An analysis of investing in U.S. equities with the application of quantitative factor portfolios

Sergei Lazarevich

Bachelor's thesis
October 2019
Technology and Finance
Degree Programme in International Business

Jyväskylän ammattikorkeakoulu
JAMK University of Applied Sciences

Description

Author(s) Lazarevich, Sergei	Type of publication Bachelor's thesis	Date October 2019
		Language of publication: English
	Number of pages 100	Permission for web publication: x

Title of publication **An analysis of investing in U.S. equities with the application of quantitative factor portfolios**
Degree programme Degree programme in International Business
Supervisor(s) Hundal, Shabnamjit
Assigned by JAMK Centre for Competitiveness

This study aimed to examine the potential of applying factors of the modern asset pricing models to automated long-term portfolio management in the U.S. context. Specifically, the factors of Fama and French's Three-factor model and Five-factor models were used. The analysis was performed on a period from 2003 to 2018 based on firm-specific secondary data. The primary goal was to discover whether the factor models could be used by a retail investor to implement portfolios that could outperform the market on a risk-adjusted basis.

The secondary data included corporate fundamentals provided by Morningstar and pricing data sourced from the Quantopian database. The usage of the Quantopian software enabled performing simulation and sensitivity analyses that provided a series of descriptive statistics about the robustness of the factor-based strategies and strength of the factors' predictive qualities. Additionally, the cost simulation analysis revealed the impacts of the portfolio size and the invested capital on the performance of the strategies. Those methods served to test the asserted hypotheses and to answer the research questions.

The empirical findings suggested that the portfolios based on a combination of factors tended to outperform single-factor portfolios on a risk-adjusted basis. In their turn, the single-factor portfolios achieved a higher risk-adjusted return than the S&P 500, RSP (equal-weight S&P 500) and Russell 3000. The analysis also showed significant variability in sensitivity to factors between the sectors of the U.S. economy. Likewise, stocks in different sectors demonstrated diverse factor sensitivity patterns during the three sub-periods: pre-crisis, crisis, and post-crisis. Furthermore, the results revealed that - given sufficient capital - it should be possible for a retail investor to outperform the market using the factor portfolios.

Keywords/tags (subjects)
CAPM, Factor asset pricing model, Beta, Fama and French, factor investing, quantitative portfolio, momentum, value, profitability, smart beta.

Miscellaneous (Confidential information)

Contents

Figures

Tables

1 Introduction

Managing personal finances has become an increasingly relevant problem over the years. Many people seek to make financial investments for the preservation and accumulation of their wealth over time.

The primary goal of this study was to examine the potential of applying factors of modern asset pricing models to automated long-term portfolio management. Specifically, this work attempts to evaluate the effectiveness of systematic investing in portfolios based on the quantitative selection rules derived from the factors of popular asset pricing models. This work was largely based upon the efforts of Eugene Fama and Kenneth R. French (1993, 2015), namely their Three-Factor Model (TFM) and Five-Factor Model (FFM) as well as on works of other scholars contributing to the field of factor-based asset pricing models. These include Mark M. Carhart (1997), Titman, Wei, and Xie (2004), and Novy-Marx (2013). One of the major concerns in this work was to deliberately simulate the impact of the market frictions in the historical simulations of the strategies. Such frictions could include but are not limited to the transaction costs, the commissions and fees charged by the broker, bid-ask spread or the difference between the prices for the buyers and sellers, order execution latency and other factors. In order to arrive at somewhat realistic expectations about the performance of the strategies in the retail investor setting, it is crucial to take such effects into account.

Among the works on the topic, it could be appropriate to highlight Wouter J. Keller, Adam Butler, and Ilya Kipnis (2015) who employed a practically oriented approach in order to test strategies based on the combination of the return persistence phenomenon described by Jegadeesh and Titman (1993) and Mean-Variance Optimization (MVO) as introduced by Markowitz (1952). While the work produced optimistic findings for one of the proposed strategies, it did not specifically suggest any capital requirements and lacked the consideration of transaction costs. The research was based on the work of Jegadeesh and Titman (1993), specifically a tendency of stocks with relatively high historic return or “winners” to keep delivering a high relative return over a certain time with the opposite logic for the underperforming, so-called “losers” stocks. This effect was the given name of “momentum” and was frequently researched by academics ever since (e.g. Louis K.C. Chan, Narasimhan Jegadeesh, and Josef Lakonishok (1996), Daniel Kent, David Hirshleifer, and Avanidhar Subrahmanyam (1998), Bruce D. Grundy, J. Spencer Martin (2001), etc.). According to Jegadeesh and Titman (1993), the premium associated with momentum can be extracted by taking long positions in “winners” and short ones in “losers”, the so-called “up minus down” portfolio or UMD for short.

David Blitz and Pim Van Vliet (2008) were the other contributors to the field of factor-based investment strategies. In their work, they described the strategy based on the “value” and “momentum” factors and applied it globally across different asset classes. They considered the impacts of transaction costs, but their analysis did not propose capital requirements, and the paper did not include the source code to their strategy that would allow a retail investor to replicate the approach. The “value” factor used by David Blitz and Pim Van Vliet (2008) was based on the effort of Eugene Fama and Kenneth R. French (1993) where the value of a company was measured as a ratio of a firm’s book value to the market value of the equity. In their work, Fama and French demonstrated that the returns of the stocks trading at a higher book-to-market ratio on average would be higher than the ones with a lower ratio. Thus, according to their evidence, an investor should be able to isolate the value premium for holding a portfolio with the long positions in high-value stocks and short ones in low-value stocks, the so-called “high minus low” portfolio or HML for short.

As observed during the review of the literature, there were no formal studies that would address the perspective of a retail investor on the problem of investing in the U.S. equities using factor-based strategies. Specifically, based on the existing empirical observations, it was neither clear whether diversified factor portfolio strategies could be successfully applied in a personal investment account considering the relatively high transaction costs for a retail investor, nor was it clear what the adequate amount of capital required to break even would be. With the aforementioned in mind, it was decided to conduct a comprehensive analysis of the factors and their potential to be used as a long-term investment vehicle in a retail investor context. The research employed several factors including value (HML) and operating profitability, the so-called "robust minus weak profitability" portfolio or RMW for short, which are the factors that have been demonstrated by Eugene Fama and Kenneth R. French (2015) to have the highest performance premium over the long-term. The "profitability" factor, as defined by the scholars, was measured as the gross profit of a firm divided by its book equity. Similarly to the value (HML) factor, companies with robust profitability were expected to yield higher returns on average than the ones with weak profitability. It was decided to include the momentum (UMD) as a complementary factor as it was shown by Jegadeesh and Titman (1993) to deliver a comparable premium to the ones of the other factors mentioned earlier. Additionally, to enrich the analysis done in this work, it was decided to study the performance of different factor combinations along with the single-factor portfolios. To ensure relevance, the study was conducted from the perspective of a retail investor as the main consumer of the research paper.

1.1 Motivation for the research

The author's main motivation was in learning about the practicality of using established factor models for hands-off long-term investing. One of the main concerns of the study was to determine whether basic factor portfolios would be capable of delivering risk-adjusted returns superior to one of the U.S. market indices, such as the S&P 500. There are a few reasons why one would attempt to implement these strategies as opposed to investing in an index fund following a similar security selection pattern. For instance, such funds could be simply inaccessible in one's particular situation or part of the world due to financial regulations. Similarly, one might prefer keeping full control over the asset selection process over the alternative of holding shares of an index fund. Thus, one might keep the ability of tweaking and "adding new flavours" into the strategy. Some of the main setbacks in the progress towards this goal are the costs known as active investment management costs. These are the costs associated with transaction costs incurred by an active manager. To study the problem efficiently, it was decided to apply modern tools, such as Quantopian and Python used by quantitative investment managers to simulate the realistic outcomes of such investment strategies. Based on the aforesaid, the research could be beneficial to a retail investor looking for an introduction to implementing an automated long-term investment strategy or for a student seeking to familiarise himself or herself with the field of quantitative investment management. Additionally, this topic is of the author's interest since he is interested in pursuing a career in financial data science.

1.2 Research questions

The study examined the following list of questions:

1. Does a portfolio based on multiple factors provide a better risk-adjusted return than a single factor portfolio and a market portfolio?
2. What is the extent of the variation of the sensitivity of the sectors of the U.S. economy in response to each factor?
3. Are factor-based portfolios the expedient alternative to a market portfolio for a retail investor?

To answer the questions above, the author conducted the simulations for a list of algorithmic strategies holding and rebalancing different sets of the U.S. equities for the period from 01.01.2003 to 31.07.2018. Each strategy selected securities based on specific criteria in an attempt to capture the effects of factors described by Fama and French (1993, 2015) and Jegadeesh and Titman (1993). All simulations were conducted with the assistance of Quantopian (www.quantopian.com), a platform for developing and testing algorithmic trading strategies. All of the historical pricing data used in the simulations were sourced from the platform's free database. Fundamental data including data from the financial statements were provided by Morningstar. The results produced by all simulations were carefully examined. The key performance metrics used for analysis and comparison of the strategies were cumulative returns for the period, volatility, Beta, Sharpe ratio and maximum drawdown.

Apart from testing the historical performance of the strategies, the predictive characteristics of each factor were examined. To make it easier to identify the potential patterns in the factors' predictive characteristics during different states of the market, the analysis was done separately for the three sub-periods, pre-crisis (01.01.2003 – 01.01.2007), crisis (01.01.2007 – 01.01.2010) and post-crisis (01.01.2010 – 01.01.2018). The study also addressed the impacts of the portfolio size and the amount of invested capital on the performance of the strategies.

The following results were revealed as the answers to the research questions: first, it was observed that the portfolios based on a combination of factors tended to outperform single-factor portfolios on a risk-adjusted basis. In their turn, single-factor portfolios achieved a higher risk-adjusted return than the S&P 500, RSP (equal-weight S&P 500) and Russell 3000. The analysis also showed significant variability in sensitivity to factors between the sectors of the U.S. economy. Likewise, the sectors demonstrated diverse factor sensitivity patterns during the three sub-periods: pre-crisis, crisis, and post-crisis (see Tables 8 – 16). Further, the results revealed that given sufficient capital, it should be possible for a retail investor to outperform the market using the factor portfolios.

1.3 Structure of the thesis

To introduce the topics addressed in the research, the paper provides the academic background on asset pricing models in the chapter "Theoretical background of asset pricing models." Next, the "Methodology" chapter familiarises the reader with the research design and approach used in testing the hypotheses. The following chapter presents the analysis of the descriptive statistics and gives a graphical illustration of the findings. Lastly, the "Conclusion" chapter summarises the answers to the research questions and outlines the practical implications of the findings. Additionally, it discusses the research limitations and presents recommendations for future studies.

2 Theoretical background of asset pricing models

This chapter defines the conceptual framework of the study. Most concepts introduced in the work are built upon the foundation of the Capital Asset Pricing Model (CAPM) which is discussed in sub-chapter 2.1. In brief, the CAPM states that the return on an asset is a function of the asset's specific risk as well as its exposure to the systematic risk. More recent advancements of CAPM-based models are also discussed within the sub-chapters 2.2 – 2.4. Sub-chapter 2.2 describes the conceptual background of Fama and French's (1993) Three-Factor Model (TFM) and defines the factors proposed by the scholars while explaining how they enhanced the formulation of CAPM. The Three-Factor Model was an attempt to increase the capacity of CAPM to explain the return of the market by adding "value" and "size" factors to the original expression. Sub-chapter 2.4 continues by introducing the Five-Factor Model (FFM), a modification of the Three-Factor Model (TFM) with two additional factors "profitability" and "investment" described by Fama and French (2015) based on the contributions of Novy-Marx (2013) and Sheridan Titman, K.C. John Wei and Feixue Xie (2003). Mark M. Carhart's (1997) Four Factor Model is discussed in brief in sub-chapter 2.3 as it introduced the "momentum" factor that was not used by Fama and French as part of their models, however, was a persistent phenomenon as demonstrated by Narasimhan Jegadeesh and Sheridan Titman (1993). Sub-chapter 2.5 outlines the works on the factor portfolios in different market states. Namely, Kent Daniel's (2013) effort on describing the "momentum crashes" and their predictability.

As this study prioritises the examination of the potential of applying the factors as the building paradigms of a retail investor's portfolio, it was decided to focus on a select number of factors that demonstrated the highest long-term earning potential based on the historical data as per Kenneth R. French's data (mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). For the graphical representation, refer to Figures 2 and 3.

As the following chapter provides the foundational notion of the author's rationale, it is essential for understanding the context of the research.

2.1 Capital asset pricing theory

History denotes the development of the Capital Asset Pricing Model (CAPM) as an independent effort of three scholars William Sharpe (1964), John Lintner (1965) and Jan Mossin (1966). Building upon the foundation of the works on portfolio selection theory (1952) and diversification (1959) proposed by Harry Markowitz, the model introduced a theoretical capital asset valuation framework that could be used for pricing individual securities or a portfolio. The model considers asset's sensitivity to two types of risk, one of which is a diversifiable or company-specific risk and the other is a non-diversifiable risk also known as systematic risk attributed to the fluctuations of the market. Each security, as defined by the model, is naturally expected to have a certain rate of return due to its unique risk sensitivity characteristics.

CAPM is in widespread use in corporate finance for computing the cost of capital and calculating the market risk of an enterprise as well as in portfolio valuation. The formula is given below:

$$R_i = R_f + \beta_i(R_m - R_f)$$

where: R_i — the rate of return of security i predicted by the model;

R_f — the risk-free rate;

β_i — the beta coefficient of security i;

R_m — the return of the market (Denzil Watson, 2007, p. 242).

The predicted rate of return on a capital asset (R_i) implies the percentage income the security is expected to generate. A risk-free rate (R_f) is a theoretical rate of return on a risk-free investment. While no investments are risk-free, bonds issued by the governments of politically and economically stable countries are generally considered to be free from the risk of default. Therefore, the risk-free rate can be approximated by taking the current rate of return or yield on short-dated government bonds (Denzil Watson, 2007, p. 247). Beta coefficient (β_i) denotes the extent to which the security is a subject to unsystematic risk. Specifically, it indicates the relation between the market return and the asset's. Market return (R_m) in CAPM is the return of the entire market. Typically, analysts estimate the equity risk premium for the national equity market of the issues being analysed (but if a global CAPM is being used, a world equity premium is estimated that considers the totality of equity markets) (Jerald E. Pinto, 2010, p. 45). For this purpose, depending on the application an equity index can be used. For instance, Standard & Poor's 500 Index in a case of the U.S. market or OMX Helsinki 25 in the case of Finland. The choice of the index would depend on the goals pursued by the analyst.

It is necessary to mention that CAPM requires several assumptions:

- Investors are rational and want to maximise their utility;
- All information is freely available to investors and, having interpreted it, investors arrive at similar expectations;
- Investors can borrow and lend at the risk-free rate;
- Investors hold diversified portfolios, eliminating all unsystematic risk;
- Capital markets are perfectly competitive. The conditions required for this are: a large number of buyers and sellers; no one participant can influence the market; no taxes and transaction costs; no entry or exit barriers to the market; and securities are divisible;
- Investment occurs over a single, standardised holding period (ibid., p. 242.).

Although these limitations are important to consider when using the model, they are not unacceptable to the point of making CAPM useless. As reasonably noted by William F. Sharpe (1964): "the proper test of a theory is not the realism of its assumptions but the acceptability of its implications."

The CAPM itself might not reflect the focal point of this research specifically as it in its original form was not applied within the study. Nonetheless, as CAPM is the common conceptual framework of the newer theories used in this thesis, understanding the model "as is" is essential for following the thought process of the author.

2.2 The Three-factor asset pricing model

A modification designed to improve on the CAPM approach was proposed by two scholars Eugene Fama and Kenneth French in 1993 in their publication for the Journal of Financial Economics, “Common risk factors in the returns on stocks and bonds.” Fama and French began by observing that two classes of stocks tended to do better than the market as a whole: small market capitalisation stocks and high book-to-market stocks (commonly referred to as value stocks, as opposed to growth stocks) (Eugene F. Fama, 1993). These factors have been given names representing respective long-short diversified portfolios: Small-minus-big (SMB) meaning long positions in small caps and short in big caps, and High-minus-low (HML) or long positions in high book-to-market stocks and short in low book-to-market.

In their paper, Fama and French demonstrated that their model did a better job in explaining the U.S. stock returns as compared to CAPM when the size (SMB), value (HML) and market risk (Beta) combined. Their Three-Factor Model (TFM) had the lowest R^2 of 0.83 for the portfolio in the largest-size and highest-BE/ME quintiles routinely reaching over 0.90 for the period from 1963 to 1991. It is significantly larger than the 0.69 generated by the market alone (CAPM) (Eugene F. Fama, 1993, p. 19). The model is represented by the formula below:

$$R_i = R_f + \beta_i\left(R_m - R_f\right) + b_s * SMB + b_v * HML + \alpha_i$$

Where: R_i — the expected rate of return of the portfolio i;

R_f — the risk-free rate of return;

R_m — the return of the market portfolio;

β_i — the beta coefficient of security i. β_i in this model is comparable to the classical β_i however not equal to it since there are now two additional factors (SMB and HML) that partially absorb it. The factors measure the historic excess returns of small-cap stocks over big caps and of value stocks over growth stocks;

b_s — the coefficient of exposure of the portfolio to size factor;

b_v — the coefficient of exposure of the portfolio to value factor;

SMB — the difference in returns between diversified small cap and big cap stock portfolios. Expressed as $(R_{small} - R_{big})$;

HML — the difference in returns between diversified high book-to-market and low book-to-market stock portfolios. Expressed as $(R_{high} - R_{low})$;

α_i — the portfolio's alpha (abnormal return).

The factors in the model are calculated by regressing the historical excess returns of the portfolio onto the returns on the portfolios constructed from stocks ranked by their market capitalisation and book-to-market ratio. The data containing the premiums on these factor-based portfolios can be found on Kenneth French's website (mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). Importantly, the corresponding exposure coefficients b_s and b_v produced by the regression can take negative as well as positive values.

Size and value premiums isolation

The method of isolating the factor premiums described in "Common risk factors in stock and bond returns" (1993) implied constructing six portfolios. First, the universe has been defined as all CRSP firms incorporated in the U.S. and listed on the NYSE, AMEX, or NASDAQ. Typically, a universe of securities refers to a set of securities that share a common feature. Security universes can be used for different purposes. Institutionally, investment managers typically specify a universe of securities that defines some of the investing parameters for a managed fund. Broadly, investors may choose to allocate different portions of their portfolio based on various security universes with different risk-reward characteristics (Investopedia, 2018). Then, after the universe was defined all stocks were ranked based on their size and split into two groups: "small" and "large". Due to significant tilt towards small caps in the sample used by Fama and French (1993), mainly caused by the stocks on AMEX and NASDAQ (3,616 small caps out of 4,797 in 1991), it was decided to use median NYSE size as a breakpoint. Then, each group was separately ranked by book-to-market and split into three groups: bottom 30%, middle 40%, and the top 30% (this grouping method was referred to as 2x3 sorts). Although the scholars acknowledged the choice of these specific breakpoints to be arbitrary, they saw no reason why the tests could be sensitive to these choices. As the result, six portfolios were created from the intersection of the two size groups and three book-to-market groups: Small/Low book-to-market (S/L), Small/Medium book-to-market (S/M), Small/High book-to-market (S/H), Big/Low book-to-market (B/L), Big/Medium book-to-market (B/M), and Big/High book-to-market (B/H). According to the paper, the size premium (SMB) is the difference in simple average returns of 3 small and 3 big cap portfolios each month. Book-to-market premium (HML), however, was the difference between the simple average of the returns on the two high book-to-market portfolios (S/H and B/H) and the average of the returns on the two low book-to-market portfolios (S/L and B/L) (Eugene F. Fama, 1993, p. 7).

2.3 Momentum and four-factor asset pricing model

After Fama and French publication in 1993 Mark M. Carhart proposed a modification to the model that was designed to improve on the previous findings (1997).

Before describing the four-factor model, however, it is necessary to introduce the early works on the so-called “momentum” phenomenon on which Mark M. Carhart based his work. One of the first efforts on the topic belongs to Bondt and Thaler (1985) who were among the first to introduce a strategy based on past stock returns. In the study, they demonstrated the performance of portfolios consisting of past winners and losers (the stocks with high and low past returns). In their sample, they used NYSE listed stocks between 1926 and 1982. According to the paper, the portfolio of losers has outperformed the market by 19.6% on average while the winner portfolio underperformed the market by 5% on average for 36 months holding period. De Bondt and Thaler attributed the findings to the behavioural bias of investors’ overreaction to unforeseen events. An alternative explanation was given by K. C. Chan in 1988 in which he argued that the estimates of the returns in the strategy proposed by De Bondt and Thaler (1985) were sensitive to the methods used since the risks of losers and winners were not constant. Thus, controlling for the systematic risk using CAPM would significantly decrease the returns generated by the strategy leaving only small abnormal returns. In 1990, Paul Zarowin discussed another explanation of the “overreaction” phenomenon arguing that the major part of it could be explained by the losers being small-cap stocks and winners being large caps. Therefore, any return anomalies could be justified by controlling for the size effect. Later, Narasimhan Jegadeesh and Sheridan Titman (1993) revealed a few interesting findings in their study of relative strength strategies. They examined the mid-term (3 – 12 months) performance of zero-cost portfolios constructed on a basis of the past returns using the sample that included all NYSE and AMEX listed stocks from 1965 to 1989. 16 different variations of the strategy have been used where they altered the period over which the returns on the stocks were calculated, referred to as *J* number of months as well as the holding period or *K*-month (ranging from 3 to 12 months for *J* and *K*). In each case, they used equal weight portfolios rebalanced monthly. Another set of 16 similar strategies were different by having a week gap between the end of the portfolio formation period *J* and the beginning of the holding period *K*. By skipping a week, they avoided some of the bid-ask spread, price pressure, and lagged reaction effects that underlie the evidence documented in Jegadeesh (1990) and Lehmann (1990) (Narasimhan Jegadeesh, 1993, p. 83). Jegadeesh and Titman (1993) concluded that strategies buying past winners and selling past losers generated a significant abnormal return over the period from 1965 to 1989. They argued that the evidence was consistent with the delay of a stock price adjustment to the firm-specific information. The most successful strategy selected stocks based on their returns over the previous 12 months and then held the portfolio for 3 months. This strategy yielded 1.31% per month when there was no time lag between the portfolio formation period and the holding period and 1.49% per month when there was a 1-week lag between the formation period and the holding period (ibid., p. 69).

In 1997 Mark M. Carhart applied momentum in his four-factor model (FFM) which was an alteration of the three-factor (TFM) model proposed by Fama and French (1993). In his model, the momentum factor mimicking portfolio was defined as the equal-weight average of firms with the highest 30% eleven-months returns minus the equal-weight average of firms with the lowest 30% eleven-months returns lagged one month. The portfolios included all NYSE, AMEX, and NASDAQ stocks and were rebalanced monthly (Carhart, 1997, p. 61). Using the factor, Carhart demonstrated monthly excess returns that were larger than those of the other factors in Fama-French TFM for his sample (July 1963 to December 1993). Later, Clifford S. Asness et al. (2013) studied the effects of value and momentum factors in other markets and asset classes. They gathered evidence for a presence of value and momentum factor premiums in the markets including European equity, commodities, fixed income, and currency markets.

2.4 The Five-factor asset pricing model

In 2015 E. Fama and K. French published an updated version of their three-factor model to which they added two new factors. They relied on the evidence of Novy-Marx (2013), and Titman, Wei, and Xie (2004) arguing that the three-factor model ignores much of the variation in the average returns related to profitability (RMW) and investment style (CMA). Despite the two additional factors, the new model had the same approach to defining the factors as in Fama and French (1993). However, in Fama and French (2015), more bucketing frequencies were used including 2x2, 5x5, 2x2x2x2, etc. For instance, 5x5 would split the universe into five equal size groups (SMB) and then each sub-group into five value (HML) or profitability (RMW) groups. 2x2x2x2, on the other hand, was a multilevel sorting that was used to control for other factors while isolating the single factor premium. This allowed for more precise estimates of the factor premiums and was generally an advantage to using just one sorting technique.

It was noticed by E. Fama and K. French that the “value” factor (HML) had present a size dependency with the small stocks demonstrating significantly higher book-to-market premium for their sample (the period from July 1963 to December 2013). A similar but weaker tendency was present for “profitability” (RMW). Nevertheless, one of the work results was that for portfolios formed on “size”, “value”, “profitability”, and “investment”, the five-factor model provided better descriptions of average returns than the three-factor model (Eugene F. Fama, 2014, p. 4). The formula is given below:

$$R_i = R_f + \beta_i(R_m - R_f) + b_s * SMB + b_v * HML + b_p * RMW + b_c * CMA + \alpha_i$$

Where: R_i, R_f, R_m, β_i, b_s, b_v, SMB, HML, and α_i — as in the three-factor model;

b_p — the coefficient of exposure of the portfolio to profitability factor;

b_c — the coefficient of exposure of the portfolio to investment factor;

RMW — the difference in returns between diversified high operating profitability and low operating profitability stock portfolios. Expressed as $(R_{high\_op} - R_{low\_op})$;

CMA — the difference in returns between diversified portfolios of stocks that have conservative investment style and stocks that invest aggressively. Expressed as $(R_{conservative} - R_{aggressive})$.

According to Fama and French (2015), both of their models captured more of the market variation than CAPM with the five-factor model (FFM) capturing the most variation. This meant that it should be technically possible to use the factor premiums to generate an additional premium on top of the market premium. The priority of this thesis was to assess the practical achievability of this implication by simulating the factor-based investment strategies and taking the transaction costs and capital requirements into account.

Profitability and investment factors

In 2012, Robert Novy-Marx published an article in which he claimed profitability measured by gross profits-to-assets to have roughly the same power as book-to-market predicting the cross-section of average returns (Novy-Marx, 2012, p. 1).

As a proxy to Novy-Marx's measure, Fama and French used operating profitability in their work that was slightly different from his in a way that they would also subtract interest expense from operating income. For instance, for portfolios formed in June of year *t*, profitability (measured with accounting data for the fiscal year ending in *t-1*) is annual revenues minus cost of goods sold, interest expense, and selling, general, and administrative expenses, all divided by book equity at the end of fiscal year *t-1* (Eugene F. Fama, 2014, p. 7).

The fifth factor in Fama and French five-factor model is investment style (CMA). The original idea for using it was suggested by Sheridan Titman, K.C. John Wei and Feixue Xie in 2003 in their publication “Capital investments and stock returns”. According to the paper, companies that significantly increase capital investments tend to have negative benchmark-adjusted performance. To measure how conservative a firm is with its capital investments the measure of abnormal capital investment has been proposed which was calculated using the following formula:

$$CI_{t-1} = \frac{CE_{t-1}}{(CE_{t-2} + CE_{t-3} + CE_{t-4})/3} - 1$$

Where: CI_{t-1} — the capital expenditure of a firm scaled by its sales in year t-1; $(CE_{t-2} + CE_{t-3} + CE_{t-4})/3$ – the average capital expenditure of a firm for the three years scaled by the respective annual sales.

Using sales as deflator implicitly assumes that capital expenditure would grow proportionally with the company’s sales. According to the paper, CI could be viewed as a measure of abnormal capital expenditure or an aggressive investment style.

2.5 Factor portfolios and market states

Before studying the potential of factor portfolios as the long-term automated investment vehicles, one may rightfully question how these portfolios would perform in stressed market conditions. One of the notable works on the topic belongs to Kent Daniel (2013). K. Daniel studied the performance of the momentum-based portfolio in different states of the market. Based on his data, he claimed that the momentum crashes that were observed during market downturns were partly forecastable. Kent argued that such crashes occur during periods of panic following market declines where the volatility is high. Using the momentum portfolios, he demonstrated that the momentum crashes such as the one in June 1932 and the one in March 2009 were followed by a significant outperformance of the losers and underperformance of the winners with respect to the market, the opposite to the performance of the portfolios in a bullish market. Kent showed that ex-ante hedging for the momentum crashes allows generating slightly better returns on a long term as compared to the unhedged portfolio. For a graphical representation of this idea see Figure 1.

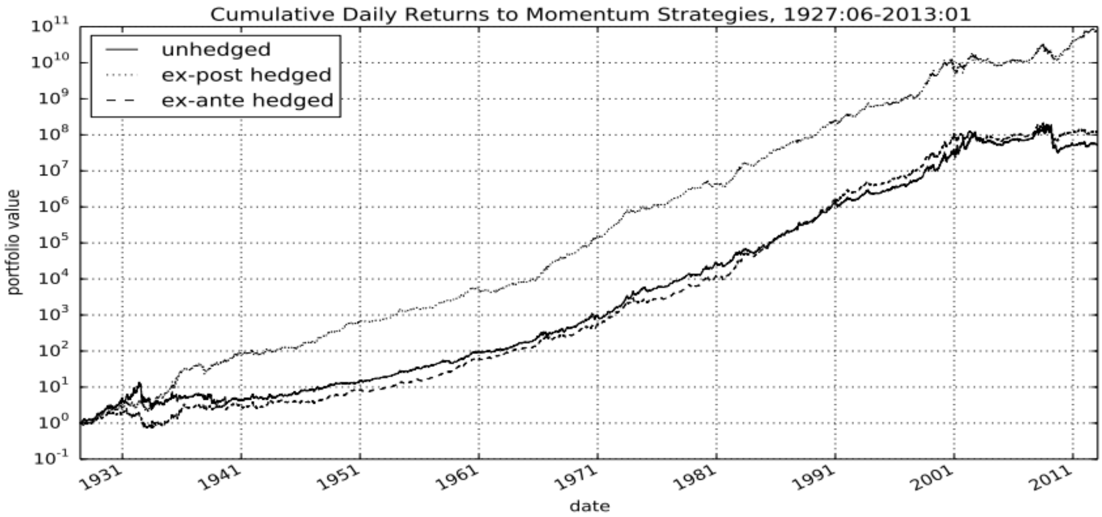


Figure 1 Cumulative daily returns to momentum strategies 1927 – 2013 (Kent Daniel, 2013)

There was not much academic coverage of the reaction of the other specific factors to the market states. However, a limited judgement could be made based on the data on the performance of the factor portfolios provided by Kenneth R. French on his website (mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). Figure 2 illustrates the cumulative performance of the factor portfolios over the period from 1963 to 2018 based on that data. Firms in the low prior return portfolio are below the 30th NYSE percentile. Those in the high portfolio are above the 70th NYSE percentile. Although all factor portfolios have demonstrated positive cumulative returns over the long term, momentum (UMD) though, having the strongest drawdown, had the highest return as compared to the other factors.

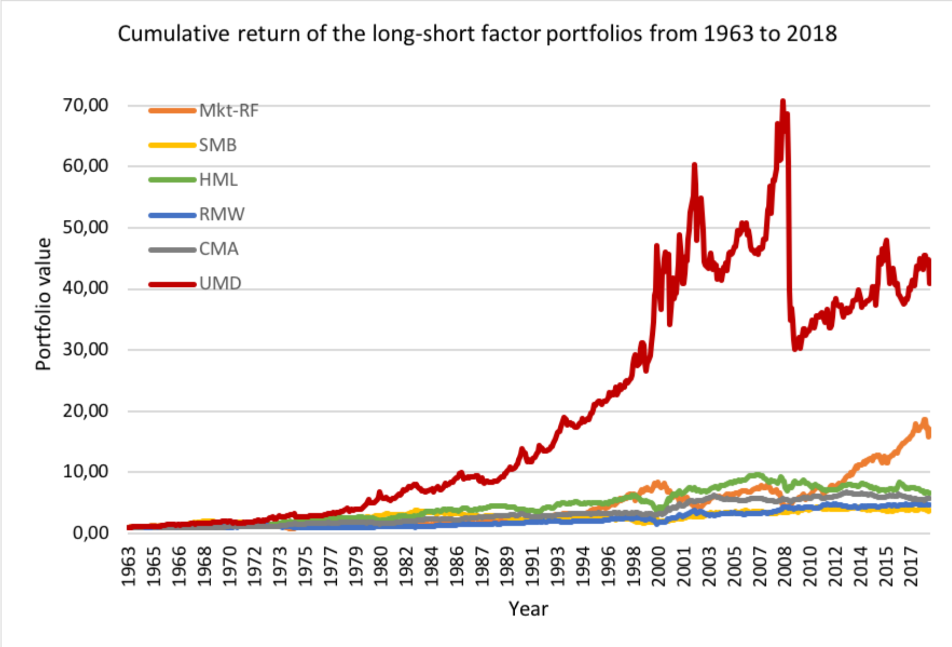


Figure 2 Cumulative return of the 2x3 factor portfolios from 1963 to 2018 (French, 2019)

For a clearer comparison of the other factor portfolios, Figure 3 represents the same graph without momentum (UMD).

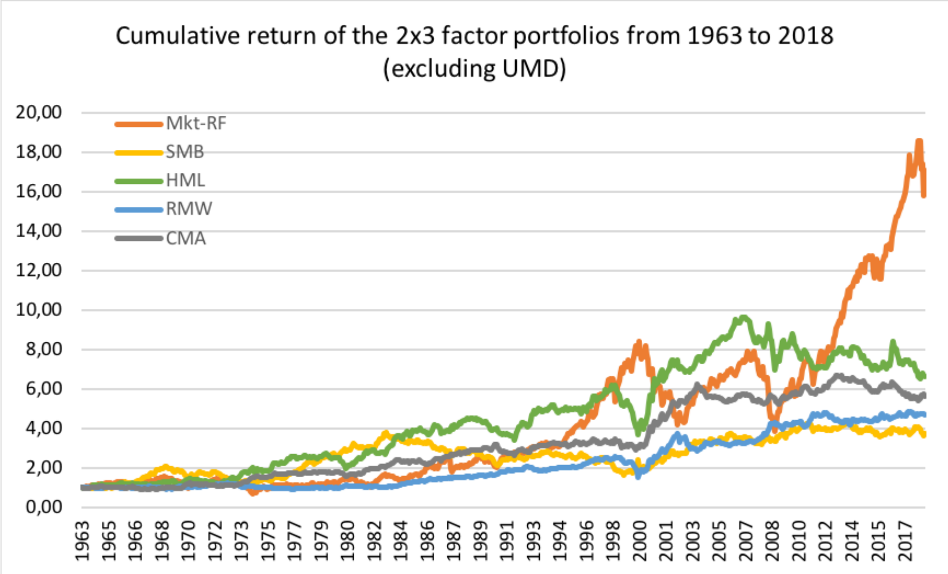


Figure 3 Cumulative return of the 2x3 factor portfolios from 1963 to 2018 (excluding UMD) (ibid.)

2.6 Hypothesis development

Macleod Clark J and Hockey L (1981) define a hypothesis as a statement or explanation that is suggested by knowledge or observation but has not yet been proved or disproved. To classify a hypothesis as a scientific hypothesis, the scientific method requires it to be testable. To test the research hypothesis, one may conduct an experiment aimed at the hypothesis to be either proved or falsified. Falsifiability is the principle that a proposition or theory admits the possibility of being shown false by the authentic data (National Academy of Sciences, 1998). Therefore, this study aims to provide credible and relevant data for hypotheses testing, thus ensuring the acceptable level of objectivity.

Scientists generally base scientific hypotheses on previous observations that cannot satisfactorily be explained with available scientific theories (Paul G. Hewitt, 2013). In the process of investigating the literature the following hypotheses were defined:

H_1: Investing in a portfolio based on multiple factors provides a better risk-adjusted return than a portfolio based on a single factor and a market portfolio.

H_2: Different sectors of the U.S. economy have different sensitivity to factors.

H_3: Factor-based portfolios are less expedient for a retail investor than a market portfolio due to high transaction costs in a retail investor's account.

3 Methodology

In traditional definition, a methodology is a system of principles and approaches to research on which one relies during the gathering and developing of knowledge about the discipline. The methodology could be viewed as the primary research strategy that outlines the means by which research is to be undertaken and, among other things, identifies the methods to be used in it. Research methods described in the methodology define the means or modes of data collection or sometimes how a specific result is to be calculated (Howell, 2013). This chapter is aimed at describing the steps and choices made by the author in the process of the research. The provision of the insight into the decision-making process behind the study should allow the reader to evaluate the critical thinking of the author as well as check the validity of the research design and implications.

3.1 Research design

The definition of the research philosophy is an essential step in the early stage of any scientific research. According to Saunders (2009) research philosophy is an over-arching term related to the development of knowledge and the nature of that knowledge. The philosophy of the research encompasses the scientist's view of reality as well as reflects the nature of the phenomena studied and the goals pursued in the study. Saunders et al. (2009) define positivism, realism, interpretivism, and pragmatism as the main research philosophies. This thesis was rather concerned with studying the measurable facts about observable reality and generalisation of the results that could be replicated in the same manner at any given time. Taking the aforementioned into account, this thesis is consistent with the positivist research tradition.

The original objective of this research was to determine whether factor strategies could be a better alternative to a market portfolio. To answer the research questions, the appropriate type of research has to be adopted. Saunders et al. (2009) define three main research categories - descriptive, explanatory and exploratory. Exploratory research includes relatively unstudied areas or new topics and generates ideas and hypotheses for future research. It evaluates phenomena in a new light and almost exclusively based on a qualitative approach. Looking for an explanation of a situation or problem, explanatory research finds out the answers to *"why"* questions. Descriptive research defines and describes social phenomena (Sarma, 2012, p. 3). In order to better serve the research objective, this study has been designed to provide descriptive and explanatory components.

Saunders and colleagues (2009) state that in deductive research one develops a conceptual framework based on existing theories, which is subsequently tested using data. Besides, an important characteristic of the deduction is that concepts need to be operationalised in a way that enables facts to be measured quantitatively (ibid., p. 125). Since the questions in this research were aimed at studying the objective measurable facts about the factor-based investment strategies, quantitative techniques were employed. These traits are intrinsic to the deductive approach.

3.2 Methods of data collection

Secondary data was studied in this thesis for the purpose of answering the research questions. Unlike in primary data analysis, in an analysis of secondary data a study typically employs the data or information that was gathered by someone else (e.g., researchers, institutions, other NGOs, etc.) for some other purpose than the one currently being considered (Cnossen, 1997).

Most of the secondary data was sourced from the Morningstar database, which includes corporate fundamentals such as book equity of a firm, gross profit, etc. The dataset covered over 8,000 companies traded in the U.S. with over 670 metrics (Quantopian Inc., 2019). Historical returns on stocks were accessed via Quantopian's database. As a universe of securities, all public companies from NYSE, AMEX, or NASDAQ listed between 01.01.2003 and 31.07.2018 were selected. These exchanges should encompass most of the publicly traded companies in the U.S. Since a similar universe was defined by Fama and French (2015), it was deemed reasonable for this work as well. However, for the results to be realistic, it was necessary to exclude several categories of firms from the universe. QTradableStocksUS is one of the default universes provided in the Quantopian API to simplify the exclusion of untradable or illiquid securities. Here are the filters applied to the universe:

- **Market capitalisation over $500M.** This restriction eliminates many undiversifiable risks like low liquidity and difficulty in shorting;
- **Median daily dollar volume of $2.5m or more over the trailing 200 days.** This ensures that stocks in the universe are relatively easy to trade when entering and exiting positions;
- **Prior day's close higher than $5.** In cases where the price is lower, the bid-ask spread becomes larger relative to the price, thus making the transaction cost too high;
- **200 days of price and volume data in place.** If a stock has missing data from the previous 200 days, the company is excluded. This targets stocks with trading halts, IPOs, and other situations that make them harder to access;
- **Primary/Common share.** The QTradableStocksUS chooses a single share class for each company. The criteria are to find the common share with the most dollar volume;
- **ADRs, Limited Partnerships.** QTradableStocksUS excludes ADRs and LPs (Payne, Working On Our Best Universe Yet: QTradableStocksUS, 2017).

The abovementioned limitations were put in place to secure the ability of unobstructed trade preserving the adequate transaction costs as well as risks. For instance, an American depositary receipt (ADR) is a negotiable certificate issued by a U.S. bank representing a specified number of shares in a foreign stock traded on a U.S. exchange (Chen, American Depositary Receipt - ADR, 2018). ADR being dollar-denominated security that trades in the United States but represents a share of a foreign corporation can be a subject to underlying currency risk (Merjan, 2018). Hence, excluding ADRs from the universe may be a reasonably simple measure to decrease the overall portfolio's currency risk which is intrinsic to this type of securities.

This thesis covers the sample from 01.01.2003 to 31.07.2018. The simulations of the strategies were carried out over the whole period. The factor sensitivity analysis (Alphalens analysis), however, was done separately for each economic sector during the three sub-periods: "pre-crisis" (01.01.2003 – 01.01.2007), "crisis" (01.01.2007 – 01.01.2010) and "post-crisis" (01.01.2010 – 31.07.2018). This was done to capture the changes in stocks' sensitivity to factors under conditions of a normal market as well as in the stressed conditions observed during the recession after the subprime mortgage crisis. According to data, major financial markets lost more than 30% of their value during the period (Kosakowski, 2017).

3.3 Definition of key variables

The description of the key variables used in this work is given in Table 1

Table 1 Definition of key variables

Variable	***Description***	***Calculation***	***Source***
Information coefficient (IC)	The IC shows how closely the factor's financial forecasts match actual financial results. The IC can range from 1 to -1, with -1 indicating the forecasts bearing no relation to the actual results, and 1 indicating that the forecasts perfectly matched actual results (Kenton, 2018).	$IC = (2 * pc) - 1$ pc — the proportion of the correct forecasts made by the factor. For example, if there are 100 forecasts made in total and 68 were directionally correct *pc = 0.68.*	Quantopian
Risk-adjusted information coefficient (RAIC)	Information coefficient (IC) adjusted for its standard deviation over the period of calculation.	$RAIC = \frac{IC}{\sigma_{IC}}$	Quantopian
Portfolio return ($\boldsymbol{R_p}$)	The total percentage return of the portfolio from the start to the end of the backtest.	$R_p = \frac{P_t}{P_{t-1}} - 1$ P_t — the dollar value of the portfolio at time t; P_{t-1} — the dollar value of the portfolio at time *t-1.*	Quantopian
Volatility (σ)	The standard deviation of the portfolio's returns. For the purpose of this work, average annual volatility values were used.	$\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$ x — the value of the portfolio at a moment in time; $\bar{x}$ — mean value of the portfolio in the testing period; n — the number of observations.	Quantopian

Sharpe ratio (SR)	A measure of risk-adjusted performance of the portfolio. The portfolio's excess return minus the risk-free rate divided by the portfolio's standard deviation.	$$SR = \frac{R_p - R_f}{\sigma_p}$$ R_p — portfolio return; R_f — the risk-free rate; σ_p — the portfolio standard deviation.	Quantopian
Sortino ratio (Srt)	A modified version of the Sharpe ratio that differentiates the portfolio's harmful volatility (downward deviation) from the overall volatility of portfolio returns as measured by the standard deviation.	$$Srt = \frac{R_p - R_f}{\sigma_d}$$ $R_m \sigma_d$ — the standard deviation of the downside.	Quantopian
Portfolio Beta (β)	A measure of the portfolio's exposure to systematic risk (market risk). The Beta of 1 indicates that the portfolio on average would tend to experience a 1% increase in value for the same increase in the return of the market portfolio and a 1% decrease given the market portfolio lost 1% of its value and vice versa.	$$\beta = \frac{Covariance(R_p, R_m)}{Variance(R_m)}$$ R_m — return on the market portfolio.	Quantopian
Maximum drawdown (MDD)	Maximum drawdown is a measure of the maximum loss from a peak reached by the portfolio before a new peak value is attained.	$$MDD = \frac{P - L}{P}$$ P — peak value before the biggest drop; L — the lowest value of the portfolio before the new peak was attained.	Quantopian
Book-to-market (BM)	A company's book value of the equity (as per balance sheet) divided by its market capitalisation. The ratio is used as a proxy of value (HML) and can be applied to determine undervalued and overvalued firms.	$$BM = \frac{Book\ value}{Market\ cap}$$ $Book\ value$ — total book value of the firm's equity;	Morningstar

		$Market\ cap$ — market capitalisation of the firm.	
Operating profitability ratio (OP)	A measure of the company's profitability defined as operating income minus interest expense divided by the firm's total equity. The ratio is the proxy of the firm's profitability (RMW) that can be used to differentiate highly profitable companies from the least profitable ones.	$$OP = \frac{Operating\ income - Interest\ expense}{Total\ equity}$$	Morningstar
Momentum (UMD)	A ratio of the stock's price at t-1 divided by its price at t-12 where 1 and 12 being the number of months. The returns for the abovementioned period serve as a proxy for momentum where the stocks are classified as winners and losers according to the respective returns.	$$UMD = \frac{P_{t-1}}{P_{t-12}}$$ P_{t-1} — the price of the stock one month ago; P_{t-12}— the price of the stock twelve months ago.	Quantopian
Portfolio turnover (To)	Turnover represents the rate at which assets are being bought and sold within the portfolio. A turnover of 100% would imply that the portfolio positions have all been replaced within the period of time in question. For the purpose of this work, average daily turnover values were used.	$$TO = \frac{1}{T - \tau - 1} \sum_{t=\tau}^{T-1} \sum_{j=1}^{N} (\lvert w_{j,t+1}^{i} - w_{j,t+}^{i} \rvert)$$ $w_{j,t}^{i}$— the portfolio weight in asset *j* chosen at time *t* under strategy *i;* $w_{j,t+}^{i}$— the portfolio weight before rebalancing but at *t+1*; $w_{j,t+1}^{i}$— the desired portfolio weight at time t+1 (after rebalancing); The definition above implies that the turnover is equal to the sum of the absolute value of the rebalancing trades across the *N* available assets and over the $T - \tau - 1$ trading dates normalised by the total number of trading dates (Victor DeMiguel, 2009).	Quantopian

Gross leverage (Lv)	Portfolio gross leverage is a fraction of the total invested funds that are used by the investor (own + borrowed) at a moment in time.	$$Lv = \frac{MV_{longs} + \lvert MV_{shorts} \rvert}{Net\ liquidation\ value}$$ MV_{longs} — the total dollar market value of long positions; $\lvert MV_{longs} \rvert$ — the absolute total dollar market value of short positions; $Net\ liquidation\ value$ — the sum of net portfolio value ($MV_{longs} - MV_{shorts}$) and cash on hand.	Quantopian

3.4 Methods of data analysis

This chapter describes the methods of the data analysis applied in this paper. First, the work introduces a series of tables summarising the key variables computed based on the secondary data as well as characteristics of the tested portfolios. For each factor or a combination of factors, portfolios with different numbers of positions were defined: 100, 200 and 500. Although the choice of these particular numbers was rather arbitrary, the main purpose of using three different values was in capturing the variation in returns and the volatility of the factor strategies with the change of the weights/size of the portfolios in question. Based on empirical tests, it was noticed that having a highly diversified portfolio of 500 positions would typically result in not all orders being executed. This was likely caused by the unavailability of sufficient liquidity due to a low trading volume (simulated by Quantopian’s slippage model) as well as due to the economic sector exposure constraint (must not exceed $\pm$10% to any given sector). 100 and 200, on the other hand, were used to demonstrate the performance achieved by more concentrated factor portfolios with the assumption of 1% being the highest acceptable level of position concentration. Additionally, each portfolio was subject to an economic sector constraint with maximum exposure to each sector between -10% and 10%. This allowed controlling for the bias arising from the tilt to a certain economic sector.

It was decided to keep a similar grouping as in Fama and French (2015) to ensure capturing the size effect. Therefore, every simulation was run separately for the small and the large-cap groups, resulting in 42 portfolios in total including single-factor portfolios and combined factor portfolios.

3.4.1 Asset sorting

For sorting the assets in the universe, each stock was given a rank within the universe based on a relative value associated with the factor being tested. For instance, for simulating the value factor (HML) portfolio, the ranks were assigned based on the book-to-market ratio. Then, long positions were taken in undervalued stocks. A similar approach was applied to profitability (RMW) and momentum (UMD). To automate the process of ranking and sorting, the Quantopian Pipeline API was utilized, which allowed the user to define the factor (also referred to as Alpha-factor) arithmetically, according to which the securities are ranked and sorted. For the sake of the data completeness, all securities with missing data points were dropped from the universe before the simulation.

3.4.2 Simulation and backtesting

To test the first hypothesis, simulations of several factor portfolios were carried out. For this purpose, Quantopian's backtesting environment was used. There are multiple reasons in favour of such a choice. First, Quantopian provided free access to a variety of reliable financial data that included ongoing corporate fundamentals from 2002. Secondly, the software has demonstrated superior performance over similar software solutions when working on large chunks of data, allowing it to execute algorithms in a time-efficient manner. Thirdly, the Quantopian engine automatically deals with the calculations of the trade commissions and slippage, making it possible to account for the transaction costs and other frictions experienced by an investor on a live account. Finally, the platform provides the researcher with versatile output of descriptive statistics on backtests that include graphs and useful variables that allow the researcher to save time while decomposing the results produced by the strategy. The timeframe selected for all tests was between 01.01.2003 and 31.07.2018 since the fundamental data dated back to 2002. An important characteristic of this particular sample was that it included the period of the subprime mortgage crisis. Thus, the sample allowed us to observe the effect of the recession on the performance of the strategies. All backtests executed in this works were accompanied by metrics such as portfolio returns for the period, volatility, Sharpe ratio, Sortino ratio, daily turnover, Beta, maximum drawdown and gross leverage.

Portfolio weights

In this work, three types of equal-weight portfolios were used (100, 200 and 500 positions) as opposed to Fama and French (2015) where capitalisation-weighted indexes had been used. A capitalisation-weighted index is a type of market index with individual components that are weighted according to their total market capitalisation. The larger components carry a higher percentage weighting, while the smaller components in the index have lower weights. This type of index is also known as a market value-weighted index (Chen, 2018). Although hypothetical value-weighted portfolios are useful in explaining market returns, using them as an investment vehicle has a few major setbacks. Foremost, a market portfolio tends to have larger positions in stocks with a high market capitalisation, and hence, asymmetric exposure to economic sectors. For example, in recent years, certain sectors and industries have performed better than others, and that is now reflected in the makeup of the S&P 500. It also means that many sectors will be underrepresented in the index (Lemke, 2018). For instance, purchasing a share of the S&P 500 in November 2018 would imply investing about 20% in technology, 15.8% in healthcare and only 2.6% in materials. Furthermore, empirical evidence suggests that the equal-weighted version of an index tends to outperform its capitalisation-weighted version over time. For example, the Invesco S&P 500 Equal Weight ETF (ticker: RSP) launched in 2003, which is an equal-weighted version of the S&P 500, has consistently outperformed the index throughout its lifetime. Figure 4 is the visual representation of this observation:

Figure 4 SPY vs RSP (Alden, 2018)

This effect goes in line with Fama and French (1993) since the equal-weighted fund should be more exposed to the size effect holding equally large positions in the small-cap equities. It is worth mentioning, however, that value-weighted portfolios typically would have a reduced tracking error with respect to the market (MSCI Inc., 2018). This allows achieving the lower deviation of returns from those of the benchmark (if the benchmark is a value-weighted index), which could be beneficial to a certain type of investors.

Initial capital balance

In order to ensure sufficient liquidity as well as smooth order execution, the initial capital balance was set to $1,000,000 for all simulations. However, since this work is aimed at studying the factor strategies from the perspective of a retail investor, a selection of strategies has also been tested with $100,000, $50,000 and $10,000 as the initial capital (see sub-chapter 4.4).

Rebalancing

The studies like Elton (1993) and French (2008) suggest that higher portfolio turnover, ceteris paribus, leads to a lower overall portfolio performance due to the increase in costs of managing such a portfolio. In factor strategies based on corporate fundamentals, the rebalancing frequency would likely depend on the availability of the new data that typically comes from the financial statements. Other factors such as the appearance of new securities and exchange withdrawals should make less difference when it comes to rebalancing frequency and turnover. Regardless, these issues are taken care of automatically by the testing software. Although the rebalancing was set to monthly for all fundamental strategies, the positions are altered only when a more attractive option comes around or when the current position becomes irrelevant (e.g. security getting withdrawn from the exchange). Therefore, the asset turnover in such strategies is lower than in strategies relying on momentum.

In momentum portfolios, however, the rebalancing frequency might have a stronger impact on the performance. For consistency, it was decided to rely on the findings of Mark M. Carhart (1997) when constructing momentum portfolios which implied keeping the rebalancing monthly.

Benchmarking

According to Investopedia, a benchmark is a standard against which the performance of a security, mutual fund or investment manager can be measured. Generally, broad market and market-segment stock and bond indexes are used for this purpose. When assessing the success of an investment strategy it is essential to consider whether the right benchmark is in place to compare it against.

Typically, investment managers would select the benchmark against which prospective investors will evaluate their performance. Some managers choose a narrow benchmark for evaluation purposes — even though they may invest in a broader group of securities than the benchmark includes (Wespath, 2016).

For the benchmark to be appropriate, it has to satisfy certain criteria. Those include:

- Transparency — the contents and weights of individual investments are known;
- Investability — it has to be possible to invest in the benchmark;
- Measurability — the performance should be possible to calculate on a sufficiently frequent basis;
- Appropriateness — investment style, security types, and components of the benchmark are consistent with the portfolio being measured;
- Specified in advance — the benchmark is selected before the beginning of an evaluation period;
- Public — an investor should be able to verify performance using third-party data (ibid., p. 1).

A benchmark serves two purposes. First, there is a backward-looking function that allows one to compare the performance delivered by an investment strategy against an alternative investment (often a passive management fund). Second, a forward-looking function helps the investor to build expectations with regards to the potential performance of the benchmarked investment strategy based on the benchmark's risk and performance characteristics.

Apart from the qualities described above, it was important for this study that the benchmark had equal exposure to small, mid-sized, and large firms. Another requirement was that the benchmark must be a passive management fund and have a low expense ratio. The expense ratio is the annual fee that all funds or ETFs charge their shareholders. It expresses the percentage of assets deducted each fiscal year for fund expenses, including 12b-1 fees, management fees, administrative fees, operating costs, and all other asset-based costs incurred by the fund (Morningstar, Inc., 2019).

Based on the abovementioned criteria, it was decided to use three funds as benchmarks. First, S&P 500 (ticker: SPY) as being the most liquid and accessible proxy for the U.S. equity market. Second, Russell 3000 (ticker: RUA) as it had exposure to firms with small and medium market capitalisation representing approximately 98% of the investable U.S. equity market (Bloomberg, 2019). Furthermore, since this work studies equal-weighted portfolios and both S&P 500 and Russell 3000 are the value-weighted indices, equal-weighted S&P 500 (ticker: RSP) has been added as the third benchmark.

3.4.3 Forward testing

It is typically believed that in order to provide a measure of protection against overfitting, an investment approach is required to be tested on out-of-sample data (David E. Rapach, 2005). Nonetheless, it was decided to leave forward testing out of scope for this particular work due to time limitations. The tests described in the report, however, could be replicated and extended further with no programming skills as all the source code along with the instructions were published in Appendices 10 – 15. Additionally, source code with instructions is publicly available at a GitHub repository: https://github.com/slazarevich/fama_french_quantopian.

3.4.4 Costs simulation analysis

One of the major concerns in this work was to deliberately simulate the impact of the slippage and commissions to arrive at somewhat realistic approximations about the performance of the strategies built on the foundation of these studies. Therefore, an analysis designed to estimate the impact of the market frictions, such as transaction costs and capital requirements, was undertaken. As mentioned earlier, one of the main strengths of Quantopian lies within its sophisticated slippage model which was designed to account for the impact on a security price that order makes once sent to market. For instance, one’s “buy” order drives prices up, and the “sell” order drives prices down; this is generally referred to as the “price impact” of the trade. The size of the price impact is driven by how large the order is compared to the current trading volume (Quantopian Inc., 2019).

For the tests conducted in this thesis, the default slippage model was applied. This is a fixed slippage of 5 basis points on the price of the order. A buy order for a stock currently selling at $100 per share would fill at $100.05 (100 + (0.0005 * 100)), while a sell order would fill at $99.95 (100 - (0.0005 * 100)).

There is also a volume cap of 10% that limits the proportion of volume one order may take up per bar. For instance, suppose one wants to place an order, and 1000 shares trade in each of the next several minutes, and the volume cap is 10%. If an order is placed for 220 shares then the order will be split into three transactions (100 shares, 100 shares, and 20 shares) (ibid., p.1).

The commission structure used in this work was based on a commission structure offered by a discount brokerage firm Interactive Brokers. Therefore, there were two types of trading commissions — commission per share of $0.005 and a fixed per order fee of $1. Since Quantopian is based on Zipline API, which at the moment of this study supported Interactive Brokers, it was deemed reasonable to use their commission structure as a reference.

In addition, as it appeared during the preliminary testing, the performance of a strategy may vary depending on the amount of capital under management. Therefore, to estimate the capital required to breakeven, several strategies were run with $10,000, $50,000, and $100,000 as initial capital (see sub-chapter 4.4).

Each simulation in this study was intentionally run with maximum leverage equal to 100% of the invested capital. Having this restriction in place ensures that only the capital itself is used for trading. This allows avoiding the potential inaccuracies due to additional costs and risks associated with leverage.

3.4.5 Factor analysis with Alphalens

To test the hypothesis that different sectors of the U.S. economy have different sensitivity to factors, an Alphalens analysis was undertaken. Alphalens is a regression-based tool that is commonly used for measuring the sensitivity of securities' forward returns to a factor defined by the analyst. Thus, it allows studying the predictive characteristics of a factor before simulating an investment strategy based on it. Generally, one would run a regression based on given parameters such as asset universe, arithmetical declaration of the factor (e.g. book-to-market ratio), and a look-ahead window that could be thought as the assumed longevity of the forecast relevance (30, 60, and 90 days in the case of this study). The output produced by Alphalens is information coefficient *(IC)*, which is also a correlation of the factor value of a security at time *t* and the returns on that security at *t + 1* (see Table 1). Information coefficient shows how closely the factor's financial forecasts match actual financial results. The IC can range from 1 to -1, with -1 indicating the forecasts bearing no relation to the actual results, and 1 indicating that the forecasts perfectly matched actual results (Kenton, 2018). The Alphalens analysis (see sub-chapter 4.3) was undertaken for each of the factors included in the portfolio simulation analysis (see sub-chapter 4.1). These included value (HML), profitability (RMW), and momentum (UMD). For convenience, the Alphalens analysis was done separately for the three sub-periods, pre-crisis (01.01.2003 – 01.01.2007), crisis (01.01.2007 – 01.01.2010) and post-crisis (01.01.2010 – 01.01.2018). The results are presented in sub-chapter 4.3 with the relevant source code along with the comments of the author provided in Appendix 15.

3.5 Validity and reliability

According to Greener (2008), research validity refers to the degree to which the research method measures what it was intended to measure. In other words, it determines whether the combination of the data and techniques utilised in the research leads to the evidence of the causal relationship presented in the study. Malcom (2003) defines two types of research validity – internal and external. Internal validity describes the issues concerning research design and implementation. A common example falling under this classification would be an error caused by the choice of instruments (e.g. incompatible versions of tools used in the research). Similarly, flaws in the process of sample definition could lead to the sample being unrepresentative. This and other inconsistencies in the research implementation are referred to as internal validity issues. External validity, however, requires that the outcomes of the research have generalisable implications for the population studied. In a number of cases, the purpose of the research will not be to produce a theory that is generalisable to all populations. Instead, the researcher's task would be to simply try to explain what is going on in his particular research setting (Mark Saunders, 2009).

This study followed a list of strategies to ensure the research validity. First, to study the research questions, a sample between 01.01.2003 and 31.07.2018 was defined. This sample includes all firms incorporated in the U.S. and listed on the three major exchanges (NYSE, AMEX, or NASDAQ) throughout the whole period. The sample intentionally excluded several types of stocks that could be difficult to trade due to the variety of reasons described in sub-chapter 3.2. Figure 5 illustrates the change in the number of firms in the studied universe over time.

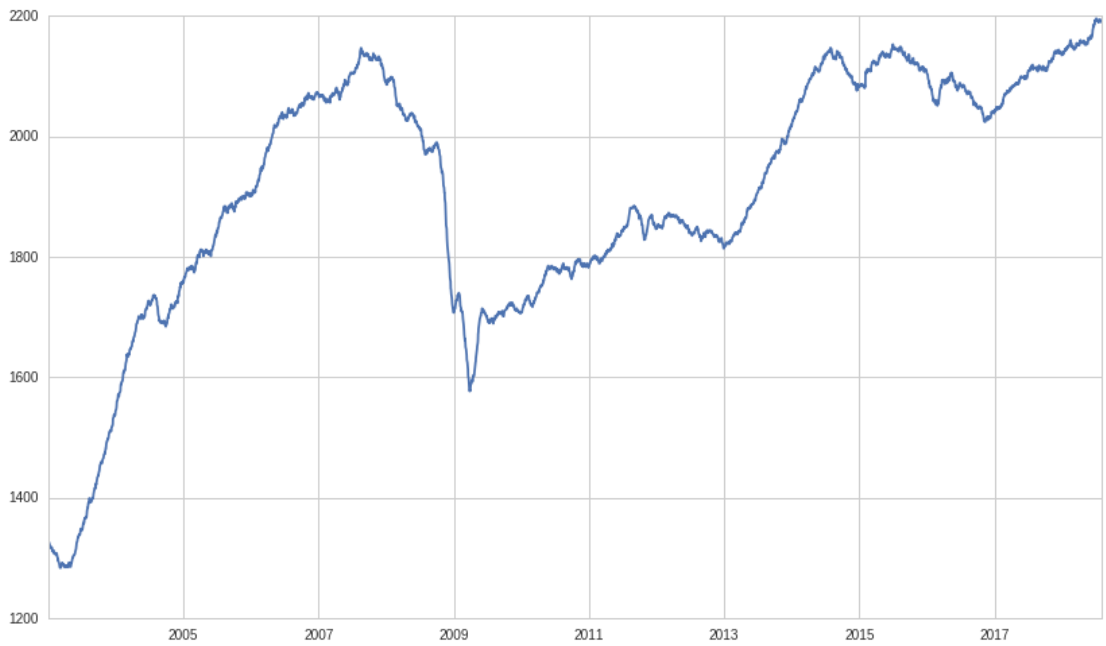


Figure 5 The number of securities in the tradable universe (Payne, 2017)

A downfall in the number of securities during the subprime mortgage crisis (2007 – 2009) could be partly attributed to the large portion of companies going bankrupt during the sub-period. As could be seen from Figure 6 the sample represents various business fields facilitating the study to be consistent and resumptive, thus preventing the results from being discreet to a particular industry or sector.

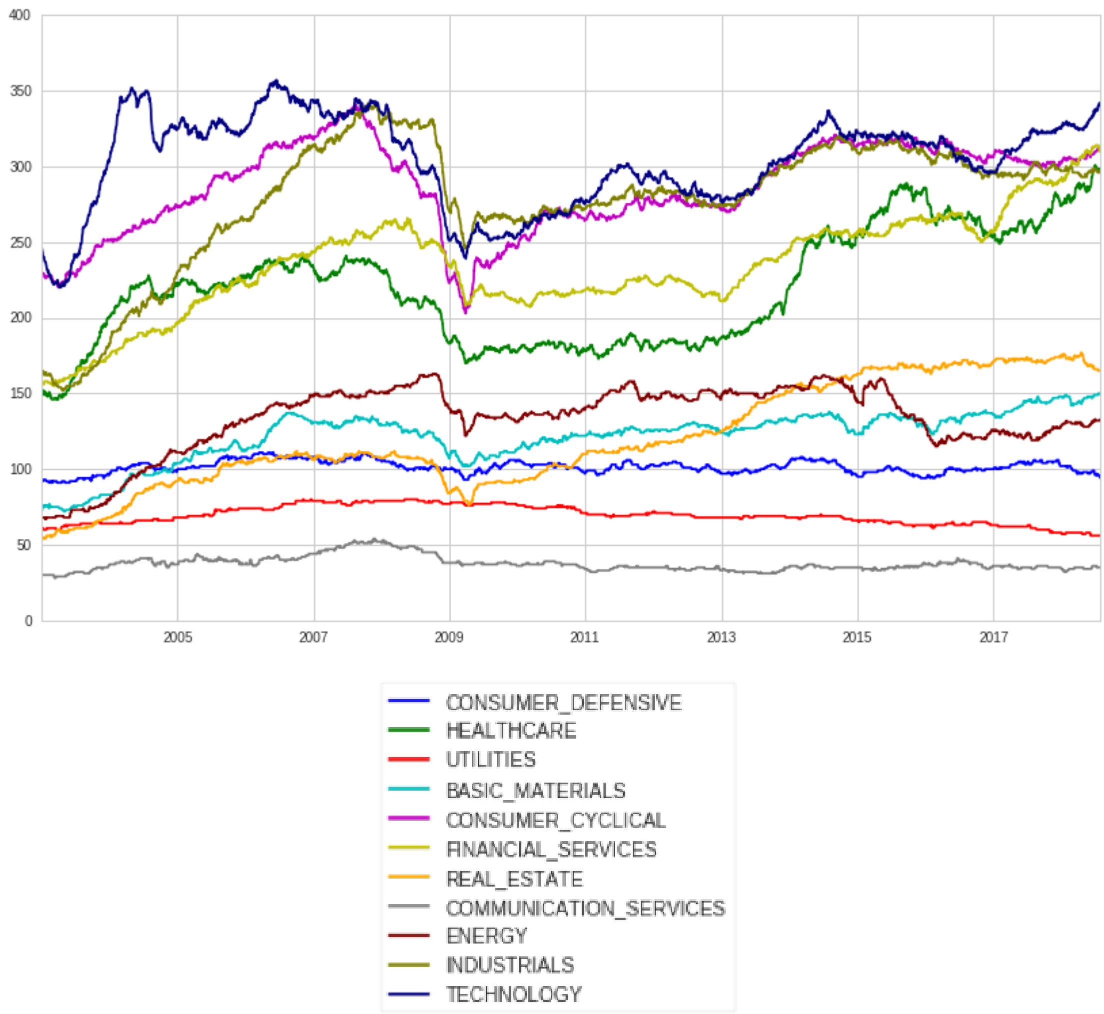


Figure 6 Number of securities in the tradable universe by sector (ibid., p.1)

In order to secure internal validity, the data used in the research was sourced from Morningstar and Quantopian databases which are considered reliable sources. The ways in which the key variables were measured as well as the tools used were kept unchanged over the course of the research.

To secure external validity and to avoid ambiguous notation, the variables operated in the work were treated in accordance with the previous studies on similar topics in the field.

Saunders et al. (2009, p.156) define the reliability as the extent to which the results produced in the research could be replicated with a similar set of data by another observer in a different occasion. This work was designed in a way to encourage the reader to undertake an independent validation by replicating the presented results. The data produced by the tests were taken "as is" and carefully assessed from a critical standpoint. Furthermore, the methods used in the research were based on the methods described by similar studies in the area. Additionally, the process of results generation was reported in reasonable detail while the input data used in the study was taken from sources known for reliability. To maintain the academic structure of the thesis, all necessary references to the works done by other researchers were provided. Hence, the research can be considered to provide an acceptable level of reliability.

4 Empirical findings

This chapter presents the empirical findings produced by the study. The data given here was used as the basis in the preparation of the research conclusions and implications. For understandability, the chapter contains the descriptive statistics in sub-chapter 4.1 as well as graphical analysis in sub-chapter 4.2. The statistics on the factors' sensitivity patterns produced by an Alphalens analysis are presented in the sub-chapter 4.3. For convenience, the Alphalens analysis results were categorised according to the three sub-periods (pre-crisis, crisis, post-crisis).

4.1 Descriptive statistics analysis

Based on the secondary data, several variables were calculated for 42 factor-based portfolios. These variables included cumulative returns, annual returns, maximum drawdown, annual volatility, Sharpe ratio, Sortino ratio, gross leverage, daily turnover, and Beta. The simulations were done over the period between 01.01.2003 and 31.07.2018. The results of the computation are given in tables 2 – 7. For convenient comparison, the portfolios were grouped based on market capitalisation (small and large) and the number of stocks held (500, 200 and 100).

During the testing, it was noticed that concentrated factor-based portfolios tended to provide better risk-return profiles as per Sharpe ratio (Tables 3,4, 6, and 7), while the portfolios of 500 stocks performed relatively worse (Tables 2 and 5). The highest Sharpe ratio (0.73) was demonstrated by the large-cap RMW portfolio of 100 stocks (Table 7). Over the period, the portfolio earned 542.8% with an annual volatility of 18.9% and a maximum drawdown of 54.7%. The highest return among the portfolios was the HML-RMW portfolio of 100 stocks with a cumulative return of 765.4%, annual volatility of 23.8%, Sharpe ratio of 0.7 and maximum drawdown of 61.3%. Such maximum drawdown figures were typical among the tested portfolios due to their exposure to the market during the subprime mortgage crisis. A higher daily turnover rate was generally associated with portfolios of fewer stocks where the momentum (UMD) was used, whereas portfolios based on value (HML) and profitability (RMW) with 500 stocks tended to have a relatively low daily turnover. For all 42 portfolios, the daily turnover averaged at 1.94% with the highest of 3.5% for the HML-RMW-UMD portfolio holding 100 small-cap stocks and the lowest of 0.9% for HML portfolio of 500 large-cap stocks.

Table 2 Performance of small market capitalisation 500 stocks long-only factor portfolios against Russell 3000, RSP and S&P 500

SMALL 500 ($1,000,000)	*RUSSELL 3000*	*RSP (from 30.04.2003)*	*S&P 500*	*HML*	*RMW*	*UMD*	*HML-RMW*	*HML-UMD*	*RMW-UMD*	*HML-RMW-UMD*
Portfolio return (R_p)	312.4%	395.1%	315.6%	289.7%	259.4%	216.7%	438.5%	363.0%	383.3%	391.4%
Annual return (R_{1y})	9.5%	11.1%	9.6%	9.1%	8.6%	7.7%	11.4%	10.4%	10.7%	10.8%
Maximum drawdown (MDD)	-56.4%	-60.1%	-54.9%	-53.9%	-44.0%	-47.7%	-60.1%	-59.1%	-57.6%	-58.7%
Annual volatility (σ)	18.2%	19.6%	18.0%	18.5%	14.7%	16.5%	22.1%	22.0%	21.3%	21.5%
Sharpe ratio (SR)	0.59	0.63	0.60	0.57	0.63	0.53	0.60	0.56	0.58	0.58
Sortino ratio (Srt)	0.83	0.89	0.85	0.80	0.89	0.74	0.85	0.79	0.82	0.82
Gross leverage (Lv)	1.0	1.0	1.0	0.75	0.67	0.72	0.93	0.93	0.93	0.93
Daily turnover (To)	0.1%	0.1%	0.1%	1.4%	1.7%	2.4%	1.5%	1.7%	1.6%	1.7%
Portfolio Beta (β)	1.0	1.0	1.0	0.92	0.74	0.82	1.12	1.12	1.09	1.10

Table 3 Performance of small market capitalisation 200 stocks long-only factor portfolios against Russell 3000, RSP and S&P 500

SMALL 200 ($1,000,000)	*RUSSELL 3000*	*RSP (from 30.04.2003)*	*S&P 500*	*HML*	*RMW*	*UMD*	*HML-RMW*	*HML-UMD*	*RMW-UMD*	*HML-RMW-UMD*
Portfolio return (R_p)	312.4%	395.1%	315.6%	590.7%	487.4%	350.4%	674.8%	456.9%	456.9%	533.3%
Annual return (R_{1y})	9.5%	11.1%	9.6%	13.2%	12.1%	10.2%	14.1%	11.7%	11.7%	12.6%
Maximum drawdown (MDD)	-56.4%	-60.1%	-54.9%	-65.6%	-51.9%	-58.2%	-60.1%	-61.8%	-58.3%	-58.8%
Annual volatility (σ)	18.2%	19.6%	18.0%	25.0%	20.1%	22.4%	23.4%	22.7%	22.1%	22.4%
Sharpe ratio (SR)	0.59	0.63	0.60	0.62	0.67	0.54	0.68	0.60	0.61	0.64
Sortino ratio (Srt)	0.83	0.89	0.85	0.89	0.94	0.76	0.97	0.84	0.86	0.91
Gross leverage (Lv)	1.0	1.0	1.0	0.98	0.90	0.95	1.0	1.0	1.0	1.0
Daily turnover (To)	0.1%	0.1%	0.1%	1.8%	2.1%	2.9%	2.3%	2.7%	2.5%	2.8%
Portfolio Beta (β)	1.0	1.0	1.0	1.23	1.02	1.11	1.18	1.15	1.11	1.13

Table 4 Performance of small market capitalisation 100 stocks long-only factor portfolios against Russell 3000, RSP and S&P 500

SMALL 100 ($1,000,000)	***RUSSELL 3000***	***RSP (from 30.04.2003)***	***S&P 500***	***HML***	***RMW***	***UMD***	***HML-RMW***	***HML-UMD***	***RMW-UMD***	***HML-RMW-UMD***
Portfolio return ($\boldsymbol{R_p}$)	312.4%	395.1%	315.6%	661.5%	675.8%	333.2%	765.4%	461.8%	466.1%	546.7%
Annual return ($\boldsymbol{R_{1y}}$)	9.5%	11.1%	9.6%	13.9%	14.1%	9.9%	14.9%	11.7%	11.8%	12.7%
Maximum drawdown (MDD)	-56.4%	-60.1%	-54.9%	-70.3%	-57.7%	-61.2%	-61.3%	-63.3%	-57.1%	-59.4
Annual volatility (σ)	18.2%	19.6%	18.0%	27.1%	22.6%	24.3%	23.8%	22.8%	22.6%	22.5%
Sharpe ratio (SR)	0.59	0.63	0.60	0.62	0.70	0.51	0.70	0.60	0.61	0.65
Sortino ratio (Srt)	0.83	0.89	0.85	0.88	0.99	0.71	1.0	0.84	0.85	0.91
Gross leverage (Lv)	1.0	1.0	1.0	1.0	0.98	0.99	1.0	1.0	1.0	1.0
Daily turnover (To)	0.1%	0.1%	0.1%	2.2%	2.4%	3.4%	2.9%	3.3%	3.1%	3.5%
Portfolio Beta (β)	1.0	1.0	1.0	1.31	1.14	1.18	1.19	1.14	1.12	1.13

Table 5 Performance of large market capitalisation 500 stocks long-only factor portfolios against Russell 3000, RSP and S&P 500

LARGE 500 ($1,000,000)	***RUSSELL 3000***	***RSP (from 30.04.2003)***	***S&P 500***	***HML***	***RMW***	***UMD***	***HML-RMW***	***HML-UMD***	***RMW-UMD***	***HML-RMW-UMD***
Portfolio return ($\boldsymbol{R_p}$)	312.4%	395.1%	315.6%	230.7%	250.2%	235.9%	403.7%	404.6%	443.4%	430.8%
Annual return ($\boldsymbol{R_{1y}}$)	9.5%	11.1%	9.6%	8.0%	8.4%	8.1%	11.0%	11.0%	11.5%	11.3%
Maximum drawdown (MDD)	-56.4%	-60.1%	-54.9%	-54.6%	-43.2%	-47.5%	-58.3%	-55.9%	-52.8%	-53.8%
Annual volatility (σ)	18.2%	19.6%	18.0%	16.5%	13.4%	15.0%	19.2%	19.2%	18.6%	18.7%
Sharpe ratio (SR)	0.59	0.63	0.60	0.55	0.67	0.59	0.64	0.64	0.68	0.67
Sortino ratio (Srt)	0.83	0.89	0.85	0.76	0.92	0.82	0.89	0.89	0.95	0.93
Gross leverage (Lv)	1.0	1.0	1.0	0.80	0.72	0.77	0.99	0.99	0.99	0.99
Daily turnover (To)	0.1%	0.1%	0.1%	0.9%	1.2%	2.0%	1.2%	1.4%	1.3%	1.4%
Portfolio Beta (β)	1.0	1.0	1.0	0.88	0.72	0.79	1.4	1.03	1.0	1.01

Table 6 Performance of large market capitalisation 500 stocks long-only factor portfolios against Russell 3000, RSP and S&P 500

LARGE 200 ($1,000,000)	***RUSSELL 3000***	***RSP (from 30.04.2003)***	***S&P 500***	***HML***	***RMW***	***UMD***	***HML-RMW***	***HML-UMD***	***RMW-UMD***	***HML-RMW-UMD***
Portfolio return ($\boldsymbol{R_p}$)	312.4%	395.1%	315.6%	347.8%	428.1%	422.9%	438.1%	370.1%	456.0%	434.9%
Annual return ($\boldsymbol{R_{1y}}$)	9.5%	11.1%	9.6%	10.1%	11.3%	11.2%	11.4%	10.5%	11.7%	11.4%
Maximum drawdown (MDD)	-56.4%	-60.1%	-54.9%	-66.9%	-52.3%	-54.0%	-62.0%	-57.9%	-54.8%	-56.0%
Annual volatility (σ)	18.2%	19.6%	18.0%	21.8%	17.3%	19.7%	19.9%	19.3%	18.9%	19.2%
Sharpe ratio (SR)	0.59	0.63	0.60	0.55	0.70	0.64	0.64	0.61	0.68	0.66
Sortino ratio (Srt)	0.83	0.89	0.85	0.77	0.98	0.89	0.90	0.85	0.94	0.91
Gross leverage (Lv)	1.0	1.0	1.0	1.0	0.92	0.97	1.0	1.0	1.0	1.0
Daily turnover (To)	0.1%	0.1%	0.1%	1.3%	1.6%	2.6%	2.1%	2.4%	2.3%	2.7%
Portfolio Beta (β)	1.0	1.0	1.0	1.15	0.93	1.01	1.06	1.03	1.0	1.02

Table 7 Performance of large market capitalisation 100 stocks long-only factor portfolios against Russell 3000, RSP and S&P 500

LARGE 100 ($1,000,000)	***RUSSELL 3000***	***RSP (from 30.04.2003)***	***S&P 500***	***HML***	***RMW***	***UMD***	***HML-RMW***	***HML-UMD***	***RMW-UMD***	***HML-RMW-UMD***
Portfolio return ($\boldsymbol{R_p}$)	312.4%	395.1%	315.6%	342.7%	542.8%	493.3%	497.5%	368.0%	520.0%	408.7%
Annual return ($\boldsymbol{R_{1y}}$)	9.5%	11.1%	9.6%	10.0%	12.7%	12.1%	12.2%	10.4%	12.4%	11.0%
Maximum drawdown (MDD)	-56.4%	-60.1%	-54.9%	-70.0%	-54.7%	-56.2%	-63.4%	-57.3%	-55.1%	-60.5%
Annual volatility (σ)	18.2%	19.6%	18.0%	22.9%	18.9%	21.3%	20.5%	19.5%	19.6%	19.6%
Sharpe ratio (SR)	0.59	0.63	0.60	0.53	0.73	0.65	0.66	0.61	0.70	0.63
Sortino ratio (Srt)	0.83	0.89	0.85	0.74	1.02	0.89	0.93	0.84	0.97	0.87
Gross leverage (Lv)	1.0	1.0	1.0	1.0	0.99	1.0	1.0	1.0	1.0	1.0
Daily turnover (To)	0.1%	0.1%	0.1%	1.6%	1.9%	3.1%	2.6%	3.0%	2.8%	3.3%
Portfolio Beta (β)	1.0	1.0	1.0	1.19	1.01	1.06	1.08	1.03	1.02	1.03

4.2 Graphical analysis

Figure 7 represents the 5 highest performing long-only strategies for the large market capitalisation portfolios according to their Sharpe ratios. 100 RMW portfolio had the highest risk-adjusted return within the large-cap group with a Sharpe ratio of 0.73 and the annual return of 12.7%. The other four portfolios had weaker overall performance yet still outperformed the three benchmarks (Russell 3000, RSP and S&P 500) almost over the whole sample period (01.01.2003 – 31.07.2018). Notably, each of the 5 highest performing strategies had exposure to the profitability factor (RMW).

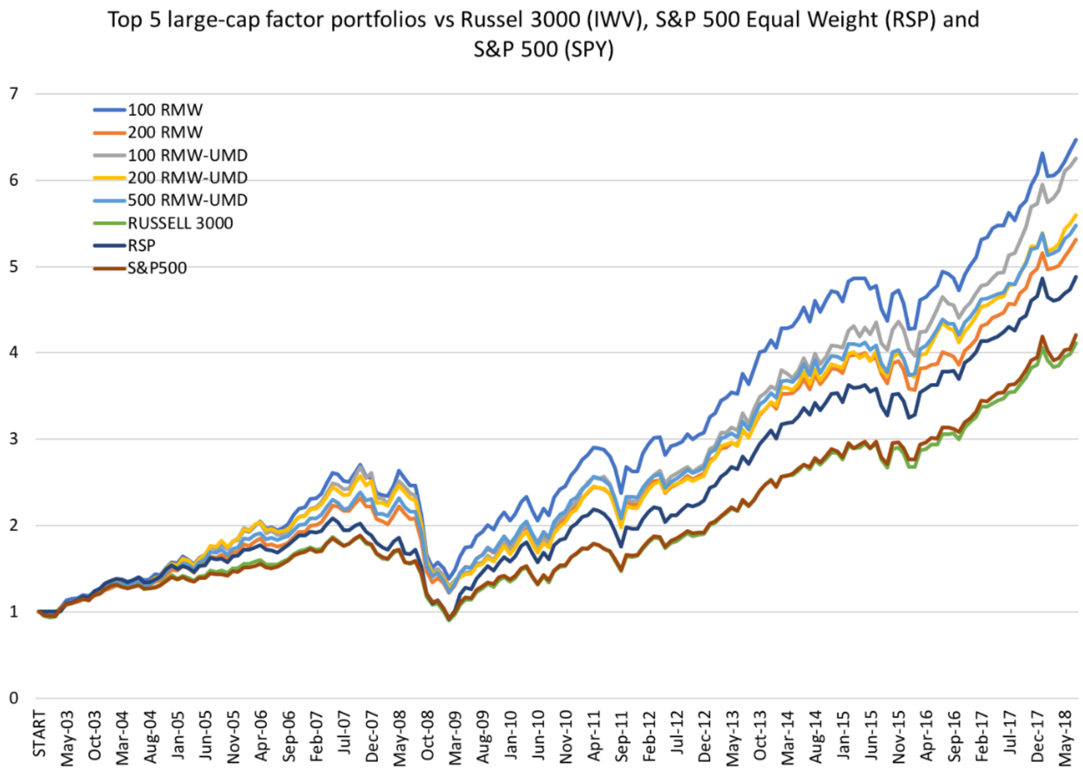


Figure 7 Top 5 large-cap factor portfolios vs Russell 3000 (IWV), S&P 500 (SPY) and S&P 500 Equal Weight (RSP)

Similarly, Figure 8 demonstrates the 5 highest performing small market capitalisation portfolios. Small-cap strategies appeared to have significantly higher volatility as compared to their large-cap counterparts. The leading strategy in the group with a Sharpe ratio of 0.7 and an annual return of 14.9% was a portfolio of 100 stocks selected based on the combination of the value and profitability factors (HML-RMW).

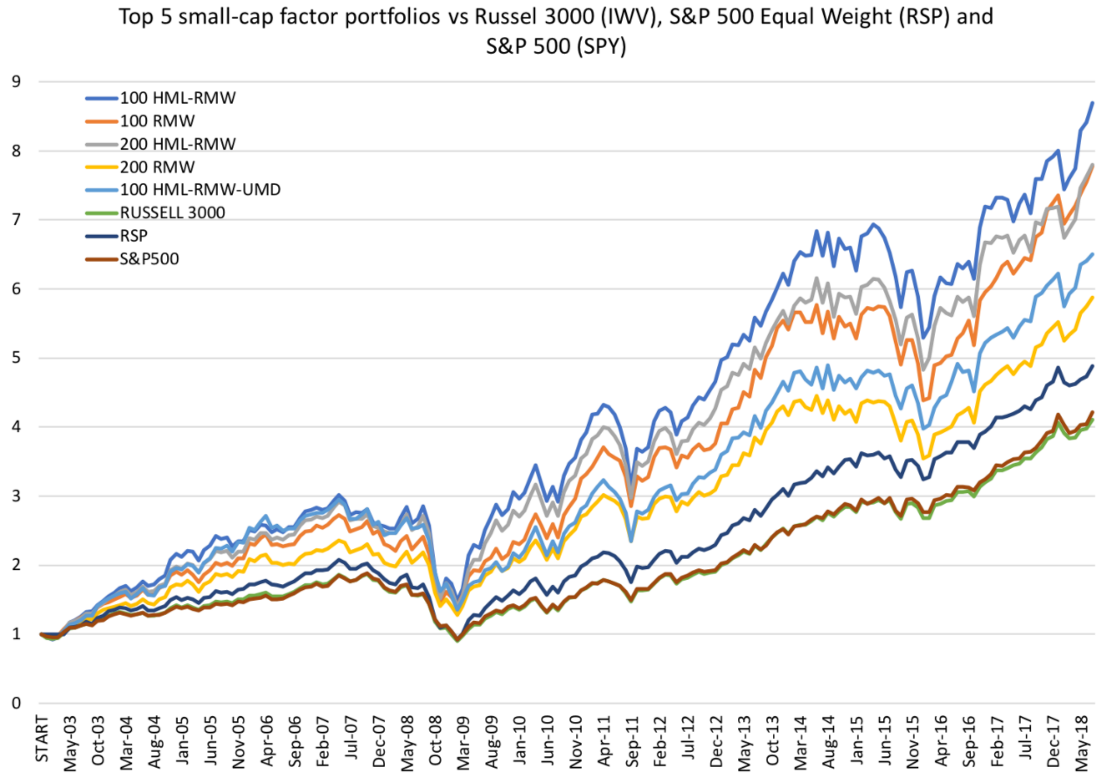


Figure 8 Top 5 small-cap factor portfolios vs Russell 3000 (IWV), S&P 500 (SPY) and S&P 500 Equal Weight (RSP)

4.3 Factor analysis with Alphalens

This sub-chapter reports the sensitivity of the sectors in the U.S. economy to the factors covered in this study, namely value (HML), profitability (RMW), and momentum (UMD). The results of the computation were grouped into three sub-periods: pre-crisis (01.01.2003 – 01.01.2007), crisis (01.01.2007 – 01.01.2010), and post-crisis (01.01.2010 – 31.07.2018). Tables 8 – 16 depict the mean information coefficients (IC) of each of the factors for each of the sub-periods. The IC shows how closely the factor's financial forecasts match actual financial results. The IC can range from 1 to -1, with -1 indicating the forecasts bear no relation to the actual results, and 1 indicating that the forecasts perfectly matched actual results (Kenton, 2018). A common practice is to associate a higher information coefficient with stronger predictive power of the factor. This computation was made based on the forward returns (FR) generated by each factor assuming no transaction costs. An example of a forward return would be if one took a position in the asset *A* and held it for *n* days based on the factor value computed on the first day. In order to give a broader perspective on how the predictive qualities of factors decay over time, the periods of 30, 60 and 90 days of forward returns were calculated. The number of days in the context of this analysis means *n* days of holding the security after the factor value was calculated.

The data in this section was augmented with the graphical depictions of the information coefficients for each factor and sector in Appendices 1 – 9.

4.3.1 Value (HML)

During the pre-crisis, sub-period small-cap stocks had stronger average information coefficients for value (HML) than the large caps for all three forward return periods (30, 60 and 90 days). The highest IC values among the group were the ones with utilities (0.135, 0.202 and 0.255), basic materials (0.074, 0.106 and 0.134) and health care (0.055, 0.096 and 0.115). Leaders of the large-cap group during pre-crisis were basic materials (0.066, 0.103 and 0.139), consumer defensive (0.041, 0.069 and 0.094) and health care (0.044, 0,066 and 0,083). Financial services, industrials, and technology all showed positive sensitivity towards value (HML) among both size groups with slightly higher IC for the small caps (Appendix 1a, 1b, and Table 8 - 10). The sectors with the negative value (HML) sensitivity in the sub-period were real estate and communication services with the lowest IC values in both the small and the large-cap groups. When looking at the small and the large caps together, mean IC tended to deviate from 0 more with the longer period taken for forward return computation. This was the case for all sectors except for the energy (Appendix 1c).

During the crisis sub-period, both the small and the large-cap groups experienced significant changes in the average ICs as compared to the pre-crisis. ICs of the small-cap utilities dropped from 0.135, 0.202 and 0.255 to 0.022, 0.018 and 0.041 for 30, 60 and 90 days forward returns respectively (Table 8, 9 and 10). ICs of real estate in the large-cap group, on the other hand, increased to 0.036, 0.058 and 0.058 for 30, 60 and 90 days almost inverting the IC values as compared to the pre-crisis sub-period (Table 8, 9 and 10). Such a change in IC of the sector might be explained by the strong direct connection of the businesses to the real estate market, which was one of the most affected during the sub-period. Basic materials and technology seemed to have lost most of the sensitivity to the value factor (HML) in both the small and the large size groups during the crisis. Large-cap utilities, however, had one of the highest HML sensitivities (0.04, 0.06 and 0.077) after having slightly negative ICs in the preceding sub-period (-0.007, -0.017 and -0.032). Industrials was the only sector that kept having the positive sensitivity to HML in both the small and the large size groups during the crisis although with the large caps having a slightly higher ICs this time (Tables 8 - 10).

During the post-crisis, basic materials, energy, industrials, and technology held on to negative ICs towards HML, never returning to the positive sensitivity to the factor. The communication services sector had an IC reversal in the small caps as well in the post-crisis as compared to both previous sub-periods. This time the small caps had a positive ICs of 0.03, 0.072 and 0.076 as opposed to -0.014, -0.044 and -0.101 before the crisis. Financial services, real estate, and utilities were the only sectors that re-turned to the similar HML sensitivity patterns they had before the crisis (Table 8, 10).

Generally, the factor has been the strongest throughout the pre-crisis sub-period with its ability to predict returns deteriorating after the crisis. The small-cap group was more sensitive to HML over the whole period covered in the study.

Table 8 Mean information coefficient by sector (HML 30 days)

HML (30 days)	*PRE*		*CRISIS*		*POST*	
Market cap	**Small**	**Large**	**Small**	**Large**	**Small**	**Large**
Basic materials	0.074 (12.327)	0.066 (11.670)	0.004 (0.488)	-0.010 (-1.156)	-0.026 (-5.477)	-0.060 (1.805)
Consumer cyclical	-0.003 (-0.940)	0.034 (8.156)	-0.018 (-3.140)	-0.019 (-2.609)	-0.008 (-3.245)	0.006 (-11.018)
Financial services	0.056 (10.628)	0.029 (6.468)	0.018 (2.105)	-0.010 (-1.301)	0.010 (2.698)	0.017 (4.314)
Real estate	-0.076 (-12.277)	-0.087 (-10.682)	-0.004 (-0.410)	0.036 (4.897)	-0.006 (-1.173)	-0.022 (-4.668)
Consumer defensive	-0.021 (-4.654)	0.041 (7.943)	0.016 (2.397)	-0.011 (-1.433)	0.000 (0.034)	0.008 (1.704)
Health care	0.055 (14.163)	0.044 (9.426)	0.057 (10.031)	0.018 (2.756)	0.010 (2.880)	0.000 (0.099)
Utilities	0.135 (16.041)	-0.007 (-1.086)	0.022 (2.054)	0.040 (4.969)	0.062 (9.543)	0.028 (7.245)
Communication services	-0.014 (-1.660)	-0.020 (-2.385)	-0.027 (-3.153)	0.025 (3.057)	0.030 (4.305)	-0.105 (-16.477)
Energy	-0.005 (-1.019)	0.044 (7.788)	0.022 (3.006)	0.008 (1.467)	-0.010 (-2.488)	-0.010 (-2.278)
Industrials	0.041 (10.045)	0.015 (3.637)	0.029 (5.867)	0.032 (6.060)	-0.027 (-10.494)	-0.023 (-6.156)
Technology	0.042 (14.217)	0.026 (6.062)	0.000 (0.073)	-0.003 (-0.534)	-0.010 (-2.933)	-0.017 (-4.673)

Note: t-statistics for IC appear in parentheses. [blue] – 3 highest IC values in the sub-period. [yellow] – 3 lowest IC values in the sub-period. [green] – positive IC values in the sub-period. [pink] – negative IC values in the sub-period. [white] – neutral IC values in the sub-period.

Table 9 Mean information coefficient by sector (HML 60 days)

HML (60 days)	*PRE*		*CRISIS*		*POST*	
Market cap	**Small**	**Large**	**Small**	**Large**	**Small**	**Large**
Basic materials	0.106 (18.951)	0.103 (17.187)	0.030 (3.989)	-0.002 (-0.242)	-0.041 (-9.514)	-0.069 (2.620)
Consumer cyclical	0.000 (-0.061)	0.050 (12.211)	-0.022 (-3.653)	-0.022 (-3.127)	-0.009 (-3.581)	0.008 (-13.268)
Financial services	0.073 (12.849)	0.045 (9.816)	0.016 (1.822)	-0.019 (-2.322)	0.014 (3.479)	0.022 (5.663)
Real estate	-0.082 (-12.726)	-0.119 (-14.405)	0.009 (0.877)	0.058 (10.665)	-0.012 (-2.536)	-0.037 (-7.771)
Consumer defensive	-0.011 (-3.042)	0.069 (13.714)	0.012 (1.733)	-0.016 (-2.001)	0.006 (1.419)	0.012 (2.503)
Health care	0.096 (23.666)	0.066 (14.313)	0.078 (16.374)	0.003 (0.568)	0.023 (7.219)	0.000 (0.122)
Utilities	0.202 (23.362)	-0.017 (-2.723)	0.018 (1.528)	0.060 (7.038)	0.098 (17.116)	0.040 (9.943)
Communication services	-0.044 (-5.767)	-0.019 (-2.246)	-0.072 (-9.923)	0.012 (1.322)	0.072 (10.491)	-0.126 (-21.088)
Energy	-0.024 (-4.593)	0.051 (9.530)	0.033 (4.246)	0.022 (4.510)	-0.004 (-0.879)	-0.010 (-2.349)
Industrials	0.064 (14.843)	0.031 (7.637)	0.040 (8.515)	0.056 (11.741)	-0.039 (-14.800)	-0.028 (-7.610)
Technology	0.064 (21.369)	0.039 (8.834)	0.005 (1.046)	-0.004 (-0.678)	-0.012 (-3.331)	-0.018 (-4.478)

Note: t-statistics for IC appear in parentheses. [blue] – 3 highest IC values in the sub-period. [yellow] – 3 lowest IC values in the sub-period. [green] – positive IC values in the sub-period. [pink] – negative IC values in the sub-period. [white] – neutral IC values in the sub-period.

Table 10 Mean information coefficient by sector (HML 60 days)

HML (90 days)	***PRE***		***CRISIS***		***POST***	
Market cap	**Small**	**Large**	**Small**	**Large**	**Small**	**Large**
Basic materials	0.134 (21.921)	0.139 (24.127)	0.054 (6.696)	0.013 (1.664)	-0.052 (-13.424)	-0.083 (2.808)
Consumer cyclical	-0.002 (0.664)	0.063 (18.281)	-0.028 (-4.703)	-0.017 (-2.414)	-0.016 (-6.278)	0.009 (-16.305)
Financial services	0.097 (16.985)	0.060 (13.239)	0.020 (2.266)	-0.021 (-2.500)	0.016 (3.919)	0.026 (6.519)
Real estate	-0.079 (-13.543)	-0.136 (-18.963)	0.029 (2.623)	0.058 (11.567)	-0.012 (-2.520)	-0.042 (-8.651)
Consumer defensive	-0.011 (-2.819)	0.094 (21.041)	-0.008 (-1.061)	-0.025 (-3.109)	0.006 (1.458)	0.012 (2.409)
Health care	0.115 (29.081)	0.083 (19.328)	0.097 (21.083)	-0.001 (-0.092)	0.036 (12.419)	-0.003 (-0.974)
Utilities	0.255 (32.124)	-0.032 (-4.739)	0.041 (3.508)	0.077 (8.408)	0.121 (22.658)	0.036 (9.293)
Communication services	-0.101 (-14.499)	0.012 (1.365)	-0.076 (-11.052)	0.004 (0.477)	0.076 (11.263)	-0.141 (-27.316)
Energy	-0.043 (-7.985)	0.061 (10.871)	0.041 (5.157)	0.033 (6.615)	-0.008 (-1.864)	-0.008 (-1.963)
Industrials	0.088 (20.491)	0.045 (12.696)	0.050 (11.180)	0.071 (14.362)	-0.047 (-18.162)	-0.032 (-8.509)
Technology	0.078 (25.170)	0.049 (10.858)	0.011 (2.068)	-0.006 (-1.132)	-0.018 (-4.903)	-0.013 (-3.163)

Note: t-statistics for IC appear in parentheses. — 3 highest IC values in the sub-period. — 3 lowest IC values in the sub-period. — positive IC values in the sub-period. — negative IC values in the sub-period. — neutral IC values in the sub-period.

4.3.2 Profitability (RMW)

During the pre-crisis sub-period, small-cap stocks demonstrated stronger average ICs for RMW. The sectors with the highest RMW sensitivity among the small caps were financial services (0.022, 0.053 and 0.088), real estate (0.077, 0.101 and 0.113), and health care (0.058, 0.071 and 0.09). Among the large caps, energy had the highest IC (0.047, 0.062 and 0.081) during the sub-period (Tables 11 – 13). The sectors with the strongest negative RMW sensitivity in the sub-period were basic materials, utilities, and industrials (both small and large caps), and consumer defensive (large-cap group).

During the crisis, energy sector experienced the largest sensitivity decline for both size groups with the small caps' IC dropping to -0.013, -0.038 and -0.065 and the large caps' falling to -0.004, -0.03 and -0.044 for the 30, 60 and 90 days respectively.

Among the other sizable changes were the large-cap consumer cyclical and consumer defensive (both size groups). Both sectors demonstrated positive ICs during the sub-period contrary to the tendency of the previous sub-period. Technology (both size groups) and communication services (large caps) also experienced a positive change in their IC values as compared to the pre-crisis.

After the crisis, most sectors had positive RMW sensitivity (Appendix 8a – 8c, Table 11 – 13). The exceptions were consumer defensive (both size groups) and utilities (large caps). This time, the strongest ICs were in large-cap basic materials (0.072, 0.093 and 0.11) and small-cap communication services (0.117, 0.156 and 0.179).

During the whole period, the factor's IC was predominantly in favour of the small-cap stocks. After the crisis, the effect of the RMW factor was noticeably more widespread across all the sectors. This might indicate the factor's universality which could be practical for designing diversified strategies.

Table 11 Mean information coefficient by sector (RMW 30 days)

RMW (30 days)	***PRE***		***CRISIS***		***POST***	
Market cap	**Small**	**Large**	**Small**	**Large**	**Small**	**Large**
Basic materials	0.006 (0.913)	-0.030 (-5.210)	-0.033 (-5.441)	-0.021 (-2.697)	0.062 (13.067)	0.072 (15.832)
Consumer cyclical	0.020 (6.248)	-0.006 (-1.205)	-0.002 (-0.458)	0.021 (4.078)	0.008 (2.837)	-0.008 (-3.545)
Financial services	0.022 (3.835)	0.001 (0.301)	0.026 (3.001)	0.038 (4.146)	-0.001 (-0.234)	0.028 (7.085)
Real estate	0.077 (12.881)	0.026 (3.891)	0.033 (4.449)	0.016 (2.347)	0.015 (3.296)	0.019 (5.356)
Consumer defensive	0.027 (4.507)	-0.016 (-3.635)	0.052 (8.152)	0.008 (1.294)	-0.018 (-4.435)	-0.006 (-1.772)
Health care	0.058 (8.591)	0.007 (1.710)	0.007 (1.776)	-0.004 (-0.810)	0.034 (9.934)	0.012 (3.705)
Utilities	-0.022 (-2.846)	-0.008 (-1.549)	0.005 (0.506)	-0.012 (-2.035)	0.025 (4.237)	-0.008 (-2.333)
Communication services	0.041 (4.731)	0.003 (0.346)	0.031 (3.006)	0.051 (4.496)	0.117 (17.827)	0.030 (5.412)
Energy	0.044 (7.095)	0.047 (6.082)	-0.013 (-1.952)	-0.004 (-0.712)	0.053 (13.301)	0.040 (10.936)
Industrials	0.005 (1.118)	0.014 (4.497)	-0.026 (-7.052)	-0.016 (-5.548)	0.035 (14.431)	0.023 (7.554)
Technology	0.007 (1.575)	0.002 (0.417)	0.008 (1.418)	0.015 (2.843)	0.018 (7.118)	0.021 (7.202)

Note: t-statistics for IC appear in parentheses. ☐ – 3 highest IC values in the sub-period. ☐ – 3 lowest IC values in the sub-period. ☐ – positive IC values in the sub-period. ☐ – negative IC values in the sub-period. ☐ – neutral IC values in the sub-period.

Table 12 Mean information coefficient by sector (RMW 60 days)

RMW (60 days)	***PRE***		***CRISIS***		***POST***	
Market cap	**Small**	**Large**	**Small**	**Large**	**Small**	**Large**
Basic materials	-0.020	-0.043	-0.049	-0.015	0.077	0.093
	(-3.048)	(-7.671)	(-8.160)	(-2.078)	(17.014)	(23.862)
Consumer cyclical	0.037	-0.011	0.012	0.019	0.019	-0.003
	(11.119)	(-0.398)	(2.460)	(4.330)	(7.105)	(-1.108)
Financial services	0.053	0.004	0.035	0.040	0.004	0.041
	(9.722)	(0.808)	(3.692)	(4.355)	(1.034)	(9.777)
Real estate	0.101	0.029	0.053	0.014	0.013	0.024
	(16.990)	(4.004)	(7.637)	(2.203)	(2.741)	(6.240)
Consumer defensive	0.019	-0.026	0.058	0.007	-0.021	-0.013
	(3.598)	(-5.540)	(8.141)	(0.978)	(-5.009)	(-3.884)
Health care	0.071	0.014	0.007	-0.016	0.052	0.020
	(11.143)	(2.967)	(1.544)	(-3.184)	(16.731)	(6.236)
Utilities	-0.012	-0.032	-0.030	-0.026	0.061	-0.017
	(-1.404)	(-5.608)	(-2.965)	(-4.228)	(10.421)	(-5.525)
Communication services	0.011	-0.010	0.006	0.052	0.156	0.034
	(1.264)	(-1.194)	(0.580)	(4.661)	(22.957)	(5.912)
Energy	0.049	0.062	-0.038	-0.030	0.082	0.052
	(8.156)	(8.530)	(-6.684)	(-6.409)	(21.725)	(13.806)
Industrials	-0.011	0.002	-0.038	-0.033	0.057	0.037
	(-2.287)	(0.560)	(-9.522)	(-9.061)	(23.442)	(11.275)
Technology	0.003	0.008	0.021	0.027	0.025	0.021
	(0.640)	(1.508)	(3.576)	(5.359)	(9.922)	(7.801)

Note: t-statistics for IC appear in parentheses. – 3 highest IC values in the sub-period. – 3 lowest IC values in the sub-period. – positive IC values in the sub-period. – negative IC values in the sub-period. – neutral IC values in the sub-period.

Table 13 Mean information coefficient by sector (RMW 90 days)

RMW (90 days)	***PRE***		***CRISIS***		***POST***	
Market cap	**Small**	**Large**	**Small**	**Large**	**Small**	**Large**
Basic materials	-0.038	-0.028	-0.031	-0.003	0.084	0.110
	(-5.882)	(-5.049)	(-5.589)	(-0.476)	(18.442)	(29.344)
Consumer cyclical	0.051	-0.016	0.040	0.023	0.027	0.006
	(15.655)	(-3.784)	(7.587)	(5.650)	(10.657)	(2.623)
Financial services	0.088	0.009	0.040	0.052	0.025	0.046
	(15.739)	(1.631)	(4.101)	(5.774)	(7.470)	(11.604)
Real estate	0.113	0.024	0.071	-0.005	0.018	0.025
	(19.969)	(3.420)	(8.841)	(-0.715)	(4.036)	(6.798)
Consumer defensive	0.002	-0.036	0.087	0.016	-0.023	-0.019
	(0.429)	(-7.975)	(13.219)	(2.532)	(-4.973)	(-5.298)
Health care	0.090	0.020	0.006	-0.009	0.071	0.029
	(14.746)	(4.035)	(1.375)	(-2.169)	(25.221)	(9.082)
Utilities	0.001	-0.038	-0.004	-0.032	0.056	-0.015
	(0.114)	(-6.547)	(-0.432)	(-4.893)	(9.648)	(-4.692)
Communication services	0.007	-0.026	-0.007	0.093	0.179	0.047
	(0.835)	(-3.519)	(-0.678)	(7.862)	(26.348)	(8.436)
Energy	0.050	0.081	-0.065	-0.044	0.095	0.058
	(8.570)	(12.872)	(-11.629)	(-9.045)	(25.562)	(15.443)
Industrials	-0.013	-0.015	-0.035	-0.037	0.068	0.050
	(-2.934)	(-5.003)	(-8.392)	(-10.235)	(27.892)	(16.020)
Technology	-0.003	0.022	0.034	0.037	0.030	0.026
	(-0.730)	(4.102)	(5.577)	(6.676)	(13.052)	(10.498)

Note: t-statistics for IC appear in parentheses. – 3 highest IC values in the sub-period. – 3 lowest IC values in the sub-period. – positive IC values in the sub-period. – negative IC values in the sub-period. – neutral IC values in the sub-period.

4.3.3 Momentum (UMD)

During the pre-crisis, sub-period energy (0.089, 0.117 and 0.149) and real estate (0.076, 0.076 and 0.081) had the highest ICs for UMD among the small caps. The small-cap technology (-0.061, -0.099 and -0.121), utilities (-0.04, -0.066 and -0.082) and basic materials (-0.035, -0.056 and -0.08), on the other hand, had the lowest ICs to the factor (Appendix 3a). In the large-cap group, basic materials had the highest IC for the holding periods of 30 and 60 days while communication services and energy had the highest ICs for 90 days of forward returns (Table 14 – 16, Appendix 3b). It is worth noting that unlike other sectors, real estate and energy had positive UMD sensitivity for both size groups throughout the sub-period.

The crisis brought some drastic changes for UMD. With the exception of the large-cap real estate stocks (0.029, 0.049 and 0.06) a consistent negative sensitivity pattern could be seen among all the sectors throughout the sub-period (Table 14 – 16, Appendix 6a – 6c). The strongest negative IC for the small caps was one of the utilities (-0.131, -0.175 and -0.205) and energy (-0.1, -0.152 and -0.182). In the large-cap group consumer defensive (-0.086, -0.127 and -0.142) and energy (-0.067, -0.099 and -0.111) had the lowest ICs in the sub-period.

During the post-crisis, all sectors, even those that used to have negative UMD sensitivity in the preceding sub-periods, had positive ICs, a pattern contrary to the one seen during the crisis. Financial services and communication services, however, returned to a similar to the pre-crisis sensitivity patterns. Basic materials had the strongest UMD sensitivity in both the small and large-cap groups (Appendix 9a – 9c, Table 14 – 16).

Table 14 Mean information coefficient by sector (UMD 30 days)

UMD (30 days)	***PRE***		***CRISIS***		***POST***	
Market cap	**Small**	**Large**	**Small**	**Large**	**Small**	**Large**
Basic materials	-0.035 (-5.030)	0.060 (6.520)	-0.059 (-6.096)	0.018 (1.634)	0.057 (10.098)	0.060 (10.024)
Consumer cyclical	0.036 (7.622)	0.004 (0.788)	-0.012 (-1.480)	0.001 (0.104)	0.040 (12.802)	0.030 (7.344)
Financial services	-0.006 (-0.898)	-0.004 (-0.634)	-0.050 (-5.423)	0.007 (0.656)	-0.009 (-2.106)	-0.005 (-0.916)
Real estate	0.076 (10.674)	0.055 (7.848)	-0.003 (-0.265)	0.029 (2.174)	0.044 (9.253)	0.033 (5.857)
Consumer defensive	0.014 (2.017)	-0.031 (-5.829)	-0.058 (-7.808)	-0.086 (-9.566)	0.039 (7.887)	-0.004 (-0.913)
Health care	-0.011 (-2.598)	-0.018 (-3.205)	-0.057 (-8.574)	-0.032 (-4.307)	0.018 (6.025)	0.008 (2.000)
Utilities	-0.040 (-4.211)	-0.001 (-0.159)	-0.131 (-12.282)	-0.021 (-2.238)	0.023 (3.564)	0.043 (8.432)
Communication services	-0.034 (-4.058)	0.039 (5.224)	-0.004 (-0.318)	-0.001 (-0.062)	0.013 (1.761)	0.061 (10.271)
Energy	0.089 (13.388)	0.028 (3.464)	-0.100 (-13.294)	-0.067 (-7.701)	0.048 (10.130)	0.031 (6.223)
Industrials	-0.033 (-5.850)	0.014 (2.329)	-0.040 (-4.957)	-0.035 (-4.295)	0.026 (7.933)	0.027 (6.410)
Technology	-0.061 (-14.156)	-0.029 (-6.088)	-0.056 (-9.732)	-0.016 (-2.518)	0.012 (4.447)	0.019 (5.062)

Note: t-statistics for IC appear in parentheses. ─ 3 highest IC values in the sub-period. ─ 3 lowest IC values in the sub-period. ─ positive IC values in the sub-period. ─ negative IC values in the sub-period. ─ neutral IC values in the sub-period.

Table 15 Mean information coefficient by sector (UMD 60 days)

UMD (60 days)	***PRE***		***CRISIS***		***POST***	
Market cap	**Small**	**Large**	**Small**	**Large**	**Small**	**Large**
Basic materials	-0.056 (-8.952)	0.068 (7.870)	-0.113 (-11.366)	-0.012 (-1.134)	0.080 (13.844)	0.070 (11.733)
Consumer cyclical	0.041 (10.960)	-0.004 (-0.704)	-0.023 (-2.519)	-0.015 (-1.462)	0.050 (16.730)	0.032 (8.319)
Financial services	-0.026 (-4.442)	-0.029 (-4.735)	-0.041 (-4.696)	0.010 (0.860)	-0.021 (-4.830)	-0.023 (-4.203)
Real estate	0.076 (10.552)	0.055 (8.380)	-0.008 (-0.570)	0.049 (3.801)	0.050 (11.469)	0.026 (4.684)
Consumer defensive	0.018 (2.658)	-0.045 (-8.740)	-0.071 (-11.484)	-0.127 (-16.279)	0.060 (12.599)	-0.005 (-1.120)
Health care	-0.024 (-6.694)	-0.024 (-5.494)	-0.079 (-11.580)	-0.051 (-6.767)	0.020 (7.260)	0.005 (1.385)
Utilities	-0.066 (-6.799)	0.003 (0.340)	-0.175 (-17.253)	-0.046 (-5.538)	0.039 (6.120)	0.055 (10.898)
Communication services	-0.056 (-7.012)	0.066 (8.755)	-0.038 (-3.470)	-0.011 (-1.185)	-0.013 (-1.688)	0.070 (12.602)
Energy	0.117 (20.166)	0.039 (5.596)	-0.152 (-21.812)	-0.099 (-13.608)	0.067 (14.479)	0.042 (8.785)
Industrials	-0.055 (-10.129)	-0.002 (-0.424)	-0.058 (-7.741)	-0.056 (-7.490)	0.030 (9.551)	0.038 (9.159)
Technology	-0.099 (-25.001)	-0.062 (-13.525)	-0.082 (-15.858)	-0.022 (-3.592)	0.006 (2.638)	0.007 (2.031)

Note: t-statistics for IC appear in parentheses. ─ 3 highest IC values in the sub-period. ─ 3 lowest IC values in the sub-period. ─ positive IC values in the sub-period. ─ negative IC values in the sub-period. ─ neutral IC values in the sub-period.

Table 16 Mean information coefficient by sector (UMD 90 days)

UMD (90 days)	*PRE*		*CRISIS*		*POST*	
Market cap	**Small**	**Large**	**Small**	**Large**	**Small**	**Large**
Basic materials	-0.080 (-14.173)	0.055 (6.384)	-0.143 (-12.719)	-0.059 (-5.173)	0.090 (15.488)	0.097 (16.066)
Consumer cyclical	0.049 (15.552)	0.003 (0.660)	-0.030 (-2.877)	-0.025 (-2.182)	0.061 (20.992)	0.032 (9.425)
Financial services	-0.045 (-6.556)	-0.033 (-5.829)	-0.046 (-4.772)	0.012 (0.987)	-0.018 (-4.194)	-0.023 (-4.433)
Real estate	0.081 (12.036)	0.079 (15.060)	-0.039 (-2.676)	0.060 (4.819)	0.061 (14.991)	0.035 (7.139)
Consumer defensive	0.037 (5.789)	-0.044 (-8.358)	-0.073 (-9.590)	-0.142 (-18.517)	0.066 (14.383)	-0.010 (-2.380)
Health care	-0.031 (-10.301)	-0.022 (-5.617)	-0.091 (-11.909)	-0.058 (-7.304)	0.019 (7.377)	0.007 (1.732)
Utilities	-0.082 (-8.127)	0.005 (0.637)	-0.205 (-19.223)	-0.055 (-7.483)	0.046 (7.390)	0.057 (11.832)
Communication services	-0.027 (-3.350)	0.070 (8.388)	-0.065 (-5.969)	-0.037 (-3.671)	-0.017 (-2.239)	0.085 (15.762)
Energy	0.149 (31.352)	0.062 (8.969)	-0.182 (-25.678)	-0.111 (-17.483)	0.082 (17.379)	0.051 (10.732)
Industrials	-0.068 (-14.220)	-0.009 (-1.585)	-0.081 (-10.421)	-0.082 (-10.940)	0.038 (12.380)	0.045 (11.633)
Technology	-0.121 (-29.781)	-0.089 (-18.843)	-0.090 (-16.640)	-0.032 (-5.144)	0.003 (1.456)	-0.002 (-0.661)

Note: t-statistics for IC appear in parentheses. ─ 3 highest IC values in the sub-period. ─ 3 lowest IC values in the sub-period. ─ positive IC values in the sub-period. ─ negative IC values in the sub-period. ─ neutral IC values in the sub-period.

4.4 Costs simulation analysis

For the cost simulation analysis, the strategies with the highest Sharpe ratios were selected with the following conditions fulfilled: first, there must be at least one portfolio based on 1, 2 and 3 factors. Second, there must be a single factor portfolio employing each factor (HML, RMW, and UMD). Third, the list must include at least one portfolio holding 100, 200 and 500 positions. Although the choice of these particular conditions could be deemed arbitrary, using them should encompass strategies of different kinds. Thus, the researcher should be able to capture the relationship between the transaction costs and the different factors and portfolio sizes used.

After analysing the strategies, five portfolios were selected: the large-cap profitability (RMW 100), the small-cap value and profitability (HML-RMW 100), the large-cap value, profitability, and momentum (HML-RMW-UMD 500), the small-cap value (HML

200) and the large-cap momentum (UMD 100). The numbers after the factor codes indicate the number of stocks held in the portfolio.

To approximate the impact of the trade commissions and the other market frictions, the backtests were run with the appropriate considerations. There were two key assumptions in the analysis: first, the investor had an account in a discount brokerage firm with the variable commission of $0.005 per share and the fixed cost of $1 per trade. It is also assumed that the reasonable initial capital of a retail investor account could be $100,000, $50,000 or $10,000. The simulations of each strategy are presented in tables 17 – 21.

The large-cap RMW 100 (Table 17) and the small-cap HML-RMW 100 (Table 18) demonstrated superior performance to all the 3 benchmarks for $100,000 and $50,000 as initial capital as measured by the Sharpe ratio. The large-cap UMD 100 portfolio (Table 21), however, managed to outperform the benchmarks with $100,000 and $50,000 as initial capital from the total return standpoint. However, the strategy had a comparable Sharpe ratio to the benchmarks due to its relatively high volatility. The large-cap HML-RMW-UMD 500 (Table 19) and the small-cap HML 200 (Table 20) delivered lower risk-adjusted returns for all three balances ($100,000, $50,000 and $10,000) as compared to the benchmarks. The results of the cost simulation analysis were mostly in favour of the concentrated portfolios, i.e. the ones with fewer stocks.

Table 17 RMW 100 large-cap portfolio performance comparison (Long-only)

LARGE 100	***RUSSELL 3000***	***RSP (from 30.04.2003)***	***S&P 500***	***RMW (w/o costs)***	***RMW (with costs)***	***RMW (with costs)***	***RMW (with costs)***	***RMW (with costs)***
Initial capital	***$1,000,000***	***$1,000,000***	***$1,000,000***	***$1,000,000***	***$1,000,000***	***$100,000***	***$50,000***	***$10,000***
Portfolio return ($\boldsymbol{R_p}$)	312.4%	395.1%	315.6%	552.8%	542.8%	502.1%	463.4%	283.1%
Annual return ($\boldsymbol{R_{1y}}$)	9.5%	11.1%	9.6%	12.8%	12.7%	12.2%	11.8%	9.0%
Maximum drawdown (MDD)	-56.4%	-60.1%	-54.9%	-54.6%	-54.7%	-55.0%	-55.1%	-56.5%
Annual volatility (σ)	18.2%	19.6%	18.0%	18.9%	18.9%	18.9%	18.9%	19.0%
Sharpe ratio (SR)	0.59	0.63	0.60	0.73	0.73	0.71	0.68	0.55
Sortino ratio (Srt)	0.83	0.89	0.85	1.02	1.02	0.98	0.95	0.76
Gross leverage (Lv)	1.0	1.0	1.0	0.99	0.99	0.99	0.99	0.99
Daily turnover (To)	0.1%	0.1%	0.1%	1.9%	1.9%	1.9%	1.9%	1.9%
Portfolio Beta (β)	1.0	1.0	1.0	1.01	1.01	1.01	1.01	1.01

Table 18 HML-RMW 100 small-cap portfolio performance comparison (Long-only)

SMALL 100	***RUSSELL 3000***	***RSP (from 30.04.2003)***	***S&P 500***	***HML-RMW (w/o costs)***	***HML-RMW (with costs)***	***HML-RMW (with costs)***	***HML-RMW (with costs)***	***HML-RMW (with costs)***
Initial capital	***$1,000,000***	***$1,000,000***	***$1,000,000***	***$1,000,000***	***$1,000,000***	***$100,000***	***$50,000***	***$10,000***
Portfolio return ($\boldsymbol{R_p}$)	312.4%	395.1%	315.6%	799.2%	765.4%	727.5%	674.0%	312.7%
Annual return ($\boldsymbol{R_{1y}}$)	9.5%	11.1%	9.6%	15.2%	14.9%	14.6%	14.1%	9.5%
Maximum drawdown (MDD)	-56.4%	-60.1%	-54.9%	-61.2%	-61.3%	-61.5%	-61.8%	-64.5%
Annual volatility (σ)	18.2%	19.6%	18.0%	23.8%	23.8%	23.8%	23.8%	24.0%
Sharpe ratio (SR)	0.59	0.63	0.60	0.71	0.70	0.69	0.67	0.50
Sortino ratio (Srt)	0.83	0.89	0.85	1.01	1.0	0.98	0.95	0.70
Gross leverage (Lv)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Daily turnover (To)	0.1%	0.1%	0.1%	2.9%	2.9%	2.9	2.9%	2.9%
Portfolio Beta (β)	1.0	1.0	1.0	1.19	1.19	1.19	1.19	1.20

Table 19 HML-RMW-UMD 500 large-cap portfolio performance comparison (Long-only)

LARGE 500	*RUSSELL 3000*	*RSP (from 30.04.2003)*	*S&P 500*	*HML-RMW-UMD (w/o costs)*	*HML-RMW-UMD (with costs)*	*HML-RMW-UMD (with costs)*	*HML-RMW-UMD (with costs)*	*HML-RMW-UMD (with costs)*
Initial capital	***$1,000,000***	***$1,000,000***	***$1,000,000***	***$1,000,000***	***$1,000,000***	***$100,000***	***$50,000***	***$10,000***
Portfolio return ($\boldsymbol{R_p}$)	312.4%	395.1%	315.6%	450.0%	430.8%	310.8%	244.4%	-40.0%
Annual return ($\boldsymbol{R_{1y}}$)	9.5%	11.1%	9.6%	11.6%	11.3%	9.5%	8.3	-3.2%
Maximum drawdown (MDD)	-56.4%	-60.1%	-54.9%	-53.6%	-53.8%	-55.7%	-56.8%	-55.5%
Annual volatility (σ)	18.2%	19.6%	18.0%	18.7%	18.7%	18.8%	18.7%	12.5%
Sharpe ratio (SR)	0.59	0.63	0.60	0.68	0.67	0.58	0.52	-0.20
Sortino ratio (Srt)	0.83	0.89	0.85	0.95	0.93	0.80	0.72	-0.27
Gross leverage (Lv)	1.0	1.0	1.0	0.99	0.99	0.99	0.99	0.49
Daily turnover (To)	0.1%	0.1%	0.1%	1.4%	1.4%	1.5%	1.5%	2.0%
Portfolio Beta (β)	1.0	1.0	1.0	1.01	1.01	1.01	1.01	0.61

Table 20 HML 200 small-cap portfolio performance comparison (Long-only)

SMALL 200	*RUSSELL 3000*	*RSP (from 30.04.2003)*	*S&P 500*	*HML (w/o costs)*	*HML (with costs)*	*HML (with costs)*	*HML (with costs)*	*HML (with costs)*
Initial capital	***$1,000,000***	***$1,000,000***	***$1,000,000***	***$1,000,000***	***$1,000,000***	***$100,000***	***$50,000***	***$10,000***
Portfolio return ($\boldsymbol{R_p}$)	312.4%	395.1%	315.6%	617.8%	590.7%	518.1%	439.6%	59.9%
Annual return ($\boldsymbol{R_{1y}}$)	9.5%	11.1%	9.6%	13.5%	13.2%	12.4%	11.4%	3.1%
Maximum drawdown (MDD)	-56.4%	-60.1%	-54.9%	-65.3%	-65.6%	-66.2%	-67.1%	-74.1%
Annual volatility (σ)	18.2%	19.6%	18.0%	25.0%	25.0%	25.0%	25.0%	25.6%
Sharpe ratio (SR)	0.59	0.63	0.60	0.63	0.62	0.59	0.56	0.25
Sortino ratio (Srt)	0.83	0.89	0.85	0.90	0.89	0.84	0.79	0.34
Gross leverage (Lv)	1.0	1.0	1.0	0.98	0.98	0.98	0.98	0.99
Daily turnover (To)	0.1%	0.1%	0.1%	1.8%	1.8%	1.8%	1.8%	1.8%
Portfolio Beta (β)	1.0	1.0	1.0	1.23	1.23	1.23	1.24	1.25

Table 21 UMD 100 large-cap portfolio performance comparison (Long-only)

LARGE 100	*RUSSELL 3000*	*RSP (from 30.04.2003)*	*S&P 500*	*UMD (w/o costs)*	*UMD (with costs)*	*UMD (with costs)*	*UMD (with costs)*	*UMD (with costs)*
Initial capital	***$1,000,000***	***$1,000,000***	***$1,000,000***	***$1,000,000***	***$1,000,000***	***$100,000***	***$50,000***	***$10,000***
Portfolio return ($\boldsymbol{R_p}$)	312.4%	395.1%	315.6%	506.3%	493.3%	449.6%	403.7%	142.6%
Annual return ($\boldsymbol{R_{1y}}$)	9.5%	11.1%	9.6%	12.3%	12.1%	11.6%	11.0%	5.9%
Maximum drawdown (MDD)	-56.4%	-60.1%	-54.9%	-56.1%	-56.2%	-56.5%	-56.9%	-59.7%
Annual volatility (σ)	18.2%	19.6%	18.0%	21.3%	21.3%	21.3%	21.3%	21.3%
Sharpe ratio (SR)	0.59	0.63	0.60	0.65	0.65	0.62	0.60	0.37
Sortino ratio (Srt)	0.83	0.89	0.85	0.90	0.89	0.86	0.82	0.51
Gross leverage (Lv)	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Daily turnover (To)	0.1%	0.1%	0.1%	3.1%	3.1%	3.1%	3.1%	3.1%
Portfolio Beta (β)	1.0	1.0	1.0	1.06	1.06	1.06	1.06	1.06

5 Conclusion

This chapter intends to clarify the results of the examination performed to answer the research questions and test the predetermined hypotheses. This section also summarises the practical implications of the research along with its limitations. Additionally, a few suggestions for future research have been provided.

5.1 Summary of key findings

The primary goal of this study was to examine the potential of applying factors of modern asset pricing models for automated long-term portfolio management. The theoretical and empirical analysis assisted the author in achieving the main goals of the work. To summarise the key findings of the study, the research questions are individually answered in this sub-chapter.

1. Does a portfolio based on multiple factors provide a better risk-adjusted return than a single factor portfolio and a market portfolio?

To answer this question, a series of portfolio simulations were conducted that produced the descriptive statistics for each of the strategies. Since the risk-adjusted return is represented by the Sharpe ratio of the portfolio, it was used for comparison. The statistics suggested that the portfolios based on factors can indeed generate higher risk-adjusted returns since they were able to beat the performance of the benchmark in the majority of observations. The results also indicated a positive relationship between the Sharpe ratio and the number of factors used in the strategy. Thus, the portfolios based on multiple factors on average tended to outperform the strategies based on a single factor that in their turn on average tended to outperform the benchmark.

2. What is the extent of variation of the sensitivity of the sectors of the U.S. economy in response to each factor?

Alphalens analysis showed that the sectors' sensitivity as per the information coefficient varied significantly from one economic sector to another. Different sectors demonstrated diverse factor sensitivity patterns during the three sub-periods: pre-crisis, crisis, and post-crisis. The majority of the sectors had positive sensitivity to value and profitability during the pre-crisis sub-period with a few sectors being negatively sensitive. Sensitivity to these factors later dropped during the crisis with a more noticeable dip in the large-caps. During the post-crisis, however, sensitivity to profitability recovered for most of the sectors showing a stronger information coefficient than before the crisis. On the other hand, sensitivity to value weakened after the crisis for both size groups. The sectors showed mixed sensitivity to momentum during the pre-crisis sub-period independent of the size. Almost all sectors had negative sensitivity to momentum during the crisis sub-period with a sharp sensitivity reversal during post-crisis for all sectors except for the financial services and small-cap communication services. Generally, consistent sensitivity for the majority of sectors has only been seen with profitability (post-crisis) and momentum (crisis and post-crisis).

3. Are factor-based portfolios the expedient alternative to a market portfolio for a retail investor?

To answer this question, costs simulation analysis was performed. The portfolios were compared based on their Sharpe ratios after consideration of the transaction costs and capital invested. It was discovered that a select number of portfolios were able to deliver superior risk-adjusted performance to one of the benchmarks after the costs of transactions. However, portfolios of fewer stocks tended to win over the portfolios with a larger number of positions. A retail investor should be able to outperform the market on the risk-adjusted basis using the factor-based portfolios with $50,000 or more at disposal.

Below are the hypotheses tested in this research. The first hypothesis was concerned with the factors ability to synergise, thus making multiple factor portfolios better than the single factor portfolios and a market portfolio:

H_1: Investing in a portfolio based on multiple factors provides a better risk-adjusted return than a portfolio based on a single factor and a market portfolio.

According to the data produced by the portfolio simulation analysis, portfolios based on multiple factors tended to outperform the portfolios based on a single factor. Portfolios based on a single factor in their turn tended to outperform the benchmarks. Therefore, the first hypothesis should be accepted.

The second hypothesis presumed that the U.S. economic sectors had asymmetric sensitivity to the factors:

H_2: Different sectors of the U.S. economy have different sensitivity to factors.

According to the Alphalens sensitivity analysis data, the sectors had indeed shown significant variability in the sensitivity patterns for the different sectors during the same sub-periods. Thus, the hypothesis has to be accepted.

The last hypothesis suggested that a retail investor would not be able to achieve the superior risk-adjusted return as compared to the benchmarks using factor portfolios:

H_3: Factor-based portfolios are less expedient for a retail investor than a market portfolio due to high transaction costs in a retail investor's account.

The data produced by the cost simulation analysis suggested that a retail investor should be able to achieve the superior risk-adjusted performance to one of the benchmarks using the factor portfolios. This means that the hypothesis should be rejected.

To conclude, this research studied the portfolios based on factors, namely value, profitability, and momentum. The analysis had shown that the factors may have good potential to be used as the basis of a viable investment alternative to a market portfolio. Nonetheless, the actual performance of such strategies will depend on the factors being used, cost structure as well as capital at disposal. The results of this work should help one approach investing in the factor-based portfolios. The findings could also be used for further developments on the topic either practically or academically.

5.2 Practical implications

Managing personal finance has been of rising relevance in recent years. With the increase in popularity, a vast amount of academic studies have been developed. However, many prioritised theoretical aspects over practical implications, leaving the community of retail investors with less actionable information. Developing sound empirical knowledge is of particular importance especially when advanced technology becomes more accessible to the public and not only industry professionals.

This study was specifically designed to be beneficial to the investment practitioners interested in applying quantitative factor models in their portfolios as well as anyone intending to develop knowledge about quantitative investment management practices. In order to facilitate this purpose, this study includes a reasonably detailed description of the approach as well as all relevant source code with the author's comments (see Appendices 10 – 15).

Although this particular work studied the context of the U.S., a similar approach could be used for other countries' markets as well. Broadly, this thesis should help the reader familiarise with the topic of quantitative investment management providing him or her with a ground for improving the ideas presented here. For instance, learning about factor combinations not discussed in the work or attempting hedging for certain sectors or factors could be the further areas of inquiry. Generally, this study intended to broaden the perspective of empirical research on the factor investment strategies by including retail investors in a group of parties interested in such research being done.

5.3 Limitations and recommendations

This sub-chapter presents the limitations of the work and introduces some of the recommendations that could be of service for further studies. First, this thesis is limited to the sample of 42 factor-based portfolios that was formed from U.S. equities. Therefore, the research results could not be generalised for the global market. Nonetheless, given access to data similar studies, it should be possible to conduct this study in the other countries' markets as well. Besides, the tests employed in this thesis were conducted on past data within a fixed timeframe. This means that testing on the newer data may suggest different results for the strategies in question. Likewise, the researcher may observe divergent behaviour when testing the strategies on a live data feed.

The limitations of the research, however, broaden the opportunities for future studies. Thus, in this work, factors were used "as is" in their simplest form. This suggests that the performance of the strategies could be significantly improved with additional conditions and constraints. For example, instead of using the single most recent value for profitability factor, one may attempt to consider the profitability growth over a specified time to improve the quality of stock ranking. Similarly, there is more than one way of combining factors in a portfolio. Instead of finding the highest sum of the factors' values one may prioritise certain factors over the others. For instance, one could filter the universe by the top profitability percentile and search for the highest value stocks within the sample. One could also apply the trend filter in a form of positive momentum similar to the previous example.

Furthermore, one could extend the scope of this work by trying to hedge factor portfolios against sectors of the economy insensitive to certain factors based on the data produced by the Alphalens analysis. A few more options could be attempting to optimise the portfolios' rebalancing frequencies and other parameters and broadening the research by studying the other countries' markets.

References

Alden, L. (2018). *Equal Weighted Index Funds: The Pros and Cons and Which to Pick*. Retrieved from www.lynalden.com: https://www.lynalden.com/equal-weighted-index-funds/

Bloomberg. (2019, February 3). *Russell 3000 Index*. Retrieved from www.bloomberg.com: https://www.bloomberg.com/quote/RAY:IND

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*.

Chen, J. (2018, July 22). *American Depositary Receipt - ADR*. Retrieved from www.investopedia.com: https://www.investopedia.com/terms/a/adr.asp

Chen, J. (2018, January 18). *Capitalization-Weighted Index*. Retrieved from www.investopedia.com: https://www.investopedia.com/terms/c/capitalizationweightedindex.asp

Cnossen, C. (1997). *Secondary Research: Learning Paper 7.* Robert Gordon University.

David E. Rapach, M. E. (2005, February 1). *In-Sample vs. Out-of-Sample Tests of Stock Return Predictability in the Context of Data Mining .* Retrieved from citeseerx.ist.psu.edu: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.561.2002&rep=rep1&type=pdf

Denzil Watson, A. H. (2007). *Corporate Finance Principles & Practice.* Harlow: Pearson Education Limited.

Eugene F. Fama, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*.

Eugene F. Fama, K. R. (2014, September). A Five-Factor Asset Pricing Model. *Journal of Financial Economics*.

French, K. R. (2019, March 12). *Current Research Returns*. Retrieved from http://mba.tuck.dartmouth.edu: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Howell, K. E. (2013). *An Introduction to the Philosophy of Methodology.* London: Sage Publications Ltd.

Investopedia. (2018, May 31). *Universe of Securities*. Retrieved from www.investopedia.com: https://www.investopedia.com/terms/u/universeofsecurities.asp

Jerald E. Pinto, E. H. (2010). *Equity Asset Valuation.* Hoboken: John Wiley & Sons, Inc.

Kent Daniel, T. J. (2013, December 24). *Momentum Crashes.* Retrieved from ssrn.com: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2371227

Kenton, W. (2018, January 27). *Information Coefficient - IC*. Retrieved from www.investopedia.com: https://www.investopedia.com/terms/i/information-coefficient.asp

Kosakowski, P. (2017, May 8). *The Fall of the Market in the Fall of 2008*. Retrieved from www.investopedia.com: https://www.investopedia.com/articles/economics/09/subprime-market-2008.asp

Lemke, T. (2018, December 6). *What is the Weighting of the S&P 500?* Retrieved from www.thebalance.com: https://www.thebalance.com/what-is-the-sector-weighting-of-the-s-and-p-500-4579847

Mark Saunders, P. L. (2009). *Research methods for business students (fifth edition).* Harlow: Pearson Education.

Merjan, D. (2018, May 24). *ABCs of ADRs*. Retrieved from blog.williamblairfunds.com: https://blog.williamblairfunds.com/david-merjan/abcs-of-adrs/

Morningstar, Inc. (2019). *Expense Ratio*. Retrieved from www.morningstar.com: http://www.morningstar.com/invglossary/expense_ratio.aspx

MSCI Inc. (2018). *THE MSCI VALUE WEIGHTED INDEXES.* Retrieved from www.msci.com: https://www.msci.com/documents/10199/cbdd6492-831c-479f-ba15-e9981a1c7644

Narasimhan Jegadeesh, S. T. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, 83.

National Academy of Sciences. (1998). *Teaching About Evolution and the Nature of Science.* Washington: The National Academies Press.

Novy-Marx, R. (2012, June). The other side of value: The gross profitability premium. *Journal of Financial Economics*.

Paul G. Hewitt, J. S. (2013). *Conceptual Physical Science - Explorations.* Cram101 Publishing.

Payne, J. (2017, November 8). *Working On Our Best Universe Yet: QTradableStocksUS*. Retrieved from www.quantopian.com: https://www.quantopian.com/posts/working-on-our-best-universe-yet-qtradablestocksus

Payne, J. (2017, November 8). *Working On Our Best Universe Yet: QTradableStocksUS*. Retrieved from www.quantopian.com: https://www.quantopian.com/posts/working-on-our-best-universe-yet-qtradablestocksus

Quantopian Inc. (2019, January). *Fundamental Data*. Retrieved from www.quantopian.com: https://www.quantopian.com/data/morningstar/fundamentals

Sarma, R. (2012). RESEARCH: IN A METHODOLOGICAL FRAME. *Journal of Arts, Science & Commerce*.

Victor DeMiguel, L. G. (2009). A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms. *© 2009 INFORMS*, 9.

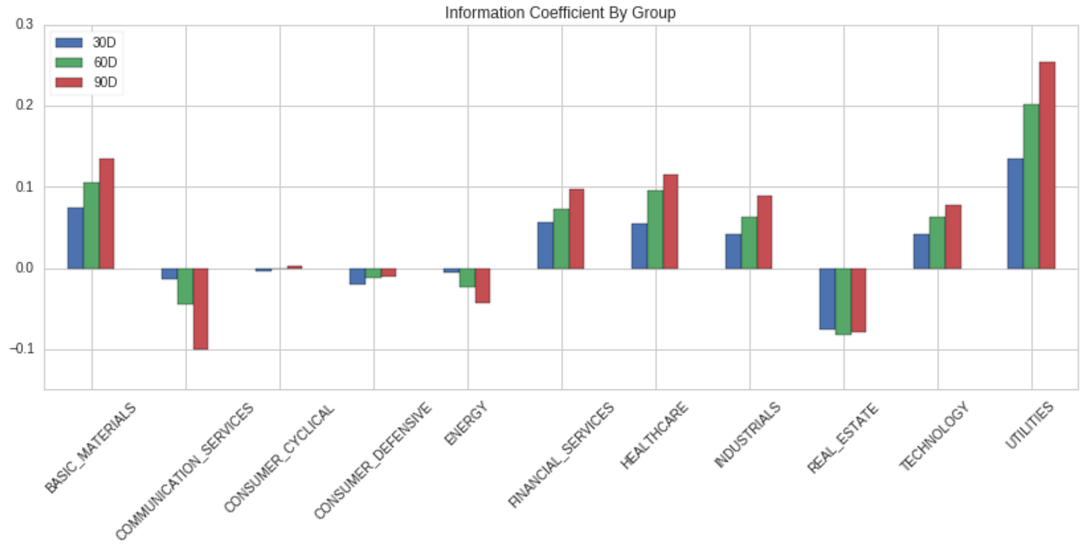
Wespath. (2016, October). *Did Your Asset Manager Choose an Appropriate Benchmark?* Retrieved from www.wespath.com: https://www.wespath.com/assets/1/16/4320.pdf

Appendices

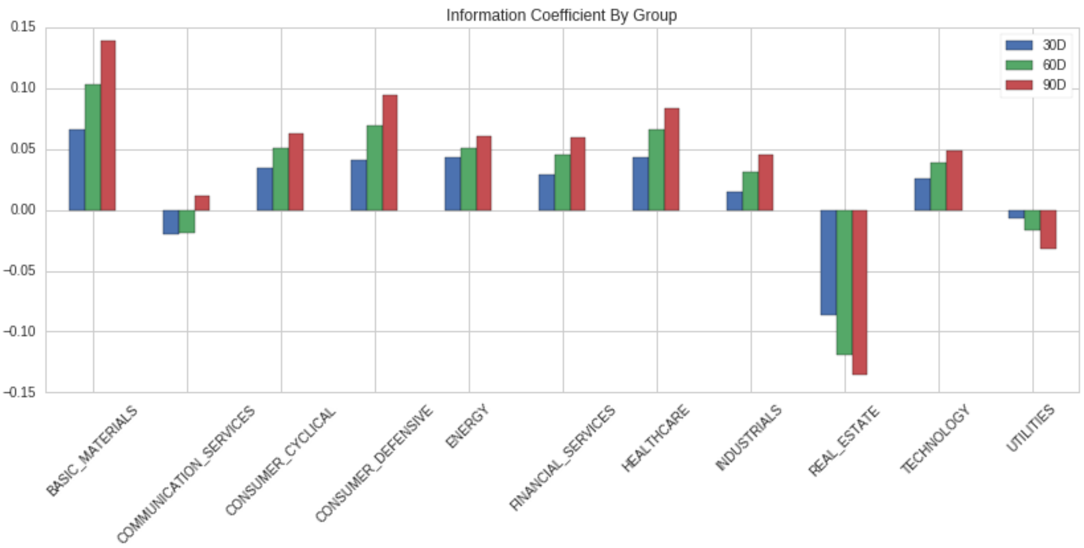
Appendix 1. HML: Information coefficient by sector (Pre-crisis)

Figures below represent the mean Information coefficient of the HML factor for the period between 01.01.2003 and 01.01.2007 as measured for 30 (blue), 60 (green) and 90 days (red) forward for each sector.

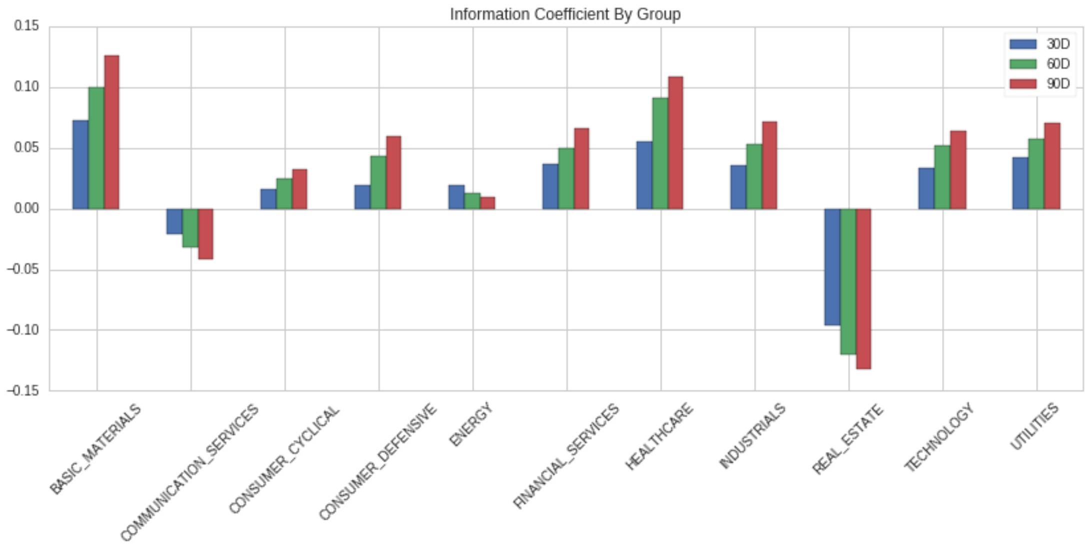
- Appendix 1a – Small market capitalisation (1904 firms)



- Appendix 1b – Large market capitalisation (1305 firms)



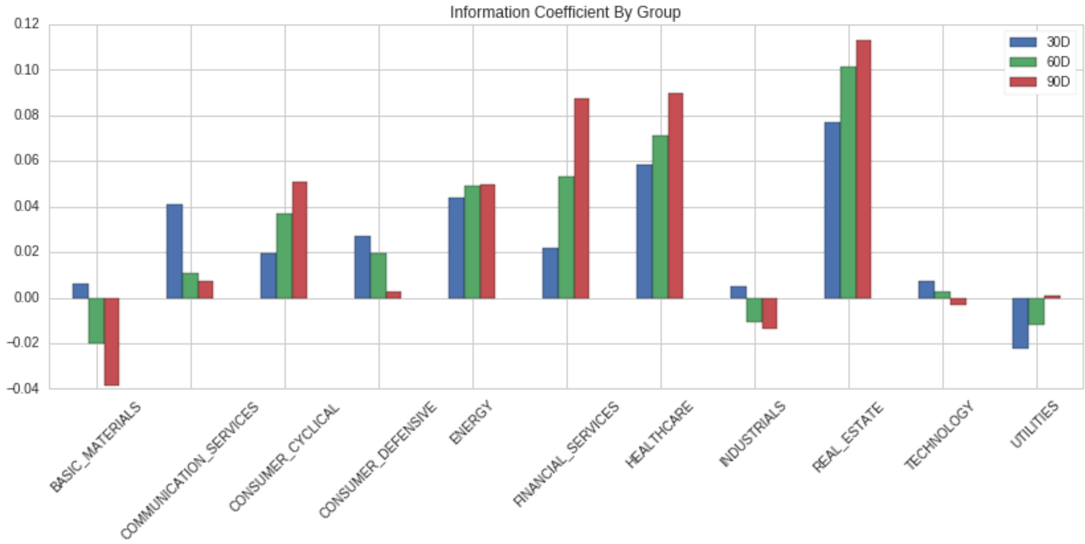
- Appendix 1c – No market capitalisation restrictions (2568 firms)



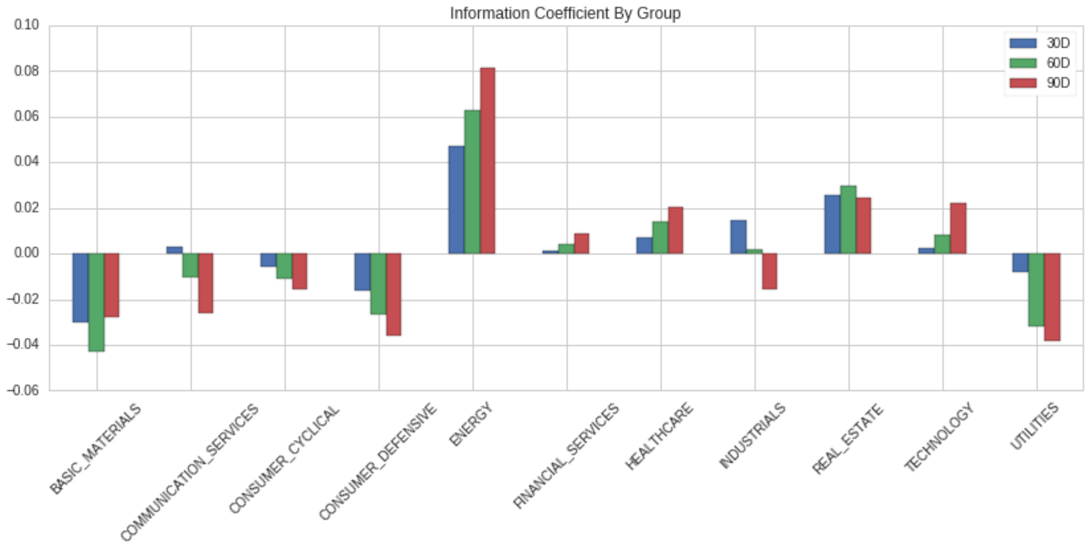
Appendix 2. RMW: Information coefficient by sector (Pre-crisis)

Figures below represent the mean Information coefficient of the RMW factor for the period between 01.01.2003 and 01.01.2007 as measured for 30 (blue), 60 (green) and 90 days (red) forward for each sector.

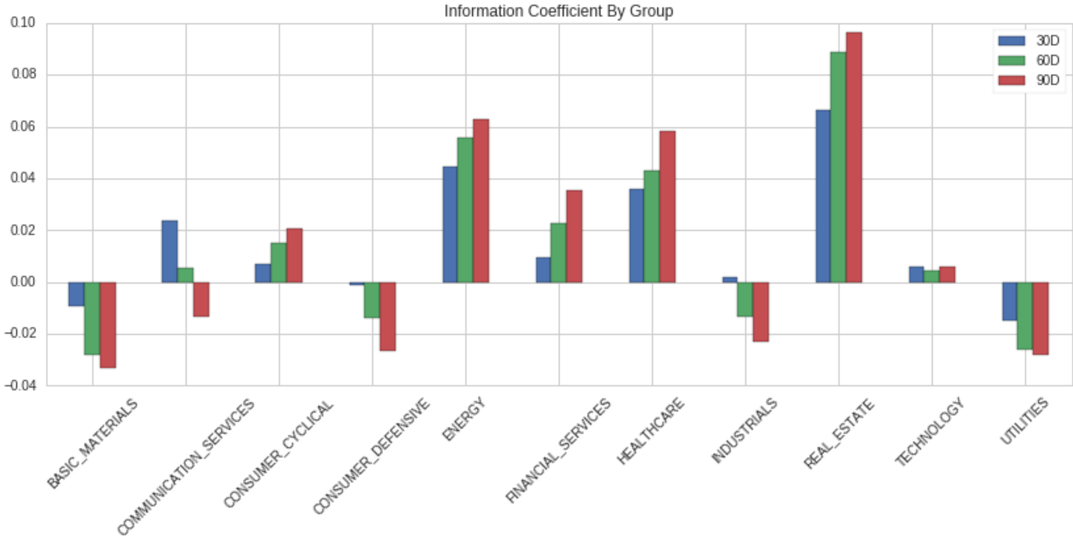
- Appendix 2a – Small market capitalisation (1738 firms)



- Appendix 2b – Large market capitalisation (1192 firms)



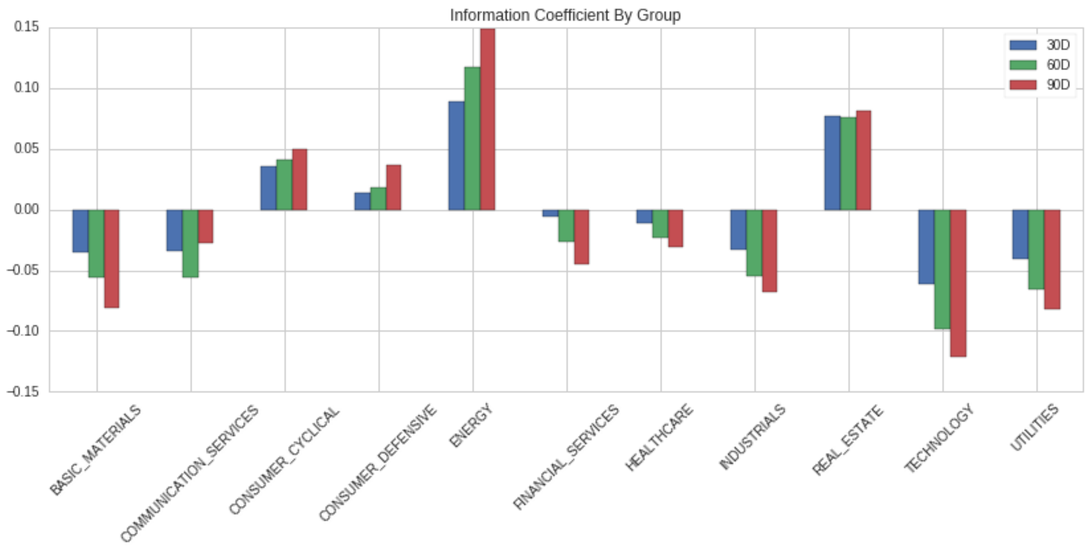
- Appendix 2c – No market capitalisation restrictions (2347 firms)



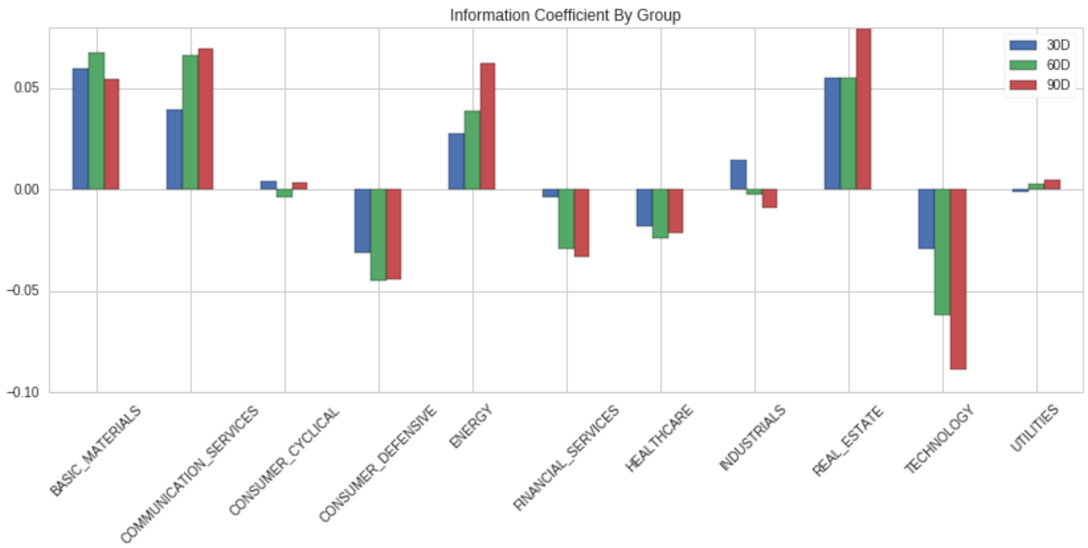
Appendix 3. UMD: Information coefficient by sector (Pre-crisis)

Figures below represent the mean Information coefficient of the UMD factor for the period between 01.01.2003 and 01.01.2007 as measured for 30 (blue), 60 (green) and 90 days (red) forward for each sector.

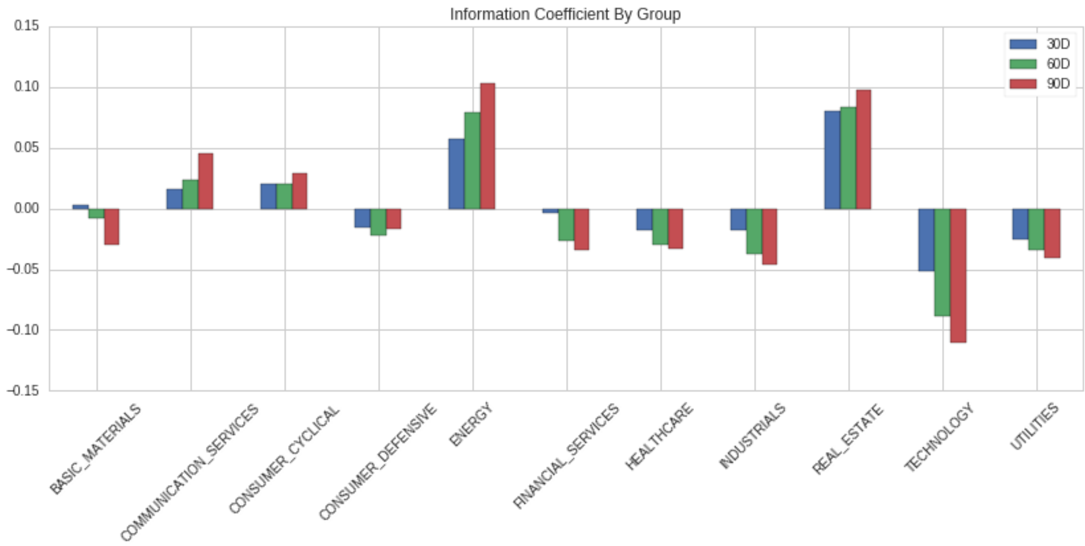
- Appendix 3a – Small market capitalisation (1898 firms)



- Appendix 3b – Large market capitalisation (1313 firms)



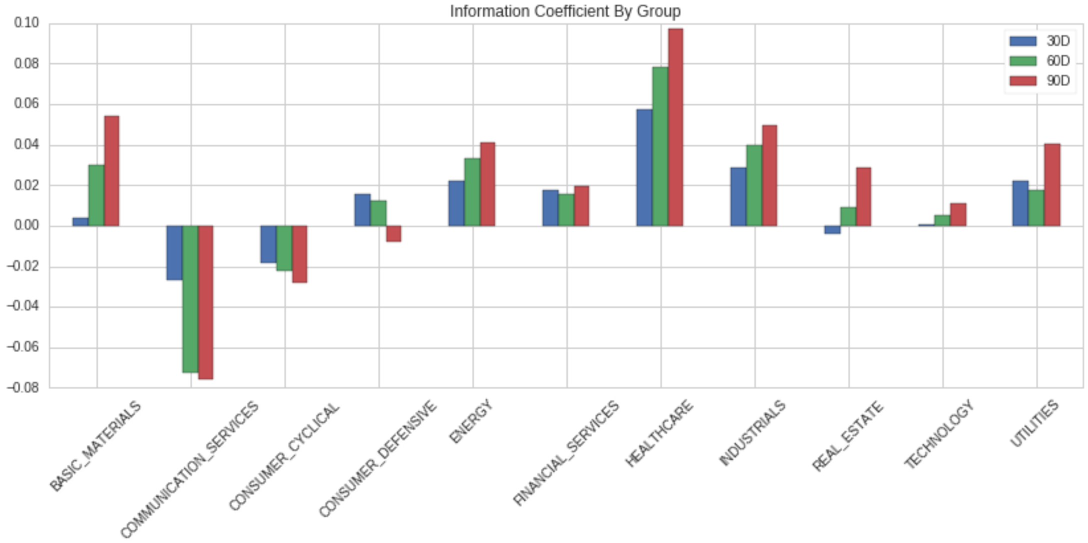
- Appendix 3c – No market capitalisation restrictions (2583 firms)



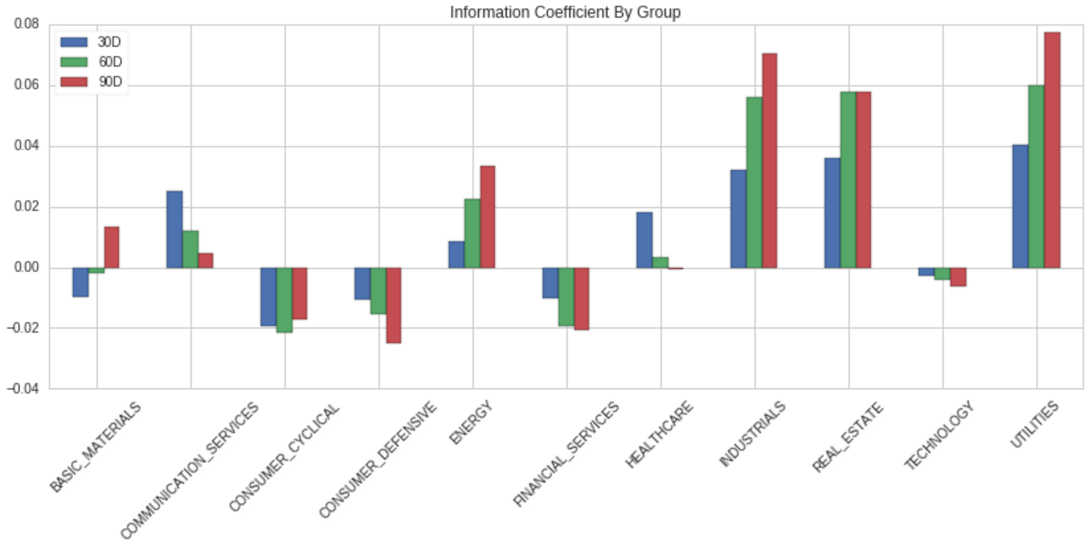
Appendix 4. HML: Information coefficient by sector (Crisis)

Figures below represent the mean Information coefficient of the HML factor for the period between 01.01.2007 and 01.01.2010 as measured for 30 (blue), 60 (green) and 90 days (red) forward for each sector.

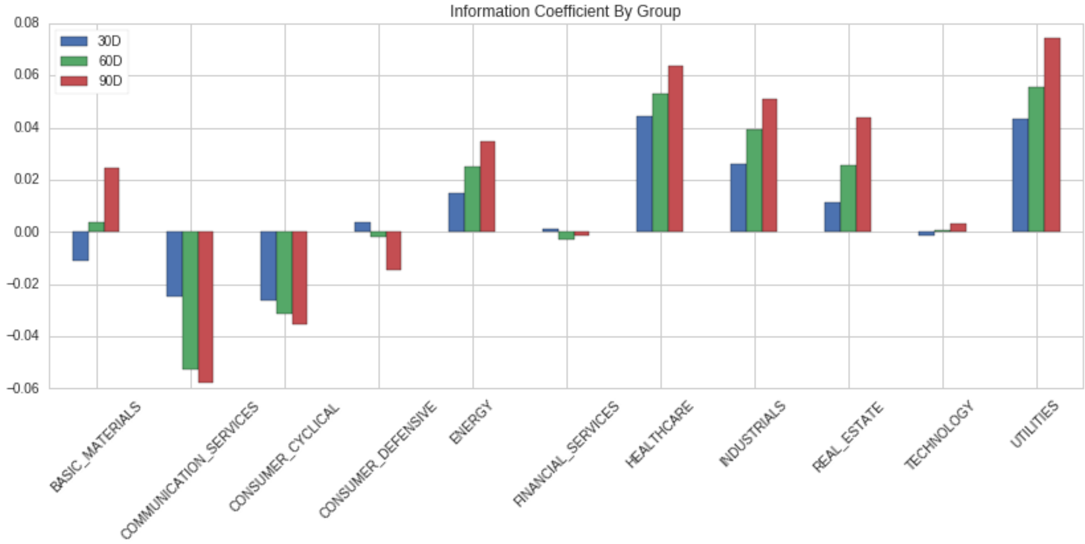
- Appendix 4a – Small market capitalisation (1909 firms)



- Appendix 4b – Large market capitalisation (1392 firms)



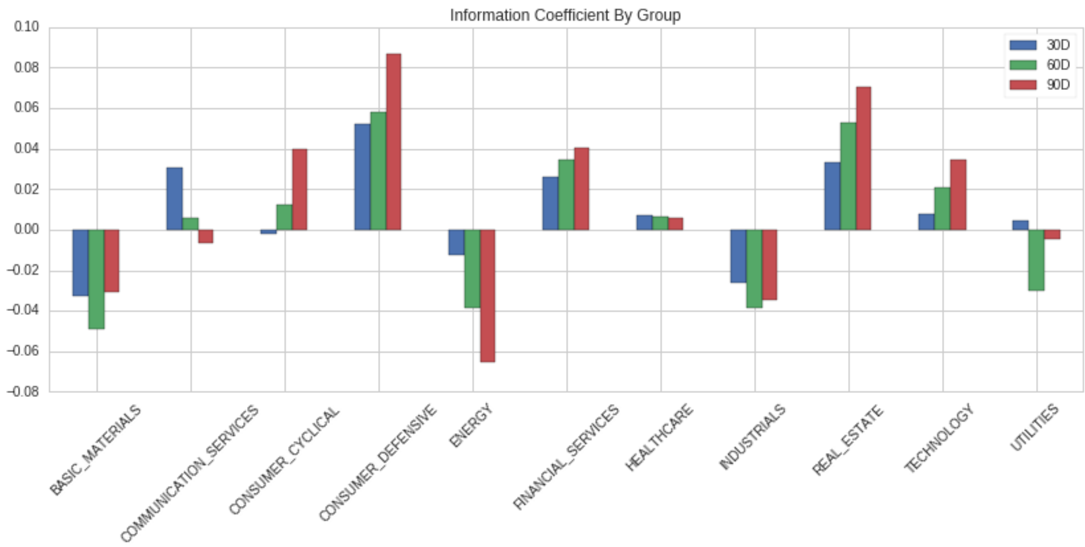
- Appendix 4c – No market capitalisation restrictions (2660 firms)



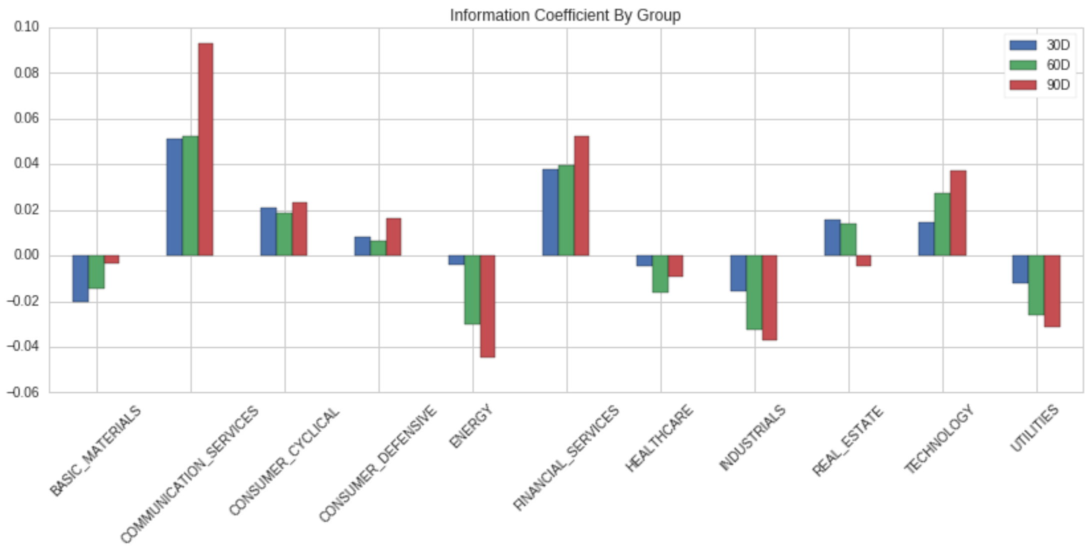
Appendix 5. RMW: Information coefficient by sector (Crisis)

Figures below represent the mean Information coefficient of the RMW factor for the period between 01.01.2007 and 01.01.2010 as measured for 30 (blue), 60 (green) and 90 days (red) forward for each sector.

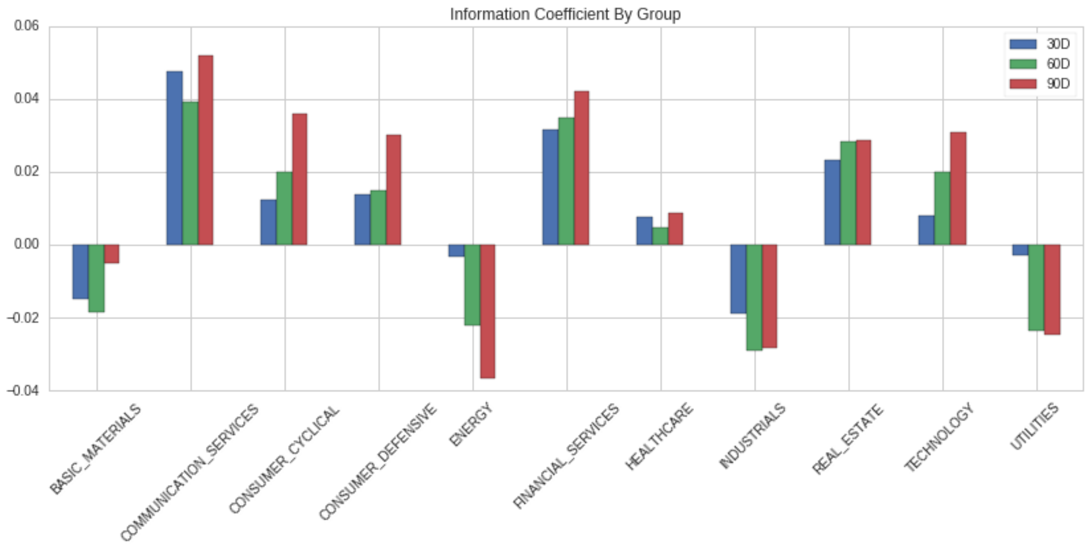
- Appendix 5a – Small market capitalisation (1805 firms)



- Appendix 5b – Large market capitalisation (1335 firms)



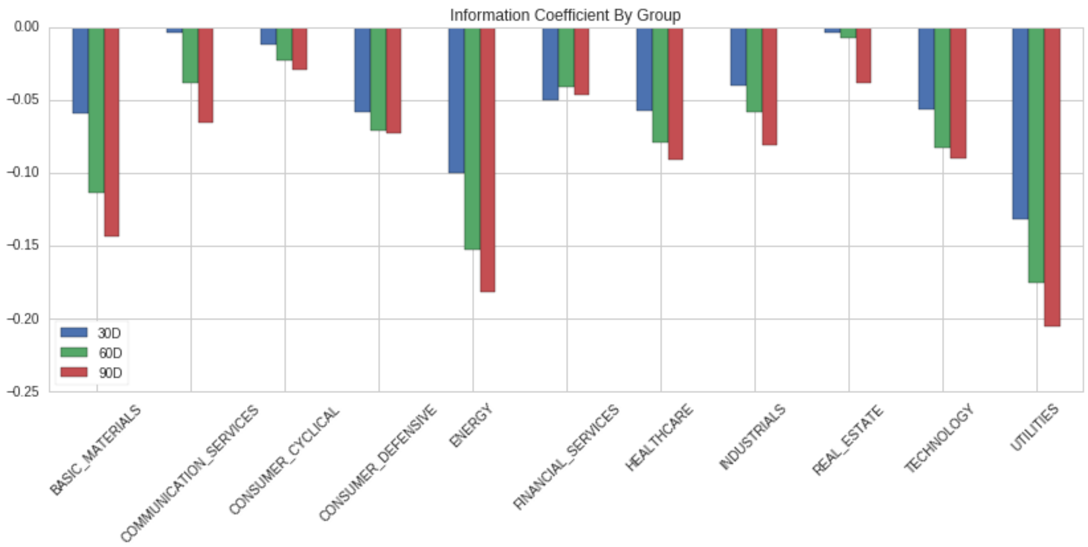
- Appendix 5c – No market capitalisation restrictions (2528 firms)



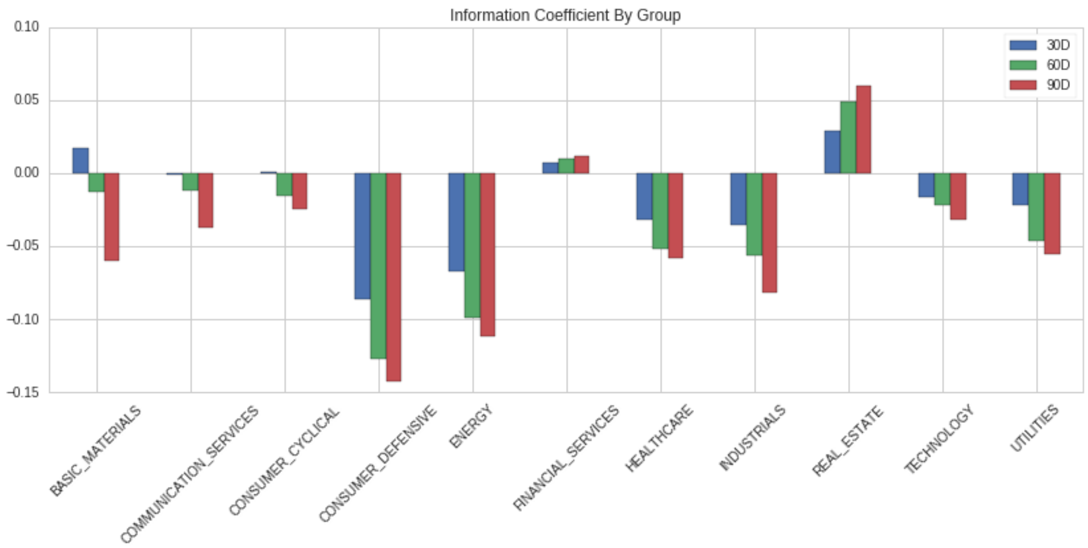
Appendix 6. UMD: Information coefficient by sector (Crisis)

Figures below represent the mean Information coefficient of the UMD factor for the period between 01.01.2007 and 01.01.2010 as measured for 30 (blue), 60 (green) and 90 days (red) forward for each sector.

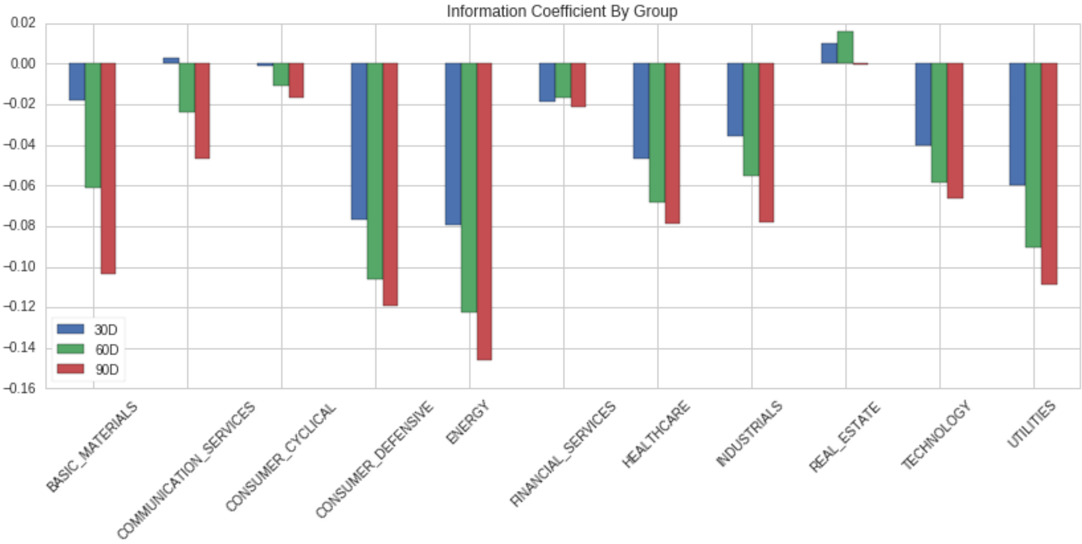
- Appendix 6a – Small market capitalisation (1992 firms)



- Appendix 6b – Large market capitalisation (1413 firms)



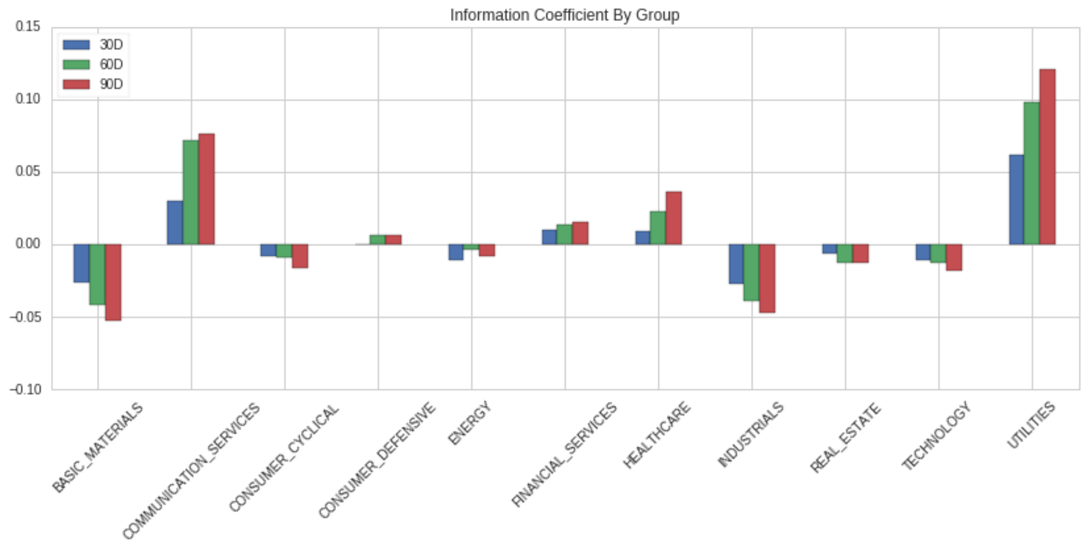
- Appendix 6c – No market capitalisation restrictions (2689 firms)



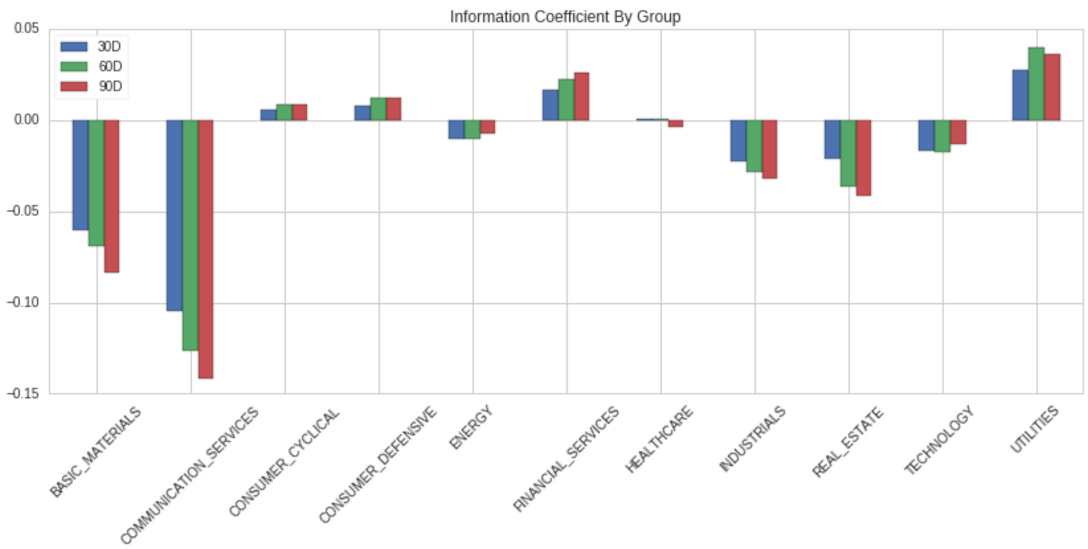
Appendix 7. HML: Information coefficient by sector (Post-crisis)

Figures below represent the mean Information coefficient of the HML factor for the period between 01.01.2010 and 01.01.2018 as measured for 30 (blue), 60 (green) and 90 days (red) forward for each sector.

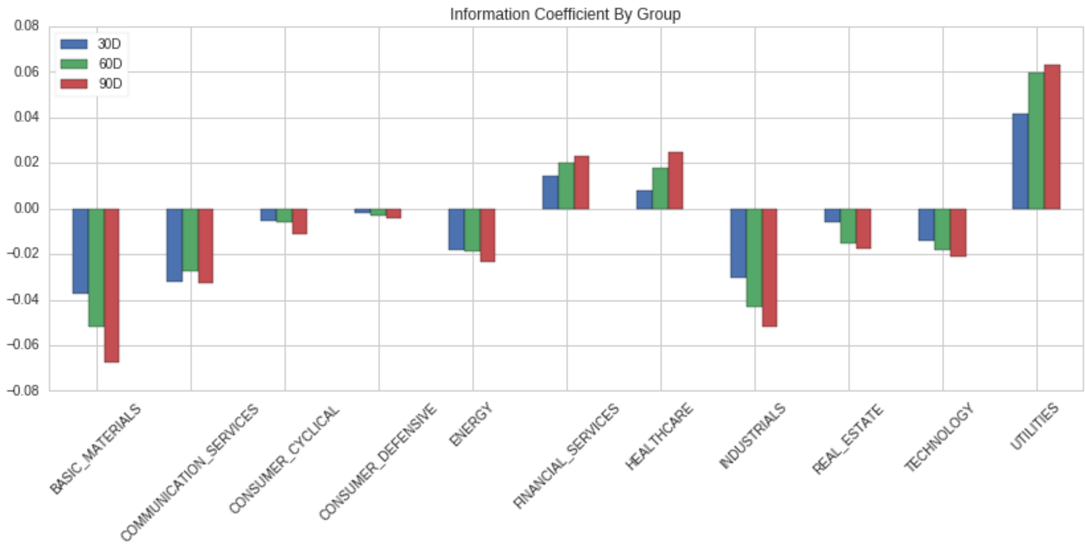
- Appendix 7a – Small market capitalisation (2529 firms)



- Appendix 7b – Large market capitalisation (1750 firms)



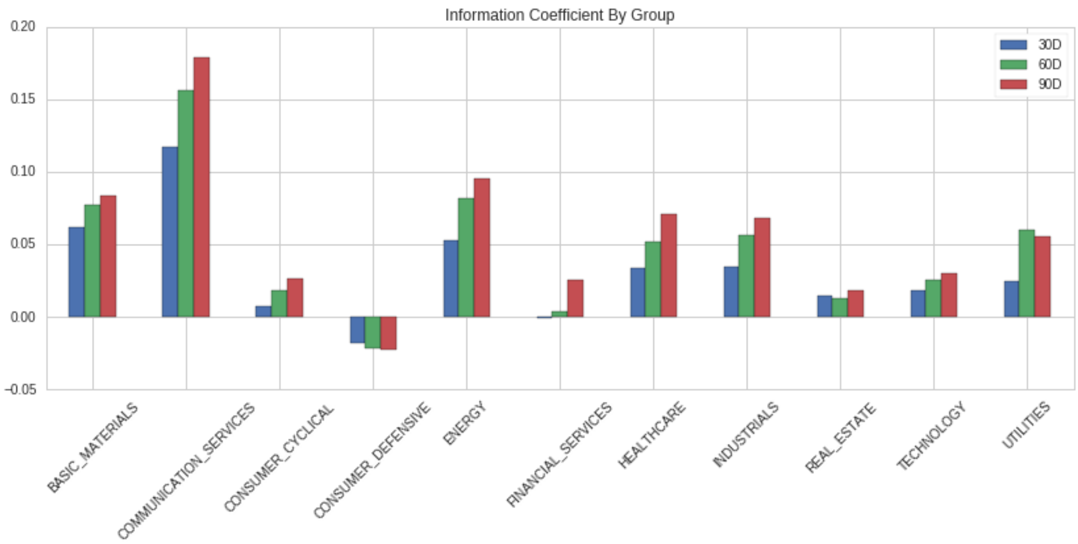
- Appendix 7c – No market capitalisation restrictions (3320 firms)



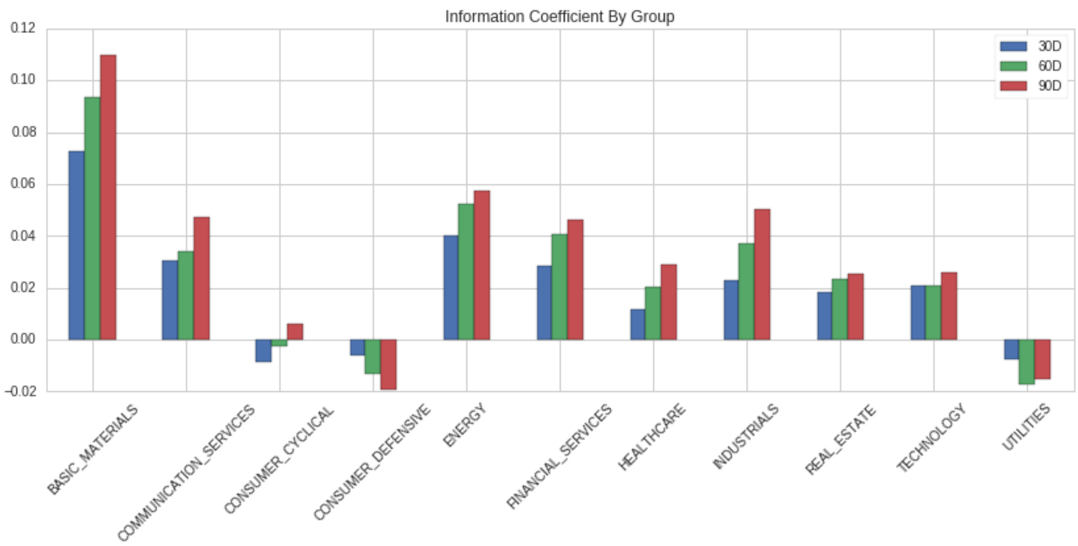
Appendix 8. RMW: Information coefficient by sector (Post-crisis)

Figures below represent the mean Information coefficient of the RMW factor for the period between 01.01.2010 and 01.01.2018 as measured for 30 (blue), 60 (green) and 90 days (red) forward for each sector.

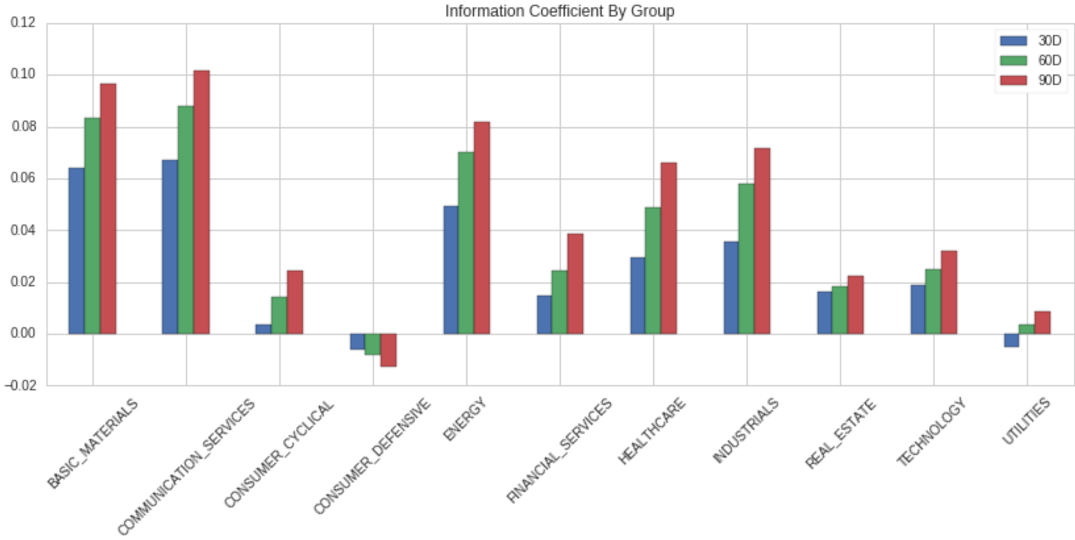
- Appendix 8a – Small market capitalisation (2363 firms)



- Appendix 8b – Large market capitalisation (1631 firms)



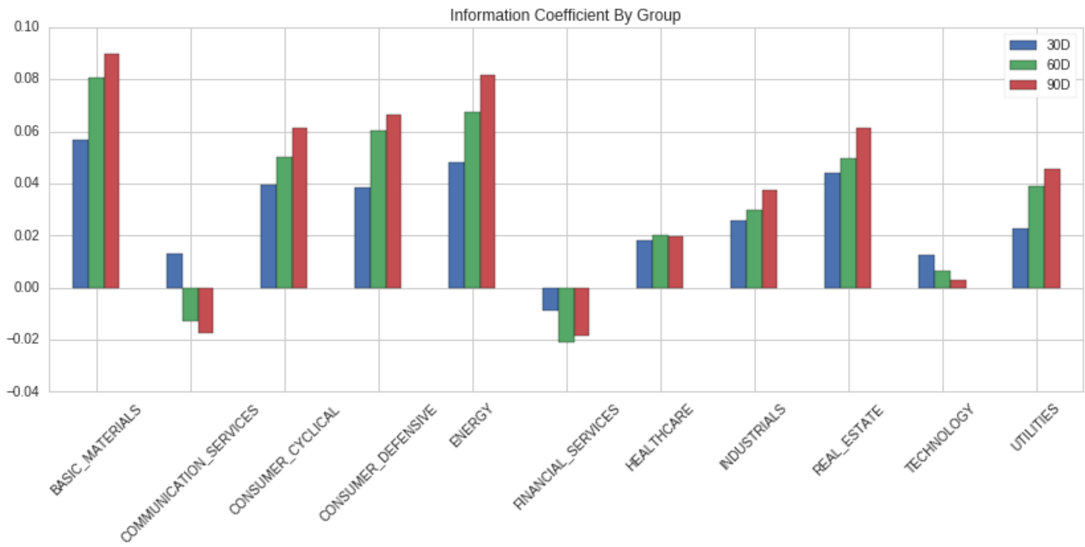
- Appendix 8c – No market capitalisation restrictions (3125 firms)



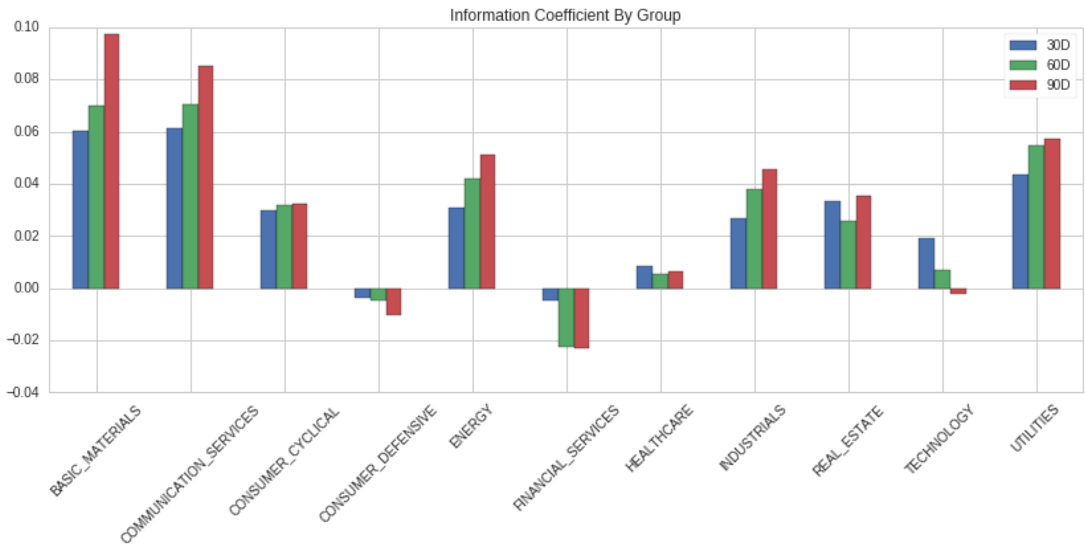
Appendix 9. UMD: Information coefficient by sector (Post-crisis)

Figures below represent the mean Information coefficient of the UMD factor for the period between 01.01.2010 and 01.01.2018 as measured for 30 (blue), 60 (green) and 90 days (red) forward for each sector.

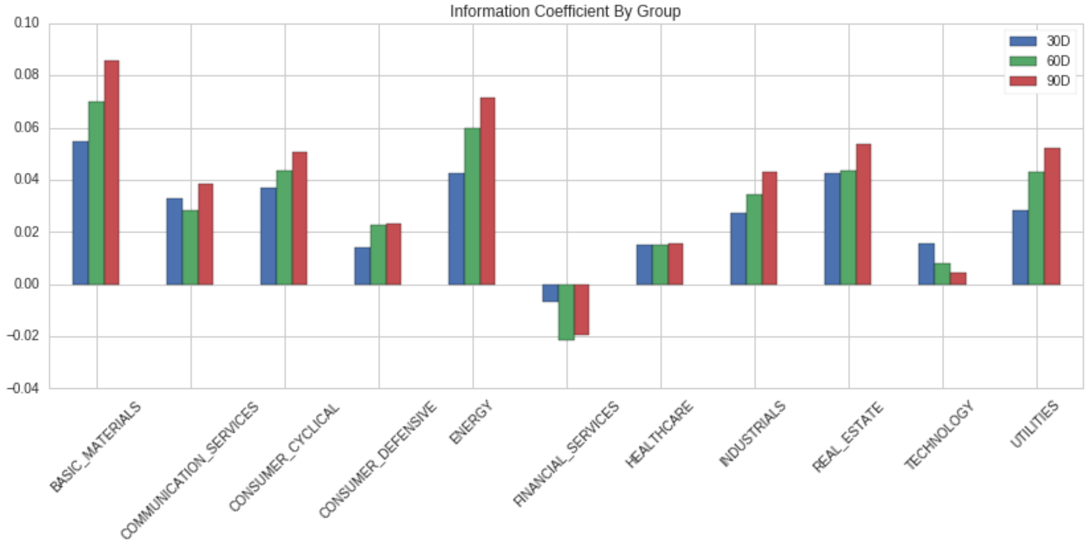
- Appendix 9a – Small market capitalisation (2527 firms)



- Appendix 9b – Large market capitalisation (1746 firms)



- Appendix 9c – No market capitalisation restrictions (3345 firms)



Appendix 10. Definition of the HML factor

```
# The HML factor is defined inside a function initialize
# expected to be defined by default in any Quantopian algorithm.
def initialize(context):

    # Commission is set to be $0.005 per share and $1 per trade.
    set_commission(us_equities=commission.PerShare(cost=0.005,
                                                   min_trade_cost=1))

    # Exchange code of a firm.
    exchange = mstar.share_class_reference.exchange_id.latest

    # A filter rule is created that returns True only for
    # the stocks from the exchanges listed.
    my_exchanges = exchange.element_of(['NYSE','NYS','NAS','ASE'])

    # Total equity of a firm as per latest balance sheet.
    total_equity = mstar.balance_sheet.total_equity.latest

    # Market capitalisation of a firm.
    market_cap = MarketCap()

    # Sector code of a firm.
    sector = Sector()

    # The trading universe is defined as QTradableStocksUS that falls into
    # my_exchanges and has data for total_equity, market_cap and sector.
    universe_exchange = QTradableStocksUS() & my_exchanges
                        & total_equity.notnull() & market_cap.notnull()
                        & sector.notnull()

    # Small and large market cap groups specified as percentile.
    small = (MarketCap(mask=universe_exchange).percentile_between(0, 50))
    large = (MarketCap(mask=universe_exchange).percentile_between(50, 100))

    # Here the universe redefined as universe_exchange that falls into either
    # small or large market cap group as defined above.
    universe = universe_exchange & small

    # Book to market is defined as total_equity divided by the market_cap.
    book_to_market = total_equity / market_cap

    # Book to market values are normalised and ranked in an ascending order.
    book_to_market_rank = book_to_market.rank(ascending=True, mask=universe)
    factor = book_to_market_rank.demean()

    # The Pipeline object is defined and filled with the data defined above.
    pipe = Pipeline(
        columns={
            'alpha': factor,
            'bm': book_to_market,
            'exchange': exchange,
            'market_cap': market_cap,
            'sector': Sector(),
        },
        # Screen out all the data points lacking any of the specified values.
        screen = universe & factor.notnull() & Sector().notnull(),
    )
```
Note: the fragment of code above is the part of another code fragment. For the full reference see Appendix 14.

Appendix 11. Definition of the RMW factor

```
# The RMW factor is defined inside a function initialize
# expected to be defined by default in any Quantopian algorithm.
def initialize(context):

# Commission is set to be $0.005 per share and $1 per trade.
set_commission(us_equities=commission.PerShare(cost=0.005,
                                               min_trade_cost=1))

    # Exchange code of a firm.
    exchange = mstar.share_class_reference.exchange_id.latest

    # A filter rule is created that returns True only for
    # the stocks from the exchanges listed.
    my_exchanges = exchange.element_of(['NYSE','NYS','NAS','ASE'])

    # Defining total_equity, operating_income and interest_expense as
    # corresponding values in the latest income statement and balance sheet.
    operating_income = mstar.income_statement.operating_income.latest
    interest_expense = mstar.income_statement.interest_expense.latest
    total_equity = mstar.balance_sheet.total_equity.latest

    # Market capitalisation of a firm.
    market_cap = MarketCap()

    # Sector code of a firm.
    sector = Sector()

    # The trading universe is defined as QTradableStocksUS that falls into
    # my_exchanges and has data for operating_income, interest_expense,
    # total_equity, market_cap and sector.
    universe_exchange = QTradableStocksUS() & my_exchanges
                        & operating_income.notnull()
                        & interest_expense.notnull()
                        & total_equity.notnull()
                        & market_cap.notnull()
                        & sector.notnull()

    # Small and large market cap groups specified as percentile.
    small = (MarketCap(mask=universe_exchange).percentile_between(0, 50))
    large = (MarketCap(mask=universe_exchange).percentile_between(50, 100))

    # Here the universe redefined as universe_exchange that falls into either
    # small or large market cap group as defined above.
    universe = universe_exchange & large

    # Operating profitability ratio is defined as operating_income subtracted
    # interest_expense divided by the total_equity.
    op_ratio = (operating_income - interest_expense) / total_equity

    # OP ratio values are normalised and ranked in an ascending order.
    op_ratio_rank = op_ratio.rank(ascending=True, mask=universe)
    factor = op_ratio_rank.demean()

    # The Pipeline object is defined and filled with the data defined above.
    pipe = Pipeline(
        columns={
            'alpha': factor,
            'op': op_ratio,
            'exchange': exchange,
            'market_cap': market_cap,
            'sector': Sector(),
        },
        # Screen out all the data points lacking any of the specified values.
        screen = universe & factor.notnull() & Sector().notnull(),
    )
```

Note: the fragment of code above is the part of another code fragment. For the full reference see Appendix 14.

Appendix 12. Definition of the UMD factor

```
# Momentum is defined as the return of a security over the period of the
# last 11 months with 1-month gap between the end of the 11th month and today.
class Momentum(CustomFactor):
    inputs = [USEquityPricing.close]
    window_length = 252

    def compute(self, today, assets, out, close):
        out[:] = close[-20] / close[0]

# The function initialize is expected to be defined
# by default in any Quantopian algorithm.
def initialize(context):

    # Commission is set to be $0.005 per share and $1 per trade.
    set_commission(us_equities=commission.PerShare(cost=0.005,
                                                   min_trade_cost=1))

    # Exchange code of a firm.
    exchange = mstar.share_class_reference.exchange_id.latest

    # A filter rule is created that returns True only for
    # the stocks from the exchanges listed.
    my_exchanges = exchange.element_of(['NYSE','NYS','NAS','ASE'])

    # umd is defined as price momentum.
    umd = Momentum()

    # Market capitalisation of a firm.
    market_cap = MarketCap()

    # Sector code of a firm.
    sector = Sector()

    # The trading universe is defined as QTradableStocksUS that falls into
    # my_exchanges and has data for umd, market_cap and sector
    universe_exchange = QTradableStocksUS() & my_exchanges
                        & umd.notnull() & market_cap.notnull()
                        & sector.notnull()

    # Small and large market cap groups specified as percentile.
    small = (MarketCap(mask=universe_exchange).percentile_between(0, 50))
    large = (MarketCap(mask=universe_exchange).percentile_between(50, 100))

    # Here the universe redefined as universe_exchange that falls into either
    # small or large market cap group as defined above.
    universe = universe_exchange & large

    # Price momentum values are ranked and normalised in an ascending order.
    momentum = umd.rank(ascending=True, mask=universe)
    factor = momentum.demean()

    # The Pipeline object is defined and filled with the data defined above.
    pipe = Pipeline(
        columns={
            'alpha': factor,
            'momentum': umd,
            'exchange': exchange,
            'market_cap': market_cap,
            'sector': Sector(),
        },
        # Screen out all the data points lacking any of the specified values.
        screen = universe & factor.notnull() & Sector().notnull(),
    )
```

Note: the fragment of code above is the part of another code fragment. For the full reference see Appendix 14.

Appendix 13. Definition of a combined factor

```
# Momentum is defined as the return of a security over the period of the
# last 11 months with 1-month gap between the end of the 11th month and today.
class Momentum(CustomFactor):
    inputs = [USEquityPricing.close]
    window_length = 252

    def compute(self, today, assets, out, close):
        out[:] = close[-20] / close[0]

# The function initialize is expected to be defined
# by default in any Quantopian algorithm.
def initialize(context):

    # Commission is set to be $0.005 per share and $1 per trade.
    set_commission(us_equities=commission.PerShare(cost=0.005,
                                                   min_trade_cost=1))

    # Exchange code of a firm.
    exchange = mstar.share_class_reference.exchange_id.latest

    # A filter rule is created that returns True only for
    # the stocks from the exchanges listed.
    my_exchanges = exchange.element_of(['NYSE','NYS','NAS','ASE'])

    # Market capitalisation, sector code and momentum of a firm.
    market_cap = MarketCap()
    sector = Sector()
    umd = Momentum()

    # Defining total_equity, operating_income and interest_expense as
    # corresponding values in the latest income statement and balance sheet.
    total_equity = mstar.balance_sheet.total_equity.latest
    operating_income = mstar.income_statement.operating_income.latest
    interest_expense = mstar.income_statement.interest_expense.latest

    # The trading universe is defined as QTradableStocksUS that falls into
    # my_exchanges and has data for umd, total_equity, operating_income,
    # interest_expense, market_cap and sector.
    universe_exchange = QTradableStocksUS() & umd.notnull()
                        & my_exchanges & total_equity.notnull()
                        & market_cap.notnull() & sector.notnull()
                        & operating_income.notnull()
                        & interest_expense.notnull()

    # Small and large market cap groups specified as percentile.
    small = (MarketCap(mask=universe_exchange).percentile_between(0, 50))
    large = (MarketCap(mask=universe_exchange).percentile_between(50, 100))

    # Here the universe redefined as universe_exchange that falls into either
    # small or large market cap group as defined above.
    universe = universe_exchange & large

    # Book to market is defined as total_equity divided by the market_cap.
    # The value is normalised and ranked in an ascending order.
    bm = total_equity / market_cap
    bm_rank = bm.rank(ascending=True, mask=universe)

    # Operating profitability ratio is defined as operating_income subtracted
    # interest_expense divided by the total_equity.
    # The value is normalised and ranked in an ascending order.
    op = (operating_income - interest_expense) / total_equity
    op_rank = op.rank(ascending=True, mask=universe)

    # Price momentum values are ranked and normalised in an ascending order.
    umd_rank = umd.rank(ascending=True, mask=universe)
```

```python

    # A class JoinFactors is defined that is used to combine the normalised
    # scores of the factors defined above.
    class JoinFactors(CustomFactor):
        #inputs = [factor1, factor2, ...] There can be multiple inputs.
        window_length = 1

        def compute(self, today, assets, out, *inputs):
            array = np.concatenate(inputs, axis=0)
            out[:] = np.nansum(array, axis=0)
            out[ np.all(np.isnan(array), axis=0) ] = np.nan

    # window_safe declares that scores of the factors are robust to
    # pricing adjustments from splits or dividends. In other words,
    # the value that will be the same no matter what day you are
    # looking back from. This is a required step in order to
    # use them as the input to JoinFactors.
    bm_weights.window_safe = True
    op_weights.window_safe = True
    umd_weights.window_safe = True

    # The weights of the combined factor. 1, 2, 3 or more factors can be used.
    final_weights = JoinFactors(inputs=[bm_weights, op_weights, umd_weights],
                                mask=universe)

    # Redefining the universe as universe with the items
    # where combined factor weights are present.
    universe = final_weights.notnan()

    # The Pipeline object is defined and filled with the data defined above.
    pipe = Pipeline(
        columns={
                 'umd_rank': umd_rank,
                 'bm_rank': bm_rank,
                 'op_rank': op_rank,
                 'umd_weights': umd_weights,
                 'bm_weights': bm_weights,
                 'op_weights': op_weights,
                 'alpha': final_weights,
                 'exchange': exchange,
                 'market_cap': market_cap,
                 'sector': sector,
                },
        # Screen out all the data points outside the trading universe.
        screen = universe
    )
```

Note: the fragment of code above is the part of another code fragment. For the full reference see Appendix 14.

Appendix 14. Code template implementing a factor strategy

```
# Importing objects, libraries and functions to be used in the algorithm.
import pandas as pd
import quantopian.algorithm as algo
import quantopian.experimental.optimize as opt
from quantopian.pipeline import Pipeline, CustomFactor
from quantopian.pipeline.data import builtin, morningstar as mstar
from quantopian.pipeline.factors.morningstar import MarketCap
from quantopian.pipeline.classifiers.morningstar import Sector
from quantopian.pipeline.filters import QTradableStocksUS
from quantopian.pipeline.data.builtin import USEquityPricing

# Constraint Parameters.
MAX_GROSS_LEVERAGE = 1.0
MAX_SHORT_POSITION_SIZE = 0.0 # 0.0%
MAX_LONG_POSITION_SIZE = 0.01 # 1.0%

# Scheduling Parameters. How long to wait before start after the market opens.
MINUTES_AFTER_MARKET = 10

'''
Definition of a custom factor such as Momentum happens here.
Refer to the appendix 12.
'''

# The function initialize is expected to be defined
# by default in any Quantopian algorithm.
def initialize(context):
    # To set a custom benchmark the following function can be called:
    # set_benchmark(symbol('IWV'))
    # Otherwise the default benchmark will be used (SPY).

    '''
    Factor definition logic goes in here. Refer to the appendices 10 – 13.
    '''

    # The function attach_pipeline is called
    # to load the data in defined in the pipeline.
    algo.attach_pipeline(pipe, 'pipe')

    # Schedule a function, 'do_portfolio_construction', to run once a month
    # ten minutes after market is open.
    algo.schedule_function(
        do_portfolio_construction,
        date_rule=algo.date_rules.month_start(),
        time_rule=algo.time_rules.market_open(minutes=MINUTES_AFTER_MARKET),
        half_days=False,
    )

# The function before_trading_start defines the logic
# that happens every time before the trading session begins.
# Here pipeline output is processed.
def before_trading_start(context, data):
    context.pipeline_data = algo.pipeline_output('pipe')

# Portfolio construction. Inside this function the strategy is expressed
# as a set of objectives and constraints.
def do_portfolio_construction(context, data):
    pipeline_data = context.pipeline_data
    todays_universe = pipeline_data.index

    # Objective here was to maximise alpha which is
    # our factor defined in the pipeline.
    objective = opt.MaximizeAlpha(pipeline_data.alpha)
```

```
64.
65.         # Constrain our gross leverage to 1.0 or less.
66.         # This means that the absolute value of our long and short positions
67.         # should not exceed the value of our portfolio.
68.         constrain_gross_leverage = opt.MaxGrossLeverage(MAX_GROSS_LEVERAGE)
69.
70.         # Constrain individual position size to no more than a fixed percentage
71.         # of our portfolio.
72.         constrain_pos_size = opt.PositionConcentration.with_equal_bounds(
73.             -MAX_SHORT_POSITION_SIZE,
74.             MAX_LONG_POSITION_SIZE,
75.         )
76.
77.        # Constrain ourselves to allocate the same amount of capital to
78.        # long and short positions. Not used in the simulations in this work.
79.        market_neutral = opt.DollarNeutral()
80.
81.        # Constrain the maximum average exposure
82.        # to individual sectors to -10% - 10%.
83.        sector_neutral = opt.NetPartitionExposure.with_equal_bounds(
84.             labels=pipeline_data.sector,
85.             min=-0.10,
86.             max=0.10,
87.        )
88.
89.        # Run the optimization.
90.        # This will calculate new portfolio weights and
91.        # manage moving our portfolio toward the target.
92.        algo.order_optimal_portfolio(
93.             objective=objective,
94.             constraints=[
95.                 constrain_gross_leverage,
96.                 constrain_pos_size,
97.                 # market_neutral, ---> not used in the study.
98.                 sector_neutral,
99.             ],
100.            universe=todays_universe,
101.    )
```

Note: the code template is dependent on the Zipline API and should be run within Quantopian Algorithms IDE. For testing one would want to sign up on www.quantopian.com and create a new blank algorithm. Then, the code above could be filled in with the factors defined in on the lines 20 – 23 and 32 – 34. To define the factors refer to Appendices 10 – 13. The full working factor strategy template with instructions could also be accessed on GitHub (URL: https://github.com/slazarevich/fama_french_quantopian).

Appendix 15. Factor analysis with Alphalens

The template below illustrates the source code used for generating the comprehensive factor analysis using the Alphalens API. The code should be viewed as a Jupyter Notebook file where `# In[1]:` like lines indicate notebook cells with the following after it code as input.

```
# In[1]:


# Importing objects, libraries and functions to be used in the notebook.
from quantopian.pipeline import Pipeline
from quantopian.research import run_pipeline
from quantopian.pipeline.filters import QTradableStocksUS
from quantopian.pipeline.data import morningstar, Fundamentals
from quantopian.pipeline.factors.morningstar import MarketCap
from quantopian.pipeline import CustomFactor

from quantopian.pipeline.data.builtin import USEquityPricing
from quantopian.pipeline.data import builtin, morningstar as mstar
from quantopian.pipeline.classifiers.fundamentals import Sector

from alphalens.utils import get_clean_factor_and_forward_returns
from alphalens.performance import mean_information_coefficient
from alphalens.tears import create_information_tear_sheet
from alphalens.tears import create_returns_tear_sheet

import numpy as np


# In[2]:


# Momentum is defined as the return of a security over the period of the
# last 11 months with 1month gap between the end of the 11th month and today.
class Momentum(CustomFactor):
    inputs = [USEquityPricing.close]
    window_length = 252

    def compute(self, today, assets, out, close):
        out[:] = close[-20] / close[0]

# We create a pipeline to define the factor(s).
def make_pipeline():

    # Exchange code of a firm.
    exchange = mstar.share_class_reference.exchange_id.latest

    # A filter rule is created that returns True only for
    # the stocks from the exchanges listed.
    my_exchanges = exchange.element_of(['NYSE','NYS','NAS','ASE'])

    # Market capitalisation, sector code and momentum of a firm.
    market_cap = MarketCap()
    sector = Sector()
    umd = Momentum()

    # Defining total_equity, operating_income and interest_expense as
    # corresponding values in the latest income statement and balance sheet.
    total_equity = mstar.balance_sheet.total_equity.latest
```

```
    operating_income = mstar.income_statement.operating_income.latest
    interest_expense = mstar.income_statement.interest_expense.latest

    # The trading universe is defined as QTradableStocksUS that falls into
    # my_exchanges and has data for umd, total_equity, operating_income,
    # interest_expense, market_cap and sector.
    universe_exchange = QTradableStocksUS() & umd.notnull()
                        & my_exchanges & total_equity.notnull()
                        & market_cap.notnull() & sector.notnull()
                        & operating_income.notnull()
                        & interest_expense.notnull()

    # Small and large market cap groups specified as percentile.
    small = (MarketCap(mask=universe_exchange).percentile_between(0, 50))
    large = (MarketCap(mask=universe_exchange).percentile_between(50, 100))

    # Create a filter that returns True for the assets in the universe
    # that belong to the given sector(s).
    sec = morningstar.asset_classification.morningstar_sector_code.latest
    my_sec = sec.element_of([101])

    # Here the universe redefined as universe_exchange that belongs
    # to the sector(s) in 'my_sec' and falls into either
    # small or large market cap group as defined above.
    # my_sec should be uncommented in case if a speficic sector is wanted.
    '''''
    Here are the sector codes that might be used:

     -1: 'Misc',
    101: 'Basic Materials',
    102: 'Consumer Cyclical',
    103: 'Financial Services',
    104: 'Real Estate',
    205: 'Consumer Defensive',
    206: 'Healthcare',
    207: 'Utilities',
    308: 'Communication Services',
    309: 'Energy',
    310: 'Industrials',
    311: 'Technology' ,
    '''
    universe = universe_exchange & small #& my_sec

    # Book to market is defined as total_equity divided by the market_cap.
    # The value is normalised and ranked in an ascending order.
    bm = total_equity / market_cap
    bm_weights = bm.rank(ascending=True, mask=universe)

    # Operating profitability ratio is defined as operating_income subtracted
    # interest_expense divided by the total_equity.
    # The value is normalised and ranked in an ascending order.
    op = (operating_income - interest_expense) / total_equity
    op_weights = op.rank(ascending=True, mask=universe)

    # Price momentum values are ranked and normalised in an ascending order.
    umd_weights = umd.rank(ascending=True, mask=universe)

    # A class JoinFactors is defined that is used to combine the normalised
    # scores of the factors defined above.
    class JoinFactors(CustomFactor):
        #inputs = [factor1, factor2, ...] There can be multiple inputs.
        window_length = 1

        def compute(self, today, assets, out, *inputs):
            array = np.concatenate(inputs, axis=0)
            out[:] = np.nansum(array, axis=0)
            out[ np.all(np.isnan(array), axis=0) ] = np.nan

```

```
    # window_safe declares that scores of the factors are robust to
    # pricing adjustments from splits or dividends. In other words,
    # the value that will be the same no matter what day you are
    # looking back from. This is a required step in order to
    # use them as the input to JoinFactors.
    bm_weights.window_safe = True
    op_weights.window_safe = True
    umd_weights.window_safe = True

    # The weights of the combined factor.
    # 1, 2, 3 or more factors can be used.
    final_weights = JoinFactors(inputs=[bm_weights, op_weights, umd_weights],
                                        mask=universe)
    universe = final_weights.notnan()

    # The Pipeline object filled with the data defined above is returned.
    return Pipeline(
        columns={
                 'bm_weights': bm_weights,
                 'op_weights': op_weights,
                 'umd_weights': umd_weights,
                 'final_weights':final_weights,
                 'exchange': exchange,
                 'market_cap': market_cap,
                 'sector': sector,
                },
        # Screen out all the data points outside the trading universe.
        screen = universe
    )

# Returning the data from the pipeline run over
# a period of time between 01.01.2003 and 01.01.2005.
factor_data = run_pipeline(make_pipeline(), '2003-1-1', '2005-1-1')

# We preload the pricing data for all assets in the universe
# for the time period in question.
# In order to run the analysis properly it is recommended that
# for the pricing data to cover the period in factor_data with
# at least 1 month before the start of factor_data and a few more
# months after the factor_data.
# It is also recommended to use "open_price" as the fields parameter.
pricing_data = get_pricing(factor_data.index.levels[1], '2002-12-1',
                           '2005-6-1', fields='open_price')


# In[3]:


#Show top 30 lines of the factor_data.
factor_data.head(30)


# In[4]:


# Count the unique assets in factor_data.
assets = factor_data.index.levels[1].unique()
len(assets)


# In[5]:


# Plot the mean information coefficient decay.
# The graph allows to see how long does the predicting power last.
longest_look_forward_period = 126 # week = 5, month = 21,
                                  # quarter = 63, year = 252
range_step = 5
```

```
merged_data = get_clean_factor_and_forward_returns(
    # Here the factor of choice should be specified.
    # For example, 'bm_weights',
    # ‘op_weights’, ‘umd_weights’ or ‘final_weights’.
    factor=factor_data[‘bm_weights’],
    prices=pricing_data,
    periods=range(1, longest_look_forward_period, range_step)
)

mean_information_coefficient(merged_data).plot(title=”IC Decay”)

# In[6]:

# This cell generates a set of descriptive statistics
# about the factor as well as a few useful graphs.
sector_labels, sector_labels[-1] = dict(Sector.SECTOR_NAMES), "Unknown"

merged_data = get_clean_factor_and_forward_returns(
    # Here the factor of choice should be referred.
    # For example, 'bm_weights',
    # 'op_weights', 'umd_weights' or 'final_weights'.
    factor=factor_data['bm_weights'],
    prices=pricing_data,
    groupby=factor_data['sector'],
    groupby_labels=sector_labels,
    binning_by_group=True,
    # Custom forward returns periods could be selected here.
    periods=(30,60,90)
)

create_information_tear_sheet(merged_data, by_group=True, group_neutral=True)
create_returns_tear_sheet(merged_data, by_group=True, group_neutral=True)
```

Note: the code template is dependent on the Alphalens API and should be run within Quantopian Notebooks IDE. For testing one would want to sign up on www.quantopian.com and create a new notebook. Then, the code above could be filled in (In[*] lines should be treated as breaks between the cells of the notebook. To define the factors, refer to Appendices 10 – 13. The template could be imported directly to the Quantopian Notebook research with the instructions provided on GitHub (URL: https://github.com/slazarevich/fama_french_quantopian).