***	0.716***	0.362*	0.007	1.369	0.46
Energy	*Middle*	0.385	1.162***	0.537***	0.325	-0.115	1.24	0.453
n=52	*Low*	0.362	1.021***	0.625***	0.339*	-0.154	1.132	0.486
	High-Low	0.195	0.034	0.091	0.023	0.161**	0.236	0.015
	High	0.586***	0.629***	0.106	0.166*	0.119**	1.029	0.411
Utilities	*Middle*	0.576**	0.54***	-0.011	0.133	0.158***	0.944	0.287
n=39	*Low*	0.283	0.688***	0.051	0.199*	0.063	0.755	0.379
	High-Low	0.303	-0.059	0.055	-0.033	0.057	0.274	0.018

*p<0.1; **p<0.05; ***p<0.01

* Dependent variable: Monthly excess stock returns (in percentage) on 3 portfolios sorted according the chosen industry specific quality metrics. Each quality metric is weighted based on the relative magnitude of its coefficient.

11.3 Annual Sharpe ratios of the quality sorted High-Low portfolios

Figure 6: Telecommunications annual High-Low quality portfolios Sharpe ratios

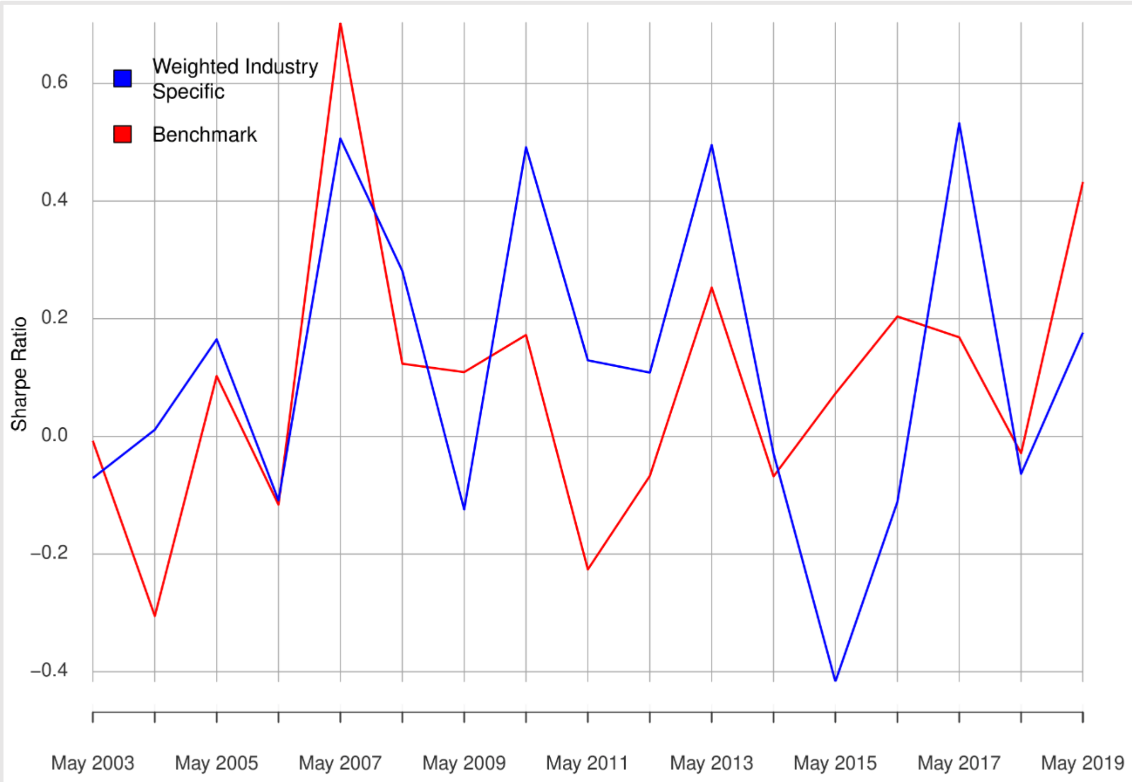


Figure 7: Health Care annual High-Low quality portfolios Sharpe ratios

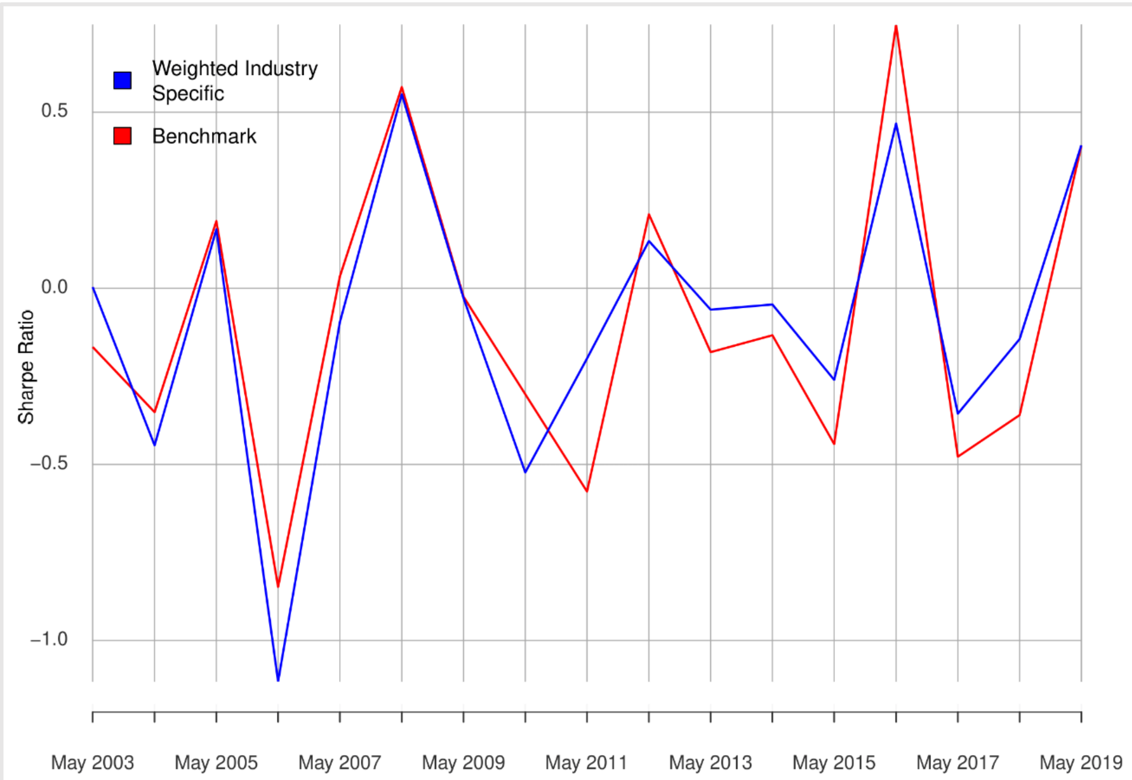


Figure 8: Financials annual High-Low quality portfolios Sharpe ratios

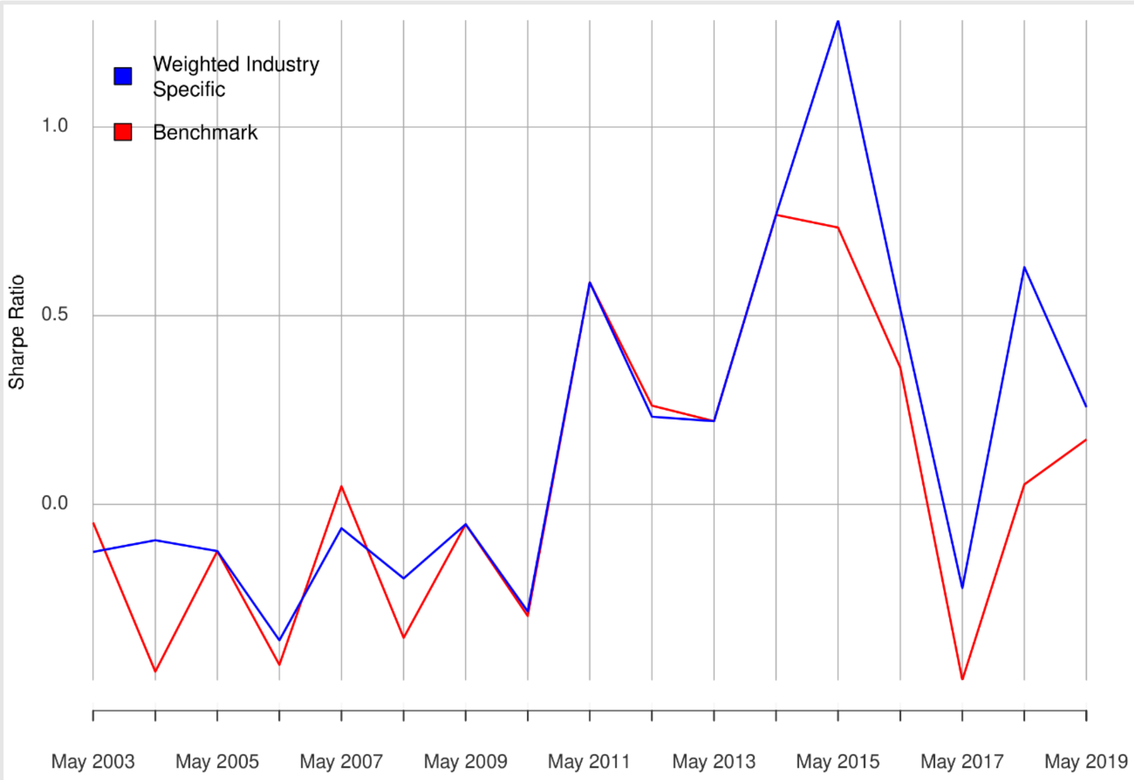


Figure 9: Real Estate annual High-Low quality portfolios Sharpe ratios

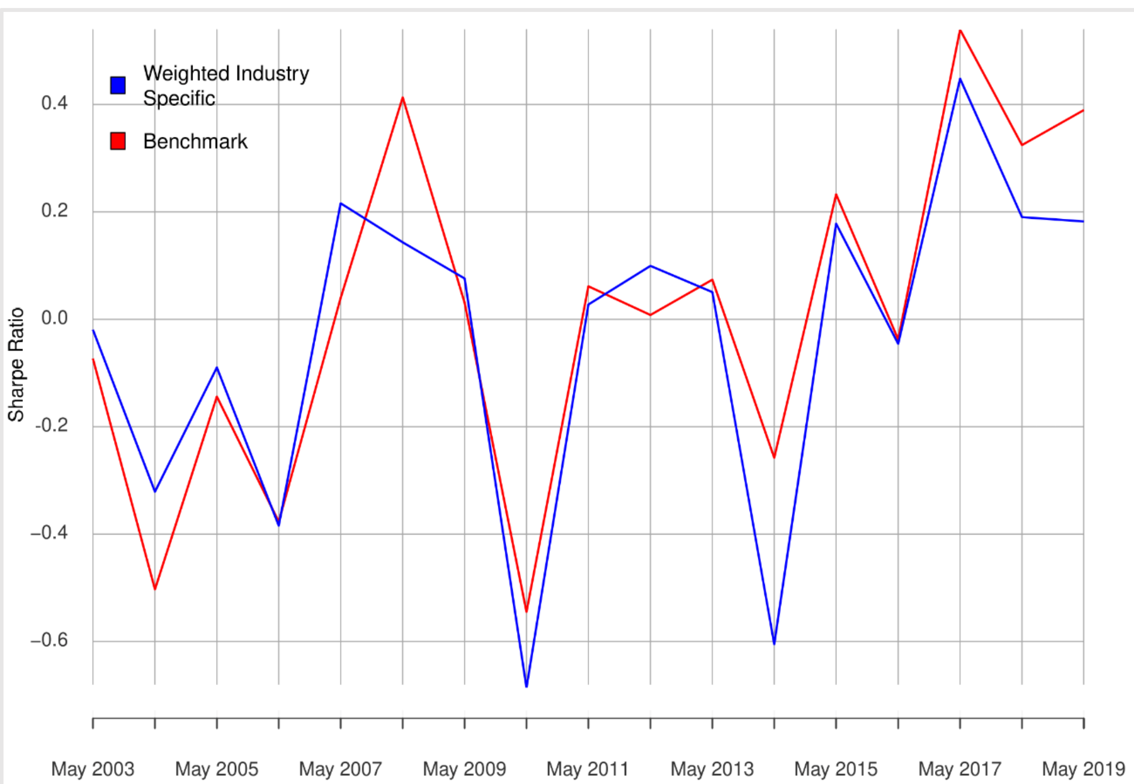


Figure 10: Consumer Discretionary annual High-Low quality portfolios Sharpe ratios

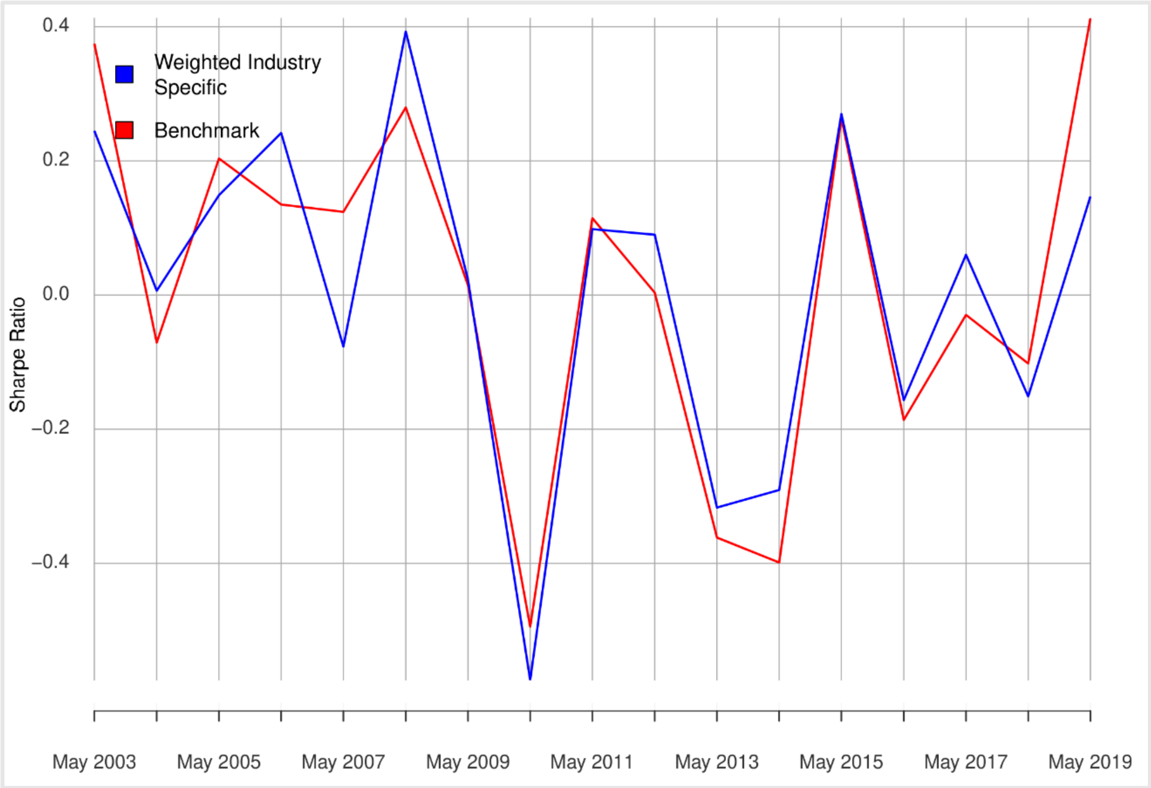


Figure 11: Consumer Staples annual High-Low quality portfolios Sharpe ratios

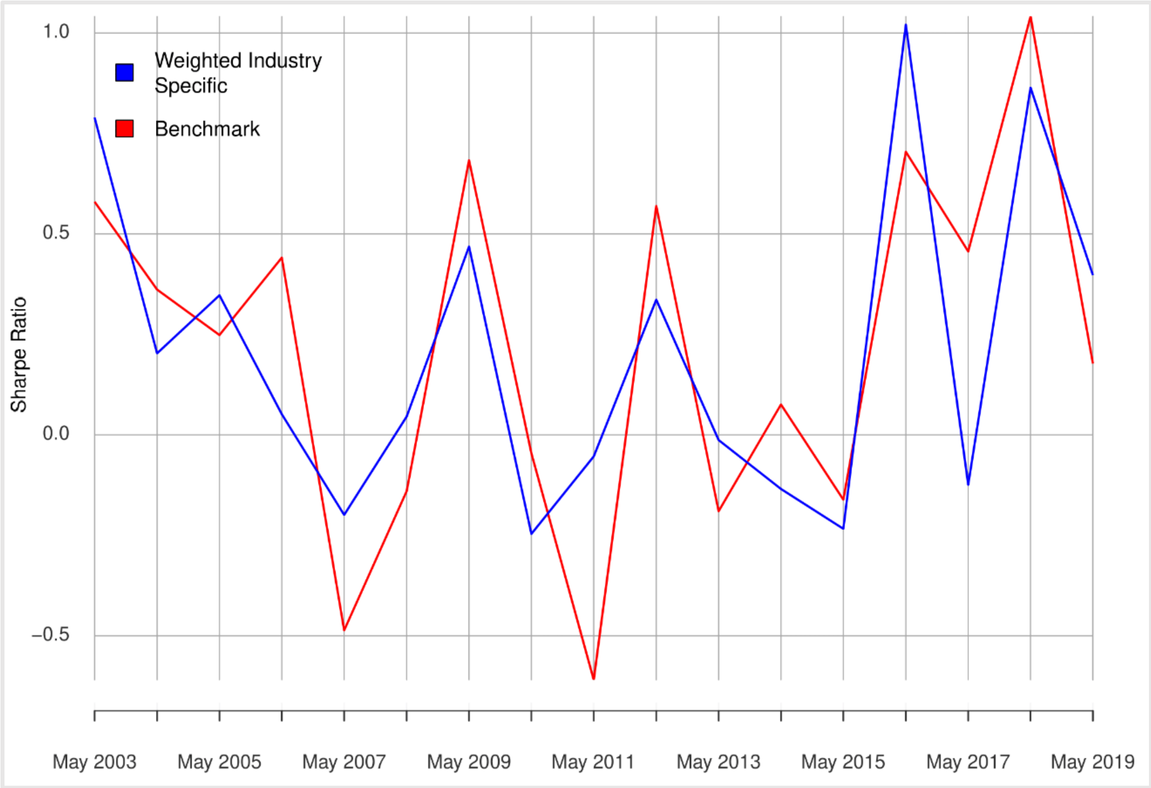


Figure 12: Basic Materials annual High-Low quality portfolios Sharpe ratios

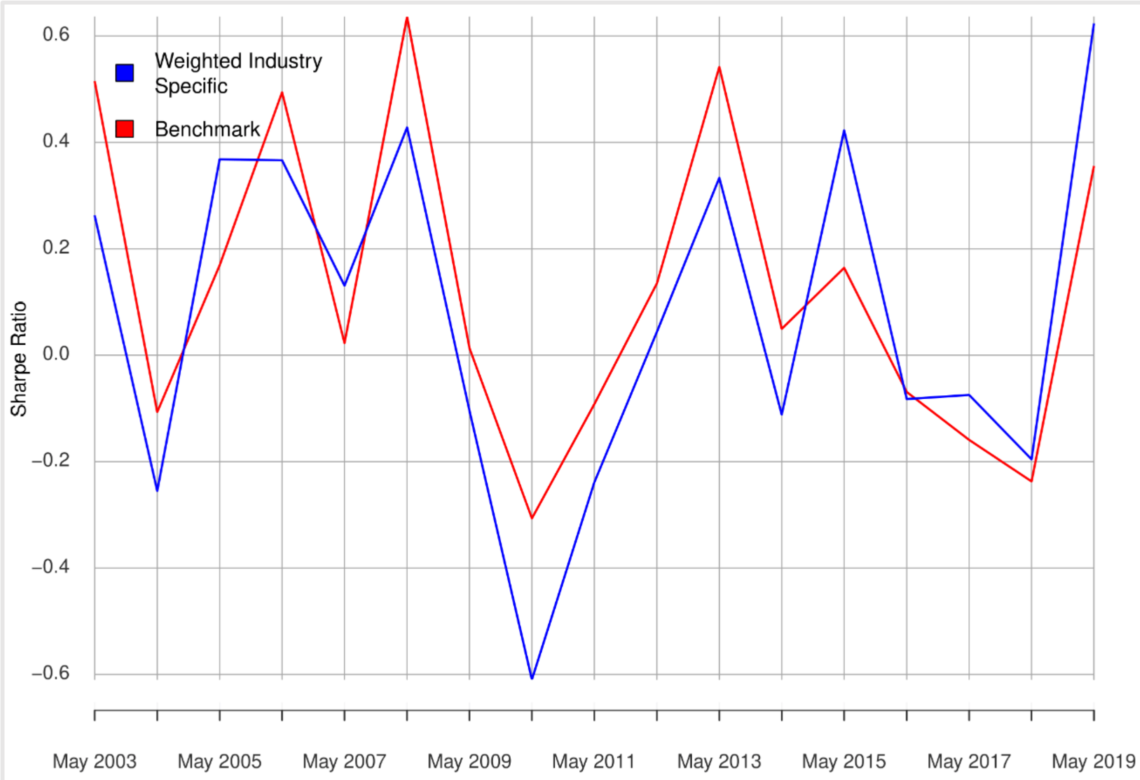


Figure 13: Industrials annual High-Low quality portfolios Sharpe ratios

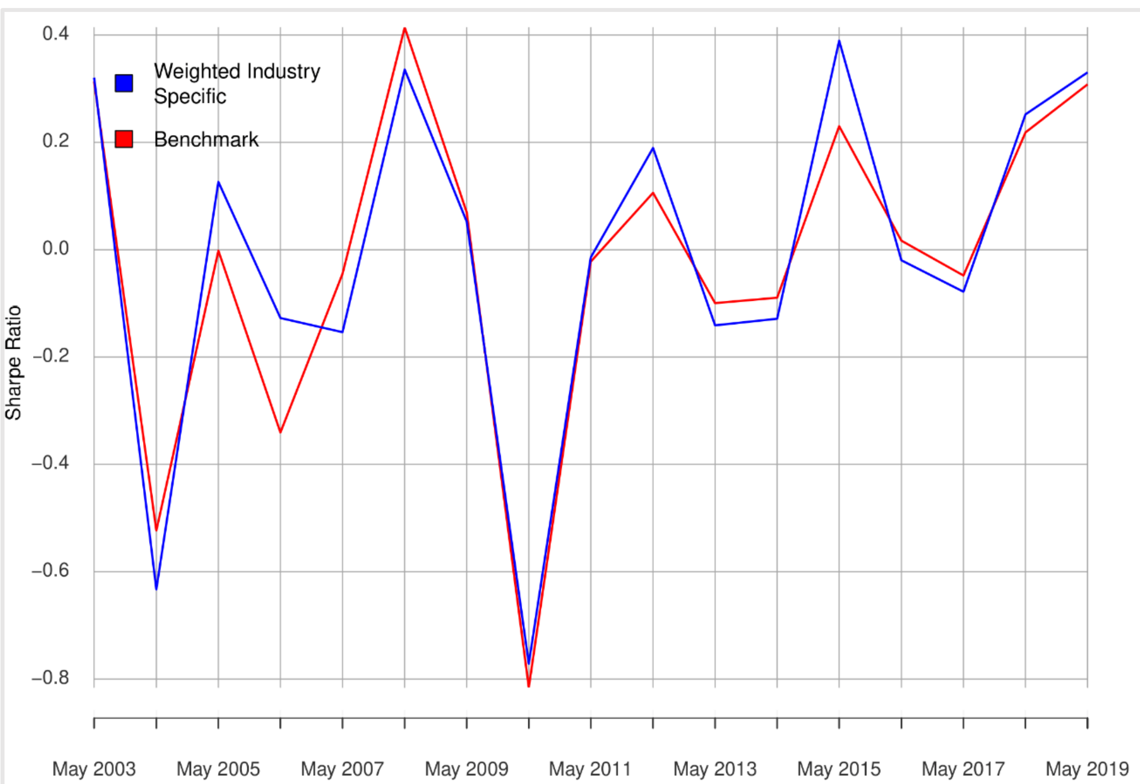


Figure 14: Energy annual High-Low quality portfolios Sharpe ratios

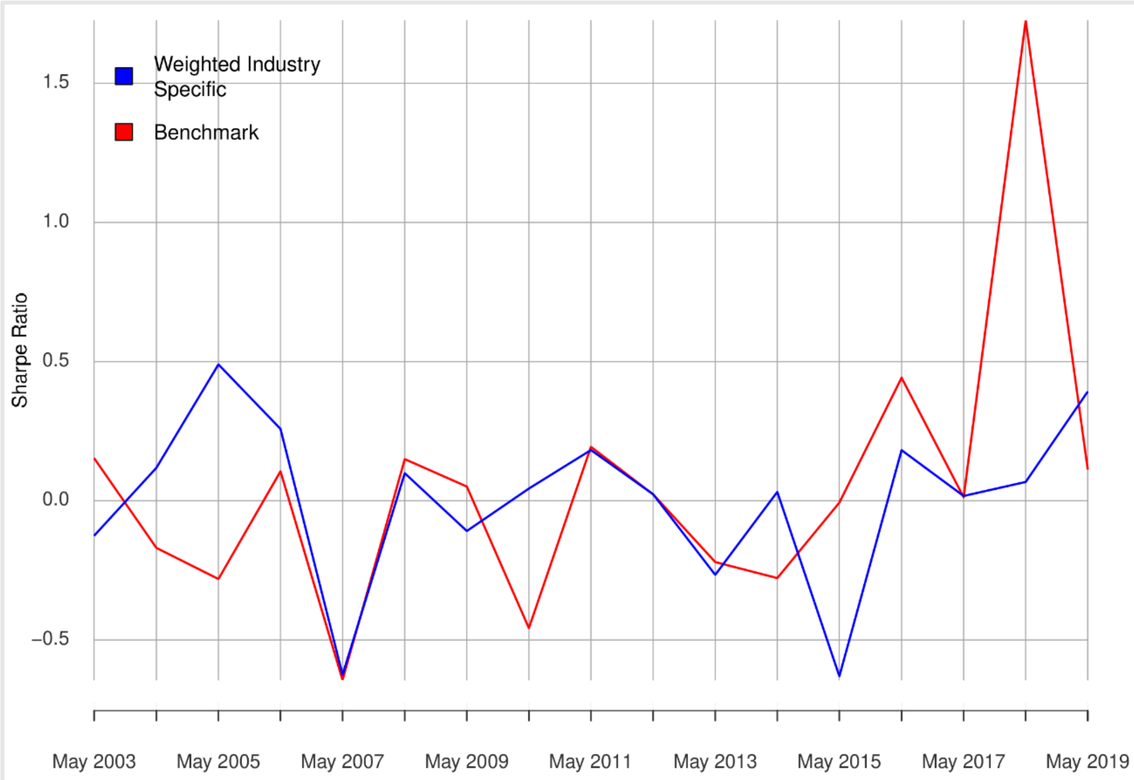


Figure 15: Utilities annual High-Low quality portfolios Sharpe ratios