Working Paper 113-2021 | **June 2021**

Revisiting Quality Investing

Confidence
must be earned

Amundi
ASSET MANAGEMENT

Revisiting Quality Investing

Frédéric LEPETIT
Quantitative Research
frederic.lepetit@amundi.com

Amina CHERIEF
Quantitative Research
amina.cherief@amundi.com

Yannick LY
Quantitative Research
yannick.ly@amundi.com

Takaya SEKINE
Quantitative Research
takaya.sekine@amundi.com

Abstract

In the field of factor investing, quality is undoubtedly the equity factor with the weakest consensus. This research investigates the best way to define it. In order to capture the multi-faceted reality of the factor depicted in academia, we address the quality factor through a multidimensional process by defining four self-reliant pillars: *profitability*, *earnings quality*, *safety* and *investment*. To better fit institutional investor's' needs, we analyze the resulting factor by focusing on the last eighteen years and on a global developed markets universe of liquid stocks (large- and mid-caps).

In a long-short framework, our quality factor delivers a statistically significant alpha that cannot be explained by loadings on conventional equity factors (market, value, size and momentum). Most regions and dimensions display positive contribution to this alpha, with the noticeable exceptions of the Eurozone region and the safety dimension. In a long-only framework, our quality factor outperforms its benchmark by 2.8% per annum over the entire analysis period, with an information ratio of 0.81. Furthermore, the outperformance has been very consistent since the 2008 Global Financial Crisis (GFC). The four dimensions are weakly correlated with each other and are therefore complementary. We show that safety is of particular importance during periods of market turmoil (GFC, Covid-19 pandemic) and that the dimension is therefore part of the quality factor in its own right. On the Eurozone side, a sector-neutral portfolio construction seems to be more suited.

We also introduce a new portfolio construction methodology by implementing a clustering approach based on the K-means algorithm to group together companies based on features that are related to both fundamentals and market characteristics. This approach allows to capture dynamic variations between fundamentals and other stock features. This fully implementable process results in better quality factor performance without impacting the associated risk measures or the portfolio's quality exposure, as measured on the unconstrained quality factor.

Keywords: Quality, Factor investing, Profitability, Earnings quality, Accruals, Safety, Capital structure, Investment, Asset pricing, Clustering, K-Means algorithm, Gaussian Mixture Models, Covid-19, Global Financial Crisis.

JEL classification: C30, G11, G12

Acknowledgement

The authors are very grateful to Thierry Roncalli, Lauren Stagnol, Edmond Lezmi, Théo Le Guenedal, Bruno Taillardat and Alessandro Russo from Amundi for their helpful comments; as well as to Henri Loustalot, Arnorld Chambon, Christelle Cottier and Liam Hynes from S&P Global Market Intelligence for their valuable support on the fundamental dataset provided by S&P.

About the authors

Frédéric Lepetit

Frédéric Lepetit was appointed Head of Equity Quantitative Research at Amundi in January 2016. He joined Société Générale Asset Management in 2006 as Quantitative analyst on the Equity side and expanded the scope of his activity to the volatility asset class area after the SGAM-CAAM merge in 2010.

Prior to joining Amundi, he was consultant to French leading Asset Management companies at FactSet Research Systems from 1999 to 2006. Frederic holds a Master degree in economics and asset management from Sorbonne University.

Amina Cherief

Amina Cherief joined the Quantitative Research team of Amundi in December 2020. Prior to that, Amina was a Cross-Asset Financial Engineer at SG CIB in New York (2019-2020), a Financial Engineer at Natixis AM (2017-2019) and an intern in Amundi's Quantitative Research dedicated to development of a multi-factor risk and performance analysis tool (2017). Her areas of research cover factor investing, sustainable investing and AI-ML in both fixed-income and equities. Amina holds a Master Degree (with honors) in Risk and Asset Management from Paris-Saclay University.

Yannick Ly

Yannick Ly joined Amundi in July 2020 in the Quantitative Research department. He worked on the quality factor and the impact of unsupervised classification algorithms in portfolio construction. Yannick graduated from ENSAE with a Master degree in Data Science. He also holds a Master degree in Quantitative Finance from Paris Dauphine University.

Takaya Sekine

Takaya Sekine joined Amundi in 2000 and is Deputy Head of Quantitative Research since July 2018. Prior to that, he was Deputy CIO at Amundi Japan (between 2011 and 2018), Head of Index and Multi-strategies at Amundi Japan (between 2010 and 2011), Fund Manager (between 2007 and 2010) and Financial Engineer (between 2001 and 2007). Takaya began his career as an IT Manager at Amundi Japan's predecessor company (between 2000 and 2001).

Takaya is a CFA charterholder since 2005, and an Associate member of the Association of Certified Fraud Examiners since 2010. He received the Ingénieur Civil des Mines degree from Ecole des Mines de Nancy in 2000. He has been involved in macroeconomic and policy related investment strategies.

Contents

1 Introduction

The quality factor has been ubiquitous in modern finance, sometimes as a stand-alone strategy, through more often a risk-factor combination, as demonstrated by the precursors Graham and Dodd (1934) who developed a strategy that exploited the quality dimension of value. Since the Global Financial Crisis (GFC), quality has experienced growing interest among both academic researchers and practitioners. High-quality companies are likely to outperform significantly in down-markets and can dampen sudden falls in market prices. These defensive characteristics are highly sought after in the uncertain environment we are experiencing following the GFC, dominated by low or negative interest rates, highly leveraged economies and structurally lower growth rates.

Despite this increasing interest, quality remains the factor with the weakest consensus among the traditional equity factors, namely size, value, momentum and low volatility. The main reason for this lies in the very roots of quality, which is entirely based on financial reporting data, whereas some other risk-factors feed on a mix of market and accounting data, and others only on market data. Since accounting records contain a wealth of data that can be combined to characterize a company's quality feature, the scope of possibilities is very broad and means that quality can be qualified in a multitude of ways. In a comprehensive survey of 150 publicly available factors, Feng *et al.* (2020) give a concrete illustration of this proliferation since more than half of these factors can be directly related to the quality factor.

Furthermore, although market participants agree on the existence of a quality premium, two views coexist and are opposed in terms of how to capture companies' quality features. On the one side, there is the purely academic approach, which relies on the largest universes and the longest possible historical data. Academic research systematically focuses on long-short strategies applied to a single metric to characterize a risk premium or a market anomaly. Usually, authors split the quality factor into different premia, such as profitability or quality of earnings for example, and treat them independently from one another. The multi-dimensional nature of quality makes it impossible to provide a unified explanation of the quality premium. While some attribute the latter to risk-factor, others to behavioral bias, we actually believe that the quality factor is a combination of both. On the other side, there are the practitioners who face multiple practical and regulatory constraints including liquidity, flexibility and the long-only framework. The two approaches both generally define quality by mixing several measures and consider the quality premium as a whole. Although the two approaches differ in many ways, they complement each other. In this research, we attempt to bring them together by mixing the strengths of academic research with the constraints of practitioners. In other words, we put into practice academic findings in investment processes.

First and foremost, we are convinced that it is unrealistic to capture the complexity of the quality factor through a single metric. Quality is multi-faceted and needs to be addressed through a multidimensional process. Therefore, we review the empirical evidence on risk premia as well as market anomalies highlighted by the academic literature that may be associated with the concept of quality. We consider quality measures that have been extensively researched and that have demonstrated their ability to predict cross-sectional differences in average stock returns. We group them into four distinctive dimensions, namely: *profitability, earnings quality, safety* and *investment*. Each of these dimensions is described by two quality metrics, carefully selected because of the intensive research they have undergone. The first part of the article shows some similarities with a couple of papers from practitioners – such as Asness *et al.* (2019) and Hsu *et al.* (2019), whose papers also analyze the quality factor through different dimensions. However, our article differs from their in that we focus only on liquid equity universes (MSCI Standard indexes) and on a recent period (from the beginning of the early 2000s value rally to May 2020). We believe these choices may be more suitable for institutional investors as they respond to scalability and liquidity constraints. By comparison, Asness *et al.* (2019) and Hsu *et al.* (2019) explore larger equity universes by covering all listed stocks for a given area, including smaller cap stocks, and go back to the early 1960s for US companies and to the mid 1980s for other developed markets. Another key difference lies in how measures are selected within the dimensions. While we aim to limit our exploration to the most documented metrics, other papers generally select a broader set of measures. By doing so, our objective is obviously not to build the best quality factor ever seen, but rather to assess the most relevant quality metrics on an investable investment universe and on a contemporary sample period.

In this study, we present new evidence on the existence of a quality premium on a large- and mid-caps investment universe. From 2003 to 2020, a long-short portfolio invested in global developed markets, long on the top 20% high quality stocks and short on the bottom 20% low quality stocks, yields a 5.0% return on an annual basis. By excluding the value rally period from the study, which by definition is less favorable to the quality risk-factor, the portfolio spread reaches 8.2%, with a risk-adjusted return of 1.34. Interestingly, the performance is driven both by the best- and the worst-ranked stocks. Also, portfolios sorted based on our quality score exhibit an almost perfectly decreasing profile all along the sorted portfolios.

We show that the performance is not driven by a single region or dimension. On the regional side, all of them contribute positively to the outperformance, with the notable exception of the EMU area, which fails to capture a quality premium during the observation period. On this specific area, profitability is the only dimension to deliver a statistically significant alpha, especially thanks to the gross profit-to-assets metric. Regarding dimensions at the world developed markets level, we find that the quality premium is especially strong on profitability and earnings quality, weaker on investment and non-existent on safety.

We also point out that the choice of metrics is paramount. A metric may have been the subject of broad academic research and have repeatedly demonstrated its ability to predict the cross-section of expected returns but may no longer be able to do so over a given period and/or on a given universe. Either because the existence of the premium repeatedly demonstrated was over very long periods that no longer reflect today's market, or because the premium was in fact concentrated on a market segment (on small caps in most cases). Of course, the explanation that the rewards are arbitraged away as the academic findings become widely known, as shown by McLean and Pontiff (2016), should not be rejected either. Among the eight well-known metrics we retain in this study, we note that the accounting accruals anomaly, as measured by Sloan (1996), fails to generate statistically significant alpha in global developed markets. The same applies to the total asset growth metric, while it benefits from a certain degree of consensus, as evidenced by its recent integration into the five-factor model of Fama and French (2015) as the investment factor. That does not mean that the premia associated with these anomalies no longer exist, but rather that it is essential to have in mind that all metrics within a given dimension are not equal and should be carefully selected according to the investment universe and period. Regarding the profitability metrics, gross profit-to-assets and cash flow return on equity fully succeed in generating statistically significant alpha in global developed markets. As part of the earnings quality dimension, accruals measured by the cash flows method are more successful at generating statistically significant alpha than the Sloan's method. We observe a similar dichotomy on the investment dimension between the capex scaled by sales metric that generates consistent excess returns and the total asset growth over one year that does not. Regarding the safety dimension, long-term debt-to-equity uniquely provides strong downside protection when markets are experiencing turbulence, while remaining insensitive in more quiet times. Finally, working capital-to-assets, which is the second metric we use to define safety, generates a strong excess return but seems to be replicable in a long-short framework using a combination of market, value and size equity factors.

To meet the needs of most practitioners and institutional investors, we also analyze the quality factor in a long-only framework. To have a better understanding of the behavior of the risk-factor, we divide our study period into two sub-periods. The first is centered on the value rally that followed the dotcom bubble, while the second starts at the beginning of the GFC. In addition to the fact that these two periods are different from an economic cycle point of view, it seemed appropriate to focus specifically on a period characterized by a paradigm shift. Indeed, the recent period is defined by low economic growth, low (or even negative) interest rates and financial markets under the perfusion of central banks. Since it appears that this environment is set to last for the long-term, it is crucial to analyze carefully this sub-period. As expected, our quality factor underperforms during the first sub-period, outperforms during the second and proves to be a good hedge against the market in times of crisis. We note in particular a consistent and smooth outperformance of 4% on an annual basis since the GFC, with an information ratio of 1.32. Beyond purely performance-related criteria, we examine how the four dimensions interact over time and show how they naturally complement each other in the various market phases.

We also investigate a new portfolio construction methodology based on unsupervised machine learning methods. In a traditional factor-based framework, investors may generally choose between unconstrained or sector-neutral portfolio constructions, depending on their mandate constraints. In the first case, no consideration is given to sector membership and stocks that are the most favorably exposed to the targeted risk-factor feed the long portfolio. This may result in strong sector biases versus the benchmark, but they are presumed to be assumed. In the second case, a given stock is systematically compared to its peers and the investor picks the companies that best fit the risk-factor targeted within each industry, in such a way that sector biases are eliminated. The notion of peers is essential here and generally relies on a business activity taxonomy that groups together companies with a common dominant sector of activity. However, there is sometimes more commonality between two companies that belong to different sectors than between companies classified in the same economic sector of activity. In this study, we explore a third solution, halfway between the unconstrained and the sector-neutral approaches. We emancipate ourselves from the traditional sector-neutral approach and seek to group companies based fundamental and market features using unsupervised learning. We focus on the K-means and Global Mixture Models algorithms and explain the different stages of our decision-making process in aiming to set up a robust investment framework. We show that the K-means algorithm is the most suitable for our exercise and we implement our solution on our multidimensional quality factor.

The clustering implementation brings a dynamic framework which is not exclusively fundamental but which includes the hybrid interdependence between fundamentals and other stock features. We analyze its impact on the composition, performance and risk statistics of our equity factor. The quality portfolio built with clusters is significantly different from the unconstrained portfolio, with an average overlap of 75% for the full sample. A large part of the composition gap is attributable to stock selection since we show that the sector allocation is very slightly impacted by the cluster implementation stage. Interestingly, these differences generate only a negligible deterioration in exposure to the aggregated quality metrics we have selected to define the quality factor, which is very satisfactory insofar as the cluster-neutral management constitutes a form of constraint comparable to that of sector-neutral management. From a performance point of view, the cluster-based quality portfolio exhibits higher performances and risk-adjusted performances than our unconstrained quality portfolio. At the global level, this translates into a performance improvement of 70 basis points on an annual basis, while leaving the associated level of risk unchanged. The region that benefits the most from the increase in performance is the Eurozone, with a meaningful +240 basis points per year.

This article is organized as follows. Section 2 provides the rationale for the investment universe and the historical depth we retain in the study. We also present the quality score and portfolio construction methodologies. In Section 3, we review academic literature on the dimensions and metrics that define our quality factor. Section 4 is dedicated to the analysis of our factor. First, we investigate the existence and robustness of the premium across regions and dimensions, before examining the portfolios sorted by each of the selected metrics. We also consider the quality risk-factor in a long-only framework. In Section 5, we propose an innovative portfolio construction methodology through the adoption of clustering techniques, and we analyze the benefits of such an approach. Finally, Section 6 offers some concluding remarks.

2 Methodology and Data

2.1 Global specifications

Investment Universe The academic standard for risk-factors studies is to encompass all listed companies from a given area in an attempt to characterize a risk premium or a market anomaly according to the size of the market capitalization of firms (large versus small companies). This implies considering securities extremely difficult to trade, that do not meet the liquidity requirements that are inherent in managing financial assets. In addition, as shown by Hou *et al.* (2015) in a review of nearly 80 anomalies, many factor premia seem exaggerated, likely by excessively weighting on microcaps. Fama and French (1992), for instance, underline that value premium in average stock returns decrease with size. To the extent that transaction costs and bid-ask spreads are inversely proportional to the size, it is likely that a large part of the premium will disappear in implementation costs. In this research, we focus our attention on equities that belong to the MSCI standard indices on developed markets, which are widely used in the asset management industry[1]. For a given market, the related index accounts for 85% of the total free float adjusted market capitalization (+/- 5%) and this roughly corresponds to the large- caps segment definition of academic papers. By doing so, we replicate the formal investment universes of the vast majority of institutional investors. Furthermore, in addition to an analysis of the MSCI World DM as a whole, we split the index into five regions: North America, EMU, Europe ex-EMU, Japan and Pacific ex-Japan to assess the robustness of our quality factor across different geographic areas. We exclude financial and real estate firms from our analysis as the fundamental metrics used to define the companies' quality characteristics are not applicable within these sectors.

Historical depth of the study Academic research generally goes back to the early 1960s for US companies and to the middle of the 1980s for other developed markets, due to the unavailability of fundamental data for the earlier period. A long history is a prerequisite to correctly ascertain the significance of a risk premium or a market anomaly and assess its robustness over time. In this article, we address the last eighteen years, from the beginning of the value rally in the early 2000s to May 2020. This historical depth matches our dual commitment: to give a contemporaneous view of the quality factor's behavior and to illustrate its emergence as an individual factor on its own right.

To shorten the observation period in this way is not without consequence as it concretely means cutting off a substantial part of the premium. In the first place because academic research can have a bias against stock return predictability, as shown by McLean and Pontiff (2016), who evaluate the post-publication decay at about 32%. When investors are informed of the benefits of a strategy from published research, the rewards are arbitraged away as the findings become widely known. Secondly, because all aspects of the asset management industry (regulation, technology, people skills, etc.) have drastically evolved over the past few decades, which has contributed to greater transparency and efficiency through increased accounting standards and publication frequencies. This continuously improving environment benefits the decision-making for all market participants and limits opportunities as well-informed investors increasingly harvest risk premia. In other words, one is more likely to generate excess returns by simulating a strategy based on a market anomaly or a risk premium which originated in the 1970s or 1980s than it is on a well-documented one in the 2010s.

Long-short vs. long-only frameworks Another common feature of most of the academic studies is that authors focus mainly on long-short strategies to analyze equity factor premia. This framework allows analyzing factor premia in a simple understandable process. If there is no absolute rule regarding the way equity factor strategies are implemented, long-short strategies are more commonly used by hedge funds and academics while institutional investors generally implement long-only frameworks, or partially-hedged by going long equity and shorting the market with futures to offset some of the market risks. The specific costs associated with shorting, the mandate constraints and the difficulty, if not the impossibility, of shorting certain stocks in practice are often a barrier for long-short implementation. To meet

[1] For more information on the construction of the MSCI indices: https://www.msci.com/eqb/methodology/meth_docs/MSCI_GIMIMethodology_Nov2020.pdf.

the needs of traditional asset managers, this research focuses on long-only strategy results to complement those obtained on the long-short framework.

2.2 Data sources

Our investment universe is based on the MSCI Developed Markets index, and the market capitalizations we consider to weight companies in our portfolio construction process are the free float-adjusted market capitalizations from the MSCI database. Stock returns and fundamental data are from the S&P Market Intelligence databases. All the fundamental metrics we consider in this article come from the S&P Global Market Intelligence's Alpha Factor Library. The library provides daily and monthly fundamental data items with point-in-time data relying on the Compustat dataset for North American companies and on the Capital IQ dataset for others regions.

2.3 Quality scores and portfolio construction

Quality scores We define quality along four dimensions: *profitability, earnings quality, safety* and *investment*. Each of these dimensions is described by two fundamental metrics, carefully selected because of the intensive research they have undergone and the robust statistical significance of their associated return premium. We will discuss the arguments behind our choices in the next section. At the end of each quarter, we convert each fundamental metric into percentiles, ensuring that the highest percentile is allocated to the highest quality company according to this metric. We then consider the obtained percentiles as probabilities and, for each stock on a given metric, we compute the test statistic for which the cumulative function of the standardized normal distribution is less than or equal to the probability argument. In other words, we convert percentiles into z-scores by using an inverse normal distribution.

At the metric level, the quality score of a given equity is equal to the z-score. At the dimension level, we combine the two quality metrics that characterize the dimension by averaging their percentiles and we compute a new percentile for the dimension. Then we convert the obtained percentile into a z-score in a similar way to what is applied at the metric level. At the multidimensional level, we average the four quality dimensions' z-scores into a single quality score.

Portfolio construction We form a set of non-sector-neutral (unconstrained) portfolios on a quarterly basis in each of our five regions. Each set consists of thirteen subsets of portfolios (eight for individual metrics, four for dimensions, and one combining the four dimensions). For each set and subset, the universe of companies is broken into quintiles from high quality (Q_1) to low quality (Q_5). With stocks in the highest-quality quintiles (Q_1), we form long-only factor-mimicking portfolios. By combining them with stocks in the lowest-quality quintiles (Q_5), we form long-short factor-mimicking portfolios that are long the stocks in Q_1 and short the ones in Q_5. Portfolios are value weighted, based on the free-float market capitalization from MSCI, and quarterly rebalanced. In addition, we compute a set of portfolios at the global level, through a region-neutral framework. The set combines each of the thirteen regional portfolios by weighting each region's portfolio by the region's free-float market capitalization in the MSCI World Developed Markets. We also construct sector-neutral versions of all these portfolios where we impose equal sectoral weights (GICS level 1) versus their respective benchmarks (versus their respective short portfolios for long-short strategies).

All returns are measured according to the region in which they occur (US dollars for North America and Pacific ex-Japan, Euros for EMU and Europe ex-EMU and Japanese Yens for Japan). Gross dividends are included. Within a region, currency effects affect performance. At the global level, we aggregate the performances of regional portfolios in their respective currencies of computation, mimicking continuously hedged portfolios. By doing so, we consider that investors accept intra-regional currency fluctuations, whereas they hedge their portfolios against this risk at the global level. We consider neither hedging nor transaction costs in the analysis. We apply the same rules for performance calculations on the regional and the global benchmarks.

Control variables We examine the robustness of our quality factors by running time series regressions with the Carhart four-factor model (Carhart, 1997). The exogenous variables are the market (MKT), value (high-minus-low, HML), size (small minus big, SMB), and momentum (up-minus-down, UMD) portfolios. We define the market factor as the MSCI World Developed Market ex-financials and real estate index for regressions on the world quality factor. Regarding regional robustness checks, we consider the associated regional benchmark. The value, size and momentum factors are constructed similarly to the quality factors. The sorting variables for value is book-to-price; for size market capitalization; and for momentum, twelve-month percentage stock price change over the prior twelve months, excluding the most recent month. All factors are value-weighted and rebalanced on a quarterly basis. The following is a description of the multiple linear regression model:

$$R_i(t) = \alpha_i + \beta_i^{MKT} R^{MKT}(t) + \beta_i^{SMB} R^{SMB}(t) + \beta_i^{HML} R^{HML}(t) + \beta_i^{UMD} R^{UMD}(t) + \epsilon_i(t) \quad (1)$$

where $R_i(t)$ is the return of asset i at a given date t, α_i is the intercept of the asset i, $R^{MKT}(t)$ is the return of the market factor, β_i^{MKT} is the systematic risk (or market beta) of stock i, $R^{SMB}(t)$ is the return of the size factor, β_i^{SMB} is the size sensitivity of stock i, $R^{HML}(t)$ is the return of the value factor, β_i^{HML} is the value sensitivity of stock i, $R^{UMD}(t)$ is the return of the momentum factor, β_i^{UMD} is the momentum sensitivity of stock i, and $\epsilon_i(t)$ is the term of the regression error or the idiosyncratic risk of stock i.

3 Quality: a multidimensional equity factor

Most of the traditional equity factors have a commonly accepted definition both by practitioners and academic researchers. Size, momentum and low volatility factors rely on stock market data and are quite intuitive. Size consists of classifying firms on their market capitalization level, momentum on their price movements' trend and low volatility on their volatility profile. Of course, slight variations may exist in choosing one metric over the other to characterize these three equity factors, but their definition is quite consensual.

Value and quality factors do not enjoy the same universality, mainly because they trace their roots in financial statements – partly for value, exclusively for quality – rather than solely on market data. Concerning value, the factor is generally defined in the literature by the book-to-market equity (B/M) (Fama and French, 1992). In practice, there are also other accepted so called valuation multiples, such as earnings to price (E/P), cash flow to price (C/P), earnings before interest, taxes, depreciation, and amortization to enterprise value (EBITDA/EV). These combinations remain however relatively limited because all are scaled versions of a firm's stock price.

With respect to the quality factor, if it is widely acknowledged that quality-based investment strategies aim to capture the premium associated with high-quality stocks over low-quality stocks, the question becomes less straightforward, however, when it comes to defining it. On the academic side, most of the research articles seek to circumcise the quality factor to a single systematic metric. This is because researchers try to limit as much as possible the interactions that may arise from the combination of several measures. For example, Sloan (1996) focuses on accruals, Novy-Marx (2013) on gross profitability and Xing (2008) on investment growth. Similarly, in their research, Fama and French (2015) do not combine the profitability and investment factors while both relate to quality. On the investor side, it is conversely common to combine several metrics into one composite quality factor. Some of them have empirical support as risk-factors that produce a benefit for investors while others are more discretionary, sometimes without any academic backing. By doing so, practitioners seek to capture all the quality dimensions; depending on the conception they precisely have from this non-consensual factor.

In the recent years, there has also been a new and noticeable trend with the emergence of research articles on the quality factor initiated by practitioners. Contrary to what is usually proposed by academic research on the factor-investing subject, they explore the quality factor through a mix of various metrics. Asness *et al.* (2019) define their quality factor (QMJ, or quality minus junk) through three dimensions – profitability, growth and safety – and use a broad set of measures within each of them. In a comprehensive review of the quality categories used by practitioners in their quality investing framework, Hsu *et al.* (2019) investigate three to nine metrics which belong each to profitability, earnings stability, capital structure, growth, accounting quality, payout/dilution, and investment, but do not combine them to create a single quality factor.

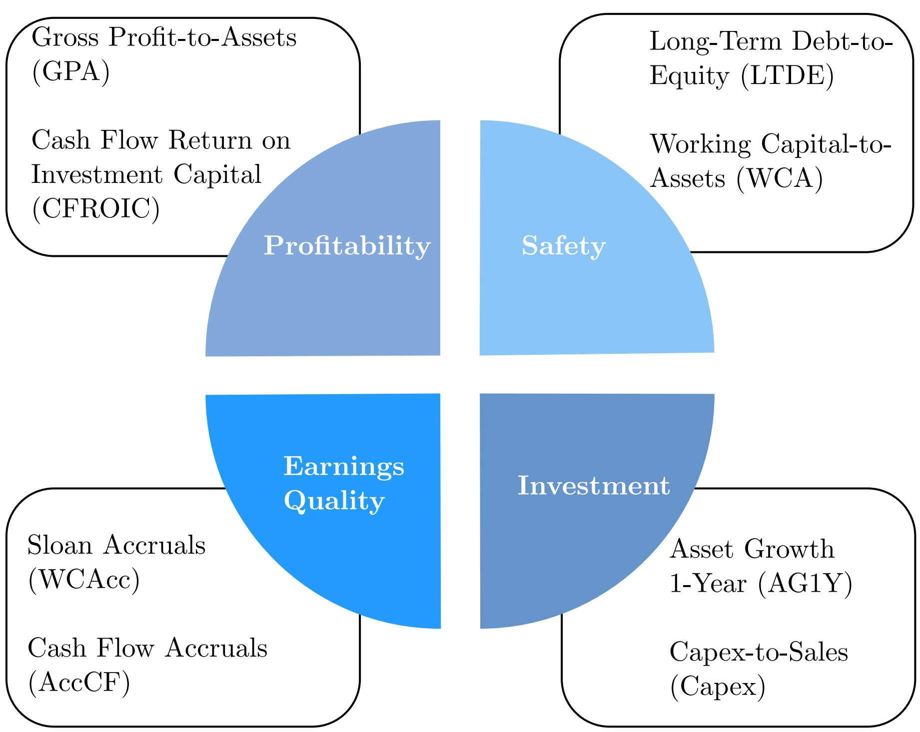
On our side, we believe that quality factor is multi-faceted and we do not want to restrict our analysis to a single metric or dimension that would lead us to cover only a partial array of quality. Assessing for instance the quality of a company under the prism of a single criterion of profitability may be misleading. Thus, while return to equity (ROE) is frequently used among practitioners for its relevance across all economic sectors, including financials, it does not provide any indication of a company's financial leverage or the quality of its earnings. A company can be very profitable based on its ROE but at the same time be very risky because it may be massively indebted, or it might lead aggressive accounting practices, distorting the true picture of its performance.

This raises the question of how we define our quality factor. We do not aim to review all metrics related to quality, and only report those that come back with significant results. It would be akin to data snooping, as pointed out by Lo and MacKinlay (1990). We do not aim either to create a quality factor that would only be a collection of fundamental criteria loosely related to the notion of quality and that would ultimately lead to the creation of a multifactor portfolio with shifting boundaries. By using only quality measures that have been subject of extensive academic research and that have demonstrated their ability to predict cross-sectional differences in average stock performances, we prevent the misuse of data analysis and we build our quality factor on a solid foundation.

Based on our understanding of the literature that we will discuss here after, quality can be defined from four perspectives: *profitability, earnings quality, safety* and *investment*. The analysis of the quality factor through complementary dimensions, such as growth, payout/dilution or earnings stability has of course been studied, but they have not been retained. While we did not find academic publications demonstrating that the growth characteristics were priced in the cross-section of expected returns, we handle the payout/dilution dimension as one of the numerous firm's investment or financing activities that are part of the broader "total asset growth" concept, as suggested by Cooper *et al.* (2008). In respect of earnings stability, we consider it as a sub-dimension of the earnings quality dimension.

As indicated previously, for each of our four quality dimensions, we select two measures derived from academic articles findings. Figure 1 below summarizes the two measures we select in each of our four quality dimensions.

Figure 1: Selected indicators per dimension



A single metric is sometimes not enough to capture the complexity of a dimension, such as the short and long-term components of safety. Moreover, *the choice to use two metrics* *lowers the probability of misclassifying stocks by using a single measure*. The decision not to use more than two metrics per dimension can be discussed but, in our opinion, retaining a wider range of metrics would distort the object of our research: to focus only on metrics that have strong empirical support as factors that produce a benefit for investors.

3.1 Profitability

Profitability is a key concept in financial analysis as it measures the company's capability to generate profits from its operations. It is certainly the quality component the most universally recognized by both researchers and practitioners. The concept covers a very wide range of metrics that roughly fall into two categories: margin and return ratios. Margin measurements give insights into a company's ability to convert sales into profits, while return metrics give insights into a company's capability to generate returns to shareholders. All these ratios offer different ways to evaluate the company's profitability, depending – for example – on the cost or profit level we consider (operational, including or excluding taxes, including or excluding financial or/and exceptional elements, etc).

Through its extensive researches on the quality factor, the academic community has largely emphasized on the profitability dimension, providing strong empirical evidence of the existence of a profitability premium. For instance, Greenblatt (2006, 2010) combines the enterprise value to earnings before interest and taxes (EV/EBIT) ratio with the return on invested capital (ROIC) to create his so-called "Magic Formula" and outperforms the market index (S&P500) by 5.7% per annum in the period 1988-2009. Novy-Marx (2013) finds that profitable firms, as measured by the gross profitability on assets, significantly outperform unprofitable firms and shows that "the metric has roughly the same power as book-to-market predicting the cross-section of average returns". In a complementary research, Novy-Marx (2014) establishes a comparative analysis of seven of the best-known quality strategies and shows that the gross profitability is the most reliable quality indicator as a stand-alone strategy, displaying the largest Fama and French (1993) three-factor alpha, especially among large-cap stocks. Fama and French (2015) finally include the profitability factor, in their five-factor model.

Among institutional investors, return metrics as return on assets (ROA) or return on equity (ROE) are popular profitability measures. They show significant advantages, such as allowing a cross-sectional comparison of all companies, including financial institutions; or capturing the net profit available to shareholders. However, these metrics do not only have benefits. Return ratios are computed on reported earnings that correspond to the bottom line of a company's income statement. In this regard, they can be highly affected by accounting choices that alter financial reporting and finally do not reflect the company's actual health. Accounting estimates for accruals and non-cash expenses such as depreciation and amortization are the most frequently used manipulations among the vast array of aggressive accounting practices.

We circumvent the issue by selecting two profitability metrics with low exposure to these aggressive accounting practices. First, we adopt the gross profitability on assets (GPA) ratio highlighted by Novy-Marx (2013) because gross profit, which accounts for cost of goods sold (COGS), is derived from the top of the income statement. Accordingly, gross profit does not suffer the potential disturbance from managements' decisions on selling, general and administrative expenses, research and development, depreciation and amortization, and capital structure mix. It is considered a cleaner measure than EBIT or net profit because it is less subject to accounting manipulations. Moreover, the outperformance of the metric with respect to other quality strategies among large-cap stocks, and its robustness over time, as demonstrated by Novy-Marx, makes it the ideal candidate. Equation 2 details the calculation of the metric.

$$GPA_{i,t} = \frac{\sum_{j=1}^{4}(SALES_{i,t-j} - COGS_{i,t-j})}{\frac{1}{4}\sum_{j=1}^{4} AT_{i,t-j}} \tag{2}$$

where the quarterly gross profit is defined as the quarterly net sales/turnover (SALES)[2] minus the quarterly cost of goods sold (COGS) and AT are the quarterly total assets.

The second ratio we elect to assess the companies' profitability dimension is cash flow return on invested capital (CFROIC). It measures the ratio of trailing four quarter operating net cash flow to average invested capital over the same period (see Equation 3). The ratio uses a cash flow based measure rather than a usual accounting based profit. Therefore, we put into practice the findings of Ball *et al.* (2016) who show that cash-based operating profitability outperforms measures of profitability that include accruals. We intend to partially mitigate the accruals-accounting issue by using the CFROIC, a more robust measure of profitability, than the return on invested capital (ROIC), its close cousin.

$$CFROIC_{i,t} = \frac{\sum_{j=1}^{4} OANCF_{i,t-j}}{\frac{1}{4}\sum_{j=1}^{4} ICAPT_{i,t-j}} \tag{3}$$

where OANCF is the quarterly net cash flow from operating activities and ICAPT is the quarterly invested capital.

3.2 Earnings quality

Earnings quality is another well documented dimension of the quality factor. It principally refers to the accrual anomaly, first documented by Sloan (1996) in his seminal study. The anomaly consists in the negative relationship between accounting accruals and subsequent stock returns. Sloan argues that market participants pay more attention to reported earning numbers and not enough to sub-components, namely accruals and real cash flows. He shows that firms with higher accounting accruals generally experience negative future abnormal stock returns. The accruals topic has quickly become a very popular subject of research that Healy and Wahlen (1999) summarize in a review of the specific accruals that can be used to manage earnings.

Many researchers have studied the anomaly and reached similar conclusion to those established by Sloan (1996), making the empirical evidence for the existence of an accrual premium

[2]For the record, the S&P Global Market Intelligence's Alpha Factor Library provides ratios relying on the Compustat dataset for North American companies and on the Capital IQ dataset for other regions. For transparency, we display the details of the Compustat fields in Appendix A on page 87. Strictly equivalent fields have been selected on the Capital IQ database for non-North American companies.

especially strong. For instance, Hirshleifer *et al.* (2004) propose to consider the net operating assets scaled by assets to assess the cumulative discrepancies between net operating income and free cash flow and conclude that the metric is a strong negative predicator of long-run stock returns. Fama and French (2008) also address the issue in a research article dissecting the most popular anomalies in the equity landscape and conclude that the accrual effect is among the most robust anomaly in the US market. In a comprehensive review of the accruals anomaly, Richardson *et al.* (2010) examine the different explanations related to the anomaly and conclude that the primary reason is that capital market participants fail to correctly utilize accrual information in their forecasts of future earnings and cash flows.

Quality of earnings does not only refer to the accrual anomaly. Other studies try to capture the quality feature of earnings by measuring their stability or their smoothness over time. These alternative measures are not only limited to earnings, but can be extended to profitability ratios as ROE, ROA or margin ratios such as gross profit margin. The underlying idea is that companies with high earnings volatility or low smoothness are more likely to be cyclical firms or firms with less robust business models that will suffer more in economic and financial storms. Conversely, low variability in earnings or high smoothness is an indicator of resilience and the prerogative of quality companies. However, there is no general consensus on these measures. Dichev and Tang (2009) find that earnings volatility negatively impacts earnings predictability, while Hsu *et al.* (2019) show that earnings stability has no empirical support as a quality metric that produce a benefit for market participants and Hsu *et al.* (2013) suggest that low earnings growth volatility is related to the low-beta anomaly.

In this article, we choose to focus only on accruals metrics. Their ability to predict future earnings and associated returns has been demonstrated by academics, regardless of the regions or periods studied. Moreover, Perotti and Wagenhofer (2014) find accruals are by far the strongest earnings quality-based measure to predict stock returns. We employ the inescapable accruals measure defined by Sloan (1996), also termed "working capital accruals" (WCAcc) as it relies on working capital components. The metric is defined as the change from four quarters ago in non-cash assets, minus the change in current liabilities (excluding short-term debt and taxes payable) minus depreciation, relative to average total assets over the past year, as shown in Equation 4.

With $\triangle X \equiv X_{i,t} - X_{i,t-4}$:

$$\begin{aligned} WCAcc_{i,t} = & \frac{\Delta\left(RECT_{i,t}\right) + \Delta\left(INVT_{i,t}\right) + \Delta\left(ACO_{i,t}\right)}{\frac{1}{4}\sum_{j=1}^{4} AT_{i,t-j}} \\ & - \frac{\Delta\left(AP_{i,t}\right) + \Delta\left(LCO_{i,t}\right)}{\frac{1}{4}\sum_{j=1}^{4} AT_{i,t-j}} \\ & - \frac{\sum_{j=1}^{4} DP_{i,t-j}}{\frac{1}{4}\sum_{j=1}^{4} AP_{i,t-j}} \end{aligned} \tag{4}$$

where RECT are the quarterly total receivables, INVT are the quarterly total inventories, ACO are the quarterly other currents assets, AT are the quarterly total assets, AP are the quarterly accounts payable, LCO are the quarterly other currents liabilities and DP are the quarterly depreciation and amortization.

The second metric we elect is a cash flow based accruals (AccCF) ratio that uses the trailing twelve months difference between net income and cash flows (operation and investment) as the measure of the accruals, which then are normalized by the average of the net operating assets (NOA), making cross-sectional comparisons possible. NOA is defined in this factor as the difference between cash free total assets and debt free total liabilities. Cash and debt are excluded because they are less subject to the management discretion. Equation 5 displays the different steps of the calculation.

$$
\begin{aligned}
AccCF_{i,t} &= \frac{\sum_{j=1}^{4}\left(IB_{i,t-j} - \left(OANCF_{i,t-j} + IVNCF_{i,t-j}\right)\right)}{\frac{1}{2}\left(NOA_{i,t} + NOA_{i,t-4}\right)} \\
where & \\
NOA_{i,t} &= \frac{1}{4}\sum_{j=1}^{4}\left(AT_{i,t-j} - CHE_{i,t-j}\right) \\
&\quad - \frac{1}{4}\sum_{j=1}^{4}\left(LT_{i,t-j} - DLC_{i,t-j} - DLTT_{i,t-j}\right)
\end{aligned}
\tag{5}
$$

where IB is the quarterly income before extraordinary items, OANCF is the quarterly net cash flow from operating activities, IVNCF is the quarterly net cash flow from investing activities, AT are the quarterly total assets, CHE are the quarterly cash and short term investments, LT are the quarterly total liabilities, DLC is the quarterly short term debt and DLTT is the quarterly long-term debt.

3.3 Safety

Safety does not enjoy the same universality as profitability, earnings quality and investment dimensions. First, academics are divided on the definition. The majority of researchers link the safety dimension to the strength of the balance sheet. Practitioners generally expect a quality company to have relatively low financial leverage and reasonable liquidity ratios. These safety features will make the company less sensitive to an unfavorable macro-economic environment, and are expected to ensure its ability to service more easily its debt obligations. Conversely, an excessive financial leverage associated to a cash flow crunch can ultimately lead to financial distress. Some authors link safety to return-based measures, such as market beta and realized volatility (Asness *et al.*, 2019) but we consider that market beta and volatility are rather related to the low beta anomaly. There are also safety metrics that are based on credit ratings, or on default risk measures (Altman *et al.*, 1968; Merton, 1974; Ohlson, 1980) but the two categories of measures embed many pieces of information and can be seen as composite multifactor models, which do no match our desire to address only pure and single fundamental based metrics.

A second contentious issue among academics lies on their definition of the relationship between leverage and return. Indeed, empirical findings are contrasted. On the one side, there are advocates of a positive relationship between leverage and equity returns. Among them, Bhandari (1988) and Fama and French (1992) show that market leverage is positively related to expected equity returns. They confirm the famous Modigliani and Miller (1958) theorems on the capital structure theory: investors tend to demand a higher return on equity to be compensated for the additional risk linked to the higher leverage level. In other words, the most distressed companies should have the highest returns. On the other side are authors who document the puzzling negative cross-sectional relation between returns and leverage. Fama and French (1992) share this view since the book leverage is used as a proxy of the market leverage. Subsequent empirical work, such as Vassalou and Xing (2004) and Campbell *et al.* (2008), confirm these findings by approaching them from the angle of the distress and conclude with the existence of a market mispricing. Others attempt to provide alternative explanations. For instance, Gomes and Schmid (2010) suggest that the link between leverage and stock return depends on the investment opportunities available to the firm. George and Hwang (2010), for their part, explain it by differences in financial distress costs from one firm to another, arguing that firms with high distress costs tend to prefer equity financing to avoid distress, but are more exposed to systematic risk than high leverage firms.

Leverage ratios mainly focus on the long-term ability of a firm to pay off its liabilities. To have a comprehensive view of the financial strength of a company, it is also necessary to assess its capacity to face short-term obligations without raising external capital. Following the pioneering work by Beaver (1966, 1968), many models have been proposed to study the likelihood of insolvency. Altman *et al.* (1968) and Ohlson (1980) are among the best known and have in common to employ the working capital-to-assets ratio as a measure of short-term insolvency risk. It shows how much short-term liquidity can be recovered as a percentage of its total assets. Other liquidity ratios include current ratio, quick ratio, and cash ratio, ranked here in ascending order of strictness. All three ratios calculate an amount of short-term assets to pay off current liabilities, but consider a more or less strict definition for liquid asset. In his seminal paper, Altman *et al.* (1968) points out the greater statistical significance of the working capital ratio compared to current and quick ratios.

In light of these empirical evidences, we elect to assess the financial strength of firms with the long-term debt-to-equity and the working capital-to-assets ratios (Equations 6 and 7, respectively). Both ratios are complementary to each other by providing a global perspective of the strength of the balance sheet. While the solvency ratio, measured by the long-term debt to equity (LTDE) metric, is calculated by dividing a company's long-term debt by its shareholders' equity, the liquidity ratio, measured by the working capital-to-assets (WCA) metric, considers the average of the differences between current assets and current liabilities over the past four quarters, divided by the average total assets over the same period.

$$LTDE_{i,t} = \frac{DLTT_{i,t}}{SE_{i,t}} \tag{6}$$

where DLTT is the quarterly long-term debt and SE is the quarterly stockholders' equity.

$$WCA_{i,t} = \frac{\frac{1}{4}\sum_{j=1}^{4}\left(ACT_{i,t-j} - LCT_{i,t-j}\right)}{\frac{1}{4}\sum_{j=1}^{4} AT_{i,t-j}} \tag{7}$$

where ACT are the quarterly total current assets, LCT are the quarterly total current liabilities and AT are the quarterly total assets.

3.4 Investment

An expanding body of academic literature has addressed the relationship between firms' investment and future stock returns. Whereas one might think that an increase in a company's investments is a positive signal for investors, empirical findings show that high-investment firms tend to underperform low investment firms.

The life of a company is marked by various events that are sometimes associated with an asset expansion (debt or equity issuance, acquisitions, bank loan initiations), sometimes with an asset contraction (spinoffs, share repurchases, debt prepayments). All these events have been the object of extensive individual research that converge on common findings: while companies experiencing growth in assets tend to produce subsequent abnormally low returns, those experiencing shrinkage in assets tend to produce subsequent abnormally high returns. To name a few, Spiess and Affleck-Graves (1995) worked on share offerings, Spiess and Affleck-Graves (1999) on debt offerings, Rau and Vermaelen (1998) on acquisitions, McConnell and Ovtchinnikov (2004) on spinoffs, Ikenberry *et al.* (1995) on share repurchases and Affleck-Graves and Miller (2003) on debt prepayments.

In their seminal paper, Cooper *et al.* (2008) suggest that all these measures quantify in a different way a firm's investment or financing activity and are in fact the individual components of a broader concept: the total asset growth. Other authors have worked on similar confluences, albeit at a lower level of aggregation. For example, Richardson and Sloan (2003) group debt and equity issuances within a larger net external financing effect, while Pontiff and Woodgate (2008) bring together equity offerings, repurchases and mergers in a broader growth in shares effect. According to Cooper *et al.* (2008) the total asset growth metric is the strongest predictor of future returns in the US market when compared with individual growth measures and is at least as powerful in explaining returns as other well-known effects such as size, book-to-market, return momentum, and accruals. Watanabe *et al.* (2013) confirm the negative relation between asset growth and average stock returns outside of the US by analyzing forty international equity markets. The total asset growth effect has since been successfully integrated into different factor models to explain the numerous anomalies and risk-factors in cross-sectional asset pricing (the q-factor model of Hou *et al.* (2015), the five-factor model of Fama and French (2015)).

As for numerous anomalies, the cause of the total asset growth anomaly remains puzzling. While some authors, as Gomes *et al.* (2003) or Li *et al.* (2009), take the risk-based explanation, others prefer the behavioral-based explanation. In such a case, several hypotheses may explain the sources of the predictability, as for instance the extrapolation of past growth by investors, or the competing interests of shareholders and the management team leading to overexpansion by managers. These hypothesis and some others are collected and analyzed by Chan *et al.* (2008) in a dedicated paper.

Regardless of its origin, the existence of the investment anomaly is nevertheless well established and we consider that it is a quality dimension in its own right. In line with the Cooper *et al.* (2008) article, we retain the one-year asset growth (AG1Y) metric, measured by the percent change in the simple average trailing four-quarter total assets (see Equation 8). As a second metric, we also refer to the academic literature by considering the capital expenditures information. Indeed, compared to the total asset growth metric, they represent a more direct way to measure management discretionary spending and Titman *et al.* (2004) find strong evidence of negative returns following large increases in capital expenditures. We measure this effect with the capital expenditures-to-sales (Capex) metric which is the ratio of the sum of capital expenditures on the trailing four quarters to the sum of sales on the trailing four quarters (see Equation 9).

$$AG1Y_{i,t} = \frac{\sum_{j=1}^{4} AT_{i,t-j}}{\sum_{j=1}^{4} AT_{i,t-j-4}} - 1 \tag{8}$$

where AT are the quarterly total assets.

$$Capex_{i,t} = \frac{\sum_{j=1}^{4} CAPX_{i,t-j}}{\sum_{j=1}^{4} SALES_{i,t-j}} \tag{9}$$

where CAPX are the quarterly capital expenditures and SALES are the net sales/turnover.

The total asset growth effect has been successfully integrated into different factor models to explain the numerous anomalies in cross-sectional asset pricing. In their q-factor model, Hou *et al.* (2015) propose a 4-factor model (market beta, size, investment and profitability) that summarizes better the cross-section of average stock returns than the Fama-French three-factor model (market beta, size and value, Fama and French (1993)) or the Carhart four-factor model (market beta, size value and momentum, Carhart (1997)). Finally, after investigation, Fama and French (2015) extend their model to five factors by adding the investment and profitability risk-factors.

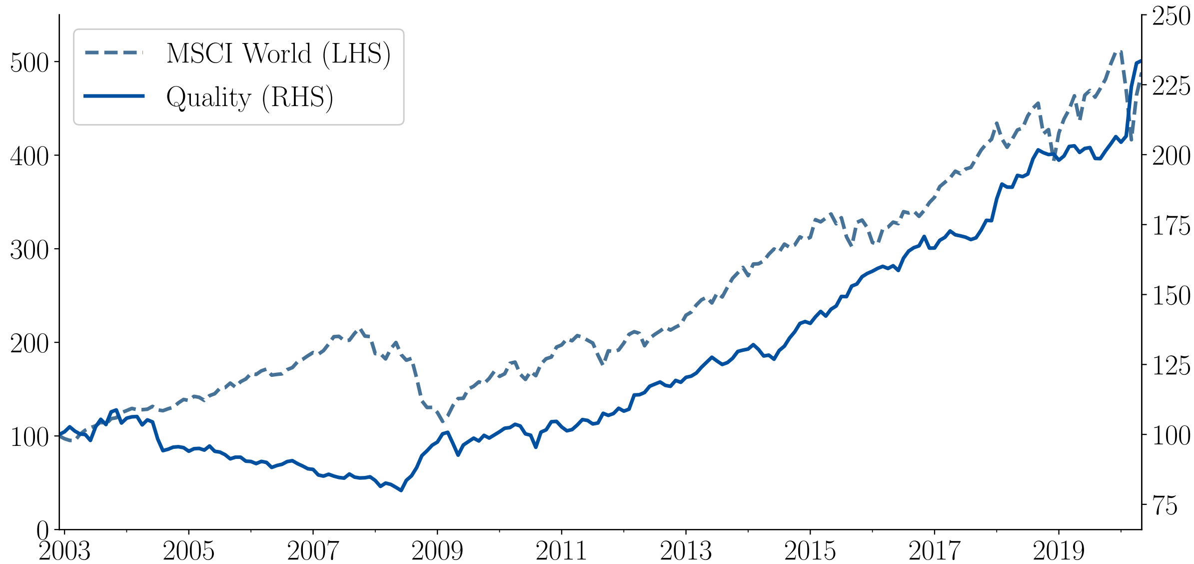
4 Analysis of the quality factor

In this section, we analyze our multidimensional quality factor. First, we scrutinize the overall behavior of our long-short quality factor, verifying that it behaves in accordance with our expectations. The long-short quality portfolio is supposed to display defensive characteristics, underperforming in expanding markets, while more immune to downturns and outperforming sharply in sudden falls in market prices. We then assess the existence of a quality premium on our investment universe and for the period of time on which we focus in this study, from December 2002 to May 2020. We also look at the quality-sorted portfolios to better identify the sources of quality premium. Finally, we investigate the multinational quality factor within a long-only framework that better fits the investor's' constraints and we also present results on a sector-neutral framework.

4.1 General overview

As an introduction, we display in Figure 2 the performance of the long-short quality factor that embeds the four aforementioned dimensions and the benchmark[3] at the world developed markets level. While the quality factor generates positive premia over the long run and especially during and after the GFC, it also goes through phases of underperformance. This is because the quality premium is time-varying. To better highlight this feature, we report performances and associated risk statistics of our long-short portfolio over different periods of time in Table 1.

Figure 2: Performance of long-short quality factor vs. benchmark at the global level



Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

The long-short quality factor underperforms during the value rally. This period corresponds to a bullish market with an increasing investor's' risk appetite, bringing with it higher investment in riskier assets, such as small capitalization and value stocks, while leaving out quality stocks. This results in a negative correlation between quality and value factors, making the quality factor an attractive hedge for value investors as pointed out by Novy-Marx (2014). In that respect, between the burst of the dotcom bubble and the start of the GFC, while the long-short value factor (defined on book-to-price) yields a 12% return on an annualized basis, the long-short multidimensional quality strategy has a negative -3.7% return. They show a risk-adjusted return of 1.76 and -0.58, respectively (see Table 1).

On the contrary, in times of crisis the quality factor proves to be a good hedge against the market. During the GFC and the Covid-19 pandemic, the long-short quality factor spikes while markets are sharply correcting. This is the reverse process to that observed during bull markets. When markets are experiencing turmoil, investors rush to less volatile and more

[3] The benchmark is calculated without the financial and real estate companies as explained in the Section 2.

Table 1: Long-short performance statistics of the multidimensional quality factor, value factor and the benchmark at the global level

	2003 - 2020			2003 - 2007			2007 - 2020		
	Quality	Value	Bench	Quality	Value	Bench	Quality	Value	Bench
Ann. Return (%)	5.0	-1.9	9.5	-3.7	12.0	17.5	8.2	-6.4	6.9
Ann. Volatility (%)	6.4	10.0	12.9	6.4	6.8	8.0	6.1	10.7	14.3
Risk-Adj. Return	0.78	-0.19	0.74	-0.58	1.76	2.19	1.34	-0.60	0.48

Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

qualitative assets as they are less sensitive to market downside risks. As a consequence, they promote higher prices for quality assets through the supply and demand mechanism.

The behavior of the long-short quality factor outside of strongly bullish or bearish markets. The period of time between the GFC and the Covid-19 pandemic was an overall bullish market, despite some major economic and financial events, such as the Euro crisis in 2010-2011 and the sharp decline of the Chinese financial markets in 2018. During this period, the long-short quality factor shows a consistent and smooth performance over time. In Table 1, quality exhibits a strong 8.2% annualized excess return on the period, along with a volatility of 6.1%, leading to an impressive risk-adjusted return of 1.34. By contrast, the benchmark shows a performance of 6.9% and a risk-adjusted performance of 0.48.

4.2 The quality premium: an empirical evidence

As described in Section 3, many researchers have demonstrated the existence and robustness of premia associated with profitability, earnings quality, safety and investment factors. Nevertheless, the existence of these premia can be questioned in recent years and in an investment universe that is more liquid and geographically broader than those usually used in the academic literature. In this section, we analyze the long-short quality portfolio at the world developed markets level, as well as its components from a regional and a dimensional point of view.

Table 2 exhibits the results of a linear regression aiming to explain the long-short returns of the quality factor in various geographical areas over the entire analysis period. The first line of the table shows the average monthly excess return we seek to explain by linear regression. We use the Carhart four-factor model (Carhart, 1997), whose explanatory variables are the market, and the value, size and momentum long-short factors, as explained in Equation 1 of Section 2. At the global level, the quality factor delivers an alpha of +36 basis points per month and is highly statistically significant, at the 99% level. However, statistics are not homogeneous in all areas. While abnormal returns are positive and highly statistically significant in North America, Europe ex-EMU and Japan, with a monthly alpha of, respectively, +34, +48 and +45 basis points, they are non-existent in EMU and not statistically significant in Pacific ex-Japan.

The quality factor appears to be neutral to the market factor at the global level. On the

Table 2: Alphas and factor loadings of the multidimensional quality factor

	Global	**North America**	**EMU**	**Europe ex-EMU**	**Japan**	**Pacific ex-Japan**
Excess return (%)	0.42	0.50	0.05	0.44	0.42	0.28
Alpha (%)	0.36*** (2.97)	0.34** (2.09)	0.00 (0.01)	0.48** (2.54)	0.45*** (2.64)	0.40 (1.62)
Market	-0.003 (-0.09)	0.051 (1.09)	0.069 (1.48)	-0.150*** (-2.64)	-0.061* (-1.72)	-0.116*** (-2.65)
Size	0.132** (2.01)	0.092 (1.24)	0.134** (2.09)	0.090* (1.79)	0.178** (2.53)	-0.052 (-0.66)
Value	-0.364*** (-5.27)	-0.440*** (-5.60)	-0.449*** (-7.71)	-0.282*** (-5.01)	-0.202*** (-3.22)	0.068 (0.92)
Momentum	-0.008 (-0.22)	-0.029 (-0.67)	-0.074* (-1.70)	0.001 (0.03)	-0.038 (-0.71)	0.035 (0.58)
Adjusted R^2	0.17	0.19	0.23	0.14	0.07	0.02

t-statistics (in parentheses) are based on White's standard error
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

regions front, loadings on the market factor are also small and not statistically significant in North America, EMU and Japan. Conversely, they are negative and highly significant in Europe ex-EMU and Pacific ex-Japan, meaning that quality shows some signs of counter cyclicality in these regions. As expected, our multidimensional quality factor exhibits strong negative and significant exposures on the value factor in all regions (except for Pacific ex-Japan). This is obviously explained by the negative correlation between both factors. We also note that the strategy loads positively on the size factor. Exposures are however quite small, suggesting that the highest quality firms have a slight bias towards mid-caps in our investment universe. The momentum factor explains only a tiny part of the long-short quality factor performances.

While results are in line with what we can expect from a quality factor, the non-existence of alpha on EMU region is quite surprising. A further examination of the data reveals that the profitability is the only dimension to deliver a statistically significant alpha of +26 basis points per month in the region, especially thanks to the gross profit-to-assets metric (see Table 19 on page 90). Earnings quality and investment bring a very low contribution to the alpha, close to zero and safety delivers a negative alpha of -23 basis points, mainly attributable to the working capital-to-assets (WCA) metric. The aggregation of the four dimensions results logically in a zero alpha on EMU.

Table 3 shows a similar multiple linear regression, but with the breakdown into the four dimensions previously described. We also display individual metrics (see Figure 1 on page 17). Profitability, earnings quality and investment deliver positive alphas of +25, +43 and +25 basis points per month, statistically significant at the 95%, 99% and 90% level, respectively. Safety does not deliver any alpha. At the metric level, we observe that both metrics (GPA and CFROIC) contribute in an equivalent manner to the profitability alpha. Regarding the earnings quality dimension, while the contribution of the accruals metric measured by the cash flow method is undeniable (+36 basis points per month, with a high statistical significance), the contribution of accruals measured by the Sloan's methodology (Sloan, 1996) seems limited (+11 basis points) and non significant. Nevertheless, the combination of the two measures at the dimension level significantly improves the alpha. For the safety dimension, both metrics fail to generate abnormal returns as part of a multi-factor regression model analysis, but for different reasons. While the long-term debt-to-equity metric does not succeed in generating excess returns over the long-run, explaining the absence of abnormal returns, the working capital-to-assets metric displays a consistent monthly excess return of +30 basis points. The absence of abnormal returns is therefore explained by the loadings on others risk-factors, meaning that the working capital-to-assets excess return can be replicated by a combination of the market and the factors selected in the regression (namely size and value). On the investment side, the total asset growth and the capex-to-sales metrics display quite small and statistically insignificant alphas, with respectively +13 and +20 basis points.

On the factor loadings side, we find large discrepancies between the dimensions. While profitability is neutral to the market factor, earnings quality and investment have a negative loading with a statistical significance at the 95% level. Conversely, safety loads positively on the market, with a statistical significance at the 99% level. In the same way, safety differs from other dimensions by loading positively on the size risk-factor with a high statistical significance, due to the working capital-to-assets metric. On the value side, most of the dimensions and their components load negatively on the factor, with a high statistical significance. Among them, profitability is clearly the more negatively exposed, with a high contribution from its gross profit-to-assets component that displays a beta coefficient of -0.73. Surprisingly, the earnings quality loads positively on value, furthermore with a statistical significance at the 95% level. It is not usual for a quality component that is expected to be negatively related to the value risk-factor. We also note this ambiguity for the asset growth metric, within the investment dimension.

Table 3: Alphas and factor loadings of individual dimensions at the global level

	Profitability			Earnings quality			Safety			Investment		
	GPA	**CFROIC**	**Dim**	**WCAcc**	**AccCF**	**Dim**	**LTDE**	**WCA**	**Dim**	**AG1Y**	**Capex**	**Dim**
Excess return (%)	0.31	0.24	0.33	0.02	0.22	0.30	0.06	0.30	0.25	-0.07	0.36	0.21
Alpha (%)	0.21**	0.23**	0.25**	0.11	0.36***	0.43***	-0.02	-0.06	-0.01	0.13	0.20	0.25*
	(2.02)	(2.39)	(2.57)	(1.04)	(2.75)	(3.55)	(-0.22)	(-0.49)	(-0.05)	(0.94)	(1.27)	(1.70)
Market	-0.020	0.036	0.021	-0.027	-0.141***	-0.097**	0.087***	0.283***	0.186***	-0.148***	0.044	-0.104**
	(-0.62)	(1.20)	(0.69)	(-0.82)	(-3.42)	(-2.58)	(2.69)	(7.07)	(4.48)	(-3.52)	(0.89)	(-2.12)
Size	0.080	-0.304***	-0.078	-0.087	-0.028	-0.075	-0.002	0.375***	0.286***	-0.160**	0.245***	0.105
	(1.39)	(-5.77)	(-1.46)	(-1.49)	(-0.39)	(-1.14)	(-0.03)	(5.33)	(3.92)	(-2.18)	(2.82)	(1.22)
Value	-0.730***	-0.244***	-0.545***	0.307***	0.174**	0.245***	-0.136**	-0.506***	-0.437***	0.370***	-0.521***	-0.053
	(-12.03)	(-4.41)	(-9.76)	(5.05)	(2.30)	(3.54)	(-2.27)	(-6.85)	(-5.70)	(4.78)	(-5.70)	(-0.58)
Momentum	-0.159***	-0.068**	-0.097***	0.163***	0.033	0.071*	0.057*	-0.035	-0.002	-0.012	-0.106**	-0.069
	(-4.85)	(-2.26)	(-3.21)	(4.91)	(0.80)	(1.89)	(1.74)	(-0.88)	(-0.06)	(-0.28)	(-2.13)	(-1.41)
Adjusted R^2	0.54	0.47	0.54	0.13	0.06	0.07	0.12	0.39	0.24	0.17	0.12	0.02

t-statistics (in parentheses) are based on White's standard error
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

As explained in Section 3, all the metrics and dimensions selected in our quality factor have been the subject of numerous publications and have shown their ability to predict cross-sectional differences in average stock performances. Profitability fully meets our expectations, thus confirming the findings of Novy-Marx (2014) who stressed in particular the reliability of the gross profit, especially among large-cap stocks. Similarly, this fully justifies the addition of the profitability factor to the Fama French three-factor asset pricing model of Fama and French (2015). Finally, it does not come as a surprise that profitability is the most commonly used characteristic in the construction of quality indices or quality portfolios in the investment community.

The earnings quality dimension also displays results in line with the academic findings. As argued by Sloan (1996), firms with high (low) reported accruals tend to have abnormally low (high) stock returns in subsequent periods. We note however that the two metrics we retain on accruals do not contribute in the same way to the alpha. While reported accruals measured by the Sloan method display a weak alpha of +11 basis points, the cash flow method shows the best statistics among all the selected metrics, with an alpha of +36 basis points. Prior research has underlined that securities with extreme values of accruals tend to be smaller, more volatile and less liquid (Mashruwala *et al.*, 2006). However, the weakness of the traditional Sloan (1996) measure cannot be explained by our investment universe, composed of larger firms than in the usual academic studies, because it would also have affected the accruals measured by the cash flow method. In our view, the explanation is more related to the accruals calculation methodology. While Sloan (1996) measures accruals by the change in non-cash working capital, the cash flow method is broader and allows to take into account accounting elements beyond the short term.

As underlined in Section 3, safety is the weakest consensual dimension among the four that we have retained in the definition of the quality factor. According to our study, safety is also by far the dimension displaying the more disappointing results with regard to the quality premium expectation. This may be due to our investment universe or/and by the shorter than usual observation period. We will analyze the dimension in further detail in the next sub-section.

On the investment dimension side, the existence of a premium is confirmed, although relatively small and statistically significant only at the 90% level. This supports the findings of Fama and French (2015) and Rizova and Saito (2020) who argue that the premium of investment factor is much stronger among small caps than large-caps. We also notice the negative excess return (-0.07%) obtained on the total asset growth (AG1Y) measured on the last twelve months.

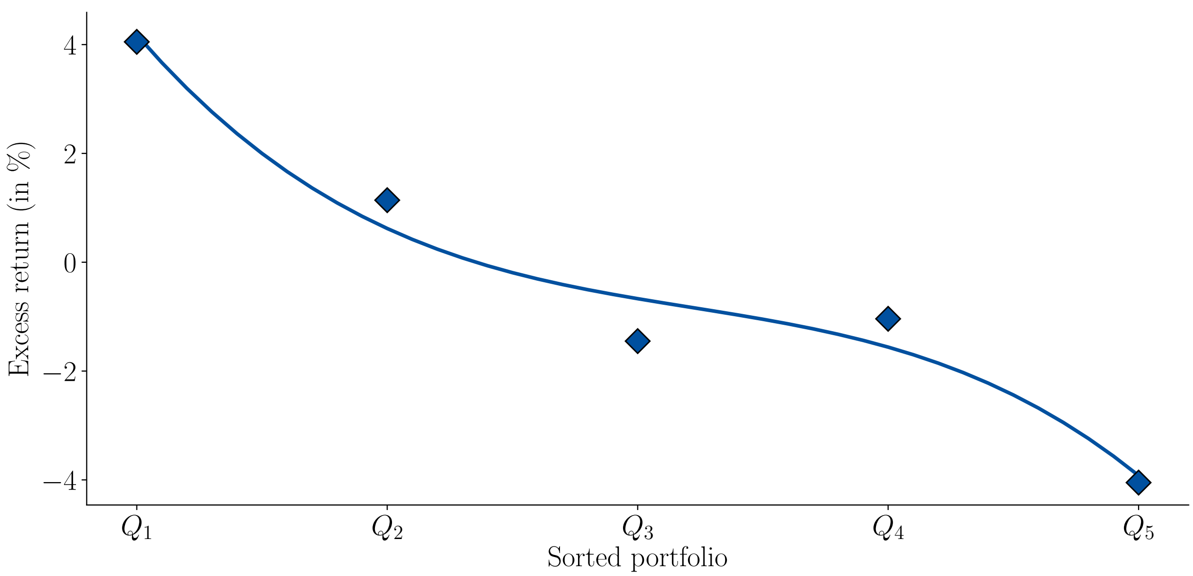
The information we derive from these two regressions is very instructive. First, our quality factor delivers a highly statistically significant alpha of +36 basis points per month at the world developed market level, meaning that it cannot be explained by loadings on conventional equity factors. Second, almost all regions display a positive contribution to this alpha, with the noticeable exception of EMU. Third, quality dimensions display positive and statistically significant four-factor alphas, with the exception of safety. The premium is especially strong on profitability and earnings quality, but appears to be more homogeneous across metrics for profitability than for earnings quality. On the other hand, while the premium is weaker on the investment dimension, mainly due to the poor premium displayed on the AG1Y metric, it is nonexistent on safety.

Finally, we confirm the existence of a quality premium associated with our multidimensional quality factor, both on a recent past and on an investment universe composed of large- and mid-caps.

4.3 Quality-sorted portfolios analysis

In this sub-section, we analyze the quality-sorted portfolios within each of the dimensions in order to better understand the premium structure of our quality factor over the post-GFC period. To this end, we split the period into two sub-periods, from December 2002 to June 2007 and from June 2007 to May 2020, and focus on the second period. While the first period refers to the value rally that followed the dotcom bubble, the second period marks the entrance into a new paradigm, born from the GFC. Undoubtedly, the GFC has opened the door to a new financial era, characterized by a lower economic growth, low or even negative interest rates and financial markets driven by central bank balance sheet expansion. It seems particularly interesting to scrutinize the premium of the quality factor in such an environment, especially with the current international health context, linked to the Covid-19 pandemic, which seems to add uncertainty to uncertainty.

Figure 3: Annualized excess return of sorted portfolios at the global level (2007 - 2020)



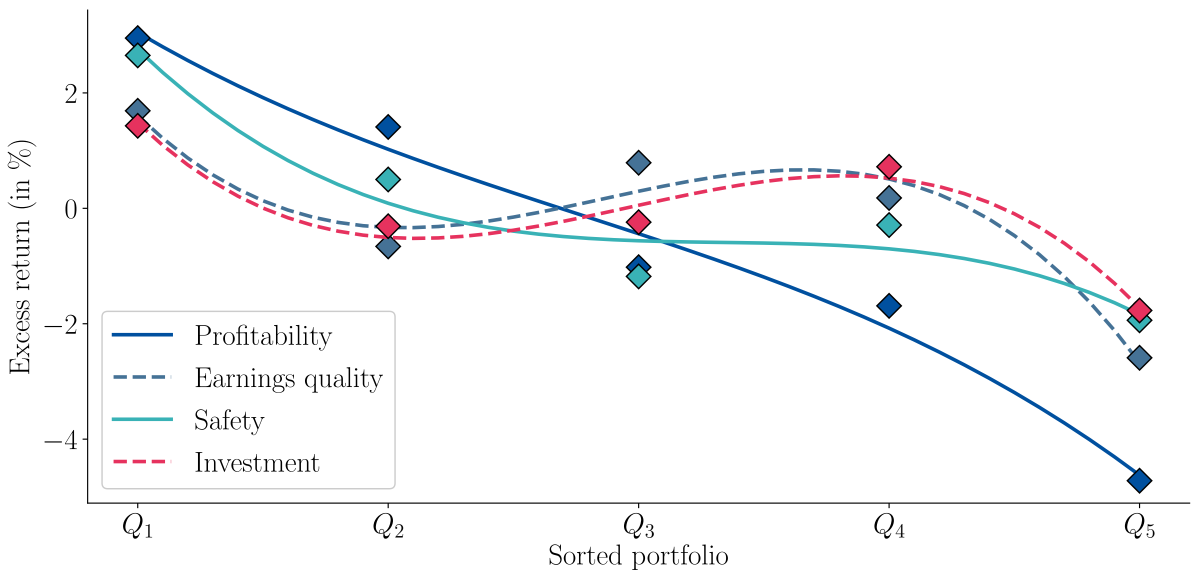
Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

In Figure 3, we have reported the annualized excess return of the five quality-sorted portfolios at the world developed markets level. Portfolios are sorted on our combined quality score, as explained in Section 2. Since all our portfolios are value weighted, they do not have pronounced size bias, as it could be in an equal weighting scheme for example, and thus can be fairly compared to their benchmark[4]. Figure 3 shows that the combined quality score al-

[4]The benchmark is calculated without the financial and real estate companies as explained in the Section 2.

lows to correctly rank stocks within their universe. Companies displaying the highest quality scores largely outperform companies with the lowest scores. We also notice that the largest performance gap mainly concerns Portfolios Q_1 and Q_5, implying that the best quality-sorted stocks are rewarded, whereas the worst quality-sorted stocks are penalized by investors. This allows the long-short strategy to display a 8.2% annual performance, with a very strong risk-adjusted return of 1.34, as shown in Table 4[5]. Regarding other portfolios, while Portfolio Q_2 outperforms portfolios with lower-rated securities, Portfolios Q_3 and Q_4 display similar performances.

Figure 4: Annualized excess return of sorted portfolios at the global level for each dimension (2007 - 2020)



Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

Figure 4 displays the same information than above, but for each of the four dimensions of our global multidimensional quality factor. We observe significant differences between the dimensions. Profitability shows a monotonically decreasing profile with a wide 7.7% spread between Portfolios Q_1 and Q_5. Earnings quality and investment display a very similar profile and have in common to both reward the best stocks while penalizing the worst. Inversely, most of the safety long-short performance seems to be built on the safest companies, the least secure companies being only marginally impacted by negative subsequent excess returns. On balance, while safety and earnings quality display akin spread between Portfolios Q_1 and Q_5, as shown in Table 4, investment displays the weakest performance spread with the lowest Portfolio Q_1 and the highest Portfolio Q_5.

In the following paragraphs, we analyze portfolios sorted on single metrics in order to better understand the performance of sorted portfolios at the dimension level. As a reminder,

[5]Tables providing similar information on 2003-2007 and 2003-2020 periods are available in Appendix B.1 on pages 88 and 89.

Table 4: Long-short performance statistics (2007 - 2020)

	Multi-dimensional	Profitability	Earnings quality	Safety	Investment
Global					
Ann. Return	8.2%	7.7%	4.2%	4.4%	2.9%
Ann. Volatility	6.1%	7.1%	6.2%	7.3%	8.0%
Risk-Adj. Return	1.34	1.08	0.68	0.60	0.36
North America					
Ann. Return	9.4%	8.7%	5.7%	5.1%	1.7%
Ann. Volatility	8.3%	9.2%	7.2%	9.5%	11.3%
Risk-Adj. Return	1.13	0.95	0.79	0.54	0.15
EMU					
Ann. Return	1.9%	5.5%	-2.1%	0.6%	2.0%
Ann. Volatility	8.9%	10.6%	12.8%	12.6%	8.1%
Risk-Adj. Return	0.21	0.52	-0.16	0.05	0.25
Europe ex-EMU					
Ann. Return	7.3%	5.2%	2.6%	2.2%	5.8%
Ann. Volatility	9.7%	10.8%	11.4%	11.7%	11.3%
Risk-Adj. Return	0.75	0.48	0.23	0.19	0.51
Japan					
Ann. Return	8.8%	4.6%	2.3%	7.0%	3.4%
Ann. Volatility	8.4%	10.2%	9.5%	10.8%	8.9%
Risk-Adj. Return	1.05	0.45	0.24	0.65	0.38
Pacific ex-Japan					
Ann. Return	1.0%	5.1%	0.4%	-2.8%	1.6%
Ann. Volatility	13.2%	11.5%	14.1%	11.1%	14.4%
Risk-Adj. Return	0.07	0.44	0.03	-0.25	0.11

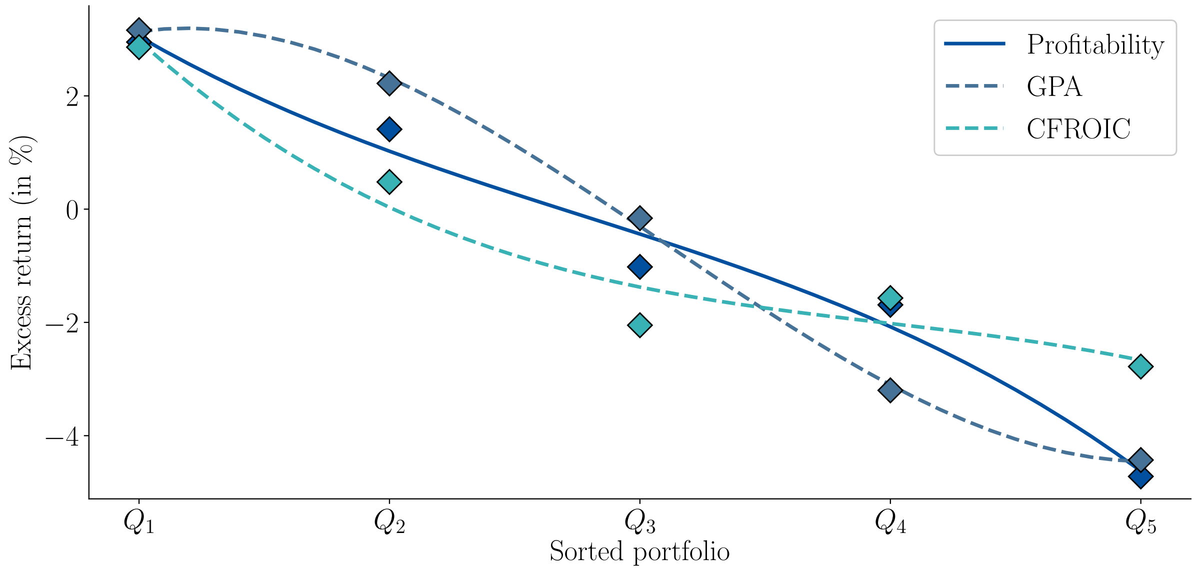
Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

dimensions and associated metrics are detailed in Figure 1 on page 17.

Profitability Over the post-GFC period, the long-short performance associated with the profitability dimension is very strong, showing a risk-adjusted return of 1.08 at the global level, as we can see in Table 4. The spread is strong across all regions, with a risk-adjusted return ranging from 0.44 in Pacific ex-Japan to 0.95 in North America.

Figure 5 exhibits the excess returns of portfolios sorted on the two single metrics related to the profitability dimension at the global level. We can observe a small difference between GPA and CFROIC-sorted portfolios. While GPA shows a decreasing profile all along the five GPA-sorted portfolios, CFROIC shows a decreasing profile only from Portfolios Q_1 to Q_3, then flatter between Portfolios Q_4 and Q_5. This confirms the Novy-Marx (2014) findings on the gross profit-to-assets robustness compared to other popular quality notions. Having said that, the combination of the two signals creates a perfectly monotonically decreasing profile.

Figure 5: Annualized excess return of sorted portfolios at the global level for the profitability dimension (2007 - 2020)



Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

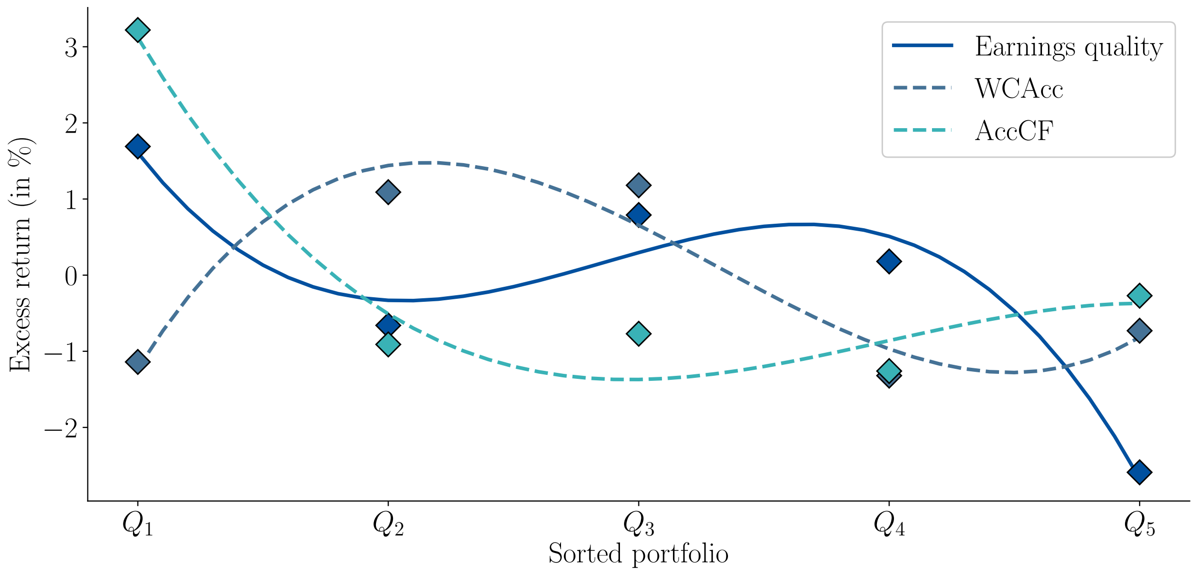
The results are comparable across all regions.

Earnings quality The long-short performance associated with the earnings quality dimension stands at the second position among the dimensions, with a risk-adjusted return of 0.68 at the world developed markets level. While North America exhibits the higher risk-adjusted return (0.79) at the regional level, the long-short performances on other regions are lower, and negative on EMU[6] (see Table 4).

As discussed in the previous section, the premium on the earnings quality dimension over the entire observation period is mainly penalized by the working capital accruals component (WCAcc). For the 2007-2020 sub-period, the situation remains the same. In Figure 6, we can observe that Portfolios Q_1 and Q_5 sorted on WCAcc display very close performances, Portfolio Q_5 being even slightly above Portfolio Q_1, and that Portfolios Q_2 and Q_3 outperform all other portfolios. The overall picture is the same for all regions except Japan, where the WCAcc metric behaves as expected. This is not the ideal configuration for a metric supposed to forecast companies' expected returns. This result may be surprising given that Sloan's working capital accruals have been the subject of extensive research as discussed in Section 3, providing strong empirical evidence of the existence of a premium associated to this measure. However, several publications help to explain such a surprising result. For instance, McLean and Pontiff (2016) study the out-of-sample and post-publication return-predictability of ninety-seven variables having demonstrated their ability to generate superior

[6]We have identified that the compounded returns for the Earnings Quality long short statistics on EMU suffer from a negative compounding effect. For our risk premium analysis we focus on the Carhart approach with averaged monthly excess returns as they are more indicative from a forward-looking perspective.

Figure 6: Annualized excess return of sorted portfolios at the global level for the earnings quality dimension (2007 - 2020)



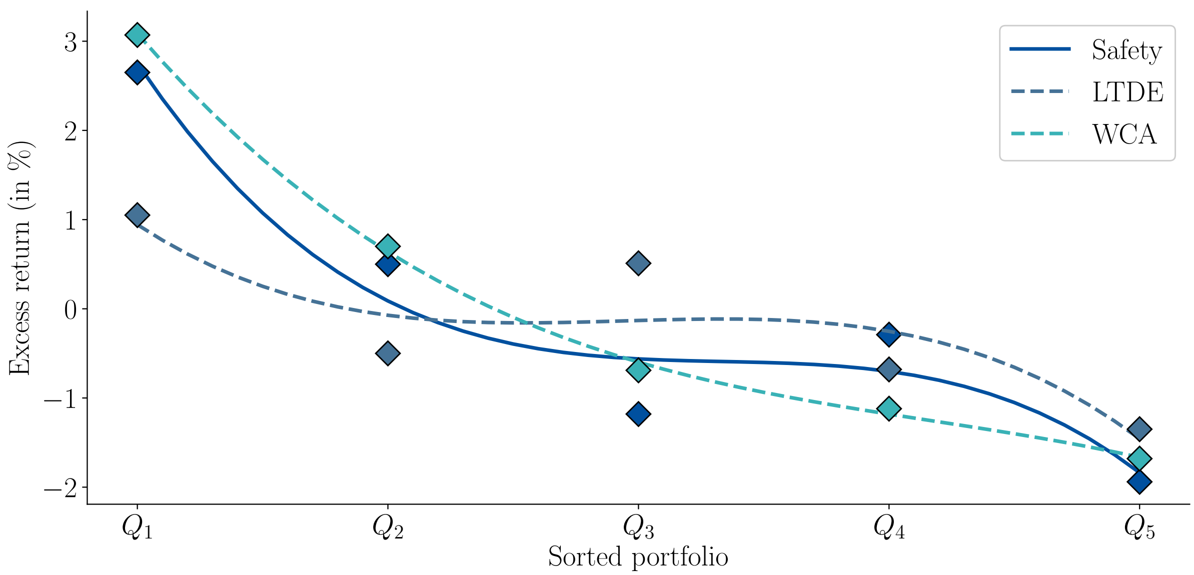
Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

returns in the cross-section and they find that portfolio returns are on average 58% lower post-publication. While a significant part of these 58% can be explained by data mining effects (26%), the remaining part (32%) would be related to a post-publication effect. Indeed, the authors suggest that investors learn about these anomalies from the publications and invest in accordance. In addition, Richardson and Sloan (2003) show in their article that the negative relation between accruals and future stocks has greatly attenuated over time. Thus, Sloan's working capital accruals may have been over-harvested by investors, leading to price corrections and subsequently lowering the expected returns of this investment strategy. However, that does not mean that investors should stay away from the accruals anomaly. This can be convincingly demonstrated by examining the excess returns of portfolios sorted on the cash flow measure of accruals (AccCF) in Figure 6. We can observe that this metric identifies very effectively the winning companies (Portfolio Q_1) from others in the cross-section of returns, while failing to differentiate them outside the Portfolio Q_1. Of course, other alternate definitions of accruals factor based on related items from financial statements may also be considered in replacement of the Sloan's method, as the balance sheet-based calculation of accruals.

Despite the disappointing behavior of portfolios sorted on the WCAcc measure, the combination of our two metrics on accruals in order to create a combined signal at the dimension level allows to significantly improve the graphic shape of the sorted portfolios. This is done at the price of a lower excess return on the Portfolio Q_1 side, but the worst quality-sorted stocks according to the combined signal are now penalized, which was not the case before, generating a higher spread between long and short portfolios.

Safety On the safety dimension side, as shown in Table 4, our global long-short portfolio shows a robust annualized excess return of 4.4% on the most recent period (2007-2020). However, the situation is not uniform across all regions. While the spread between the long and short portfolios is consistent in Japan and North America with a risk-adjusted return of 0.65 and 0.54, respectively, it is substantially lower in Europe, and negative in Pacific ex-Japan. Compared to the entire study period, the situation remains the same in terms of alpha measurement. At the world developed level, the safety long-short portfolio and its two components considered individually fail to generate abnormal returns as part of the Carhart four-factor regression model. However, Japan is the exception to the rule at the regional level, exhibiting a consistent monthly alpha of 0.53, statistically significant at the 95% level. To a lesser extent, Europe ex-EMU shows an alpha of 0.38, but not statistically significant.

Figure 7: Annualized excess return of sorted portfolios at the global level for the safety dimension (2007 - 2020)



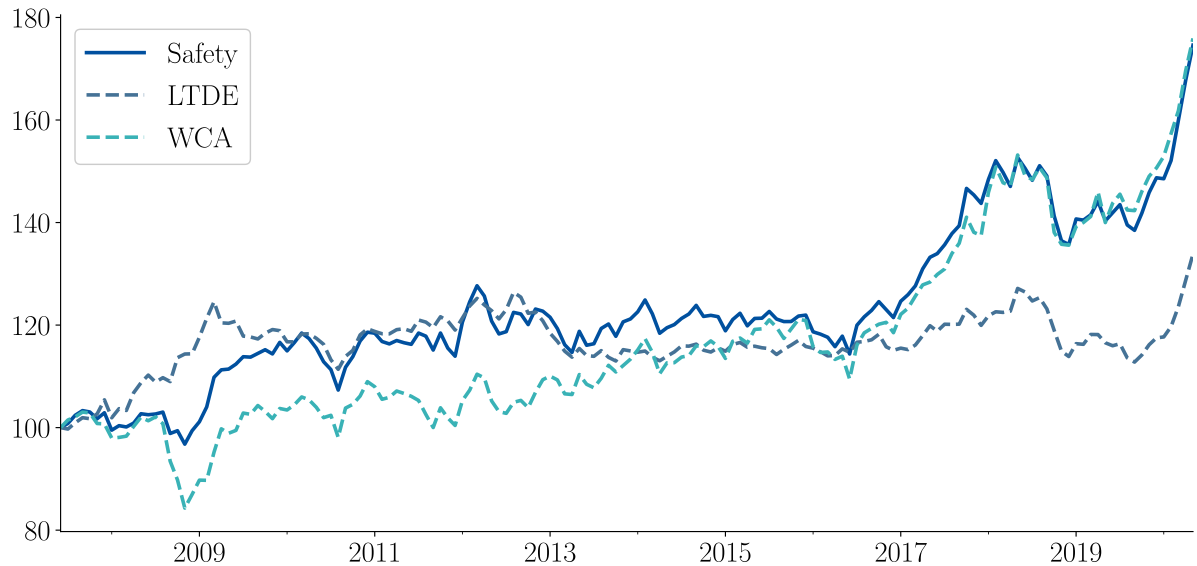
Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

In Figure 7, we observe that portfolios sorted on the working capital-to-assets (WCA) show a perfectly monotonically decreasing profile with a Portfolio Q_1 well above the Portfolio Q_5 as regards their excess return. However, as for the AccCF metric within the earnings quality dimension, we note that the spread between excess returns of Portfolios Q_3 to Q_5 is flatter than the spread of Portfolios Q_1 to Q_2, suggesting that the best WCA-sorted stocks are more rewarded than the worst WCA-sorted are penalized. As regards the long-term debt-to-equity metric (LTDE), we observe a somewhat flat shape of the sorted portfolios, but the metric succeeds nonetheless in slightly rewarding the best ranked stocks while penalizing the worst.

While we stressed the nonexistence of any abnormal return on the safety dimension as part of the Carhart four-factor regression model, it is very interesting to look in more detail at the time-series of long-short performances on the safety dimension and its components. Figure 8 shows that the portfolio that is long the low LTDE portfolio and short the high

LTDE portfolio spikes precisely when markets are experiencing extreme variations (GFC, Covid-19 pandemic) while remaining relatively flat outside of these periods of turmoil. Based on these elements, we can say that the long-short portfolio sorted on LTDE metric behaves similarly to a quality factor only when the market sharply corrects, then providing a downside protection as a market hedge. In other words, it appears that low long-term debt is primarily important to investors in times of market turbulence, while in more quiet times, investing in the most leveraged companies is no more discriminating than investing in less leveraged.

Figure 8: Performance of long-short safety dimension at the global level (2007 - 2020)

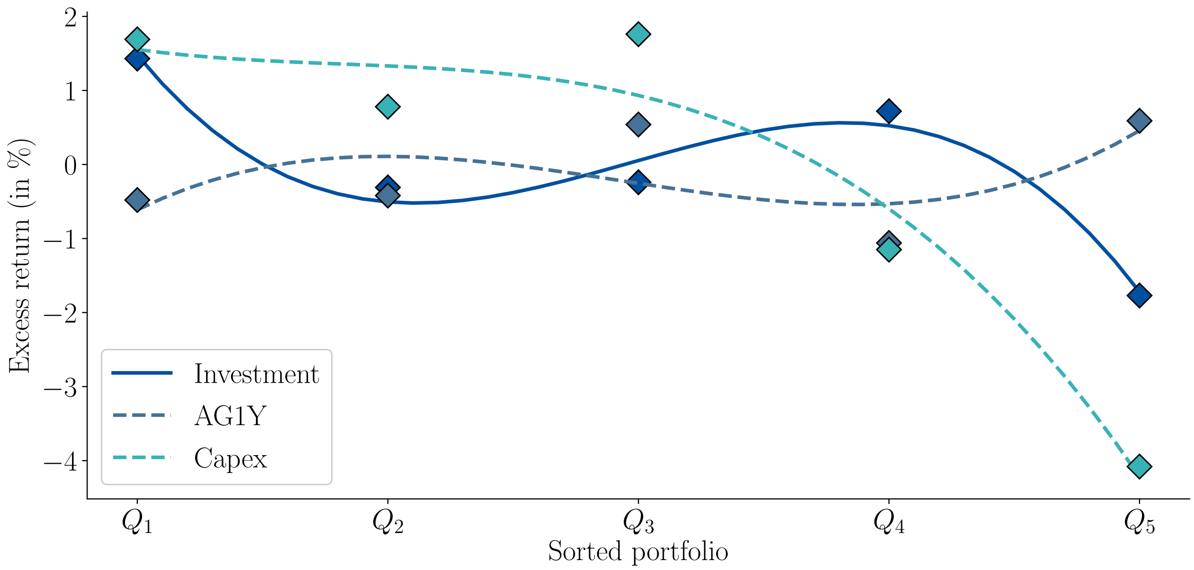


Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

By comparison, the working capital-to-assets (WCA) metric does not behave in the same way. To be invested in a portfolio that is long companies with a low working capital scaled by assets ratio while shorting the highest would have delivered a positive return during all market environments, except during the first part of the GFC. Indeed, during this sub-period, WCA strongly underperforms in absolute terms, but as well compared to LTDE, suggesting that investors choose at first to buy shares issued by firms with low long-term debt (LTDE) but not necessarily with low short-term debt (WCA). By contrast, the two metrics outperform strongly and jointly during the Covid-19 pandemic. We will discuss the topic further in the section dealing with the long-only portfolio behavior in times of market turmoil.

Investment Table 4 shows that the long-short portfolio sorted on the investment score exhibits the lowest annualized return (2.9%) among our four quality dimensions at the world developed markets level. North America and Pacific ex-Japan are the main areas where the strategy yields low returns spreads, with 1.7% and 1.6%, respectively. Results on Europe ex-EMU are on the contrary very satisfying (5.8%), thus enabling the investment dimension to be the first contributor to the long-short performance on the region, just ahead of profitability.

Figure 9: Annualized excess return of sorted portfolios at the global level for the investment dimension (2007 - 2020)



Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

Figure 9 provides some explanations of the low long-short performance on the investment dimension. The total asset growth 1-year metric (AG1Y) is weakly priced in the cross-section of expected returns, confirming the point we discussed earlier in the sub-section dealing with the analysis of the quality premium. We can observe that portfolios sorted on AG1Y exhibit a flat shape, with a Portfolio Q_5 slightly higher than Portfolio Q_1. As we briefly mentioned earlier, a possible explanation is that the premium associated with the total asset growth is mainly concentrated on the small caps universe, as pointed out by Fama and French (2015) and confirmed later by Rizova and Saito (2020). This finding is compatible with our investment universe which is focused on large- and mid-caps. Furthermore, in their article, Hsu *et al.* (2019) use the argument of McLean and Pontiff (2016) and demonstrate that asset growth has lower returns post-publication in the US market by separating the pre- and post-publication period in 2008. This may be a complementary explanation of the results we obtain on the total asset growth metric.

On the contrary, portfolios sorted on the capex-to-sales measure (Capex) display interesting features. While the metric does not reward high-ranking stocks, as we can see on Figure 9, with Portfolios Q_1 to Q_3 showing similar excess returns, it severely penalizes the low-ranking companies that show excessive capex scaled by their sales. In this regard, Portfolios Q_4 and even more Q_5 exhibit strong negative excess returns. In the end, the combination of the two signals at the dimension level is significantly affected by the shape of sorted portfolios on the asset growth metric, which explains the rather low long-short return obtained on the investment dimension.

Despite strong academic evidences on their ability to predict the cross-section of average returns, the analysis of eight individual metrics distributed within four dimensions sometimes reveals some discrepancies between academic conclusions on the one hand and our findings on the other. A number of factors may explain such differences. In the first instance, it is very common that risk premia and market anomalies identified by academic research are in fact mainly concentrated on the smallest market capitalizations. To the extent that our study focuses on large- and mid-caps, some premia may not be represented within our investment universe. The observation period may also be a source of differences. As explained in Section 2, to focus on a recent period allows concentrating on more contemporaneous topics and getting more realistic results, but it also exposes us to the post-publication decay as researchers' findings become more widely known. We summarize below the key elements based on our analysis.

- Portfolios sorted on gross profit-to-assets (GPA) show a perfect monotonically decreasing profile, making the metric the most effective to predict the cross-section of average returns all along the five sorted portfolios.
- While cash flow return on invested capital (CFRCOIC), accruals cash flows (AccCF) and working capital-to-assets (WCA) are more successful at identifying best performers, capex-to-sales (Capex) shows more ability in identifying worst performers.
- Long-term debt-to-equity (LTDE) both slightly rewards the best ranked stocks and penalizes the worst, thus showing a low spread between the two extreme portfolios. However, the metric provides a strong downside protection when markets are experiencing turbulence.
- Sloan's accruals (WCAcc) and total asset growth (AG1Y) have no empirical support as risk-factor bringing benefits for investors.
- Despite the weak or non-existent empirical support displayed by some individual metrics in this study, their combination in a global quality score allows to predict the cross-section of average returns in a very satisfying way.

4.4 Long-only framework of the quality factor

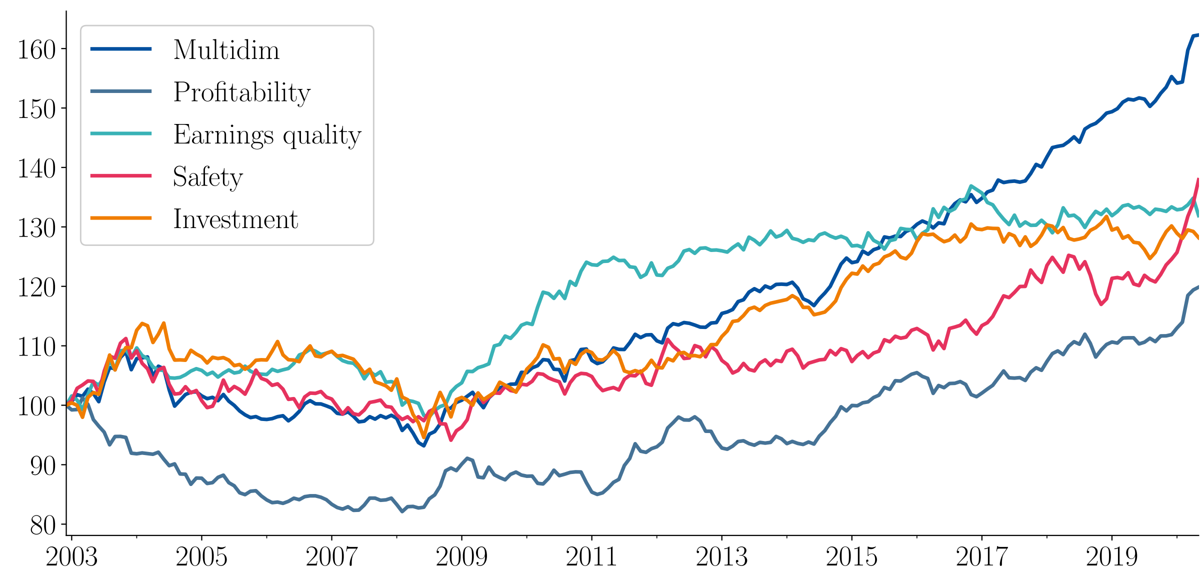
Due to constraints related to management mandate or short selling restrictions, most of institutional investors are constrained to implement factor investing solutions through a long-only framework. In this section, we specifically focus on the long portfolios that invest in high quality stocks. First, we analyze the performance of the long portfolio by focusing both on dimensions and regions. Secondly, we study the correlations between dimensions and then examine the sector allocation resulting from our construction process. Finally, we examine the long-only quality factor in specific times of market turmoil.

Long-only performance of the quality factor In Table 5, we examine the annualized excess returns and information ratios (IR) of our global multidimensional long-only quality factor, namely Q_1, over the entire analysis period[7]. The average excess return of the combined

[7] From December 2002 to May 2020.

quality factor is 2.8% per annum, showing an impressive IR of 0.81. It also outperforms by far all single quality dimensions both in absolute and risk-adjusted terms. While the average excess return is positive for all dimensions, ranging from 1.0% for profitability to 1.9% for safety, their IR are below or slightly higher to half of the multidimensional quality factor IR. To sum up, the combination of individual quality dimensions into a composite factor allows to enhance significantly the performance and risk statistics of the quality factor. In this respect, Figure 10 helps to measure these improvements by showing the relative performances of the four single dimensions and the result of their aggregation at the world developed markets level. The consistent and smooth outperformance of the multidimensional quality factor since the GFC is notable.

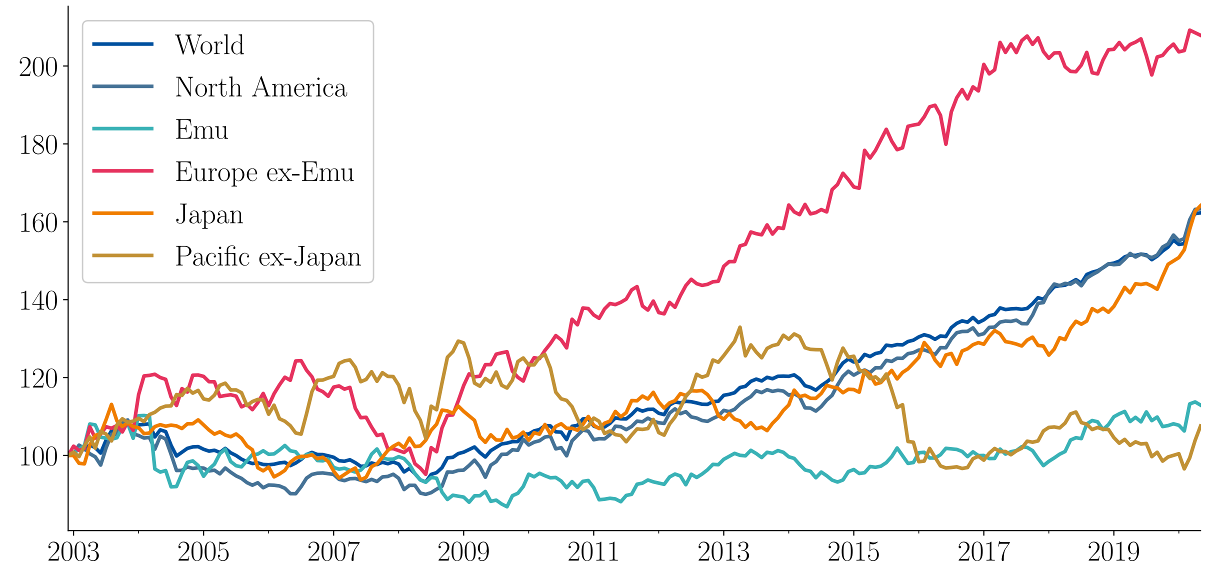
Figure 10: Excess return of the global quality factor



Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

However, a careful examination of the sub-periods and regions which compose our overall sample reveals that things are not as simple. First, we confirm the empirical evidence on the defensive characteristics of the long quality factor. While it underperforms by -0.6% during the bull market in the mid-2000s, the strategy delivers a strong positive annualized excess return of 4.0% from the initial stage of the GFC to May 2020, showing acceleration phases during the GFC and the Covid-19 pandemic (see Figure 10). Second, we note that there are some differences between regions, as shown in Figure 11. While the multidimensional quality factors on North America and Japan display excess returns similar to those of the global quality factor, Europe ex-EMU exhibits an impressive outperformance of 4.3% and performs across a broader range of market environments. In term of risk-adjusted performance, the three regions disclose appealing IR ranging from 0.53 to 0.63. In contrast, EMU shows disappointing results, failing to outperform significantly its benchmark during the period the most favorable to quality investment, confirming in this way our findings in the long-short framework section. Surprisingly, Pacific ex-Japan displays results diametrically opposed to other regions, massively outperforming during the value rally (3.9%, on an annualized basis) while slightly underperforming during the subsequent period. From this point of view, the region can be described as an outlier.

Figure 11: Excess return of the multidimensional quality factor per region



Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

Third, with respect to the individual dimensions, we also note important specificities across time and regions. Profitability stands out from others, as it is the most responsive to the market environment, both in bullish and post-GFC markets. In almost all regions, with the exception of Pacific ex-Japan, as previously mentioned, profitability exhibits both the largest negative annualized excess returns during the first sub-period (around -4.5%) and the strongest outperformance on the second, ranging from 2.1% in EMU to 3.2% in North America. Associated IR on the full sample may appear somewhat weak, but they reflect the counter-cyclicality of the profitability dimension. We also note that safety exhibits characteristics similar to those of profitability, though in a lesser extent. Conversely, the earnings quality and the investment dimensions share some commonalities across regions in their aptitude to yield positive excess returns during both the value rally and the post-GFC sub-periods. At the world developed markets level, they roughly deliver an annualized excess return of 1.5% regardless of the period considered. Among the notable elements at the regional level, we also note that the two dimensions yield negative excess returns on EMU while Europe ex-EMU records impressive excess returns over 4% during the value rally.

To summarize, the ability to discriminate out from under performers on our quality factor framework is not driven by a single dimension or region. All of them contribute positively to the outperformance. Regarding regions, while Europe ex-EMU comes first, followed by North America and Japan, EMU disappoints from a return perspective and Pacific ex-Japan exhibits contrarian statistics. On the dimension's side, we note that they display different profiles resulting in a certain degree of complementarity. While profitability and, to some extent, safety

Table 5: Long-only performance and information ratio (2003 - 2020)

	Annualized Excess Return			**Information Ratio**		
	2003-2007	**2007-2020**	**2003-2020**	**2003-2007**	**2007-2020**	**2003-2020**
Global						
Profitability	-4.2%	2.9%	1.0%	-1.28	0.82	0.29
Earnings quality	1.3%	1.7%	1.6%	0.38	0.49	0.46
Safety	-0.4%	2.7%	1.9%	-0.07	0.62	0.41
Investment	1.4%	1.4%	1.4%	0.29	0.37	0.35
Multidimensional	**-0.6%**	**4.0%**	**2.8%**	**-0.15**	**1.32**	**0.81**
North America						
Profitability	-4.5%	3.2%	1.2%	-0.99	0.72	0.25
Earnings quality	1.4%	2.4%	2.1%	0.32	0.54	0.48
Safety	0.2%	3.2%	2.4%	0.03	0.54	0.39
Investment	0.5%	1.9%	1.5%	0.09	0.37	0.29
Multidimensional	**-1.4%**	**4.4%**	**2.9%**	**-0.25**	**1.1**	**0.63**
EMU						
Profitability	-4.9%	2.1%	0.3%	-0.65	0.40	0.04
Earnings quality	0.1%	-1.1%	-0.8%	0.01	-0.15	-0.12
Safety	-3.4%	1.1%	-0.1%	-0.39	0.16	-0.01
Investment	0.9%	-0.5%	-0.1%	0.12	-0.08	-0.02
Multidimensional	**-0.9%**	**1.2%**	**0.7%**	**-0.10**	**0.24**	**0.11**
Europe ex-EMU						
Profitability	-4.1%	2.6%	0.8%	-0.69	0.42	0.13
Earnings quality	4.1%	0.8%	1.6%	0.51	0.12	0.23
Safety	-0.7%	1.1%	0.6%	-0.10	0.14	0.08
Investment	4.4%	0.7%	1.7%	0.42	0.10	0.20
Multidimensional	**2.1%**	**5.1%**	**4.3%**	**0.25**	**0.78**	**0.61**
Japan						
Profitability	-4.9%	2.8%	0.8%	-1.07	0.48	0.13
Earnings quality	-1.2%	1.0%	0.4%	-0.19	0.19	0.07
Safety	-1.9%	3.6%	2.1%	-0.32	0.56	0.34
Investment	3.0%	0.8%	1.3%	0.53	0.14	0.24
Multidimensional	**-1.4%**	**4.4%**	**2.9%**	**-0.24**	**0.84**	**0.53**
Pacific ex-Japan						
Profitability	6.3%	2.1%	3.1%	0.95	0.32	0.49
Earnings Quality	0.8%	0.1%	0.3%	0.10	0.01	0.03
Safety	-0.9%	-1.4%	-1.3%	-0.09	-0.18	-0.15
Investment	2.4%	0.2%	0.8%	0.30	0.03	0.11
Multidimensional	**3.9%**	**-0.8%**	**0.4%**	**0.59**	**-0.10**	**0.05**

Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

show counter cyclicality characteristics, earnings quality and investment outperform in both sub-periods of our sample. This allows the multidimensional factor to limit underperformance during the value rally and to outperform strongly since the beginning of the GFC, resulting in an IR of 1.32 over the second sub-period.

Correlation analysis To gain insight into how dimensions interact over time, we measure in Table 6 the correlations among their excess returns at the world developed markets level. First, we notice that the correlation levels are somewhat low, ranging from -0.22 to 0.44, meaning that dimensions do not bring the same information. It is an important feature investors should seek in a multidimensional factor framework as the lower the correlation between signals, the better the complementarity between signals will be. The correlation analysis unsurprisingly confirms what we observe on the performance side: profitability and safety on the one hand, earnings quality and investment on the other show the highest pairwise correlations across time, with 0.34 and 0.44, respectively. By contrast, when it comes to the correlations outside these two pairs, we note that they do not exceed 0.22 in absolute term. We produce similar correlation tables for all regions (see Appendix B.3 on page 91). With the exception of Pacific ex-Japan, the results are very comparable from one region to another.

Table 6: Correlation between dimensions at the global level (2003 -2020)

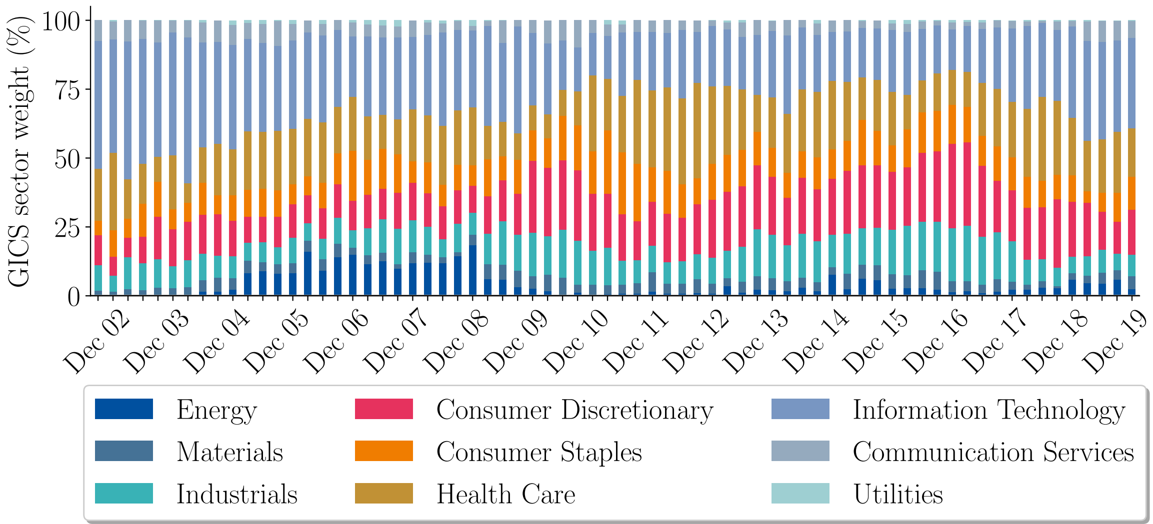
	Profitability	**Earnings quality**	**Safety**	**Investment**
Profitability	1.00			
Earnings quality	-0.22	1.00		
Safety	0.34	-0.03	1.00	
Investment	-0.14	0.44	0.21	1.00

Source: S&P Compustat, S&P Capital IQ, Authors' calculations

Sector allocation In the next two figures, we compare the sector allocation of our global quality factor (Figure 12) against the sector weights in the MSCI World DM, our benchmark (Figure 13). For the record, both are constructed on a region neutral basis (for more details, see Section 2). These figures show that information technology, consumer discretionary and healthcare sectors tend to be structurally overweighted. Conversely, sectors such as utilities, materials and communications services are underweighted. The quality feature is therefore not equitably distributed among sectors. Some are more likely to display characteristics that are more in line with the dimensions that define the quality factor. We will discuss this topic in more detail in Section 5, when we compare the sector allocations of our long-only portfolio with those obtained by adding a clustering technique to our portfolio construction process.

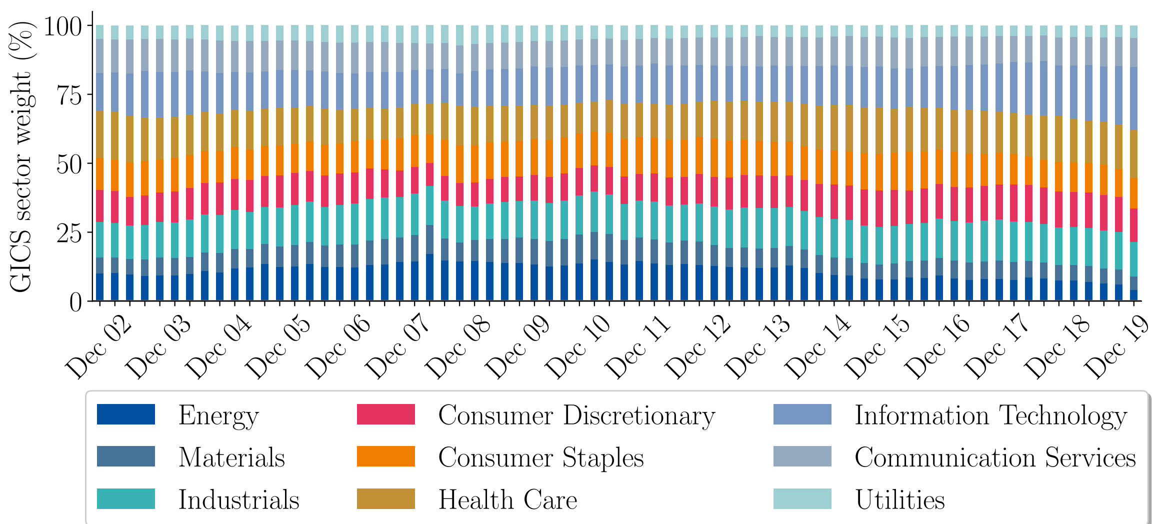
Figure 12 and Figure 13 also show that, if the quality characteristic of a sector remains rather constant over time, it can however evolve according to the economic conjuncture. For instance, the energy sector is most of the time significantly underweighted and sometimes

Figure 12: Sector allocation of the global quality factor



Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

Figure 13: Sector allocation of the benchmark



Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

even absent of the global quality factor. Nevertheless, during the period from 2005 to 2008, energy companies benefited from the rising demand from emerging markets and from the resulting price increase in commodities. Thus, they considerably improved their fundamental profile by being more profitable and reached momentarily the status of a quality sector, before being dismissed again. We observe the opposite behavior from the information technology sector in 2010 or from the healthcare one in 2009. Subsection 4.5 provides a summary of the performance and risk statistics of the sector-neutral version of our long-only global quality factor.

The long-only quality factor in times of market turmoil In the academic literature and among investors, the quality factor is usually known to underperform in bullish markets, while outperforming in times of crises. The underperformance of quality can be explained by the investor's' risk appetite which increases during rising markets, the outperformance is the consequence of a flight-to-quality phenomenon in periods of financial turmoil. Insofar as the four dimensions of our quality factor do not react in the same way to market movements, it is interesting to analyze how they interact in such market environments.

For the bullish market side, we have already provided some answers at the beginning of this section. Indeed, while profitability underperforms significantly by -4.2% per annum during the period from December 2002 to June 2007, as is expected from a traditional quality factor in this type of market environment, earnings quality and investment dimensions act as dampeners, improving considerably the relative trend and smoothing the performance over time.

On the downturn market side, the GFC and the Covid-19 pandemic are worth investigating in detail, as these are two major global crises, the first being financial while the second is a global sanitary crisis. In Table 7 we show the excess return of our multidimensional quality factor at the world developed market level by focusing on the periods from June 2008 to February 2009 and January 2020 to March 2020, respectively. We do not annualize the excess returns as the periods of observation are very short. During both crises, the multidimensional quality factor strongly outperforms the benchmark, as expected. The contribution of each dimension, however, is not the same from one crisis to another and we believe that this can be explained by the differences between their causes and consequences. On the one side, the GFC corresponds to a period of extreme stress in global financial markets and banking system that support our modern economies, raising concerns about a global credit crunch in the long-term. On the other side, the Covid-19 pandemic, although exercising a more radical and abrupt effect on the real economy, evaporating supply and demand simultaneously, was more perceived as a stop-and-go at the time it occurred, to the extent that the fundamentals of the economy were not challenged at the start of the pandemic. Moreover, major central banks have taken direct measures very quickly to support their economies, including, for example, the ECB who announced a 750 Billion EUR Pandemic Emergency Purchase Programme (PEPP) in March of 2020[8].

While all the dimensions, with the exception of the safety, contribute significantly to the excess return during the GFC, only profitability and safety contribute positively during the Covid-19 pandemic. It is a very interesting result as it shows that investors do not invest blindly in quality stocks but on the contrary make a very clear distinction between the different components of the quality factor. The GFC was perceived as a deep crisis that could jeopardize global finance in the long run. As a consequence, investors have focused on long-term quality metrics, leaving out short term ones. As a reminder, we describe the safety dimension through the long-term debt-to-equity ratio on the one side, which focuses

[8] https://www.ecb.europa.eu/press/pr/date/2020/html/ecb.pr200318_1~3949d6f266.en.html.

Table 7: Excess returns of the global quality factor during the GFC and Covid-19 pandemic

		GFC	**Covid-19**
Profitability	GPA	12.7%	6.6%
	CFROIC	4.3%	4.4%
	Dim	**9.9%**	**4.8%**
Earnings quality	WCAcc	-7.0%	0.2%
	AccCF	9.5%	0.7%
	Dim	**7.5%**	**0.6%**
Safety	LTDE	3.6%	2.2%
	WCA	-3.4%	5.4%
	Dim	**0.2%**	**4.8%**
Investment	AG1Y	0.6%	-0.2%
	Capex	4.0%	-0.6%
	Dim	**6.6%**	**0.4%**
Multidimensional		**8.9%**	**3.6%**

Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

on the long-term ability of a firm to pay off its liabilities, and the working capital-to-assets ratio on the other side, which assesses its capacity to face short-term obligations. Table 7 shows that, during the GFC, investors have undoubtedly favored the long-term ratio (+3.6% of excess return) over the short-term one (-3.4%). A similar analysis may be conducted at the earnings quality-related dimension level, confirming our reasoning. The Sloan's definition of accruals (1996) relates to short-term indicators linked to working capital and contributes very negatively (-7.0%) to excess return. On the contrary, the accruals measured by the cash flow method is the difference between the reported accruals earnings and cash earnings and therefore cannot be associated to a short-term view, resulting in a strong contribution of 9.5% to the excess return during the GFC.

As previously mentioned, the period related to the Covid-19 pandemic was perceived as a short-term crisis at the time it occurred, owing to the prompt intervention of central banks. Moreover it is possible that the social impacts have not been fully priced-in at the time of writing. As a result, investors seem to have focused primarily on profitability and safety-related dimensions rather than earnings quality and investment. This can be explained by the counter cyclicality profile of the two first dimensions, resulting in a higher capacity for the selected stocks to move in a direction opposed to general economic trend. Similar to the GFC, we note that investors have taken into account the anticipated nature of the Covid-19 pandemic by favoring this time the short-term metric of the safety dimension (excess return of 5.4%) rather than the long-term dimension (2.2%).

Additional figures Table 8 gives complementary information on our long-only quality factor. Hit ratios are defined as the number of months during which the regional quality factors outperform their respective benchmark and are expressed as a percentage of the total number of months over the study period. We also measure these hit ratios by conditioning the calculation to the months on which the performance of the benchmark is positive or negative. We observe a very high hit ratio of 62.7% at the global level, making the quality factor a robust strategy in all markets environments. Still from a global point of view, the factor has a higher ability to beat its benchmark in bullish months (65.7%) than in bearish months (56.5%). This may appear counter-intuitive as quality is known to outperform in periods of financial turmoil, but our results are explained by the fact that we do not distinguish in this analysis the magnitude of the market decline. We have indeed shown previously that the quality factor outperformed during the most turbulent market phases.

Table 8: General statistics on the multidimensional quality factor (2003-2020)

	Global	**North America**	**EMU**	**Europe ex-EMU**	**Japan**	**Pacific ex-Japan**
Max Drawdown (%)	-44.1	-43.2	-54.2	-38.8	-46.0	-51.8
Hit (%)	62.7	57.4	53.1	56.5	55.0	53.1
Hit Up Months (%)	65.7	60.3	50.4	52.2	47.2	53.0
Hit Down Months (%)	56.5	51.5	57.5	64.4	66.7	53.3
Tracking Error (%)	3.5	4.6	6.3	7.0	5.5	7.7
Beta	0.99	1.02	0.98	0.91	0.92	0.97
Turnover (%)	125.7	117.2	138.8	154.3	115.8	119.7

Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

At a regional level, we note positive hit ratios in both up and down markets, but it is interesting to observe that in Europe ex-EMU and Japan, hit rates are significantly higher during bear markets. This information suggests that, for these regions, the quality factor is more defensive than in North America. This is also reflected in the market betas that are closed to 0.90 for these two regions when others regions display market betas closed to 1. With respect to turnover, the global strategy displays an annualized turnover of 125.7%[9], which is roughly in line with an equity factor strategy rebalanced on a quarterly basis (except momentum which, by construction, generates a higher turnover).

To summarize, our results show that the four dimensions we retain to build our quality factor are very complementary in a long-only framework. They are weakly correlated with each other, and mixing them allow to bring diversification to the final portfolio and to anticipate the different phases of economic cycle. In the next subsection, we compare the unconstrained quality factor with its sector-neutral version.

4.5 Sector-neutral results

Table 9 shows the long-only performance of our multidimensional quality factor with and without sector-neutrality. The sector-neutral versions of the portfolios are constructed as explained in Section 2.

[9]We display an annualized one-way turnover.

From an investor's' perspective, the advantage of sector-neutrality is to gain better sector diversification within the portfolio compared to an unconstrained version and to neutralize the active risk coming from the sector allocation in respect to the market index. This logically results in a significant reduction of the tracking error in all areas. From a performance point of view, the annualized return at the global level decreases from 12.6% to 11.8%, while the annualized volatility lowers from 13.3% to 12.9%, resulting in a small decline of the risk-adjusted performance, from 0.95 to 0.91. North America and Europe ex-EMU display similar reductions in their performance and risk statistics, while Asian regions are very slightly affected by the sector-neutral implementation. Regarding the EMU region, unlike all other regions, it translates into an increase in both performance (from 8.5% to 9.7%) and risk-adjusted performance (from 0.54 to 0.66). More interestingly, all individual dimensions are positively affected by the sector-neutrality implementation as shown in Table 15 on page 78 within a sub-section dealing more specifically with the EMU region. We also note a significant increase in all portfolio turnovers[10], which is the direct consequence of the sector-neutrality constraint we impose on our quality portfolios.

In the next section, we explore a new portfolio construction technique that aims to group stocks on fundamental-based and market-based features rather than on a sector classification. By doing so, we test if the information we can capture in these features could help us group companies together in a more relevant way than sector classifications.

[10] We display an annualized one-way turnover.

Table 9: Long-only performance of the multidimensional quality factor with and without sector-neutrality

		Without sector neutrality	**With sector neutrality**
Global	Ann. Return	12.6%	11.8%
	Ann. Excess Return	2.8%	2.1%
	Ann. Volatility	13.3%	12.9%
	Risk-Adj. Return	0.95	0.91
	Tracking Error	3.5%	2.2%
	Turnover	125.7%	148.3%
	Alpha 4FF monthly	0.36%	0.31%
North America	Ann. Return	13.7%	12.6%
	Ann. Excess Return	2.9%	1.7%
	Ann. Volatility	14.7%	13.8%
	Risk-Adj. Return	0.93	0.91
	Tracking Error	4.6%	3.1%
	Turnover	117.2%	137.9%
	Alpha 4FF monthly	0.34%	0.24%
EMU	Ann. Return	8.5%	9.7%
	Ann. Excess Return	0.7%	1.7%
	Ann. Volatility	15.7%	14.6%
	Risk-Adj. Return	0.54	0.66
	Tracking Error	6.3%	4.1%
	Turnover	138.8%	166.0%
	Alpha 4FF monthly	0.00%	0.11%
Europe ex-EMU	Ann. Return	13.0%	11.3%
	Ann. Excess Return	4.3%	2.9%
	Ann. Volatility	12.9%	13.2%
	Risk-Adj. Return	1.01	0.86
	Tracking Error	7.0%	5.2%
	Turnover	154.3%	168.5%
	Alpha 4FF monthly	0.48%	0.52%
Japan	Ann. Return	9.5%	9.7%
	Ann. Excess Return	2.9%	3.3%
	Ann. Volatility	16.3%	16.4%
	Risk-Adj. Return	0.58	0.59
	Tracking Error	5.5%	4.5%
	Turnover	115.8%	153.9%
	Alpha 4FF monthly	0.45%	0.35%
Pacific ex-Japan	Ann. Return	10.1%	10.2%
	Ann. Excess Return	0.4%	0.6%
	Ann. Volatility	20.1%	19.7%
	Risk-Adj. Return	0.50	0.52
	Tracking Error	7.7%	5.6%
	Turnover	119.7%	153.2%
	Alpha 4FF monthly	0.45%	0.52%

Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

5 Creating a new paradigm in the sector-neutrality framework

As previously mentioned, we can adopt a portfolio construction framework which is either sector adjusted or not sector adjusted. The two approaches may result in very different portfolios, depending on the score distribution across sectors. In addition, it also serves to compare a company with its peers by introducing a best-in-class approach to identify best and worst companies in their respective industries based on the given score.

In general, sector classifications rely on widely accepted and relatively stable definitions of sectors and companies are assigned to a particular sector based on the source of the majority of their revenue. This method ensures to group together companies with a common dominant sector of business activity, but not necessarily companies with common fundamental characteristics. In other words, from a fundamental point of view, there is sometimes more commonality between two companies that belong to different sectors than between companies that belong to the same sector. In this section, we seek to emancipate ourselves from the traditional sector-neutral framework by exploring the possibility of grouping companies with similarities from a fundamental perspective. To do so, we investigate a construction technique based on unsupervised machine learning and we implement the model on our global quality factor strategy. The objective is not necessarily to improve the performance and risk statistics of our initial strategy, but rather to measure the effects of this paradigm shift in how stocks are pooled to build sorted-portfolios.

First, we present the fundamental-based and market-based features that will feed the unsupervised machine learning algorithms. Then, we introduce the mixture models we select as candidates to group stocks from our investment universe into homogeneous clusters and we explain step-by-step the approach we follow to retain the most efficient as part of this exercise. Finally, we look into the resulting clusters in the context of the quality factor portfolio construction and we analyze the backtest results.

5.1 Data methodology

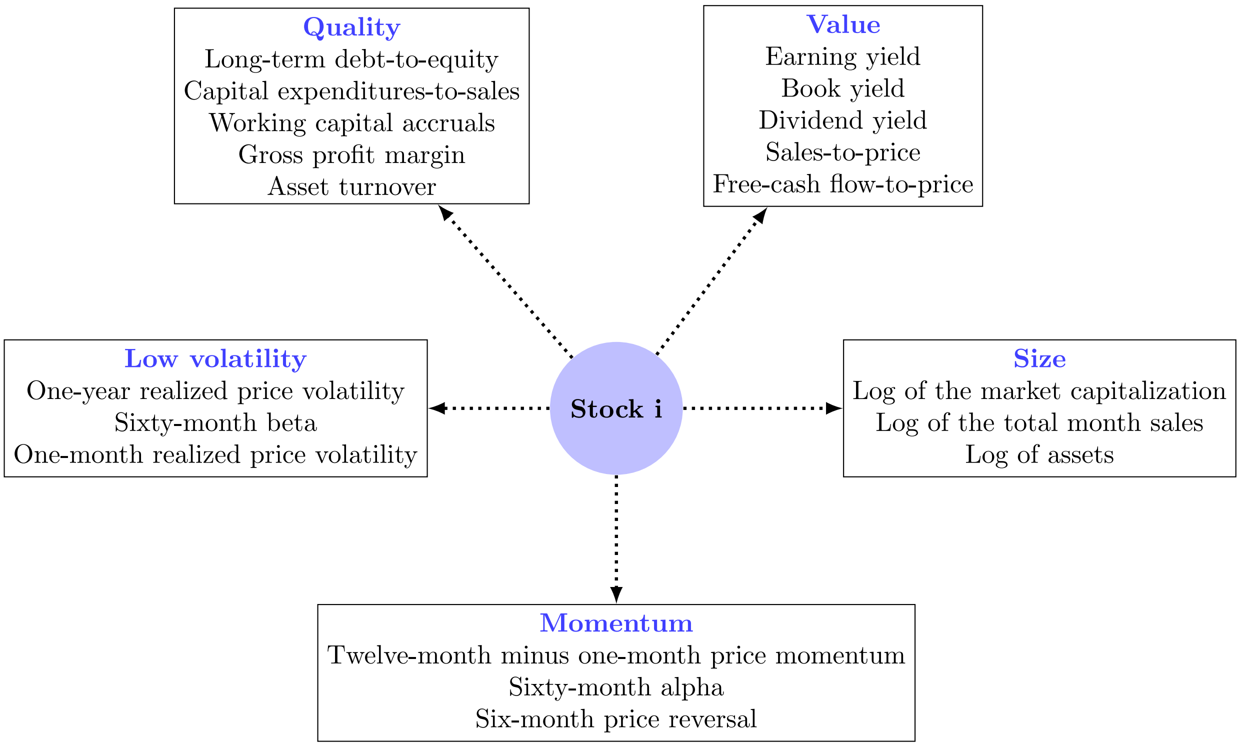
Presentation of the data Each quarter, we consider nineteen market and fundamental features[11] (see Figure 14) for all stocks that are part of our investment universe from December 2002 to March 2020. To mimic turnover controlled portfolio management approaches, we update our data on a quarterly frequency[12]. The selected data reflects stock information at the fundamental level and marginally at the market level. Indeed, we choose to elect criteria from the wide range of data listed in the S&P fundamental databases that best define, from our perspective, the main equity factors identified in the literature: quality, value, low volatility, size and momentum risk-factors in order to input all necessary stock's information to the algorithm.

We choose five criteria for quality and value risk-factors and three for the others. The

[11] From S&P Global Market Intelligence's Alpha Factor Library.

[12] By doing so, we are putting ourselves in real market conditions as the market index is re-balanced on a quarterly basis. We still exclude financial and real estate stocks from our universe (see Section 2).

Figure 14: Market and fundamental features



Source: S&P Compustat, S&P Capital IQ

selection of a larger number of criteria for quality and value risk-factors is the result of a lack of consensus between practitioners and researchers to define these risk premia, as discussed in Section 3. The greater complexity of these factors requires therefore a larger number of features to encompass their specificities.

Statistical treatment of the data First of all, we estimate missing data with the historical trend of the data and its business sector. They are substituted as explained in Equations 10 and 11.

$$\overline{S}_t = \frac{1}{N}\sum_{q=1}^{N} S_{t-q} + \Delta S_{t-q}^{GICS} \tag{10}$$

$$\Delta S_{t-q}^{GICS} = S_{t-q}^{GICS} - S_{t-q-3}^{GICS} \tag{11}$$

where N is 4, corresponding to the number of quarters in the last twelve months, S_t is the missing value of a given stock at a given date t and S_t^{GICS} is the stocks' median of the GICS sector to which the stock belongs at a given date t.

Looking at our entire sample[13], few data is missing across the different geographical areas. For instance, we observe around 2.9% of missing data for North America area, 5.4% for EMU,

[13] From December 2002 to May 2020.

4.5% for Europe ex-EMU, 4.1% for Japan and 5.4% for Pacific ex-Japan. We minimized the number of missing data by selecting accessible criteria across all sectors[14].

Then, we scale data for comparability. This step determines the stability of our algorithm. First, we clip the data at three standard deviations to manage the outliers. Then, we standardize the components, so that the mean and the standard deviation of each feature is set to zero and one respectively. Because of the non-linearity of some features, some distributions remain skewed after the standardization (see Table 25 on page 94). Indeed, by definition, some fundamental data is not normally distributed and may be concentrated. To illustrate our point, we present a few examples. With respect to the dividend yield, many companies choose not to pay a dividend or to only pay a small amount relative to their share prices, for example because they plan to reinvest their earnings, in order to fund new initiatives or acquire other companies. This affects the distribution of the statistical data set because the data of these companies are concentrated around zero. For the capital expenditures (capex) over sales, companies may invest or not on capital property during the year. For some companies, these investments may represent minor amounts compared to the sales. As a result, as for the dividend yield, this effect brings data towards zero and the distribution is distorted. On the debt front, companies may not contract debt on the long-term, which set these data also at zero. Regarding the book yield, the concentration of positive figures around zero may be explained by the companies' value that includes increasingly intangible assets when accounting rules do not acknowledge yet this shift in the valuation. This results in very low book yields for companies that invest the most in intangible assets.

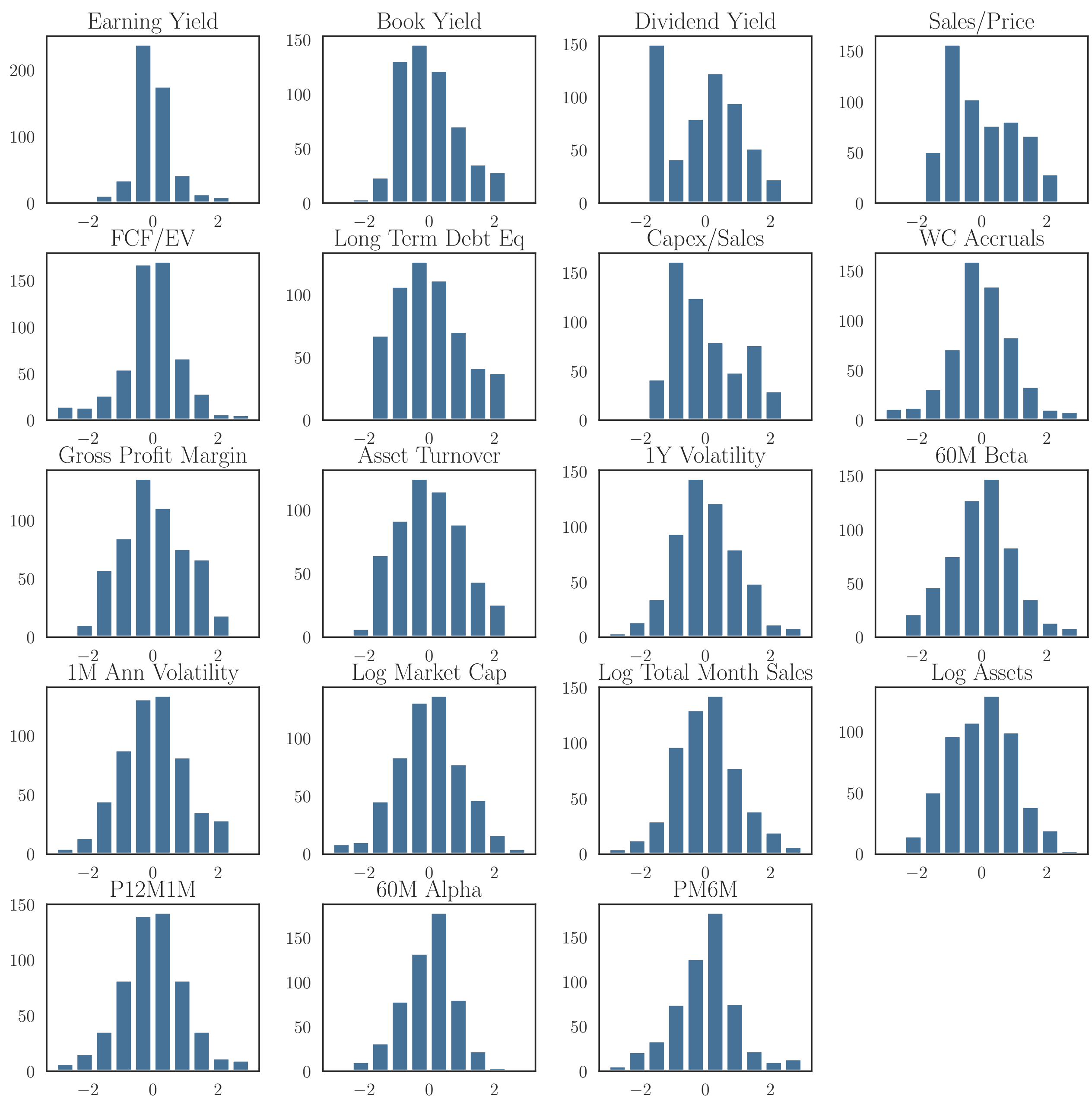
Here, we want to harmonize the distribution of each feature to prevent any concentration of data which could bias the analysis. For that, we choose to use the Power-transformation (Yeo and Johnson, 2000). This method allows transforming data into a normal distribution-like and stabilizes the variance. Moreover, it authorizes zero and negative values. Thus, Figure 15 shows the distribution of the data for North America area on March 2020 after the transformation. As shown in this Figure, most of the features' distributions become closer to a normal distribution[15]. Some features such as dividend yield, capex over sales, long-term debt-to-equity or book yield distributions remain slightly skewed (see Table 25 on page 94), because as described above, our sample contains ratios tending towards zero on these metrics.

Figure 16 shows how the standardized features are correlated between each other for the North America region on March 2020. We notice that most of the market features are highly correlated between each other. At a lower extent, some metrics relating to low volatility (60M beta, 1M volatility), value (book yield, dividend yield, sales/price) and size components (log of total month sales and log of assets) also exhibit strong correlations. According to the dendrogram above the graph, the data has been gathered into four distinct groups. The first group is made of market data, the second is only composed of the capex/sales measure, the third regroups a mix of value, volatility and size metrics and the last one brings together purely fundamental metrics. This highlights a distinction between fundamental data and market data. Indeed, as shown in this figure, market data are uncorrelated to the fundamental ones.

[14] except financial/real estate sectors, that we already exclude from our universe.

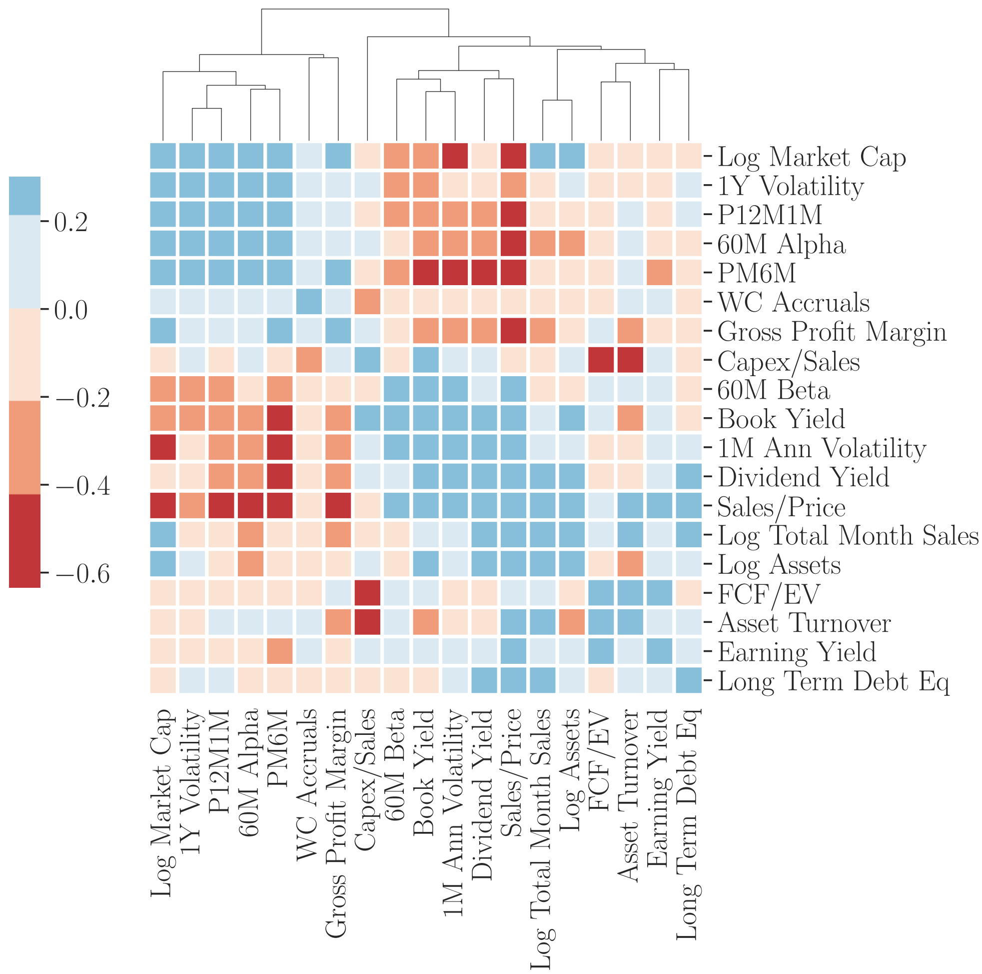
[15] see Figure 36 on page 93 to compare the change before and after the transformation.

Figure 15: Distribution of the 19 features for North America on March 2020 after Yeo-Johnson transformation



Source: MSCI, S&P Compustat, Authors' calculations

Figure 16: Correlation between standardized features in March 2020 for North America



Source: MSCI, S&P Compustat, Authors' calculations

5.2 Mixture Models

In this sub-section, we present different techniques used to build the new clusters. We focus our analysis on unsupervised methods as we would like to test an adaptive approach that would avoid our human bias. Moreover, we decide to use clustering methods because it is an interesting alternative to partition the data. Clustering is defined as a set of methods used to group in C clusters a set of N data points together. The objective is to minimize the similarity between a data point n and its own cluster C_i compared to any data points included in N of other clusters $C_{i'}$ $(i \neq i')$. We focus here, on two main algorithms that are the K-means and the Global Mixture Models.

K-means Clustering According to Rai and Singh (2010), the K-means algorithm is certainly the most commonly used and popular algorithm in the clustering methods. The K of K-means indicates the number of clusters C required to specify in the initialization phase.

Following the specification detailed in Roncalli (2020), C_i cluster defined by the index i where $i = \{1, \ldots, K\}$ must follow the following properties:

1. Clusters must be disjoint: $C_i \cap C_{i'} = \emptyset$ for $i \neq i'$;

2. Each data point has to be a member of a cluster C_i: $C_1 \cup C_2 \cup \ldots \cup C_K = \{1, \ldots, N\}$ where N is defined as the number of observations;

3. Clusters contain data points having statistically similarity;

We can write the K-means algorithm by minimizing the following objective function (Roncalli, 2020):

$$\{C^*, \mu_1^*, \ldots, \mu_K^*\} = \arg\min \sum_{i=1}^{K} n_i \sum_{C(j)=i} \|\mathbf{x_j} - \mu_\mathbf{i}\|^2 \tag{12}$$

where μ_i is defined as the centroid of C_i, and n_i is the corresponding number of observations. The algorithm iteratively does a two-step Expectation Maximization process (see Algorithm 1).

Algorithm 1: K-means

Initialize cluster centroids μ_i for each $i \in (1, \ldots, K)$

Expectation At the iteration s, map each data point to its closest centroid as follows:

$$\{C^{(s)}(j)\} = \arg\min_i \|\mathbf{x_j} - \mu_\mathbf{i}^{(\mathbf{s-1})}\|^2 \tag{13}$$

Maximization Re-define the means of all data points in each cluster C_i and affect updates mean values. The updated mean is defined as follows:

$$\mu_i^{(s)} = \frac{1}{n_i} \sum_{C^{(s)}(j)=i} x_j \tag{14}$$

where x_j represents the attribute of the observation j and n_i is the corresponding number of observations in the cluster i.

repeat

Expectation

Maximization

until Convergence

Each quarter, we iterate the method for different numbers of C_i clusters (for $i \in \{2, 3, \ldots, 10\}$) using the quarterly data introduced in Section 5.1 and we select C_i at the 'Elbow point'. The Elbow method is commonly known as the oldest method (Bholowalia and Kumar, 2014) to evaluate the optimal number of clusters for empirical K-means algorithm. It is also recognized as the most popular and widely used method. It is given by the Algorithm 2.

Algorithm 2: Elbow method to find out K of K-means

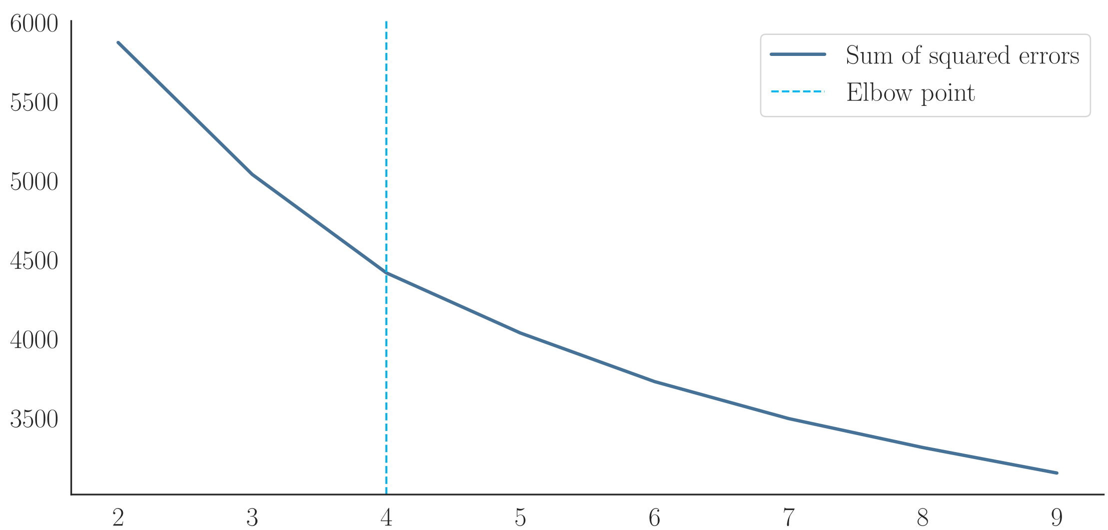
Compute K-means algorithm for different value of C_i for $i \in \{2, 3, \dots, 10\}$
Initialize C_2
compute the sum of squared errors of each data point to their closest centroid

$$S_i = \sum_{i=1}^{K} \sum_{j=1}^{N} \left(\mathbf{x_j} - \mu_\mathbf{i}\right)^2 \tag{15}$$

repeat
Initialize
until i=10

Finally, we define the 'Elbow point' as the point of inflection on the S curve. Graphically, the optimal C_i is defined as an 'elbow'. Figure 17 shows the chosen Elbow point for North America in March 2020.

Figure 17: Elbow point for North America data in March 2020



Source: MSCI, S&P Compustat, Authors' calculations

From 2002 to 2020, our algorithm selects a number of clusters K mainly concentrated around four or five, depending on the area (see Table 10). This is a very satisfactory result because it shows that the number of clusters is stable over time and across the different regions. Too much variability in the number of clusters from one period to another would have called into question the very essence of our portfolio construction strategy. First, it would have generated a very high turnover which would have penalized the performance of the portfolio. Second, beyond the turnover consideration, relative stability means that the nineteen selected features make it possible to identify long-term trends and to associate together companies which are similar based on these trends. This is a behavior sought in this type of exercise.

Table 10: Selected K for the different regions

	North America	EMU	Europe ex-EMU	Japan	Pacific ex-Japan
k=4 (%)	**67.14**	30.00	**52.86**	21.43	32.86
k=5 (%)	32.86	**70.00**	47.14	**74.29**	**65.71**
k=6 (%)	0.00	0.00	0.00	4.29	1.43

Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

Enhancing the model In a first step, we want to get a clear idea of how our clusters are represented. Due to the nature of the market and fundamental data, we expect the data points to be concentrated and the clusters difficult to visualize. To verify this assumption, we show in Figure 18 data clustering for the 19 selected features, using the optimal number of clusters (set at 4 using the Elbow method) for North America's data on March 2020. We apply the Uniform Manifold Approximation and Projection (UMAP) for Dimension Reduction method (McInnes and Healy, 2018) to plot these clusters. UMAP is a nonlinear dimensionality reduction method, very effective for visualizing clusters or groups of data points and their relative proximities. It can be applied directly to sparse matrices thereby eliminating the need to applying any dimensionality reduction such as principal component analysis (PCA) as a prior pre-processing step. In other words, it consists in taking a high dimensional data set and reducing it to a low dimensional graph that retains a lot of the original information. As expected, Figure 18 shows that data points are concentrated and we cannot distinguish the clusters visually.

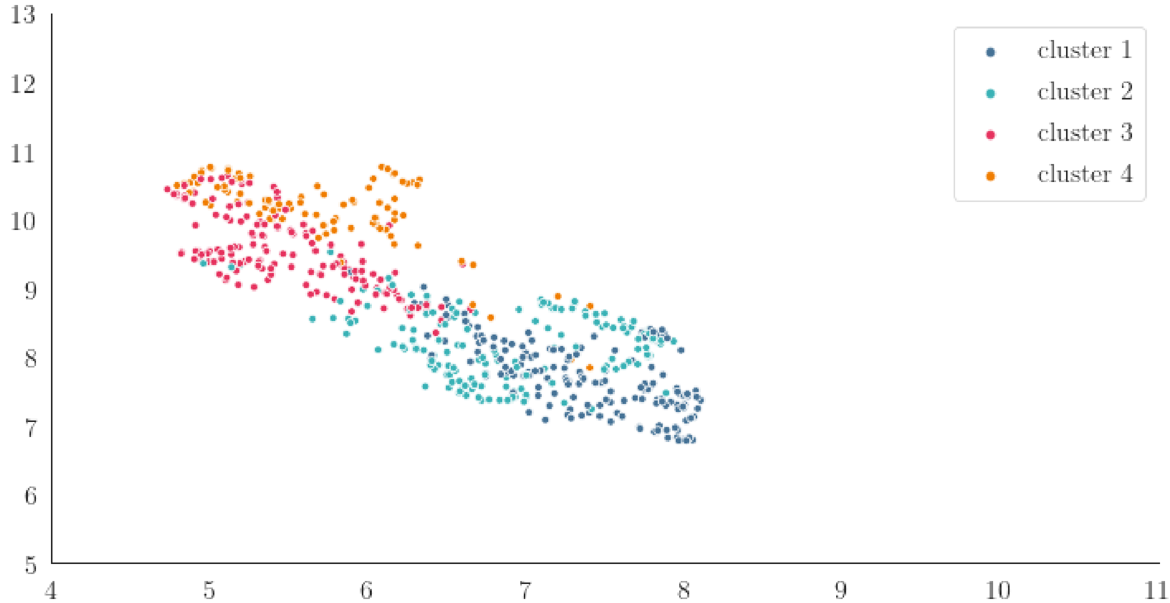
To fix this issue, we shrink the dimensionality of our model using a PCA model and we select the principal components explaining at least 70% of the variablity. According to Ding and He (2004), as it reduces the dimension, the PCA systematically succeeds at data clustering when combined with K-means algorithm. This method allows us to select new variables (between 5 and 8 depending on the geographical area) with the largest variances. The rationale behind this method is to provide coherent patterns that can be more clearly identified. Indeed, through the PCA, the data is projected into lower dimensions space and the K-means is then applied into this space. By adding this method, the data looks less concentrated and the clusters are better defined. Figure 19 shows the results for North America on March 2020.

Despite its popularity, K-means clustering faces some known weaknesses. First of all, this method is non-deterministic, meaning that the way to initialize the centroid is randomly chosen. As a consequence, the algorithm could converge to a local minimum and the listed clusters could change from one run to another one. We fix this issue by using the K-means++ approach from Arthur and Vassilvitskii (2007) during the initialization phase of the algorithm. This method allows to improve the accuracy and the speed of K-means algorithm[16].

Moreover, Barber (2012) defines this algorithm as a special case of Gaussian Mixture Models (GMM) where the variance σ^2 tends to 0. So, K-means considers only the means

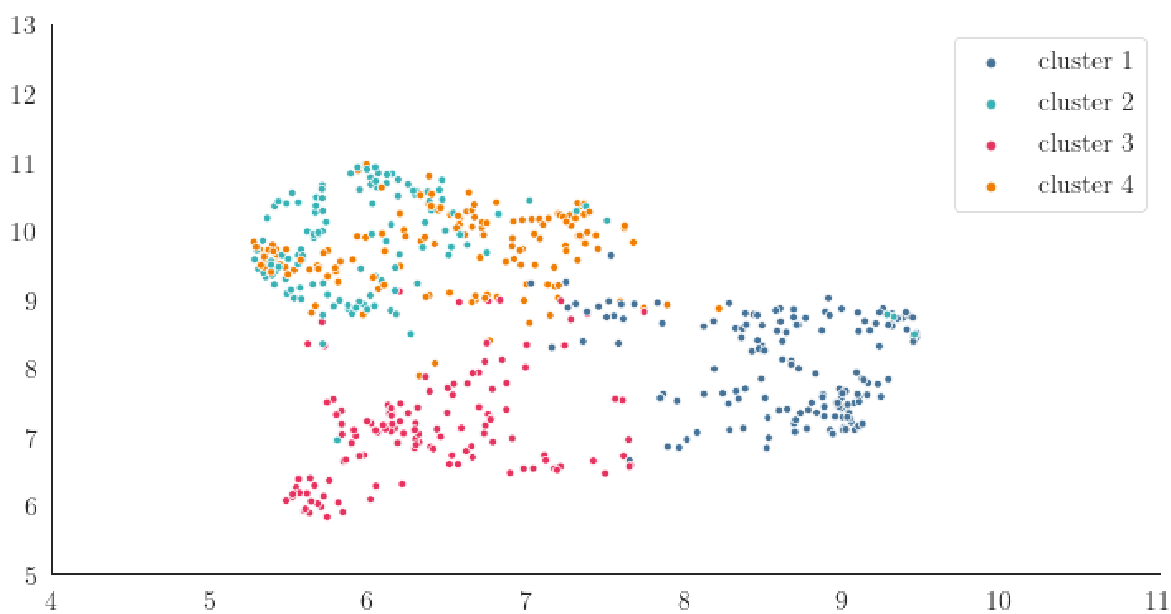
[16]The authors aim to obtain a $\Theta(logk)$ competitive solution to the optimal K-means solution.

Figure 18: Clustering of the dataset for North America on March 2020



Source: MSCI, S&P Compustat, Authors' calculations

Figure 19: Clustering using the first 6 principal components for North America on March 2020



Source: MSCI, S&P Compustat, Authors' calculations

while GMM consider the first two moments to build the clusters' shape. This implies that

K-means algorithm is efficient at capturing structure of the data if clusters have a spherical-like shape. Conversely, K-means does a poor job when the clusters have a more complicated geometric shapes. As a consequence, it is necessary to determine the shape of the clusters before proceeding further. Insofar GMM produce clusters that may be not spherical, a way to proceed is to compare the clusters resulting from the K-means algorithm on one side and from the GMM method on the other. If GMM results in non spherical clusters, we can conclude that the algorithm is better suited to our dataset. On the other hand, if K-means and GMM display comparable results, additional analyzes will have to be carried out.

Gaussian Mixture Models The Gaussian Mixture Models aim to group data points belonging to a single gaussian distribution together. They assume that each cluster is represented by a gaussian distribution. According to Lucke (2019), this algorithm is parameterized by the mixture component weights on the one hand and the means and variances on the other. The sum of the mixture component weights is by definition equal to 1 for each C_i. The probability of the components x to belong in C_i is given as follows:

$$P(x) = \sum_{i=1}^{K} r_i N\left(x|\mu_i, \sigma_i^2\right) \tag{16}$$

$$N\left(x|\mu_i^{(j)}, \sigma_i^2\right) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left(\frac{-\left(x-\mu_i\right)^2}{2\sigma_i^2}\right) \tag{17}$$

Algorithm 3 iteratively does as well as the above two-step process.

Algorithm 3: Gaussian Mixture Models

Initialize cluster centroids μ_i for each $i \in (1, \ldots, K)$ and the variance of the clusters σ^2;
Expectation compute the probability given to a data point to belong to a cluster (or to a distribution). This probability is based on the likelihood of each distribution;

$$P\left(C_i|x_j; r, \mu, \sigma^2\right) = \frac{r_i^{(j)} N\left(x_j|\mu_i, \sigma_i^2\right)}{\sum_{i=1}^{K} r_i^{(j)} N\left(x_j|\mu_i^{(j)}, \sigma_i^2\right)} \tag{18}$$

where r is the probability that x_j belongs to C_i
Maximization Re-define the means μ_i, the variance σ_i^2 of all data points based on the assigned values of the distribution. Compute the new Gaussian density and update the mean and the variance as follows:

$$r_i = \sum_{j=1}^{N} \frac{P\left(C_i|x_j; r, \mu, \sigma^2\right)}{N} \tag{19}$$

$$\mu_i = \frac{\sum_{j=1}^{N} P\left(C_i|x_j; r, \mu, \sigma^2\right) x_j}{\sum_{j=1}^{N} P\left(C_i|x_j; r, \mu, \sigma^2\right)} \tag{20}$$

$$\sigma_i^2 = \frac{\sum_{j=1}^{N} P\left(C_i|x_j; r, \mu, \sigma^2\right) \|\mathbf{x_j} - \mu_\mathbf{i}\|^2}{\sum_{j=1}^{N} P\left(C_i|x_j; r, \mu, \sigma^2\right)} \tag{21}$$

where x_j represents the attribute of the observation j.
repeat
  Expectation
  Maximization
until Convergence

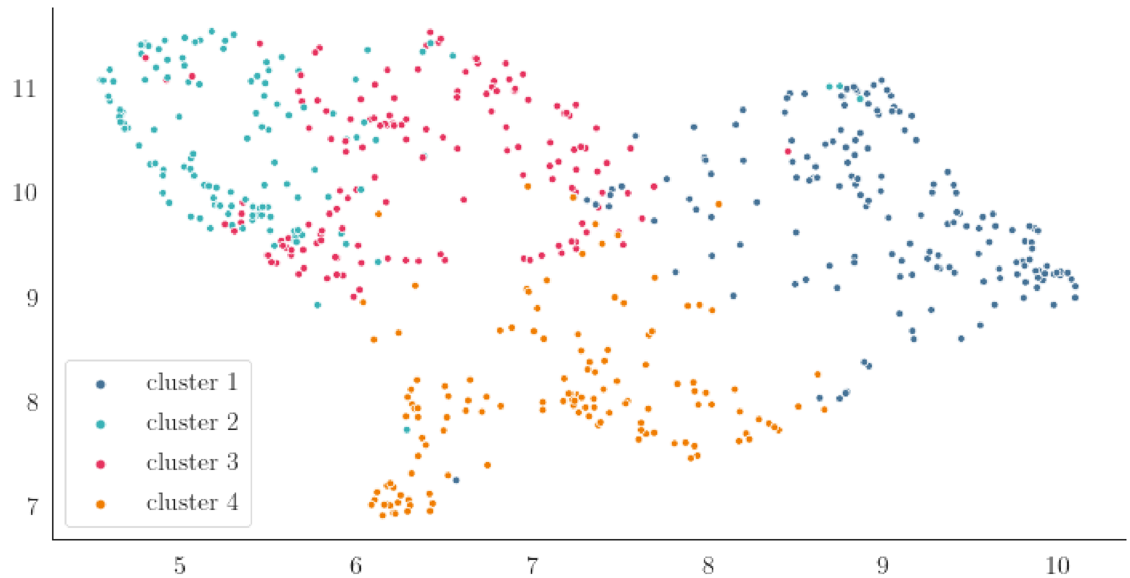
In the next two figures, we plot the clusters resulting from K-means (Figure 20)[17] and GMM (Figure 21). Still on the basis of the March 2020 data for North America region, we shrink the dimensionality of our data using PCA and we pick out the number of principal components having a cumulative sum of percentage of variability above 70%. Then, we select the optimal number of clusters C_i for K-means algorithm using the Elbow method and we launch the GMM algorithm using the same C_i.

Finally, we compare both figures using UMAP graphs as they make it possible to show two-dimensional slices of several dimensions. As it could be misleading to compare graphics with human eyes, we seek to define the best 2D viewing angle that we keep for both figures to make it easier to analyze. In the end, we confirm that clusters look very similar, which

[17] Figures 19 and 20 both display the clusters obtained by the K-means method and differ only by their viewing angle.

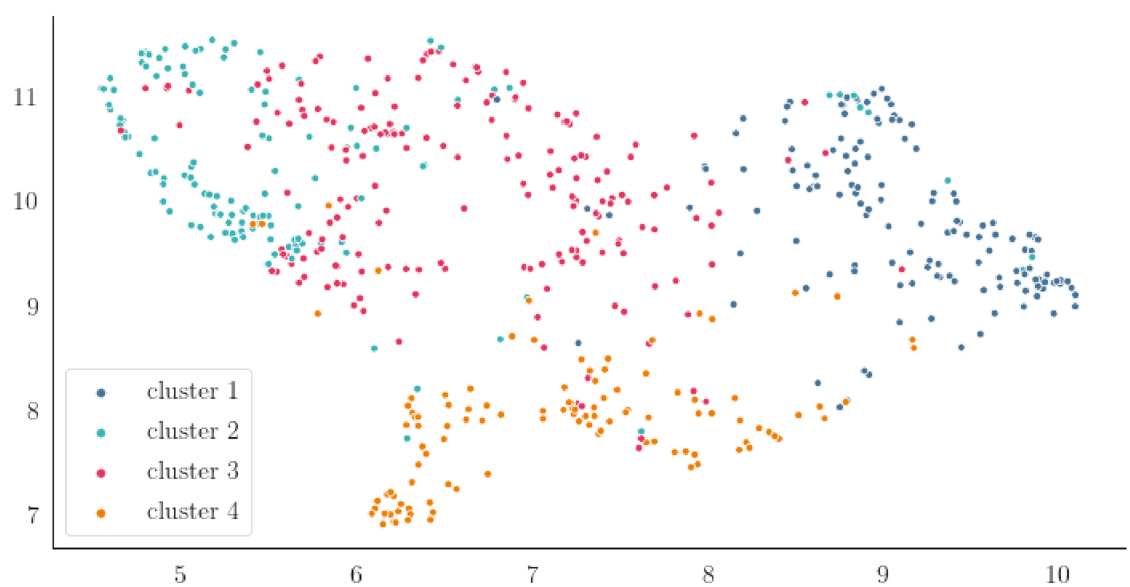
validates our cluster shape on the one hand and determines the stability and the reliability of the K-means approach for our data on the other hand. This is also probably the result of a high-quality data and highlights the impact of data transformation.

Figure 20: K-means clusters for North America on March 2020



Source: MSCI, S&P Compustat, Authors' calculations

Figure 21: Gaussian Mixture Models clusters for North America on March 2020



Source: MSCI, S&P Compustat, Authors' calculations

5.3 Model's validation

Validation methods The Elbow method is not sufficient to validate the C number of clusters. Indeed, we need a validation process testifying the quality of the clustering. To do so, we use three scores known as Silhouette score, Calinski Harabasz score and Davies Bouldin score. They are defined as follows:

- **Silhouette score**: The purpose of this approach is to measure if a data point is assigned to the correct cluster using a measure of cohesion and a measure of separation. That being said, the score computes the similarity between a data point and other points in its own cluster and the dissimilarity between a data point and other points in the closest cluster. This score ranges between -1 and 1. A score of one indicates that the data points are closer to their assignment cluster than from neighboring clusters, thereby certifying that the clusters are well disjoint. According to Rousseeuw (1987), for each observation j in the data set, the Silhouette score is defined as follows:

$$S_{SIL(j)} = \frac{\min\left(d\left(j,i'\right)\right) - d\left(j,i\right)}{\max\left(d\left(j,i\right), \min\left(d\left(j,i'\right)\right)\right)} \tag{22}$$

  where $d(j,i)$ represents the average distance between j and all other observations of cluster C_i and $\min(d(j,i'))$ the "smallest" average distance between the observation j and all other observations of cluster $C_{i'}$ where $i \neq i'$.

- **Calinski Harabasz score**: This score measured by Caliński and Harabasz (1974) is defined as the ratio between the inter cluster variance $d(\mu_i, \mu)$ and the intra cluster variance $d(j, \mu_i)$. The value of this ratio can fluctuate between 0 and $+\infty$. By definition, K-means algorithm is supposed to maximize this score. This method is not affected by the number of clusters. The Calinski Harabasz score is defined as follows:

$$S_{CH(j)} = \frac{\left(N-K\right) d\left(\mu_i, \mu\right)}{\left(K-1\right) \sum_{i=1}^{K} \left(d\left(j, \mu_i\right)\right)} \tag{23}$$

  where N is the number of observations and K, the number of clusters.

- **Davies Bouldin score**: This metric has been introduced by Davies and Bouldin (1979), a lower score suggests a better clustering. This score remains positive by construction and the minimum is reached at 0. So, clusters having a higher score will be the ones being less dispersed. The score is defined as follows:

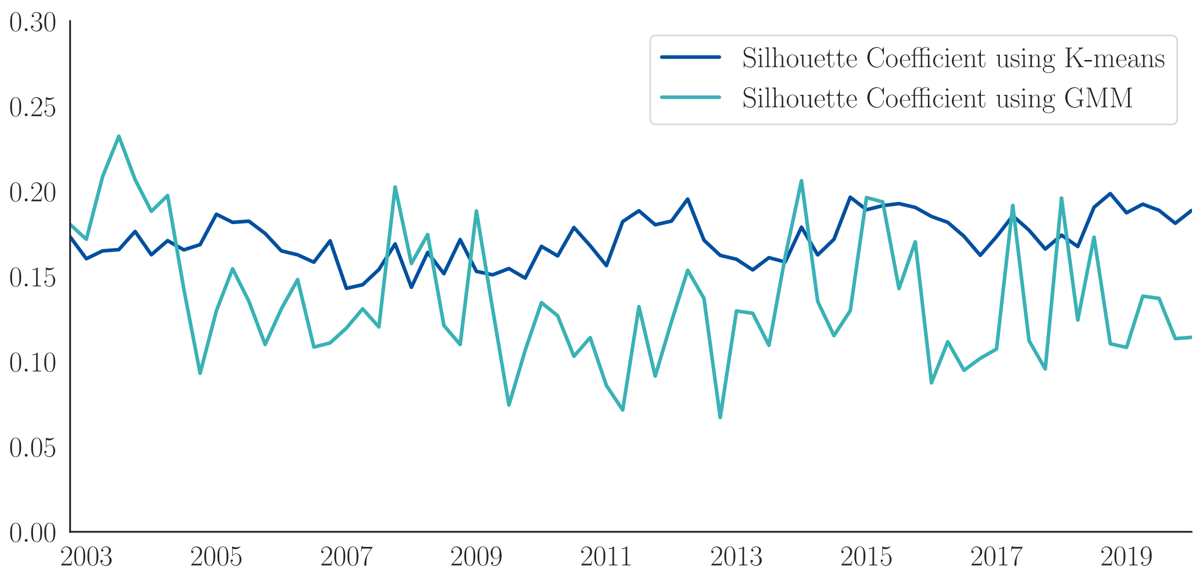
$$S_{DB(j)} = \frac{1}{K} \sum_{i=1}^{K} \max_{i \neq i'} \frac{d\left(j,i\right) - d\left(j,i'\right)}{M_{i,i'}} \tag{24}$$

  where $d(j,i)$ is measured as the average dispersion between a data point j to its cluster C_i and $M_{i,i'}$ represents the distance between vectors of clusters C_i and $C_{i'}$ where $i \neq i'$.

Choice of the algorithm GMM have been useful to determine the optimal clusters' shape for our dataset. Its characteristics also make it a robust candidate for integrating our model. To do so, each quarter we launch the GMM using different values of C_i clusters. Then, we select the optimal number using the Bayesian Information Criteria (Schwarz, 1978) (BIC). Indeed, the BIC criteria compares different models that use different number of parameters. The lower the BIC, the more the model will be able to predict the data. Here, we select the point where the BIC's curve starts flattening.

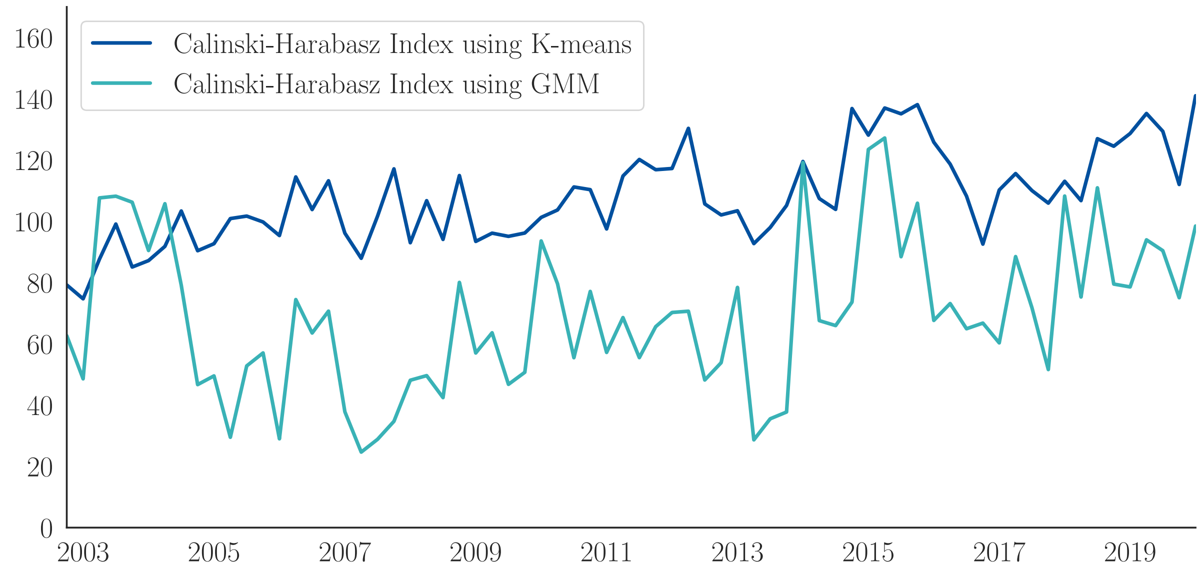
The next step consists in comparing K-means and GMM using the above validation methods. Each quarter, we compute these scores for the selected number of clusters in both algorithms for North America area in March 2020. Figures 22, 23 and 24 show that scores are more stable and higher for K-means algorithm than GMM.

Figure 22: Silhouette scores for K-means and GMM for North America (2003-2020)



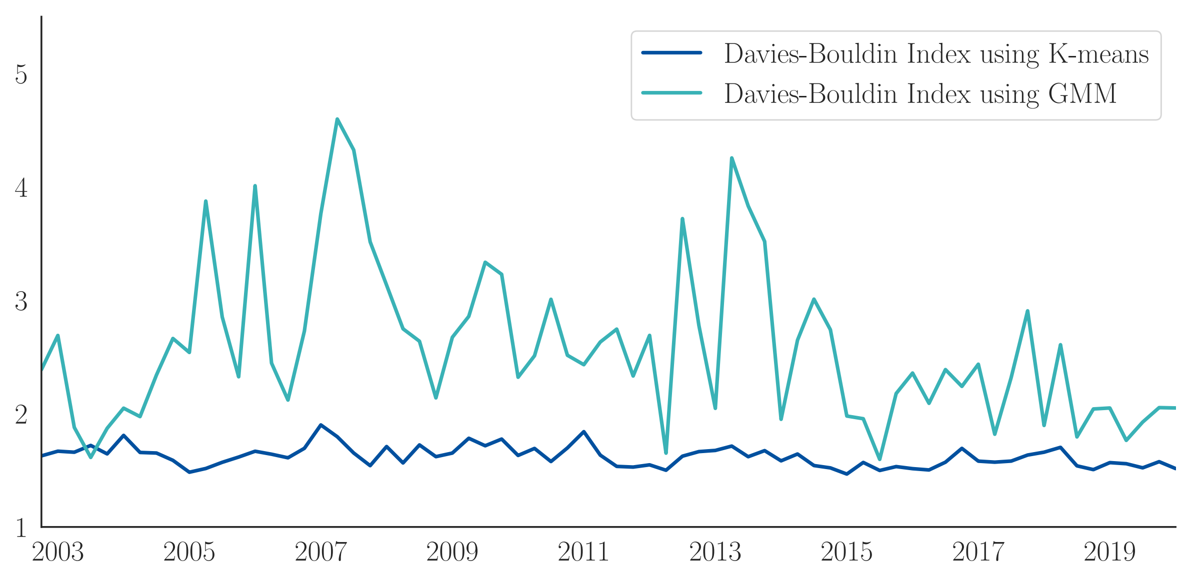
Source: MSCI, S&P Compustat, Authors' calculations

Figure 23: Calinski Harabasz scores for K-means and GMM for North America (2003-2020)



Source: MSCI, S&P Compustat, Authors' calculations

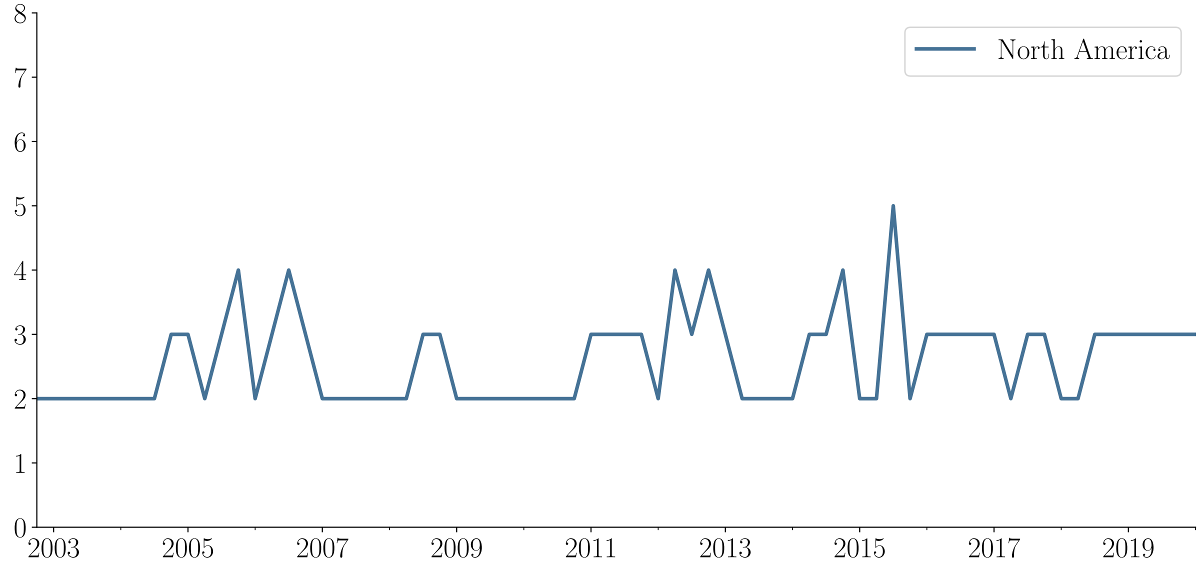
Figure 24: Davies Bouldin scores for Kmeans and GMM for North America (2003-2020)



Source: MSCI, S&P Compustat, Authors' calculations

At this stage, we may suggest to use K-means in our analysis because K-means delivers sufficient results in terms of quality of the clustering. Indeed, K-means provides higher silhouette score, higher Calinski Harabasz score and lower Davies Bouldin score than Gaussian Mixture Models over time (see Figures 22, 23 and 24).

Figure 25: Selected C_i of GMM algorithm for North America



Source: MSCI, S&P Compustat, Authors' calculations

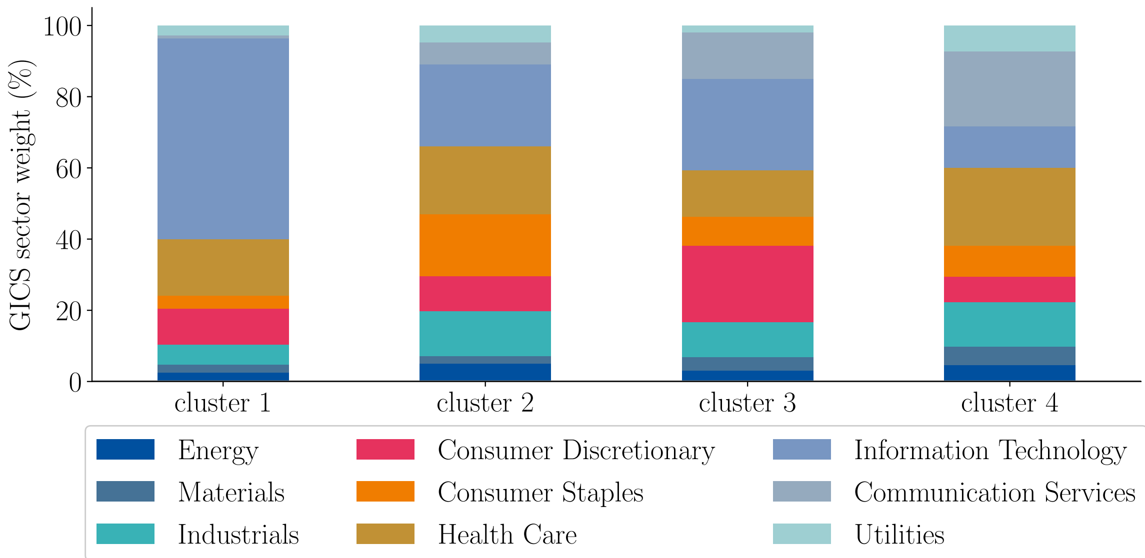
Moreover, the BIC used to select the number of C_i cluster for GMM, seems to be volatile. It means that the selected C_i using the BIC is sometimes varying a lot from one rebalancing date to another one (see Figure 25). This implies an increase of the turnover from the portfolio manager point of view, which is sufficient to exclude this method.

5.4 Quality strategy integrating unsupervised learning

Clusters analysis In this sub-section, we seek to evaluate the clusters' fundamental drivers, or, in other words, the way companies are grouped. The first finding we can highlight is that companies are well-distributed across clusters and over time, as we do not observe excessive company concentration within one or two clusters over the study period. For instance, in March 2020 for North America, our algorithm has selected 4 clusters representing respectively 22.3%, 23.9%, 25.6% and 28.1% of the universe. Figure 37 on page 95 shows the proportion of stocks in each cluster and for all regions since December 2002. The clusters are well balanced not only in North America but also across all regions.

Then, we examine how sectors are represented within clusters. We want to ensure that the clustering process does not simply result in the grouping of existing sectors within "super-sectors". Figure 26 shows how clusters are composed in terms of GICS sectors for North America as of March 2020. We note first that all sectors are represented within each cluster, thus validating that our clusters are not "super-sectors". We also notice that around 90% of our first cluster consists in information technology, healthcare and consumer discretionary companies. These three sectors are precisely the ones that are almost systematically over-represented in an unconstrained portfolio allocation, as highlighted in the Section 4, within the sector allocation of our long-only quality portfolio. This obviously gives us indications about the fundamental bias of the first cluster. We will discuss and confirm this bias in analyzes that follow.

Figure 26: Sector allocation within clusters for North America in March 2020



Source: MSCI, S&P Compustat, Authors' calculations

In addition, we verify that clusters are well diversified over time by measuring the maximum weight attributed to a sector in a given cluster. This is not an essential check insofar as if a cluster were entirely composed of a single sector, it would not call into question our approach. However, this allows us to get a more precise idea of how our clusters are constituted. We find that it exceeds 60% in only 0.29% of the time for North America and EMU zone, 2.8% times for Europe ex-EMU, 1.1% times for Japan and 4.85% times for Pacific ex-Japan. We conclude that there is no sector concentration within clusters.

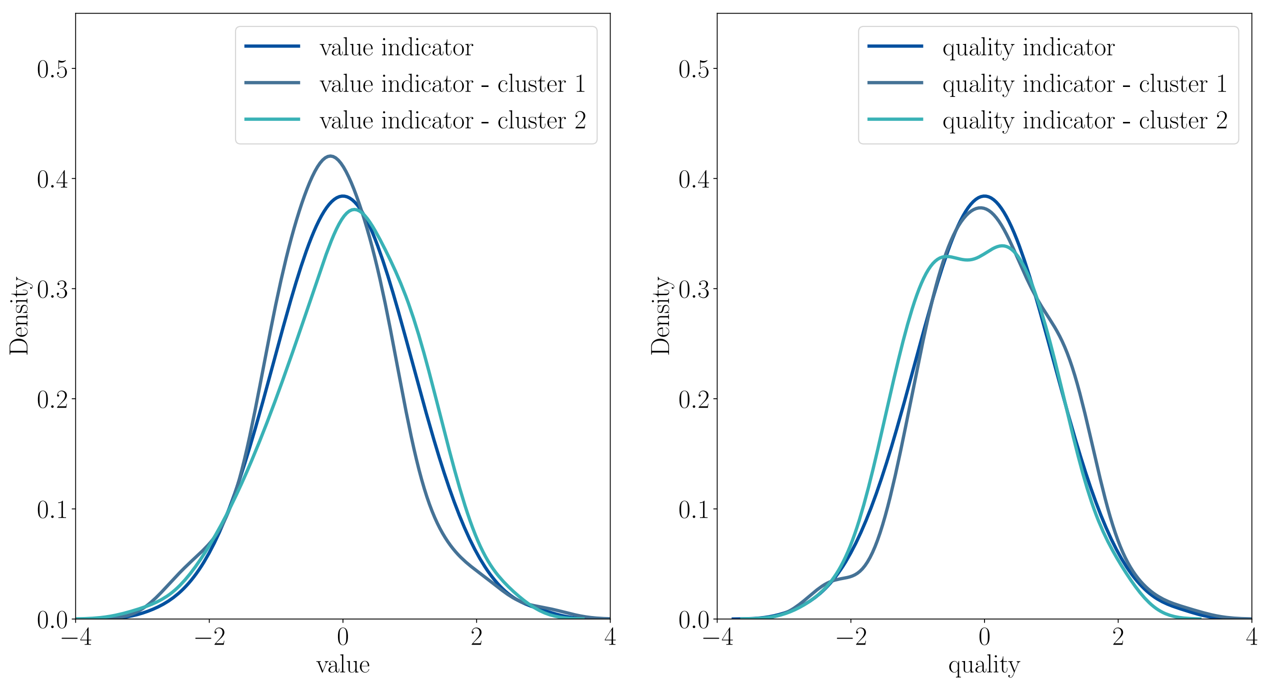
After the sector analysis, we evaluate the sensitivity of the companies to the traditional risk-factors. For that, each quarter we compute a z-score per risk-factor[18] according to the indicators we have selected in Section 5.1. Then, we display the probability density function of each indicator through the components of MSCI North America in March 2020. We expect this probability density distribution to match the normal distribution as we work with z-scores. The objective here is to compute a conditional probability density function per clusters to see how these distributions drift from the normal distribution.

Therefore, we compute for each cluster means of the selected indicators. Over this period, two clusters have particularly caught our attention. As previously assumed, results show that the first cluster is positively tilted on quality, as it displays a positive quality mean associated to a negative value mean. It also exhibits a small size, a low volatility and a positive momentum bias. On the contrary, the second cluster is positively tilted on value, with a positive mean for value indicator, a negative mean for the quality, a small size, a negative

[18]Quality, value, low volatility and momentum.

momentum and a high volatility bias. Figure 27 shows the probability density functions for these two clusters compared to the probability density function of all companies belonging to the MSCI North America index in March 2020. Therefore, we can see the small drift of the means to the left or to the right according to the above results. On the other hand, this figure validates the results of our analysis in Section 5.2: the clusters similarity between K-means and GMM works because our clusters are simply a mixture of gaussian distributions.

Figure 27: Density functions for quality and value indicators for North America in March 2020

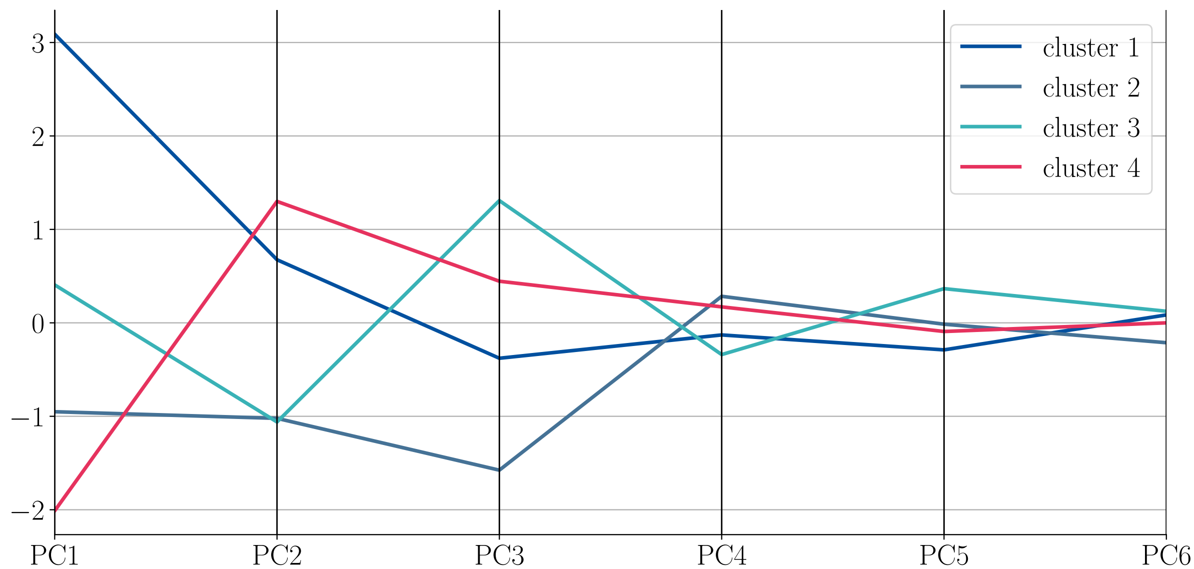


Source: MSCI, S&P Compustat, Authors' calculations

Finally, we seek to better understand the meaning of the clusters using a centroid analysis. This method allows us to capture the drift of the vector of means of each variable from zero. Figure 28 exhibits the centroid positions of each principal components for each clusters for North America area in March 2020. We display the centroid positions through a parallel coordinates graph. This method allows to visualize multivariate dimensions through axes placed in parallel.

At first glance, the analysis of this graph seems difficult to achieve because we try to evaluate the centroids of unsupervised dimension reductions. Another way of doing this analysis would have been to determinate beforehand the exposition of each principal components to the fundamental data. But this is not the objective of this study. Here we capture the principal components of a sample, independent from the other rebalancing dates or regions samples. It would not have make sense to apply it to all samples. The main result to highlight is that the distribution means drift drastically from the normal distribution mean for the first three principal components and converge to zero subsequently. We observe the same behavior for the other dates and/or regions. As a reminder, principal components are systematically ranked in descending order of explained variance. As a consequence, we can reasonably expect that a large proportion of the explained total variance is concentrated in these first three principal components. The fact that our clusters display very different exposures to them confirms that clusters are by construction very different from each other, which is obviously the expected result.

Figure 28: Centroids for North America data in March 2020



Source: MSCI, S&P Compustat, Authors' calculations

Backtest results The objective of this section is to build a quality factor strategy by grouping stocks showing fundamental-based and/or market-based similarities. Compared to an unconstrained portfolio construction (see Section 4), we therefore impose an additional step, proposing an intermediary solution between an unconstrained and a sector-neutral portfolio construction. We do not expect stocks to be gathered around a unique characteristic such as sectors. Here, we aim to capture a variety of signals but also the ones that we are not able to explicitly quantify currently. We think that fundamentals and market features catch a lot of information about companies and the whole economy. As a reminder, we do not necessarily aim to improve the in-sample backtest performance. We want to measure effects of diversifying the equity portfolio according to dimensions going beyond the sectors. From a methodological point of view, we start by running the clustering process on individual regions on a quarterly basis. Then, we select the 20% best-ranked quality stocks in each cluster to capture the quality bias through "similar" companies. For a given region, stocks are value weighted within the clusters. Finally, we aggregate the clusters in a regional portfolio by keeping the relative weight of the initially constituted clusters. As for the unconstrained quality reference strategy, the global portfolio is a combination of the regional portfolios we weight according to the region's free-float market capitalization in the MSCI World Developed Markets.

Table 11 shows the statistics of the multidimensional quality factor we presented in Section 4 with and without integrating the clustering approach. As a reminder, we also add the sector-neutral version of the long-only factor. At the global level, the annualized performance and the risk-adjusted return of the approach including clusters are up over the period (+70 basis points for the performance and +0.5 for the risk-adjusted return). Regarding the regions, adding the clustering step allows to significantly improve performances, expect for Japan (−97 basis points). The main contributor to the performance increase is EMU, with a meaningful +240 basis points per year. At the same time, the risk measure remains constant at the global and regional levels (except for EMU area where the volatility falls and Europe ex-EMU area where it increases).

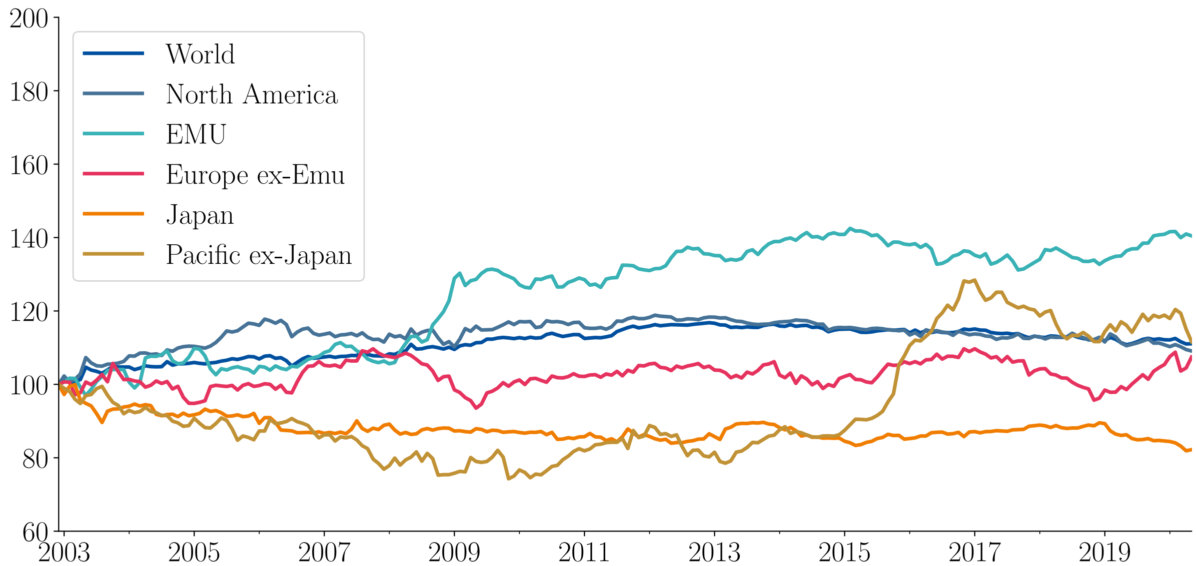
Table 11: Long-only performance of the multidimensional quality factor with and without the use of clustering approach

		Uncons-trained	**With clustering**	**Sector-neutral**
Global	Ann. Return	12.6%	13.3%	11.8%
	Ann. Excess Return	2.8%	3.5%	2.1%
	Ann. Volatility	13.3%	13.3%	12.9%
	Risk-Adj. Return	0.95	1.00	0.91
North America	Ann. Return	13.7%	14.3%	12.5%
	Ann. Excess Return	2.9%	3.4%	1.7%
	Ann. Volatility	14.7%	14.8%	13.8%
	Risk-Adj. Return	0.93	0.97	0.91
EMU	Ann. Return	8.5%	10.9%	9.7%
	Ann. Excess Return	0.7%	2.8%	1.7%
	Ann. Volatility	15.7%	15.0%	14.6%
	Risk-Adj. Return	0.54	0.73	0.66
Europe ex-EMU	Ann. Return	13.0%	13.5%	11.3%
	Ann. Excess Return	4.3%	4.9%	2.9%
	Ann. Volatility	12.9%	13.5%	13.2%
	Risk-Adj. Return	1.01	1.00	0.85
Japan	Ann. Return	9.5%	8.6%	9.7%
	Ann. Excess Return	2.9%	2.0%	3.3%
	Ann. Volatility	16.3%	16.2%	16.4%
	Risk-Adj. Return	0.58	0.53	0.59
Pacific ex-Japan	Ann. Return	10.1%	10.9%	10.3%
	Ann. Excess Return	0.4%	1.3%	0.6%
	Ann. Volatility	20.1%	20.2%	19.7%
	Risk-Adj. Return	0.50	0.54	0.52

Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

Figure 29 exhibits the outperformance of the quality strategy integrating the cluster approach versus the version we presented in the previous section. The global strategy outperformance seems to be in line with the outperformance of the North American strategy. This makes intuitive sense as North America represents in average 58% of the investment universe and is supposed to drive the overall performance. Note that generally these strategies outperform the unconstrained quality strategies until 2013 and decline slightly since. We also observe that most of the outperformance delivered by the EMU strategy integrating cluster approach occurs during the GFC. The rest of the time, the strategy still outperforms but in a much more lower way. As regards with other regions, the European ex-EMU excess return mean reverts around zero while the contribution of the cluster approach is negative for both Asian strategies during the value rally. The situation then stabilizes for Japan, while the Pacific ex-Japan outperformance rockets up in 2015 and 2016.

Figure 29: Outperformance of long-only quality strategy integrating clustering versus unconstrained quality strategy (2003 - 2020)



Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

After demonstrating that the implementation of the cluster approach on the multidimensional quality factor significantly improves the performance of the strategy without impacting its volatility, the question arises whether this improvement allows at the same time to enhance its ability to predict the cross-section of expected returns. Table 12 shows the regression statistics of the long-short quality strategy integrating clustering method to the Carhart factors. First, we note that the alpha of the global strategy remains unchanged, at 0.36% on a monthly basis, and is still highly significant. Then, and it is the most notable improvement, the EMU region delivers a positive alpha of 0.21% (against 0.0% for our unconstrained multidimensional quality factor), statistically significant at the 90% level. Results for other regions are more mixed. While both the average monthly excess return and the alpha show an improvement on Europe ex-EMU, they decrease in North America, Japan and Pacific ex-Japan. Regarding the other risk-factors, their relationship of each with the new quality factor remains substantially unchanged from our initial version. The new quality factor still loads negatively on value and shows a slight bias towards mid-caps.

Table 12: Alphas and factor loadings of the multidimensional quality factor integrating clustering

	Global	North America	EMU	Europe ex-EMU	Japan	Pacific ex-Japan
Excess return (%)	0.43	0.43	0.24	0.53	0.25	0.25
Alpha (%)	0.36***	0.25*	0.21*	0.53***	0.33*	0.31
	(2.96)	(1.47)	(1.32)	(2.97)	(1.95)	(1.49)
Market	-0.001	0.051	0.010	-0.082***	-0.092***	-0.085**
	(-0.25)	(1.05)	(0.25)	(-1.55)	(-2.60)	(-2.25)
Size	0.149**	0.138*	0.126**	0.135***	0.003	-0.058
	(2.20)	(1.80)	(2.25)	(2.83)	(0.42)	(-0.86)
Value	-0.384***	-0.511***	-0.281***	-0.227***	-0.117*	-0.002
	(-5.43)	(-5.28)	(-5.52)	(-4.28)	(-1.87)	(-0.04)
Momentum	-0.041	-0.094***	0.017	0.035	-0.018	0.038
	(-1.08)	(-2.14)	(0.44)	(1.06)	(-0.34)	(0.73)
Adjusted R^2	0.16	0.19	0.18	0.11	0.05	0.04

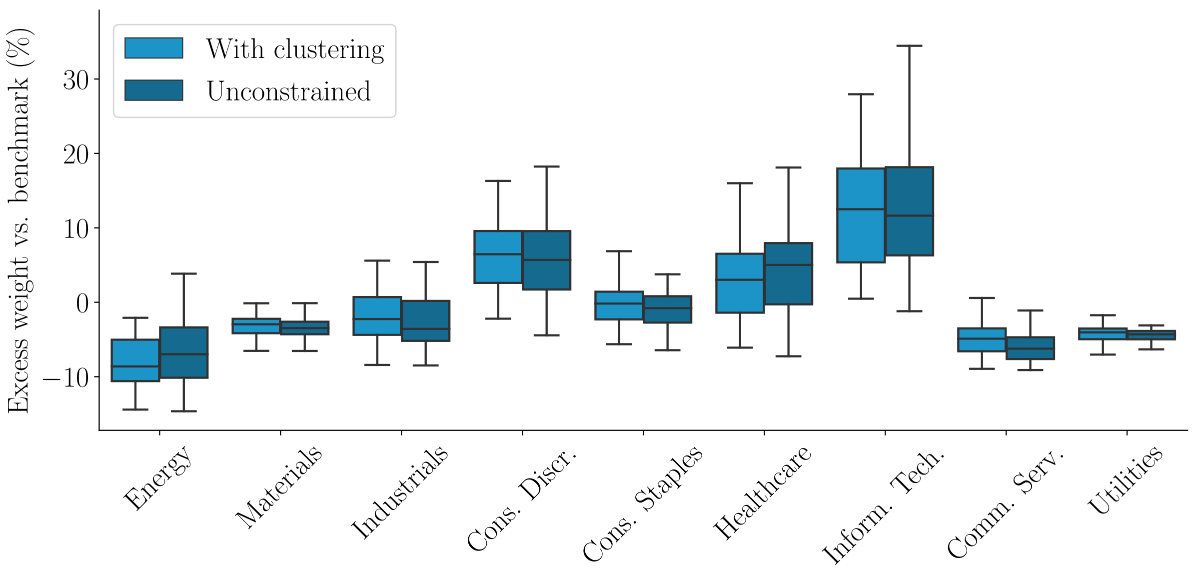
t-statistics (in parentheses) are based on White's standard error
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

In Table 11, we have highlighted the performance improvement that occurs on all regions, except Japan, by implementing the cluster approach. Thus, it may seem surprising that some excess returns we explain on Table 12, as part of the Carhart four-factor regression model, are lower than the ones displayed in a similar regression analysis, for our unconstrained quality factor (Section 4, Table 2 on page 27). This is because returns on the Portfolios Q_5 have also improved thanks to the cluster implementation, but in a higher proportion than that observed on the Portfolios Q_1.

Portfolio composition In this subsection, we analyze how the cluster technique impacts both the sector allocation and the stock selection of the global strategy. Figure 30 provides information on sector bets versus benchmark of the two global quality strategies. While we display in light blue the excess sector weights of the strategy including the cluster approach, the dark blue boxes refer to our conventional unconstrained quality factor. We choose a box plot representation as it is a standardized way of displaying the distribution of data based on a five number summary, including the minimum observed data, first (lower) quartile, median, third (upper) quartile, and maximum observed data. A given box represents the middle 50% of the distribution, showing the Q_1 and Q_3 quartiles of the dataset. The whiskers extend to show the rest of the distribution. The median is shown by the line that divides the box into two parts. Half of the observed data are greater than or equal to this value and half stands below. To consider a concrete example, the first box plot on the figure shows that the quality factor that includes the cluster approach systematically underweights the energy sector from December 2002 to May 2020. Underweightings range from -14.4% to -2.1%, and 25% of observations fall below -10.6%, while 25% are above -5.0%. Finally, half the observations are greater than or equal to -8.6%, which corresponds to the median value. Interestingly, we observe on the figure that the two versions of the global quality factor do not display fundamentally different sector bets. Values of the first and third quartiles are very close across all sectors, and medians are also very similar. We only note some slight discrepancies between medians of energy, industrials and healthcare sectors. Overall, the two strategies tend to overweight information technologies, healthcare and consumer discretionary sectors, while they underweight energy, materials, communication services and utilities companies.

Figure 30: Sector bets of the global quality factor with and without clustering (2003-2020)



Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

In Figure 31, we complete our analysis on sector allocation at the global level by providing a time-series computed on sector weight differences between the strategy including clusters on the one hand and the conventional unconstrained strategy on the other. A positive value on a given sector at a date d means that this sector, on this date, is over weighted in the strategy that includes the cluster approach. Overall, we observe that all curves display on average values close to zero, thus confirming information we collected from Figure 30. However, we also find that sector deviations can very occasionally be quite large, even reaching 10 to 15% on a few sectors. Most affected sectors are healthcare, information technology, consumer discretionary and energy. These differences in sector weights are sometimes very volatile over several consecutive rebalancing periods, suggesting a significant increase in turnover related to the cluster implementation. We will discuss the subject of turnover in the next part of this sub-section.

Figure 31: Differences in sector weight between global quality strategies (clustering version vs. unconstrained version)

Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

In Appendix E on page 96, we show complementary boxplots for the five regions that make up the global developed markets in our study. Of course, we sometimes observe larger differences by focusing at the regional level, as for the healthcare sector in Europe ex-EMU. This is due to the diversification effect: the more we focus on a concentrated area, the more we increase the probability of observing pronounced differences. However, we still note a strong similarity in sector allocation between the regional portfolios built with and without the cluster approach.

At the same time, these figures provide valuable information on how the quality factor is characterized within regions from a sector point of view. Indeed, beyond the official sector representation within each region (it is for instance well established that the technology sector is clearly over represented in North America, while financials, healthcare and industrials predominate respectively in EMU, Europe ex-EMU and Japan in recent years), we recall here that our portfolio construction process is unconstrained and region neutral. Therefore, the sector allocation resulting from the portfolio construction process gives a clear indication of the sectors where quality companies are concentrated in regions, and where they are not. To put it simply, the larger the median overweighting of a sector over our study period, the more the sector can be described as a "quality sector", and vice versa. Of course, we reason in relative terms in order to take into account the initial importance of a sector in a region as it is not equivalent to overweight by 5% a sector representing on average 25% of the index and to overweight by 5% a sector representing on average 10% of this index. Table 13 shows in a very didactic way the sectors being the most (the least) exposed to the quality factor according to our multidimensional definition. There are sometimes two signs separated by a slash that appear at the intersection of a sector and a region. It means that the construction of the portfolio integrating the clusters (to the right of the slash) displays different results from the construction of the standard unconstrained quality portfolio (to the left of the slash). We note that the information technology sector is clearly the sector the most representative of the quality concept. This is the case in all the regions that make up the developed markets, including in Europe ex-EMU and in Pacific ex-Japan, where the sector is very poorly represented in official indices. Conversely, the quality factor is structurally underrepresented, and sometimes even absent, within the energy, materials, communication services (except in Japan) and utilities sectors. If the situation does not suffer from any ambiguity within the above-mentioned sectors, it is not so clear for the others. We observe, for example, that while the quality factor is very present in the consumer discretionary sector in North America and in Europe ex-EMU, it is absent in other regions, in which the risk-factor is more associated with the consumer staples. Regarding industrial and healthcare companies, if the former seems to be solely related to the quality factor in Europe ex-EMU, the latter seems to be so both in Europe ex-EMU and in Japan.

To sum up, the cluster implementation within our quality strategy does not generate excessive sector distortions versus our initial quality factor. Both versions show that the quality characteristic is partly sector-related since some of them are systematically over / under represented in all the regions that make up the global portfolio. We also note that some specificities emerge locally. In the same way that the definition of the quality concept is not universal, the way in which the risk-factor is concretely translated within sectors in regions may sometimes differ from one to another. We think this can be explained by a variety of structural factors (economic and fiscal policy, structure of the market, number of competitors, technological advancement, cultutral specificities ...).

To complete the analysis of the composition of our quality portfolios, we assess the overlap between the two versions of our quality factor, and then we compare their quality biases. We measure overlap over a given period as the sum of the securities' weights held simultaneously in the two quality strategies. The average overlap over the study period is 75%, meaning that the cluster implementation generates significant changes in the portfolio composition. Insofar as the two portfolios are very similar from a sector allocation point of view, we can logically

Table 13: Average sector exposure to the quality factor within regions vs. benchmark (2003-2020)

	Global	North America	EMU	Europe ex-EMU	Japan	Pacific ex-Japan
Energy	−−	−−	−−	−−	−−	−−
Materials	−	−−	−	−−/−	−−	−
Industrials	−/=	−	=	++	−	−−/−
Cons. Discretionary	+	++	=	++	−	++
Cons. Staples	=	=	++	−−	+	++/+
Healthcare	+/=	=	=/−	++/=	++	−−/+
Inform. Technology	++	++	++	++	++	++
Comm. Services	−−	−−	−−	−−/−	+	−/=
Utilities	−−	−−	−−	−−	−−	−−

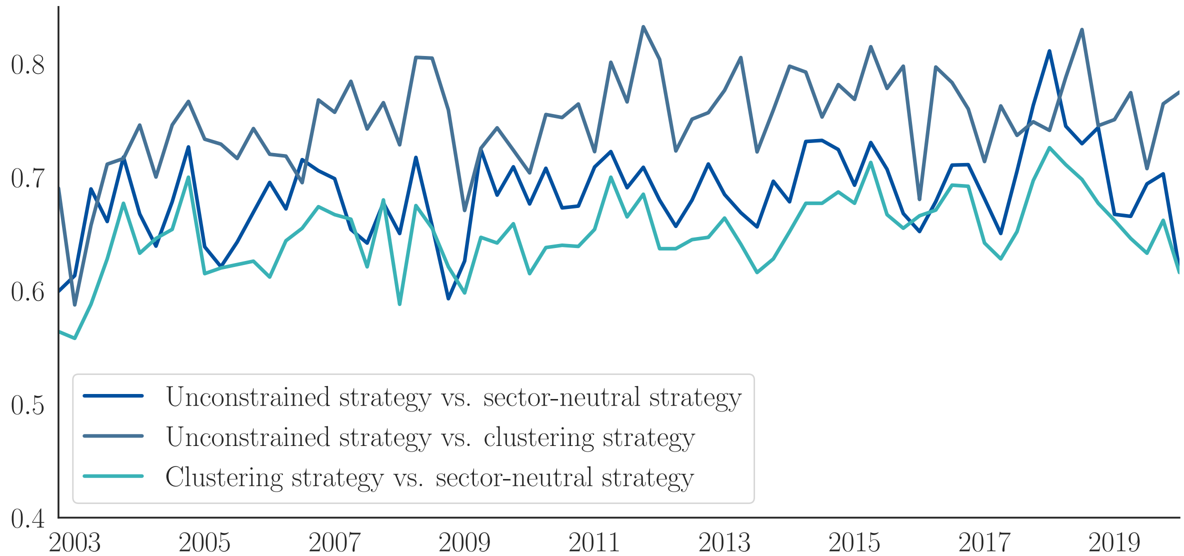
Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

deduce that these differences come from a stock selection effect.

Consequently, we can legitimately question ourselves whether these changes in composition come at the expense of the diversification of our new portfolio. To assess the effects of the cluster implementation on diversification, we also measure the overlaps between the two quality portfolios and the sector-neutral version of the quality factor. By doing so, we consider that the sector-neutral portfolio is, by construction, a well-diversified portfolio. Finally, the average overlaps versus the sector-neutral strategy stand at 69% for the initial strategy and at 65% for the strategy including clusters, meaning that the quality portfolio built with clusters deviates slightly more from the composition of the sector-neutral portfolio than does the initial quality portfolio. To make sure that these two overlap measures relate to a common base, we also measure the overlap between the three quality portfolios. Obviously, the resulting figure cannot exceed 65% and finally stands at 57%, indicating that the three portfolios share a significant common core of quality-stamped stocks. Figure 32 compares the three time-series of overlaps over time.

In Figure 33, we report the absolute frequency distribution of the aggregated z-scores computed on a quarterly basis on our three quality portfolios for the full sample. The methodology for calculating multi-dimensional z-scores is explained in Section 2. At the portfolio level, the aggregated z-score is equal to the weighted average of the individual z-scores of stocks held. The figure shows that on average, and compared to the two other strategies, the sector-neutral strategy is significantly less exposed to the eight quality indicators, with an average aggregated z-score of 1.24. This is a logical result insofar as the sector-neutral strategy is by definition a constrained strategy. The difference between the average aggregated z-score displayed by the sector-neutral portfolio and that displayed by the unconstrained portfolio corresponds to an opportunity cost. Regarding the two other strategies, they display very similar exposures to quality, with an average aggregated z-score of 1.47 for the strategy that includes clusters while the initial strategy displays a value of 1.51, meaning that the stage of cluster implemen-

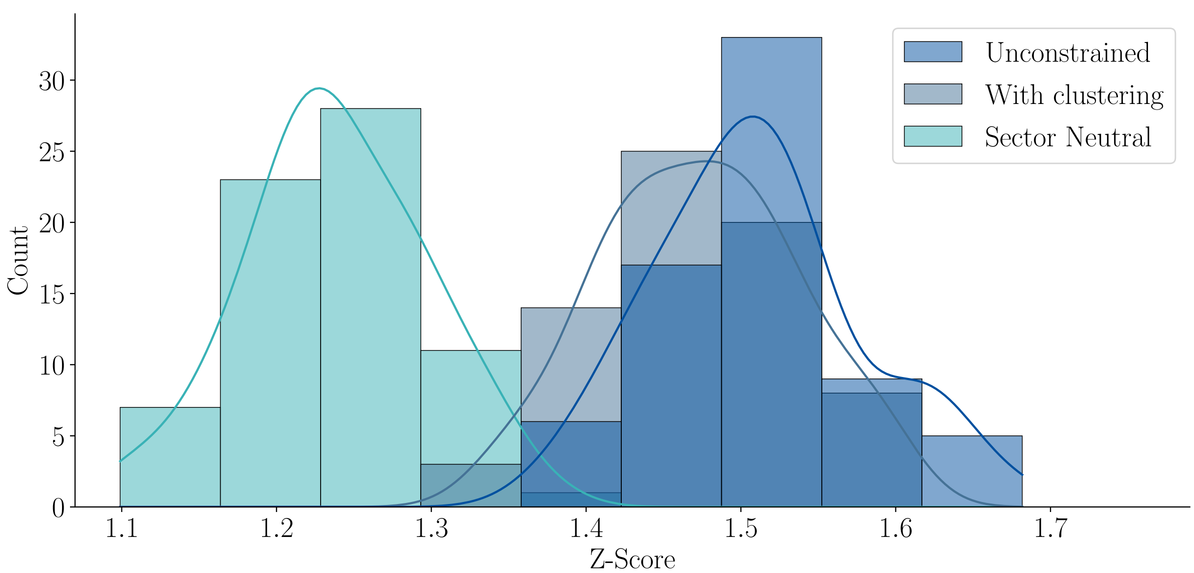
Figure 32: Overlap between quality strategies (2003 - 2020)



Source: MSCI, S&P Compustat and S&P Capital IQ, Authors' calculations

tation does not deteriorate significantly the quality bias. This is an important result because, while the strategy integrating clusters is presented as an alternative solution, halfway between the unconstrained and the sector-neutral approaches, it shows the higher risk-adjusted performance and maintains a high exposure to the quality metrics.

Figure 33: Aggregated quality z-scores of portfolios (2003 - 2020)



Source: MSCI, S&P Compustat and S&P Capital IQ, Authors' calculations

Impact of the cluster approach on turnover Table 14 exhibits the turnover[19] by regions and for the three versions of our quality factor. Unsurprisingly, the unconstrained portfolio displays the lowest level of turnover. The sector-neutral version comes in second position, with an increase in one-way turnover of around 18% for the global portfolio. Imposing sector constraints inevitably spawns costs. The quality strategy which integrates the cluster approach generates the highest turnover, up 40% compared to the global unconstrained version. A logical explanation for this is that the quarterly constitution of clusters introduces variability. It is however possible to consider different solutions to reduce this level of turnover, without affecting the overall philosophy of our investment strategy. First, as part of this theoretical exercise, we have made the very restrictive choice to reflect as closely as possible the movements of our investment universe based on the MSCI Developed Markets index. As a consequence, the clusters are constructed every three months, which constitutes a very high level of processing. We could for example consider reducing the pace of setting up clusters by conditioning it on macroeconomic signals or by implementing new clusters only if they are significantly different from the previously constituted clusters. A second way for reducing turnover could be to implement a hybrid strategy that would mix the unconstrained quality strategy with the strategy that includes the cluster implementation.

Table 14: Turnover of the quality strategies (in %)

	Uncons-trained	**Sector-neutral**	**With clustering**
Global	125.7	148.3	176.1
North America	117.2	137.9	169.8
EMU	138.8	166.0	209.4
Europe ex-EMU	154.3	168.5	212.4
Japan	115.8	153.9	181.4
Pacific ex-Japan	119.7	153.2	220.2

Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

Focus on Eurozone The Eurozone is the major beneficiary of the clusters implementation to our portfolio construction process, with a performance increase of +240 basis points per year and a volatility reduction from 15.7% to 15.0%. To better understand the reasons for such an improvement, we compare in Table 15, the performance and risk statistics of the strategy integrating the clusters in EMU to those obtained in the Section 4 by focusing on the four dimensions that compose our quality factor. In order to have a global view of the region and to make comparisons, we also include information on the sector neutral strategy. First, we note that the cluster implementation results in an increase in annualized performance in all dimensions, ranging from +30 basis points per annum for investment to +100 basis points for profitability. The safety dimension also benefits greatly from the cluster implementation (+80 basis points). To the extent that volatility decreases slightly at the same time (except

[19] We display an annualized one-way turnover.

for investment), the risk-adjusted return measure benefits from this combined effect and is also improved in all dimensions. It is also interesting to note that the sector-neutral portfolio shows better improvements in statistics for each of the dimensions considered individually but that the aggregation of the four dimensions within a composite quality factor is significantly better for the version integrating the clusters.

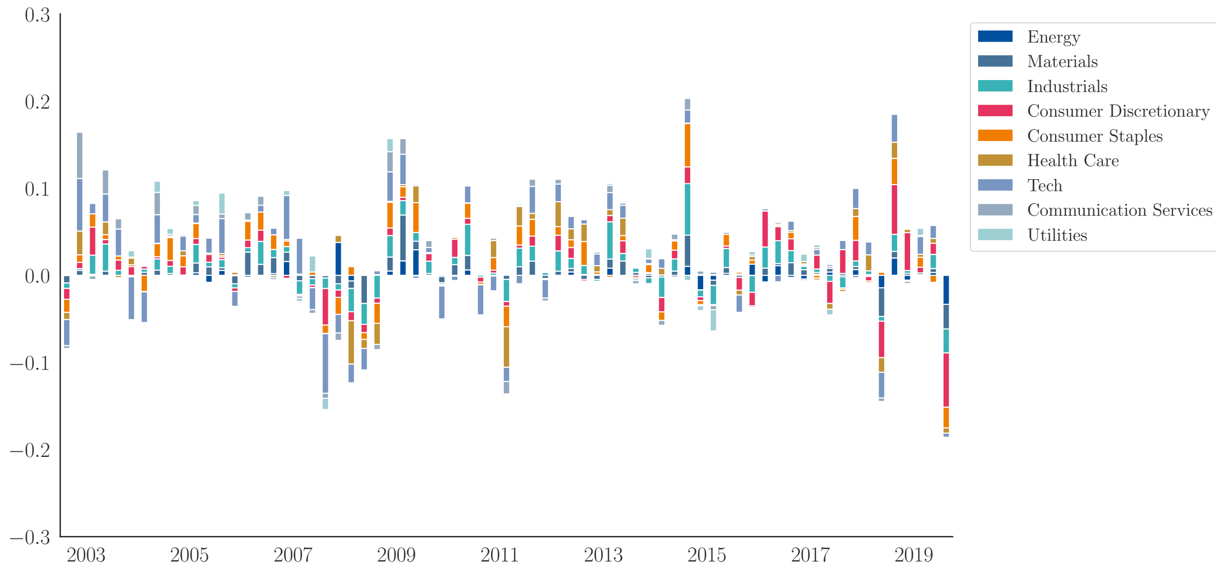
Table 15: Long-only performance of the dimensions in EMU with and without the use of clustering approach in EMU

	EMU	**Profitability**	**Earnings quality**	**Safety**	**Investment**
Unconstrained					
Ann. Return	8.5%	8.3%	7.0%	7.5%	7.5%
Ann. Excess Return	0.7%	0.3%	-0.8%	-0.1%	-0.1%
Ann. Volatility	15.7%	14.1%	15.2%	17.0%	16.7%
Risk-Adj. Return	0.54	0.59	0.46	0.44	0.45
Max. drawdown	-54.2%	-45.1%	-42.5%	-58.3%	-46.7%
With clustering					
Ann. Return	10.9%	9.3%	7.5%	8.3%	7.8%
Ann. Excess Return	2.8%	1.3%	-0.3%	0.7%	0.3%
Ann. Volatility	15.0%	14.0%	15.0%	16.8%	17.0%
Risk-Adj. Return	0.72	0.67	0.50	0.49	0.46
Max. drawdown	-43.1%	-45.9%	-54.2%	-47.2%	-47.9%
Sector-neutral					
Ann. Return	9.7%	10.1%	10.0%	8.7%	8.3%
Ann. Excess Return	1.7%	2.0%	2.1%	0.9%	0.8%
Ann. Volatility	14.6%	13.8%	15.1%	14.9%	16.6%
Risk-Adj. Return	0.66	0.73	0.66	0.59	0.50
Max. drawdown	-45.4%	-43.1%	-41.7%	-52.6%	-52.0%

Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

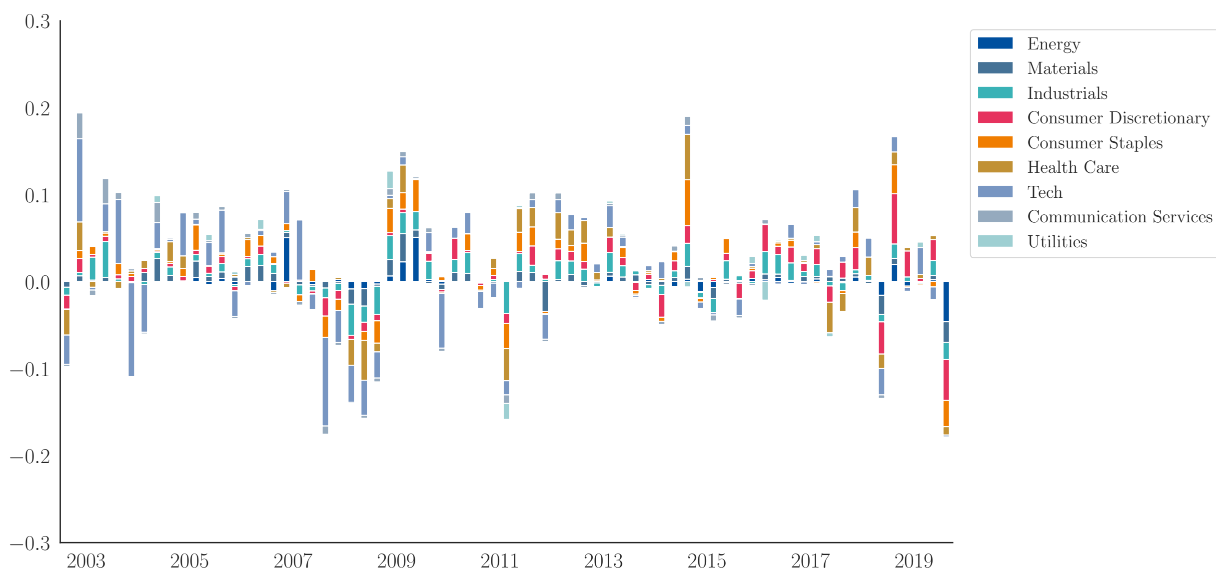
At the same time, we seek to understand the performance drivers at sector level in the EMU area. In the following two charts, we compare the contribution of sectors to the quarterly performance of portfolios. While Figure 34 shows the quality factor using clustering approach, Figure 35 exhibits our initial quality factor. We notice that the version including the clusters shows a slightly higher homogeneity in the contribution of the sectors. Over a given quarter, the performance is much more evenly distributed among the different sectors. Indeed, if we calculate the Herfindahl (Rhoades, 1993) effective number of contributors, the cluster approach gives better results than the unconstrained approach for the 2011 to 2018 period. Consequently, the implementation of clusters seems to better capture the quality component within all sectors, without however switching to a pure sector-neutrality framework, as we have previously indicated.

Figure 34: Sector contribution in the performance of quality strategy integrating clusters in EMU area



Source: MSCI, S&P Capital IQ, Authors' calculations

Figure 35: Sector contribution in the performance of unconstrained quality strategy in EMU area



Source: MSCI, S&P Capital IQ, Authors' calculations

By introducing a clustering approach with a generalist nineteen features per stock, we manage to bring a dynamic framework which is not exclusively fundamental but which includes the hybrid interdependence between fundamentals and other stock features. The clustering technique allows to embed additional information – that would not be exploited by a traditional portfolio construction process – without altering the quality profile of the resulting portfolios. At the same time, it permits to move away from the usual paradigm of relying on the sector neutrality in order to build an equity portfolio and offers an innovative alternative in portfolio construction.

From a practical point of view, the approach is fully implementable as the resulting number of clusters is stable over time and across the different regions covered in the study. Although the strategy including clusters is constrained – in the same way as the sector-neutral strategy – it shows the highest risk-adjusted performance, outperforming the unconstrained factor by 70 basis points per annum. The region most positively impacted by this implementation is the Eurozone. While turnover may be a barrier for some to implement such a strategy, we can propose implementations choices to address the issue by diluting the clustering approach with the original unconstrained factor for example.

6 Conclusion

The quality factor is undoubtedly a complex equity factor. It can be defined in many ways, from the simplest to the most sophisticated. For our part, we have chosen an approach linking the most relevant academic research findings with the daily needs of institutional investors. This results in a multidimensional definition that seeks to capture all aspects of the risk-factor that relates to *profitability, earnings quality, safety* and *investment*.

In the field of factor investing, the ability of a particular metric to predict the cross-section of expected returns is closely related to the investment universe and the study period. By focusing on large- and mid-caps in developed markets and on a contemporary analysis period, we show that not all quality metrics produce the expected results, although they have been highlighted by numerous studies. The choice of metrics therefore proves to be an essential and crucial step in the construction of a quality factor.

Beyond the very academic framework of asset pricing models, we show that the four dimensions that we have selected complement each other. They are weakly correlated with each other and display very different performance profiles. Mixing them within an overall quality factor brings diversification to the factor portfolio meaning that it can better withstand the different market phases. The performance and risk statistics displayed by our quality factor testify to the relevance of the multidimensional approach, both in long-only and long-short portfolio management.

We also introduce a new portfolio construction by implementing a clustering approach to group together companies based on features that are related to traditional equity factors (size, value, momentum, low volatility and quality) rather than on their business sector classification. This paradigm shift shows promising results. First, as expected, the resulting clusters are tilted towards fundamental and market characteristics and each cluster mixes many sectors, so they do not consist in super-sectors. Second, the number of clusters is quite stable over time in all regions, which is essential for real-life implementation. Finally, the cluster implementation results in better quality factor performance without impacting the associated risk measures or the portfolio's quality exposure. While the approach including clusters maintains a sector allocation very close to that obtained within the unconstrained portfolio framework, we manage to capture dynamic variations in the cluster structure. Insofar as quality focuses solely on fundamentals, it is a way to bring the broader market structure to this factor.

In terms of further development, an obvious area is to explore other quality measures to replace those that have demonstrated poor ability to predict cross-sectional differences in average stock returns considering both a large- and mid-caps universe in developed markets and a contemporary study period. We strongly believe that these promising results can be further improved by focusing on more adapted individual metrics. At the same time, the clustering approach sheds new light on portfolio construction by demonstrating that an intermediate path exists between unconstrained and sector neutrality techniques. This framework can be applied to factors other than quality.

References

Affleck-Graves, J. and Miller, R.E. (2003), The Information Content of Calls of Debt: Evidence from Long run Stock Returns, *Journal of Financial Research*, 26(4), pp. 421-447.

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

Arthur, D. and Vassilvitskii, S. (2007), k-means++: The Advantages of Careful Seeding, *SODA '07: Proceedings of he eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 1027-1035.

Asness, C.S., Frazzini, A. and Pedersen, L.H. (2019), Quality Minus Junk, *Review of Accounting Studies*, 24(1), pp. 34-112.

Ball, R., Gerakos, J., Linnainmaa, J. T. and Nikolaev, V. (2016). Accruals, Cash Flows, and Operating Profitability in the Cross Section of Stock Returns, *Journal of Financial Economics*, 121(1), pp. 28-45.

Banz, R. (1981), The Relationship Between Return and Market Value of Common Stocks, *Journal of Financial Economics*, 9(1), pp. 3-18.

Barber, D. (2012), Bayesian Reasoning and Machine Learning, *Cambridge University Press*, pp. 428-429.

Beaver, W.H. (1966), Financial Ratios as Predictors of Failure, *Journal of Accounting Research*, pp. 71-111.

Beaver, W.H. (1968), Alternative Accounting Measures as Predictors of Failure, *Accounting Review*, 43(1), pp. 113-122.

Bhandari, L. (1988), Debt/ Equity Ratio and Expected Stock Returns: Empirical Evidence, *Journal of Finance*, 43, pp. 507-528.

Bholowalia, P., and Kumar, A. (2014), EBK-Means: A Clustering Technique based on Elbow Method and K-means in WSN, *International Journal of Computer Applications (0975 – 8887)*, 105(9).

Caliński, T., and Harabasz, J. (1974), A dendrite method for cluster analysis, *Communications in Statistics-Simulation and Computation*, 3(1), pp. 1-27.

Campbell, J.Y., Hilscher, J. and Szilagyi, J. (2008), In Search of Distress Risk, *Journal of Finance*, 63(6), pp. 2899-2939.

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

Chan, L.K., Karceski, J., Lakonishok, J. and Sougiannis, T. (2008), Balance Sheet Growth and the Predictability of Stock Returns, *University of Illinois at Urbana-Champaign Working Paper.*

Cooper, M.J., Gulen, H. and Schill, M.J. (2008), Asset Growth and the Cross-section of Stock Returns, *Journal of Finance*, 63(4), pp. 1609-1651.

Davies, L.D., and Bouldin, W.D. (1979), A Cluster Separation Measure, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp.224-227.

Dichev, I.D. and Tang, V.W. (2009), Earnings Volatility and Earnings Predictability, *Journal of Accounting and Economics*, 47(1-2), pp. 160-181.

Ding, C., and He, X. (2004), K-Means Clustering Via Principal Component Analysis, *In Proceedings of the Twenty-First International Conference on Machine Learning*, p. 29.

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

Fama, E.F., and French, K.R. (1993), Common risk-factors in the Returns on Stocks and Bonds, *Journal of Financial Economics*, 33(1), pp. 3-56.

Fama, E.F. and French, K.R. (2008), Dissecting Anomalies, *Journal of Finance*, 63(4), pp. 1653-1678.

Fama, E.F., and French, K.R. (2015), A Five-Factor Asset Pricing Model. *Journal of Financial Economics*, 116(1), pp. 1-22.

Fama, E.F., and MacBeth, J.D. (1973), Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81.3, pp. 607-636.

Feng, G., Giglio, S., and Xiu, D. (2020), Taming the Factor Zoo: a Test of New Factors, *Journal of Finance*, 75(3), pp. 1327-1370.

Gomes, J., Kogan, L. and Zhang, L. (2003), Equilibrium Cross Section of Returns, *Journal of Political Economy*, 111(4), pp. 693-732.

Gomes, J.F. and Schmid, L. (2010), Levered Returns, *Journal of Finance*, 65(2), pp. 467-494.

George, T.J. and Hwang, C.Y. (2010), A Resolution of the Distress Risk and Leverage Puzzles in the Cross Section of Stock Returns, *Journal of Financial Economics*, 96(1), pp. 56-79.

Graham, B., and Dodd, D. (1934), *Security Analysis*, First edition, Whittlesey House.

Greenblatt, J. (2006), The Little Book That Beats the Market, John Wiley & Sons.

Greenblatt, J. (2010), The Little Book That Still Beats the Market, 29, John Wiley & Sons.

Healy, P.M. and Wahlen, J.M. (1999). A Review of the Earnings Management Literature and its Implications for Standard Setting, *Accounting Horizons*, 13(4), pp. 365-383.

Hirshleifer, D., Hou, K., Teoh, S. H. and Zhang, Y. (2004), Do Investors Overvalue Firms with Bloated Balance Sheets?, *Journal of Accounting and Economics*, 38, pp. 297-331.

Hou, K., Xue, C. and Zhang, L. (2015), Digesting Anomalies: An Investment Approach, *Review of Financial Studies*, 28(3), pp. 650-705.

Hsu, J., Kalesnik, V. and Kose, E. (2019), What is Quality?, *Financial Analysts Journal*, 75(2), pp. 44-61.

Hsu, J.C., Kudoh, H. and Yamada, T. (2013), When Sell-side Analysts Meet High-volatility Stocks: an Alternative Explanation for the Low-volatility Puzzle, *Journal of Investment Management*, 11(2), pp. 28-46.

Ikenberry, D., Lakonishok, J. and Vermaelen, T. (1995), Market Underreaction to Open Market Share Repurchases, *Journal of Financial Economics*, 39(2-3), pp. 181-208.

Li, E.X., Livdan, D. and Zhang, L. (2009), Anomalies, *Review of Financial Studies*, 22(11), pp. 4301-4334.

Lo, A.W. and MacKinlay, A.C. (1990), Data-Snooping Biases in Tests of Financial Asset Pricing Models, *Review of Financial Studies*, 3(3), pp. 431-467.

Lucke, J. and Forster, D. (2019), K-means as a variational EM approximation of Gaussian mixture models, *Pattern Recognition Letters*, 125, pp. 349-356.

Mashruwala, C., Shivaram, R. and Terry, S. (2006), Why is the accrual anomaly not arbitraged away? The role of idiosyncratic risk and transaction costs, *Journal of Accounting and Economics*, 42(1-2), pp. 3-33.

McConnell, J. and Ovtchinnikov, A. (2004), Predictability of Long-term Spinoff Returns, *Journal of Investment Management*, 2(3), pp. 35-44.

McInnes, L. and Healy, J. (2019), UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, *ArXiv preprint ArXiv*, 1802.03426.

McLean, R.D. and Pontiff, J. (2016), Does Academic Research Destroy Stock Return Predictability?, *Journal of Finance*, 71(1), pp. 5-32.

Merton, R.C. (1974), On the Pricing of Corporate Debt: The Risk Structure of Interest Rates, *Journal of Finance*, 29(2), pp. 449-470.

Modigliani, F., Miller, M.H. (1958), Dividend Policy, Growth and the Valuation of Shares, *Journal of Business*, 34, pp. 411-433.

Mossin, J. (1966), Equilibrium in a capital asset market, *Econometrica: Journal of the Econometric Society*, pp. 768-783.

Novy-Marx, R. (2013), The Other Side of Value: The Gross Profitability Premium, *Journal of Financial Economics*, 108(1), pp. 1-28.

NOVY-MARX, R. (2014), Quality Investing, Unpublished Results.

OHLSON, J.A. (1980), Financial Ratios and the Probabilistic Prediction of Bankruptcy, *Journal of Accounting Research*, pp. 109-131.

PEROTTI, P. and WAGENHOFER, A. (2014), Earnings Quality Measures and Excess Returns, *Journal of Business Finance & Accounting*, 41(5-6), pp. 545-571.

PONTIFF, J. and WOODGATE, A. (2008), Share Issuance and Cross-sectional Returns, *Journal of Finance*, 63(2), pp. 921-945.

RAU, P.R. and VERMAELEN, T. (1998), Glamour, Value and the Post-acquisition Performance of Acquiring Firms, *Journal of Financial Economics*, 49(2), pp. 223-253.

RAI, P. and SINGH, S. (2010), A survey of Clustering Techniques, *International Journal of Computer Applications (0975 – 8887)*, 7(12), pp. 1-5.

RHOADES, S.A. (1993), The herfindahl-hirschman index, *Fed. Res. Bull.*, 79, p. 188.

RICHARDSON, S.A. and SLOAN, R.G. (2003), External Financing and Future Stock Returns, *Rodney L. White Center for Financial Research Working Paper*, (03-03).

RICHARDSON, S., TUNA, I. and WYSOCKI, P. (2009), Accounting anomalies and fundamental analysis: A review of recent research advances, *Journal of Accounting and Economics*, 50(2-3), pp. 410-454.

RIZOVA, S., and SAITO, N. (2020), Investment and Expected Stock Returns, *Available at SSRN 3646575.*

RONCALLI, T. (2020), Handbook of Financial and Risk Management, pp. 944-946.

ROUSSEEUW, J.P. (1987), Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics 20*, pp. 53-65.

SCHWARZ, G. (1978), Estimating the dimension of a model, *Annals of Statistics*, 6(2), pp. 461-464.

SLOAN, R.G. (1996), Do Stock Prices Fully Reflect Information in Accruals and Cash Flows About Future Earnings?, *Accounting Review*, pp. 289-315.

SPIESS, D.K. and AFFLECK-GRAVES, J. (1995), Underperformance in Long-run Stock Returns Following Seasoned Equity Offerings, *Journal of Financial Economics*, 38(3), pp. 243-267.

SPIESS, D.K. and AFFLECK-GRAVES, J. (1999), The Long-run Performance of Stock Returns Following Debt Offerings, *Journal of Financial Economics*, 54(1), pp. 45-73.

TITMAN, S., WEI, J. and XIE, F. (2004), Capital Investments and Stock Returns, *Journal of Financial and Quantitative Analysis*, 39, pp. 677-700.

Vassalou, M. and Xing, Y. (2004), Default Risk in Equity Returns. *The journal of finance*, 59(2), 831-868.

Watanabe, A., Xu, Y., Yao, T. and Yu, T. (2013), The Asset Growth Effect: Insights from International Equity Markets, *Journal of Financial Economics*, 108(2), pp. 529-563.

Xing, Y. (2008), Interpreting the Value Effect Through the Q-Theory: An Empirical Investigation, *Review of Financial Studies*, 21(4), pp. 1767-1795.

Yeo, I. and Johnson, R. (2000), A New Family of Power Transformations to Improve Normality or Symmetry, *Biometrika*, 87(4), pp. 954-959.

A Compustat fields

Table 16: Compustat fields equivalence

Abbreviation	Description of the Compustat field	Compustat code
ACO	Quarterly other currents assets	Q39
ACT	Quarterly total current assets	Q40
AP	Quarterly accounts payable	Q46
AT	Quarterly total assets	Q44
CAPX	Quarterly capital expenditures	Q90
CHE	Quarterly cash and short-term investments	Q36
COGS	Quarterly cost of goods sold	Q30
DLC	Quarterly short-term debt	Q45
DLTT	Quarterly long-term debt	Q51
DP	Quarterly depreciation and amortization	Q5
IB	Quarterly income before extraordinary items	Q8
ICAPT	Quarterly invested capital	Q62
INVT	Quarterly total inventories	Q38
IVNCF	Quarterly net cash flow from investing activities	Q111
LCO	Quarterly other currents liabilities	Q48
LCT	Quarterly total current liabilities	Q49
LT	Quarterly total liabilities	Q54
OANCF	Quarterly net cash flow from operating activities	Q108
RECT	Quarterly total receivables	Q37
SALES	Quarterly net sales/turnover	Q2
SEQ	Quarterly stockholders' equity	Q60

Source: S&P Compustat, S&P Capital IQ

B Additional results

B.1 Long-short performance

Table 17: Long-short performance statistics (2003 - 2020)

	Multi-dimensional	Profita-bility	Earnings quality	Safety	Investment
World					
Ann. Return	5.0%	3.8%	3.5%	2.8%	2.2%
Ann. Volatility	6.4%	6.9%	6.0%	7.4%	7.8%
Risk-Adj. Return	0.78	0.55	0.58	0.38	0.28
North America					
Ann. Return	5.7%	4.7%	4.8%	3.4%	1.0%
Ann. Volatility	8.7%	9.0%	7.3%	9.7%	10.7%
Risk-Adj. Return	0.65	0.52	0.66	0.35	0.09
EMU					
Ann. Return	0.1%	1.2%	-2.1%	0.2%	1.2%
Ann. Volatility	10.0%	10.7%	12.0%	12.3%	9.1%
Risk-Adj. Return	0.01	0.11	-0.18	0.02	0.13
Europe ex-EMU					
Ann. Return	4.9%	1.9%	3.0%	0.5%	5.4%
Ann. Volatility	9.9%	10.6%	11.2%	11.0%	11.3%
Risk-Adj. Return	0.50	0.18	0.27	0.04	0.48
Japan					
Ann. Return	4.8%	0.3%	0.9%	4.1%	3.4%
Ann. Volatility	8.8%	10.4%	9.5%	10.8%	8.7%
Risk-Adj. Return	0.54	0.03	0.09	0.38	0.39
Pacific ex-Japan					
Ann. Return	2.7%	5.3%	1.0%	-2.1%	1.7%
Ann. Volatility	12.2%	11.9%	13.0%	10.7%	14.0%
Risk-Adj. Return	0.22	0.44	0.08	-0.20	0.12

Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

Table 18: Long-short performance statistics (2003 - 2007)

	Multi-dimensional	**Profita-bility**	**Earnings quality**	**Safety**	**Investment**
World					
Ann. Return	-3.7%	-6.8%	1.5%	-1.7%	0.4%
Ann. Volatility	6.4%	5.4%	5.5%	7.7%	6.9%
Risk-Adj. Return	-0.58	1.26	0.27	-0.22	0.06
North America					
Ann. Return	-4.2%	-6.1%	2.3%	-1.5%	-0.9%
Ann. Volatility	9.0%	8.0%	7.4%	10.3%	8.7%
Risk-Adj. Return	-0.47	0.76	0.32	-0.15	-0.10
EMU					
Ann. Return	-4.8%	-10.2%	-2.1%	-0.92%	-0.9%
Ann. Volatility	12.7%	10.3%	9.5%	11.5%	11.5%
Risk-Adj. Return	-0.38	-0.99	-0.22	-0.08	-0.08
Europe ex-EMU					
Ann. Return	-1.6%	-7.1%	4.1%	-4.2%	4.3%
Ann. Volatility	10.2%	9.5%	10.8%	8.8%	11.4%
Risk-Adj. Return	0.16	0.75	0.38	-0.48	0.38
Japan					
Ann. Return	-5.9%	-11.2%	-2.8%	-4.0%	3.1%
Ann. Volatility	9.2%	10.4%	9.6%	10.6%	8.2%
Risk-Adj. Return	0.64	-1.08	-0.29	-0.38	0.38
Pacific ex-Japan					
Ann. Return	7.8%	5.9%	2.9%	-0.1%	2.1%
Ann. Volatility	8.5%	13.0%	9.3%	9.4%	12.9%
Risk-Adj. Return	0.92	0.45	0.32	-0.01	0.16

Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

B.2 Multi-linear regression using Carhart pricing model - EMU

Table 19: Alphas and factor loadings of individual dimensions at EMU level

	Profitability			**Earnings quality**			**Safety**			**Investment**		
	GPA	**CFROIC**	**Dim**	**WCAcc**	**AccCF**	**Dim**	**LTDE**	**WCA**	**Dim**	**AG1Y**	**Capex**	**Dim**
Excess return (%)	0.21	-0.04	0.14	0.13	0.09	0.08	-0.17	-0.04	-0.12	-0.15	0.11	0.13
Alpha (%)	0.29**	0.11	0.26**	0.14	0.01	0.06	0.00	-0.30	-0.23	-0.10	-0.16	0.03
	(2.06)	(0.69)	(1.92)	(0.66)	(0.06)	(0.27)	(-0.03)	(-1.41)	(-1.02)	(-0.48)	(-0.89)	(0.18)
Market	-0.032	-0.08*	-0.041	-0.280***	-0.039	-0.135**	0.108**	0.333***	0.236***	-0.041	0.269***	0.074
	(-0.90)	(-1.86)	(-1.16)	(-5.14)	(-0.71)	(-2.23)	(2.04)	(6.09)	(4.04)	(-0.77)	(5.79)	(1.56)
Size	0.068	-0.252***	-0.127**	-0.310***	-0.014	-0.197**	0.214***	0.481***	0.454***	-0.104	0.309***	0.138**
	(1.37)	(-4.21)	(-2.61)	(-4.10)	(-0.18)	(-2.33)	(2.92)	(6.33)	(5.61)	(-1.39)	(4.79)	(2.10)
Value	-0.764***	-0.249***	-0.569***	0.037	-0.209***	-0.103	-0.239***	-0.194***	-0.227***	0.146**	-0.357***	-0.029
	(-17.00)	(-4.60)	(-12.91)	(0.54)	(-3.02)	(-1.35)	(-3.60)	(-2.82)	(-3.09)	(2.15)	(-6.11)	(-0.50)
Momentum	-0.152***	0.025	-0.039	-0.109**	-0.098*	-0.139**	0.101**	0.032	0.099*	0.042	0.064	0.065
	(-4.53)	(0.60)	(-1.18)	(-2.13)	(-1.89)	(-2.43)	(2.05)	(0.62)	(1.81)	(0.83)	(1.46)	(1.47)
Adjusted R^2	0.65	0.34	0.61	0.19	0.05	0.07	0.11	0.29	0.19	0.01	0.25	0.02

t-statistics (in parentheses) are based on White's standard error
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: MSCI, S&P Compustat, S&P Capital IQ, Authors' calculations

B.3 Correlations between dimensions

Table 20: Correlation between dimensions in North America (2003 - 2020)

	Profitability	Earnings quality	Safety	Investment
Profitability	1.00			
Earnings quality	-0.25	1.00		
Safety	0.41	-0.04	1.00	
Investment	-0.13	0.30	0.15	1.00

Source: S&P Compustat, S&P Capital IQ, Authors' calculations

Table 21: Correlation between dimensions in EMU (2003 - 2020)

	Profitability	Earnings quality	Safety	Investment
Profitability	1.00			
Earnings quality	0.11	1.00		
Safety	0.40	-0.06	1.00	
Investment	-0.01	0.18	0.21	1.00

Source: S&P Compustat, S&P Capital IQ, Authors' calculations

Table 22: Correlation between dimensions in Europe ex-EMU (2003 - 2020)

	Profitability	Earnings quality	Safety	Investment
Profitability	1.00			
Earnings quality	-0.16	1.00		
Safety	0.20	-0.11	1.00	
Investment	-0.05	0.28	0.17	1.00

Source: S&P Compustat, S&P Capital IQ, Authors' calculations

Table 23: Correlation between dimensions in Japan (2003 - 2020)

	Profitability	**Earnings quality**	**Safety**	**Investment**
Profitability	1.00			
Earnings quality	0.06	1.00		
Safety	0.41	-0.07	1.00	
Investment	-0.09	0.42	-0.11	1.00

Source: S&P Compustat, S&P Capital IQ, Authors' calculations

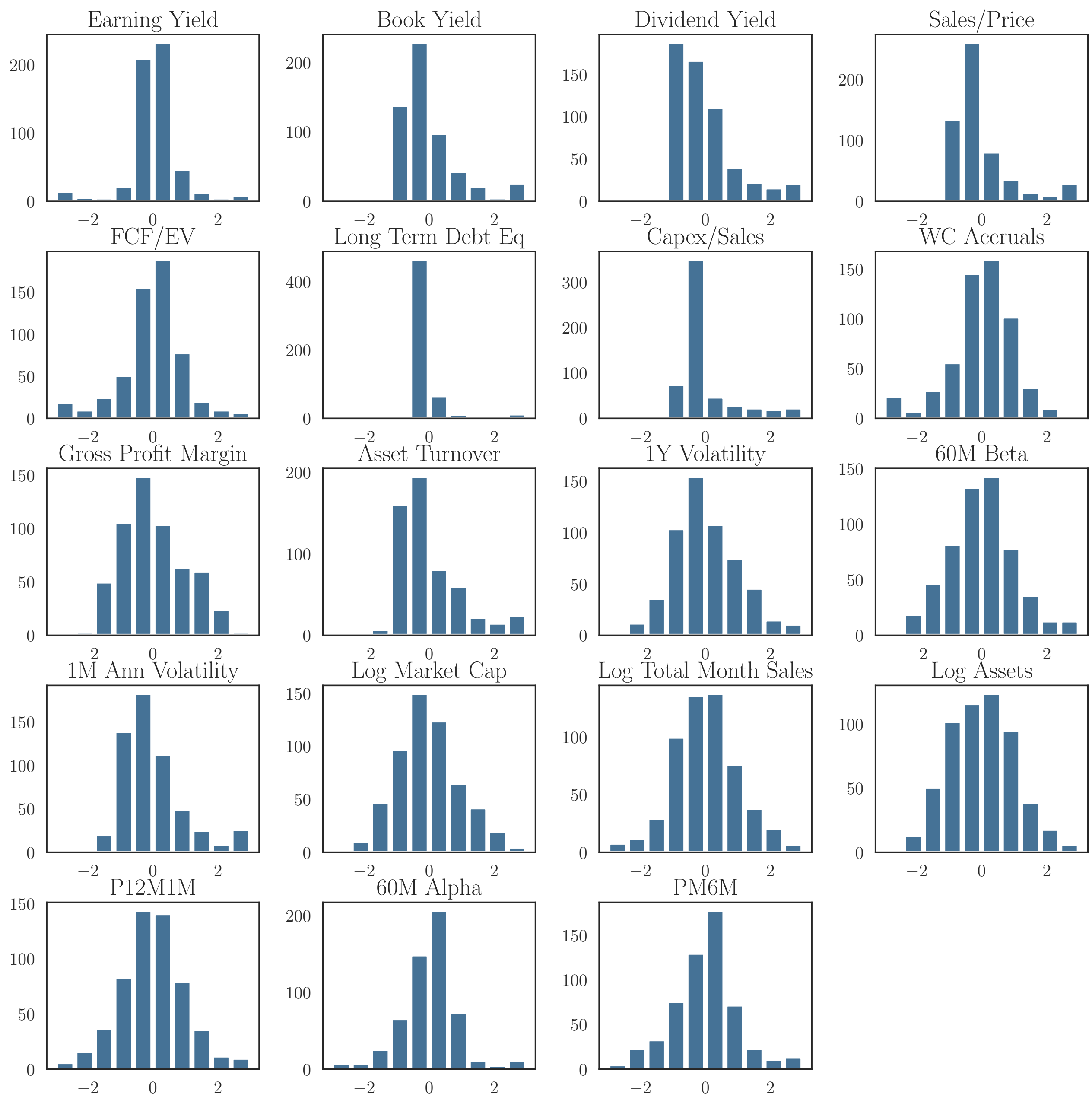
Table 24: Correlation between dimensions in Pacific ex-Japan (2003 - 2020)

	Profitability	**Earnings quality**	**Safety**	**Investment**
Profitability	1.00			
Earnings quality	-0.36	1.00		
Safety	-0.14	0.29	1.00	
Investment	-0.27	0.43	0.11	1.00

Source: S&P Compustat, S&P Capital IQ, Authors' calculations

C Data methodology

Figure 36: Distribution of the 19 features for North America on March 2020 before Yeo-Johnson transformation



Source: MSCI, S&P Compustat, Authors' calculations

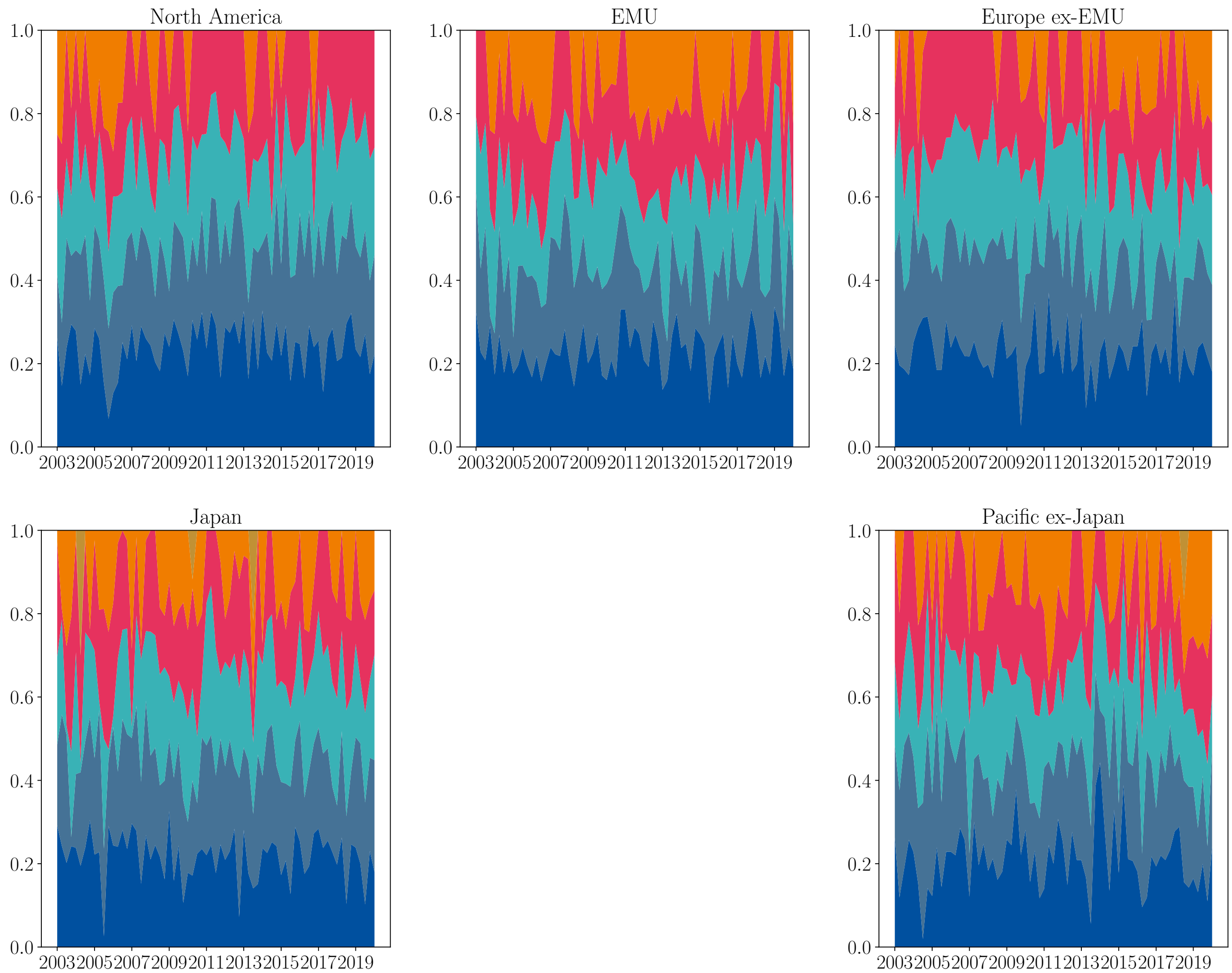
Table 25: Skewness of the 19 features for North America in March 2020

Features	**Skewness**	
	after standardization	**after transformation**
1M Ann. Volatility	1.33	-0.03
1Y Volatility	0.33	-0.04
60M Alpha	-0.07	0.07
60M Beta	0.3	0.02
Asset Turnover	1.41	0.09
Book Yield	1.71	-0.66
Capex/Sales	2.01	0.52
Dividend Yield	1.53	0.22
Earning Yield	-0.70	0.17
FCF/EV	-0.50	0.12
Gross Profit Margin	0.32	-0.03
Log Assets	0.16	0.02
Log Market Cap	0.48	-0.02
Log Total Month Sales	0.14	-0.03
Long-Term Debt/Equity	3.98	0.42
P12M1M	0.08	0.01
PM6M	0.13	0.05
Sales/Price	2.04	0.40
WC Accruals	-0.64	0.07

Source: S&P Compustat, Authors' calculations

D Cluster analysis

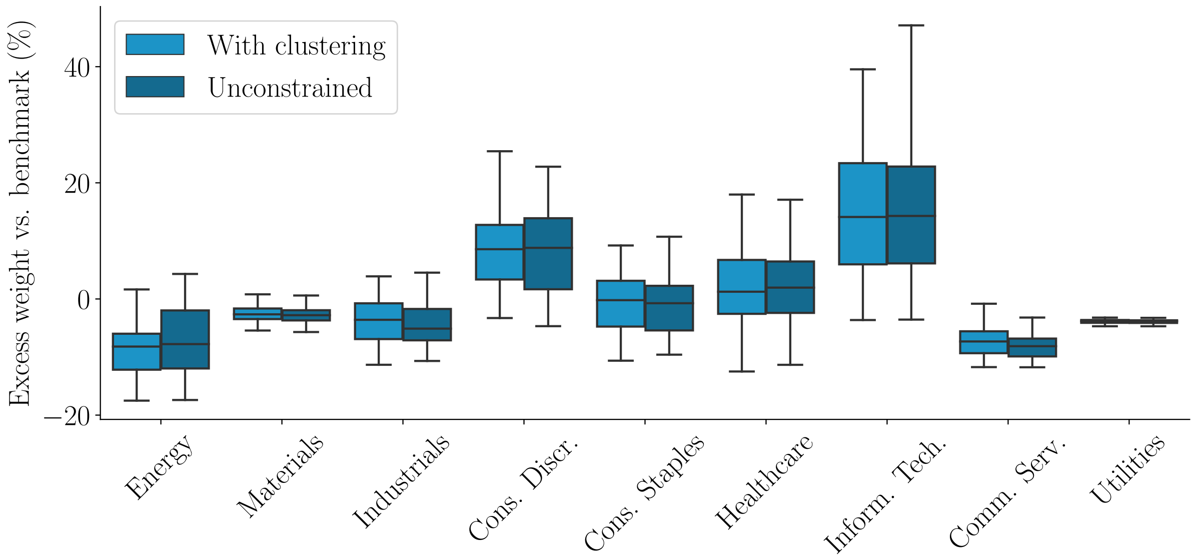
Figure 37: Proportion of stocks per cluster for the different regions



Source: S&P Compustat, S&P Capital IQ, Authors' calculations

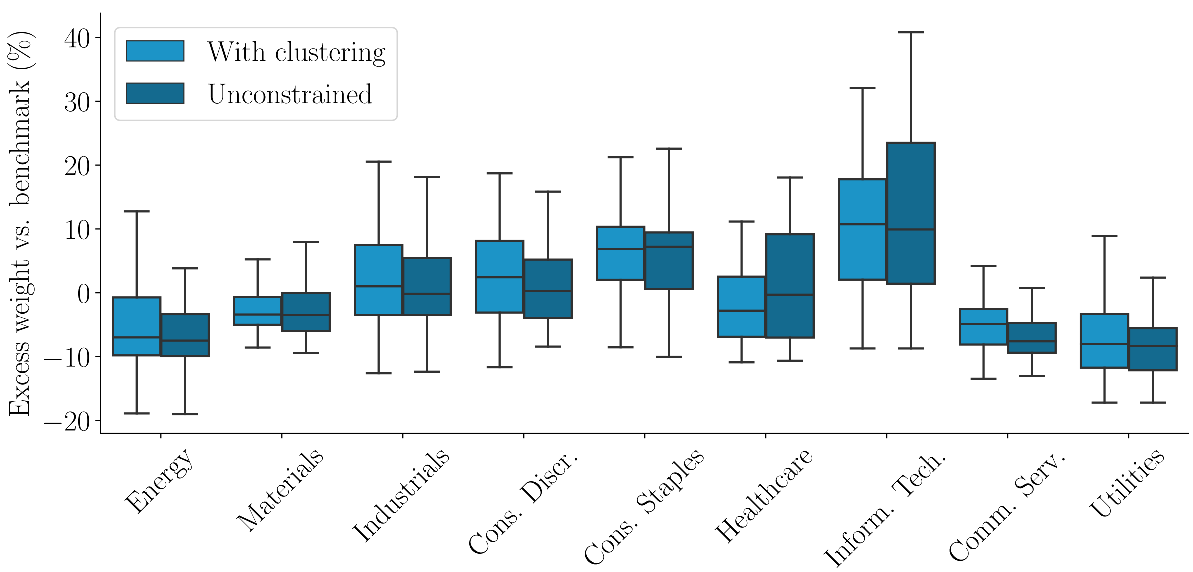
E Sector biases of the quality factor across regions

Figure 38: Sector bets of the quality factor in North America with and without clustering (2003-2020)



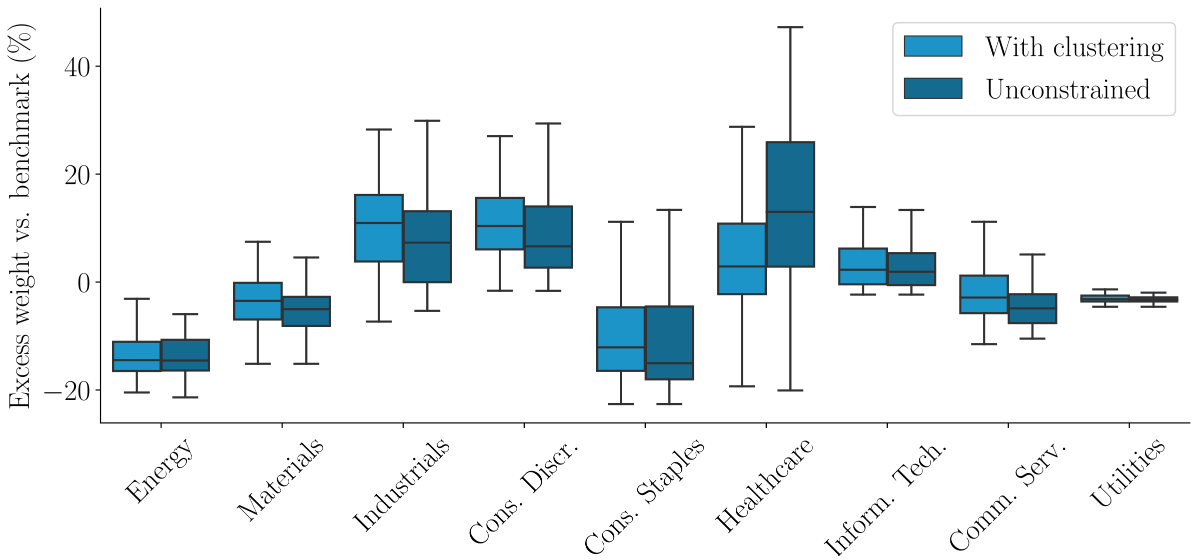
Source: S&P Compustat, S&P Capital IQ, Authors' calculations

Figure 39: Sector bets of the quality factor in EMU with and without clustering (2003-2020)



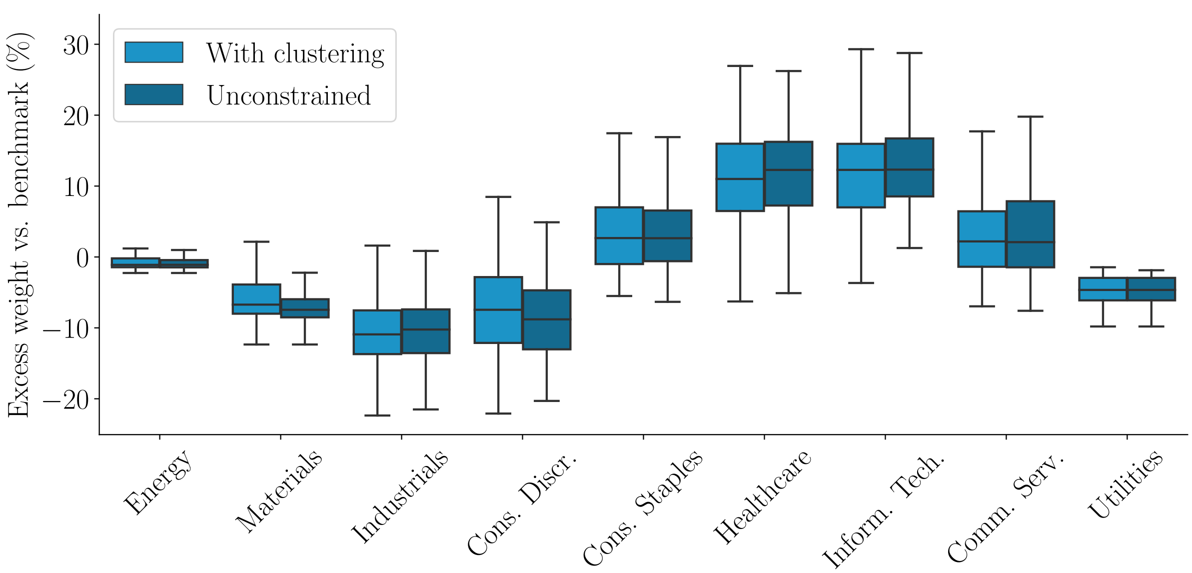
Source: S&P Compustat, S&P Capital IQ, Authors' calculations

Figure 40: Sector bets of the quality factor in Europe ex-EMU with and without clustering (2003-2020)



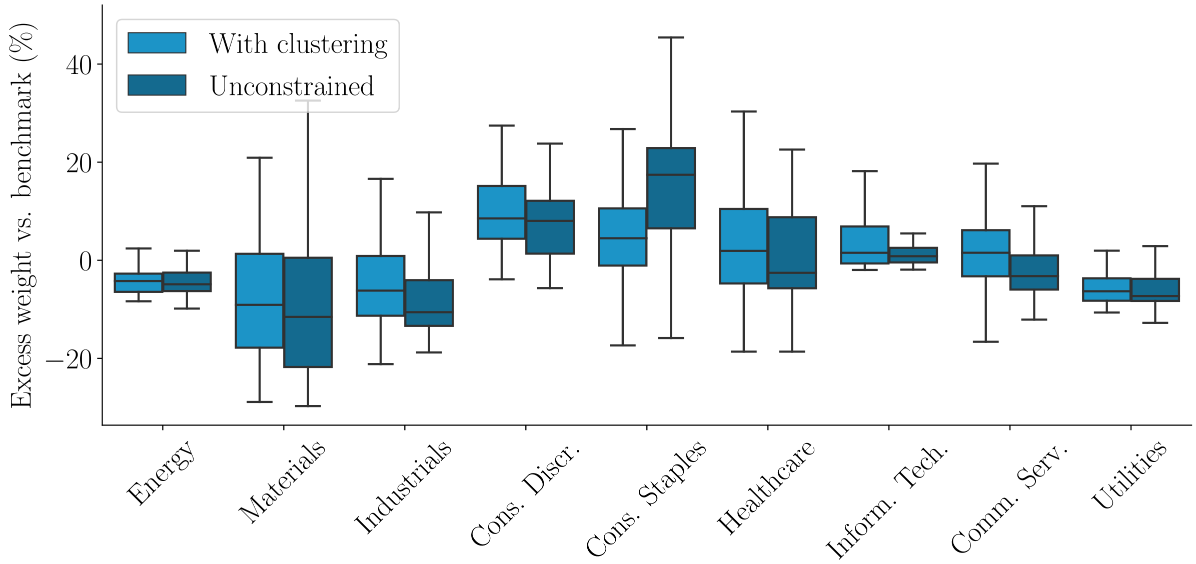
Source: S&P Compustat, S&P Capital IQ, Authors' calculations

Figure 41: Sector bets of the quality factor in Japan with and without clustering (2003-2020)



Source: S&P Compustat, S&P Capital IQ, Authors' calculations

Figure 42: Sector bets of the quality factor in Pacific ex-Japan with and without clustering (2003-2020)



Source: S&P Compustat, S&P Capital IQ, Authors' calculations

Chief Editors

Pascal BLANQUÉ
Chief Investment Officer

Philippe ITHURBIDE
Senior Economic Advisor

Amundi
ASSET MANAGEMENT

Working Paper
June 2021

DISCLAIMER

Document issued by Amundi Asset Management, "société par actions simplifiée"- SAS with a capital of €1,086,262,605 - Portfolio manager regulated by the AMF under number GP04000036 – Head office: 90 boulevard Pasteur – 75015 Paris – France – 437 574 452 RCS Paris - www.amundi.com

Photo credit: iStock by Getty Images - monsitj/Sam Edwards

Find out more about
Amundi publications

research-center.amundi.com