BI Norwegian Business School - campus Oslo

GRA 19502

Master Thesis

Component of continuous assessment: Thesis Master of Science
Final master thesis – Counts 80% of total grade

The Five-Factor Asset Pricing Model: A Corporate Finance Point of View

Navn: Dino Priganica, Ikram Ul-Haq Rashid

Start: 02.03.2018 09.00

Finish: 03.09.2018 12.00

Dino Priganica
Ikram Ul-Haq Rashid

BI Norwegian Business School

- Master Thesis -

- The Five-Factor Asset Pricing Model: A Corporate Finance Point of View -

Submission date:
01.09.2018

Supervisor:
Paul Ehling

Examination Code:
GRA 19502 – Master Thesis

Study Program:
Master of Science in Business – Major in Finance

Table of Contents

Abstract

Derivations of the 'Terminal Value of Free Cash Flow'-formula (Koller & Goedhart, 2015) show that the return on investment capital is a key driver of firm value. This implies that the investment factors from well-established asset pricing models might be mis-specified since they view the absolute level of investment undertaken as the only source of risk related to investments. This thesis suggests that there might be additional risk related to the return on the investments undertaken by firm's, which is left un-captured by five-factor asset pricing models, e.g. the Fama French Five-Factor Model (Fama & French, 2015) and the Empirical Q-Factor Model (Hou, Xue and Zhang, 2015). Our results suggest that investors should receive additional compensation for value added or destroyed through a firm`s investment policy and not only according to absolute size of investment. However, the empirical performance of our models is poor, and our results are limited to test asset portfolios based on Size-B/M.

1.0 Introduction

Ever since William Sharpe (1964), John Lintner (1965), and Fischer Black (1972) introduced the widely known capital asset pricing model (CAPM), researchers and scholars have been trying to explain the relationship between the expected return and systematic risk for financial securities. In one of their very prominent papers, Fama and French (1996) illustrates that a three-factor model, consisting of the market factor, a factor based on market equity labelled Size (SMB) and a factor based on book-to-market equity labelled as the Value factor (HML), gives a comprehensive summary of the cross-sectional average stock returns as of the mid-1990s. However, the three-factor model has over the past two decades been failing to account for a broad range of newly discovered asset pricing anomalies. The two most notable anomalies, which the three-factor model fails to capture, were based on investment and profitability. These were shown to be asset pricing anomalies in Novy-Marx (2013) and Aharoni, Grundy, and Zeng (2013). Fama and French (2015) sets out to include these new factors into a new five-factor asset pricing model and uses the well-known dividend discount model of Miller and Modigliani (1961) to provide an economically sound explanation as to why investment and profitability should have a statistically significant impact on asset returns and how factors based on these two variables can help model variations in asset returns. This thesis suggests an alternative approach to the identification of the investment and profitability factors, based on the 'Terminal Value of Free Cash Flow'-approach to valuation. We agree that the approach applied in Fama and French (1996) correctly identifies the profitability factor but argue that the investment factor might be mispecified as it potentially fails to model variations in asset returns caused by the actual returns of the investments undertaken by the firm. Using the 'Terminal Value of Free Cash Flow'-approach, we show that the return on invested capital is a key value driver of firm value and should thus have a significant impact on the variation in firm value over time, and hence also variations in asset returns. In an attempt to capture this variation and potentially improve the specification of five-factor asset pricing models, we create a factor based on the return on invested capital, which we name EMI. We define EMI as the difference between the return on a portfolio of efficient return on invested capital firms and the return on a portfolio of inefficient return on invested capital firms. In both our empirical models, we replace the original investment factors with this newly created factor, EMI. Hence, the research question of this thesis is as follows.

''Can a factor based on the return on invested capital replace the traditional investment factors and increase the performance of traditional asset pricing models?''

To test our hypothesis that the inclusion of EMI and elimination of the original investment factors can increase the performance of the five-factor asset pricing models, using the FF five-factor model as a benchmark, we rely heavily on the study of intercepts. We base the study on the notion that if a model is perfectly specified and explains all the variation in asset returns, the intercept of the estimated model should be equivalent to zero (see Gibbons, Ross and Shanken (1989), Cochrane (2005) and Barillas and Shanken (2015)). This method is widely recognized and applied when comparing the performance of different asset pricing models (see Fama and French (2015)). For robustness in test results, we apply several different techniques when testing our intercepts. An exhaustive list of methods contains; the Fama-MacBeth approach to test intercepts for statistical significance (Fama and MacBeth, 1973), the GRS-test to test if the intercepts are jointly zero (Gibbons, Ross and Shanken, 1989), a simple average of the intercepts to see which one was the closest to zero (see Fama and French (2015)), and the Barillas and Shanken Sharpe Ratio approach (Barillas and Shanken, 2016). However, before any tests can be conducted, we need to determine whether EMI is statistically significant when included in the five-factor models, or not. For this purpose, we apply the estimation- and testing techniques described in Fama and MacBeth (1973) to test the factor for statistical significance in each of the models where it is included.

Our empirical study shows some findings that confirm our hypothesis. Indeed, some of the findings from our empirical study suggests that the models containing EMI rather than CMA perform better than the benchmark model (Fama French Five-Factor Model). All the techniques and methods, with exception of the Fama-MacBeth test, used to test the intercepts of our estimated models suggests that this is the case. However, we are still sceptical of our results as we only find statistical significance for EMI on left-hand side portfolios constructed on size-B/M and therefore suggest that there is room for further research on this topic.

The rest of this thesis is organized as follows: Section (2) starts with a presentation of the empirical work done so far in the form of a literature review. Section (3) will introduce the theoretical background behind our selection of the empirical models to be examined as well as a detailed derivation of why the return on invested capital should have a statistical impact on asset value and thus also asset returns. Section (4) illustrates the procedure of constructing the factors and factor portfolios as well as methods for estimation and testing procedures used in the empirical study of this thesis. Section (5) contains a detailed description of the datasets used in this thesis and how these are collected. In section (6) we present the results and findings of our empirical study. Section (7) concludes our thesis. Section (8) contains a list of all the references used in this paper. Finally, section (9) contains the appendix which includes various derivations and all computer codes used for this thesis.

2.0 Literature Review

The literature and research in the field of finance can be dated back to the early 1950's. The earliest work on the relationship between risk and return was developed in Markowitz (1952) and Markowitz (1959) and looked at how investors can create portfolios of separate investments to optimize the risk-to-return-ratio (Perold, 2004). The main theoretical findings of these papers are that there exists an optimal risk-to-return portfolio in the intercept between the efficient frontier and the capital allocation line. This capital allocation line laid the basis for modern asset pricing theory as it shows how the excess returns on portfolios depends on the amount of risk undertaken. The now famous Capital Asset Pricing Model (‘’CAPM’’) further builds on this notion and expands the theoretical framework laid in Markowitz (1952, 1959). The CAPM, which marked the birth of modern asset pricing theory, was created in parallel by William Sharpe (1964), John Lintner (1965), and Fisher Black (1972). It suggests that excess returns are driven by the portfolios exposure to the systematic risk of the market, and not firm specific risk factors. Although it was a powerful and intuitive model, the empirical record of the CAPM is poor (Fama & French, 2004).

The CAPM was for long recognized as the most reliable asset pricing model, that is until Fama and MacBeth, (1973) used the well-known "two-parameter" portfolio model to test the relationship between the average return and risk for common stocks traded on the New York Stock Exchange. The results from this study implied that there is a linear relationship between risk and return, however, they also found no evidence to support the notion that the systematic risk factor, in addition to portfolio risk, is the only risk factor that systematically affects average asset returns. Further on, Ball (1978) found a relationship between the behaviour of stock prices and public announcements of firms` earnings. This paper suggests that securities in post earnings announcement periods, on average, yield systematic excess return, which was proven to be a consistent anomaly. Another anomaly was identified in Banz (1980) which suggests that the total market value of the common stock in question could significantly affect the risk adjusted returns. More specifically, it finds that the common stock of small firms had, on average, higher risk adjusted returns than the common stock of large firms. Banz (1980) refers to this finding as the *size effect*. Stattman (1980) finds another anomaly in asset returns, which is linked to a firm's book value of equity relative to its market value of equity (i.e. the BE/ME multiple). The paper suggests that firm's with high BE/ME multiples tend to, on average, realize higher risk adjusted returns than firms with lower BE/ME multiples. Stattman (1980) refers to high BE/ME firm's as high value firms and low BE/ME firm's as low value firms and thus dubs this finding as the *value effect*. These kinds of results were inconsistent with two-parameter model, and one potential explanation could be that it is mis-specified.

A few decades after the CAPM, Fama and French (1992) introduced two additional factors, size and book-to-market equity. This paper introduced a model which was created to capture the anomalies found in Banz (1980) and Stattman (1980), which were shown to proxy for many of the other anomalies identified up to this time period (see Jegadeesh and Titman (1993), DeBondt and Thaler (1985), Basu (1981), Rosenberg, Reid and Lanstein (1985)). This was an extension of the "two-parameter" portfolio model, i.e. the CAPM. The aim was to capture the cross-sectional variation in average stock returns associated with market beta, size, leverage, book-to-market equity, and earnings-price ratios. This *new* model including the *size factor* (SMB) and the *value factor* (HML) is referred to as the

Fama French three-factor model. To increase the explanatory power of the model, Fama and French (1993) presented another model identifying two additional risk factors related to the bond market. Fama and French (1996) continued testing the FF three-factor model and observed that the unspecified anomalies almost disappeared, except the anomaly related to the continuation of short term returns. These results imply that the factors included in the FF three-factor model correctly proxies for anomalies identified in the papers mentioned above. The results were consistent with the rational Intertemporal Capital Asset Pricing Model (''ICAPM'') and Arbitrage Pricing Model (''APT''), so possible explanations for the model not capturing the anomaly related to the continuation of short term return could have been irrational pricing and data problems. The authors admit that even though the FF three-factor model is a good model, there are anomalies that still cannot be explained e.g. the continuation of short term returns.

Other academics and researchers were also conducting tests, trying to explain anomalies using different set of factors. Chen, Novy-Marx and Zhang (2011) sets out to understand anomalies that the three-factor model failed to explain. They proposed an alternative version of the three-factor model replacing the Size- and Value factors with an 'investment'- and 'return on equity' factor. Although the investment factor played a similar role as the Value factor from the FF three-factor model, the authors concluded that the return on equity factor added a new dimension of explanatory power that was absent in the FF three-factor model. Hou, Xue and Zhang (2015) also studied the potential effects that investment and the return on equity could have on asset pricing models. Starting with a wide array of approximately 80 variables that should cover the major sorts of anomalies, they presented an empirical q factor model consisting of a market factor, a size factor, a profitability factor and an investment factor. The results from the study implied that their model in most (but not all) cases outperformed the FF three-factor model.

The latest attempt to capture the anomalies that are not explained by the CAPM is where Fama and French (2015) extends the three-factor model, introducing a five-factor asset pricing model. The three-factor model was criticized for being an incomplete model because it did not capture the variation in average returns associated with profitability and investment. The reason for the criticism had been because many researchers (e.g Novy-Marx (2013) and Aharoni, Grundy and Cheng

(2013)) had been able to identify relationships between the profitability and expected return of a firm, as well a connection between the investment and expected returns. Hence, Fama French (2015) used the dividend discount model to provide an explanation for why the factors related to profitability and investments should increase the performance of the five-factor model compared with the FF three-factor model. Their research was heavily based on the study of intercepts, in other words, the main goal was to find a model that reported an intercept equal to 0 (meaning that the model completely explains expected returns). After estimating seven different models using different set of factor combinations their results concluded that the HML factor became redundant and a four-factor model consisting of a Market, Size, Profitability, and Investment factor performed as well as the FF five-factor model. These results show that there is no clear answer to whether the FF five-factor model is a better specified model or not, and leaves room for further research on this topic.

3.0 Theory

3.1 Theory on Asset Pricing

In this section we present all the relevant theory for our thesis, ranging from the early concepts of asset pricing theory to the most modern asset pricing models available today. In the very end, we tie this asset pricing theory to a well known concept from corporate finance to derive our hypothesis and the motivations behind it.

3.1.1 Portfolio Theory

The early work on asset pricing theory was mostly based on how investors can create portfolios of separate investments to optimize the risk-to-return-ratio (Perold, 2004). The most notable theoretical framework on the subject was and still is Portfolio Theory, as discussed in Markowitz (1952, 1959). Portfolio theory is a theoretical framework that illustrates the relationship between risk and return, and how investors should allocate their resources to maximize their return given the level of risk undertaken. It starts with the notion of the *efficient frontier*. The efficient frontier contains a set of portfolios where all the portfolios included are *mean variance efficient*, i.e. they yield the highest expected return given their level of risk. However, not all portfolios are mean variance efficient and there exists portfolios that yield lower expected return than those on the efficient frontier, but for the same level of risk. Such portfolios are therefore, by definition, inefficient. In other words, any combination of assets above the efficient frontier are impossible to obtain and any combination of assets below the efficient frontier are inefficient. Somewhere along the efficient frontier, there exists a portfolio which is more efficient than any other of the portfolios. This portfolio can be recognized as the portfolio with the highest *Sharpe Ratio*, or in other words, the highest excess return given the amount of risk. Markowitz (1952, 1959) calls this the *tangency portfolio* or the *market portfolio.* Mathematically, the Sharpe Ratio can be expressed as follows.

$$\text{Sharpe ratio} = \frac{R_p - R_f}{\sigma_p} \tag{1}$$

Where R_p is the return on the portfolio; R_f is the return on a risk-free asset; σ_p is the standard deviation of the portfolio (a measure of risk). Hence, the goal of a rational and risk averse investor is to hold the portfolio with the highest Sharpe Ratio. One can find this *market portfolio* by drawing a straight line from the y-axis, starting at the risk-free rate, and up to the efficient frontier along the x-axis. At the point where this straight-line tangent the efficient frontier is where the market portfolio lies. This straight line is in portfolio theory referred to as the *capital allocation line (''CAL'')* and can be expressed as follows.

$$E(R_c) = R_f + \sigma_c \left(\frac{R_p - R_f}{\sigma_p} \right) \tag{2}$$

Where $E(R_c)$ is the expected return of a portfolio which includes the risk-free rate and a risky portfolio; σ_c refers to the risk of a portfolio which includes the risk-free rate and a risky portfolio. An important thing to take away from the CAL is that investors are only compensated with excess returns for the risk caused by the risky portfolio, and not for holding the risk-free asset. Hence, the inclusion of a risk-free asset in an overall risky portfolio of assets can help reduce the risk of the overall portfolio held by the investor. In the next part of this section, we will show how Sharpe (1964), Lintner (1965) and Black (1972) further extend this model to create the *Capital Asset Pricing Model*.

3.1.2 The Capital Asset Pricing Model (''CAPM'')

The father of asset pricing models is the Capital Asset Pricing Model (''CAPM''), developed in parallel by Sharpe (1964), Lintner (1965) and Black (1972), hence why it is often referred to as the Sharpe-Lintner-Black Model (''SLB''). According to Black (1972), the SLB-model states that any capital asset for a single period, and given certain assumptions, will satisfy the following equation:

$$E(R_i) = R_f + \beta_i\left[E(R_m) - R_f\right] + u_i \tag{3}$$

$$\beta_i = \frac{Cov(R_i, R_m)}{Var(R_m)} \tag{4}$$

Where $E(R_i)$ is defined as the expected return on asset *i* for the period; R_f is the return of a risk-free asset for the period; $E(R_m)$ is the expected return of the market portfolio (all assets taken together); β_i is the market sensitivity of asset, *i*; u_i is the error term which is referred to as the unsystematic risk factor (Black, 1972).

Even though the CAPM seems simple, it carries a critical observation about the relationship between risk and return. Moving the risk-free rate over to the left-hand side of the equation, we observe that the only two components that reward the investor with returns over and above the risk-free rate are the assets systematic- (β_i) and unsystematic risk factors ($u_{i,t}$). On one hand, Sharpe (1964) argues that unsystematic risk factors, i.e. risk factors that are only specific to the asset in question, can be diversified by holding a large enough portfolio of assets. Because this firm specific risk can essentially be eliminated through diversification, investors shall not be rewarded for their exposure to it. On the other hand, since the systematic risk factor is a risk component of the market itself and can thus not be diversified, investors shall be rewarded for their exact exposure to this risk factor. Hence, the *more correlated* asset *i* is with the market portfolio (e.g. the market index), the higher the exposure to the systematic risk for which you are rewarded with higher returns.

3.1.3 The Fama French Three-Factor Model

The original SLB-model is today still used by practitioners in finance to calculate the cost of equity of a firm, which is then used to discount the cash flows of a specific firm. This can most likely be attributed to its intuitive construction and ease of use. However, researchers have uncovered several patterns in average stock returns that are left unexplained by the SLB-model. Such patterns are referred to as *anomalies* (Fama & French, 1996). Fama and French (1992, 1993, 1996) is a sequence of papers that sets out to create a model that either proxies for or includes anomalies discovered in papers such as Jegadeesh and Titman (1993), DeBondt and Thaler (1985), Banz (1981), Basu (1983), Rosenberg, Reid and Lanstein (1985), and Lakonshik, Shleifer and Vishny (1994). The result is the *Fama French Three-Factor Model* (‘’FF3’’)

$$R_{i,t} - R_{f,t} = a_i + \beta_i\left[R_m - R_f\right] + s_i SMB_t + h_i HML_t + e_{i,t} \qquad (5)$$

Where the excess return on a portfolio, $E(R_i) - R_f$, is a function of its sensitivity to the following three factors

i. The excess return on a well-diversified market portfolio, $[E(R_m) - R_f]$
ii. The difference between the return on a portfolio of small stocks and the return on a portfolio of large stocks, i.e. small minus big (‘’SMB’’).
iii. The difference between the return on a portfolio of high BE/ME stocks and the return on a portfolio of low BE/ME stocks, i.e. high minus low (‘’HML’’).

Like the CAPM, an investor is rewarded relative to the risk he or she is exposed to through her investment. However, in the case of the three-factor model, there are two additional risk factors for which an investor should receive risk premiums.

The risk premium related to the SMB-factor is often referred to as the *size effect*, as coined by Banz (1980). The paper found that the total market value of the common stock in question could significantly affect the risk-adjusted returns. More specifically, it finds that the common stock of small firms had, on average, higher risk adjusted returns than the common stock of large firms. As such, Banz (1980) views this discovery as evidence that the CAPM indeed is mis-specified and that the size effect adds to the explanatory power of the model. However, the paper states that there is no theoretical foundation to the effect of size and that it cannot be determined whether the market value of a firm itself matters or whether it is merely a proxy for other factors correlated with the market value. Later research suggests that the earnings prospects of smaller firms are more sensitive to macroeconomic risk factors than larger firms (Chan & Chen, 1991). Hence, investors should receive higher risk premiums for holding portfolios of small firms because their earnings are more volatile than portfolios of larger firms.

The risk premium related to the HML-factor is often referred to as the *value effect*, as coined by Stattman (1980). Stattman (1980) explores the firm's BE/ME multiple (i.e. the firm's book value of equity relative to the firm's market value of equity). The paper concludes that companies with high BE/ME multiples tend to, on average, realize higher expected returns than firms with low BE/ME multiples. Penman (1991) looks at the economical meaning behind this observation and argues that high BE/ME firms realize higher expected returns because the profitability of such firms tend to be more volatile than the profitability of low BE/ME firms. Hence, because there is an increased uncertainty anchored to high BE/ME firms, investors holding such stocks should be compensated for the higher risk exposure. In other words, investors holding a portfolio of high BE/ME stocks should receive a higher risk premium than investors holding a portfolio of low BE/ME stocks.

3.1.4 The Fama French Five-Factor Model

As previously mentioned, any pattern in average stock returns not explained by a given model is referred to as an anomaly (Fama & French, 1996). This implies that once an economically sound explanation for an anomaly is identified, one can correct for this anomaly by adding another risk factor to the model. Following this trail of thought, Fama and French (2015) uses the *dividend discount model* (''DDM'') to find a sensible explanation as to why *investment* and *profitability* should have a statistically significant impact on stock returns, as shown in Novy-Marx (2013) and Aharoni, Grundy, and Zeng (2013).

The DDM states that the market value of a share of stock is determined by the discounted value of the firms expected dividends (Miller & Modigliani, 1961).

$$m_t = \sum_{t=1}^{\infty} \frac{E(d_t)}{(1+r)^t} \quad (6)$$

Where m_t is the share price at time t; $E(d_t)$ is the expected dividend payout at time t; r is the internal rate of return on the expected dividends. With a bit of manipulation, Miller and Modigliani (1961) shows that the total market value of the firm's stock can be expressed as a function of the firm's earnings and investments.

$$m_t = \sum_{t=1}^{\infty} \frac{E(Y_{t+1} - \Delta B_{t+1})}{(1+r)^t} \quad (7)$$

$$\Delta B_{t+1} = B_{t+1} - B_t \quad (8)$$

Where Y_{t+1} is the total equity earnings for time t; ΔB_{t+1} is the change in total book value of equity, i.e. the equity investment in effect of time t. From equation (7), Fama and French (2015) derives the following

1. Holding everything but m_t constant, a lower stock price implies a higher expected return
2. Holding everything but the expected future earnings and the expected return constant, higher expected earnings imply a higher expected return
3. Holding B_t, m_t and expected future earnings constant, more investment implies a lower expected return.

Having established this theoretical link between investment, profitability and expected return, Fama and French (2015) created the *Fama French Five-Factor Model* (''FF5'')

$$R_{i,t} - R_{f,t} = a_i + \beta_i[R_m - R_f] + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + e_{i,t} \quad (9)$$

As we can see, the model is a direct extension of the FF3F model with the addition of two new factors; RMW and CMA. RMW is defined as the difference between the returns on diversified portfolios of stocks with robust and weak profitability, hence r_i captures the variation in expected returns caused by profitability. CMA is defined as the difference between the returns on diversified portfolios of the stocks of low and high investment firms, hence c_i captures the variation in expected returns caused by the amount of cash invested by the firm.

3.1.5 Tobin's Q and the Q-Factor Model

The q-factor model is an investment-based asset pricing model, defined in Hou et. al. (2015), which uses the findings of Tobin (1969) as its cornerstone. According to the q-factor model, expected stock returns are driven by two factors; *the expected discounted profitability of the firm* and the *investments to assets ratio.* The basic theoretical framework of the q-factor model rests on the assumption that corporate management have incentive to maximize the net present value of their firm, which is a basic assumption derived from agency theory and relates to the concept of empire building (Baker and Wurgler, 2013). On one hand, we have the management who will only undertake new investments if it has a positive effect on firm value and ultimately the stock price. On the other hand, we have the investors who evaluate the investment projects undertaken by the management of a given firm with special interest in the expected payoff of the given project versus the risks of it. Hence, we have two parties involved, the management who calculate the costs and the investors who calculate the payoff. If the investors value the payoff higher than the expected costs of the project, then stockholders will benefit through an increased stock price as the project is expected to add to the total firm value and vice versa. Using this intuition, Brainard & Tobin (1968) concluded that the *rate of investment* should be related to the Q-value, i.e. the value of the investment relative to its adjustment cost. Based on this intuition, we can show that the Q-value from Tobin (1969) can be expressed as follows. (Full derivation can be found in Appendix 9.1)

$$1 + a\frac{I_{i0}}{A_{i0}} = E_o[M_1\pi_{i1}] \tag{10}$$

The first order condition (Euler Equation) illustrates Tobin's Q (Tobin, 1969) and states that firms will continue to invest until the marginal cost of investment is equal to the marginal benefit of investment. Rearranging the equation. we get the following equation for the expected return on stock *i*.

$$E_0[r_{i1}^s] = \frac{E_0[\pi_{i1}]}{1+a(I_{i0}/A_{i0})} \tag{11}$$

Where $E_0[r_{i1}^s]$is the time 0 expected stock return for asset *i* in period 1; $E_0[\pi_{i1}]$ is the time period 0 expected profits of asset *i* in period 1; a is a constant parameter; (I_{i0}/A_{i0}) is the ratio of investment to assets for asset *i* in time period 0. Full derivation of this expression can be found in appendix 9.1. Based on this, we see that the following two things hold true

i. High investment stocks earn higher expected returns than low investment stocks
ii. High expected profitability stocks should earn higher expected returns than low expected profitability stocks earn

We can now draw out the *Q-factor Model* as introduced in Hou et. al. (2015).

$$E[r^i] - r^f = \beta_{MKT}^i E[MKT] + \beta_{ME}^i E(r_{ME}) + \beta_{\frac{I}{A}}^i E\left[r_{\frac{I}{A}}\right] + \beta_{ROE}^i E[r_{ROE}] \tag{12}$$

Where E[MKT], $E(r_{ME})$, $E\left[r_{\frac{I}{A}}\right]$ and $E[r_{ROE}]$ are the expected factor premiums and all the betas are the factor loadings on the factors MKT, r_{ME}, $r_{\frac{I}{A}}$ and r_{ROE}. The model essentially states that the expected excess return on asset *i* is a function of its sensitivity to the excess market return (MKT), the difference between the return on a portfolio of small size stocks and a portfolio of big size stocks (r_{ME}), the difference between the return on a portfolio of low investment stocks and the return on a portfolio of high investment stocks ($r_{\frac{I}{A}}$), and the difference between the return on a portfolio of high profitability stocks and the return on a portfolio of low profitability stocks (r_{ROE}). Note that the proxy for profitability used in this model is the return on equity (‘’ROE’’).

3.2 Corporate Finance and Our Hypothesis

We have discussed a multiple of asset pricing models. Most importantly, we have highlighted the fact how many researchers have tried to uncover different asset pricing anomalies and how asset pricing theorist like Fama and French have set out to implement them into asset pricing models to increase the model's performance. An important notable fact is that most of these anomalies stem from corporate finance and are variables that have been used by professionals in corporate finance for decades. Variables such as the BE/ME-multiple, investment and profitability used in Fama and French (1992, 1993, 1996, 2008 and 2015) are indeed anomalies deduced using intuition from corporate finance. Fama and French (2015) applied the *Dividend Discount Model* as originally defined in Miller and Modigliani (1961), which is a widely applied valuation tool in finance, to uncover an economically sound explanation for the inclusion of profitability and investment to their original model. However, we will in this part suggest an alternative reasoning for the use of profitability and investment as explanatory variables in asset pricing models using the *Terminal Value of Free Cash Flow* approach for valuation.

3.2.1 The Terminal Value of Free Cash Flow Formula - Motivation

The terminal value of free cash flow (''TVM'') approach to valuation deduces the value of a firm based on its free cash flow and can be expressed as follows (Koller & Goedhart, 2015).

$$V = \frac{FCF_{t=1}}{WACC-g} \tag{13}$$

The main and most obvious difference between the TVM and the DDM, as used in Fama and French (1996), is that the value in the numerator is *free cash flow* rather than dividends. Free cash flow is a measure of cash available to all shareholders, i.e. it states how much of the cash produced through operations (less capital expenditures) is available for dividends, share buybacks, debt repayments and reinvestment. This means that the management of firms can choose to either i) distribute cash back to the shareholders through dividends or share buyback, ii) pay back on their debt obligations, iii) reinvest money into the firm, or iv) a combination of these four things. Firms have no obligation to pay dividends and there is thus no guarantee that all firms will pay dividends. Since firms have to pay debt obligations to avoid financial distress and need to reinvest to grow, dividends are not the main priority for firms (Brav et. al, 2004). Hence, not all firms choose to pay out dividends and since a steady stream of dividends are required for the DDM to be viable, we cannot use the DDM to value all the firms in a given sample of firms. However, every single firm has a free cash flow as it is simply an accounting related number. Hence, every firm that has accounting data readily available can be valued using the TVM approach. Another drawback of the DDM model is that dividend yields tend to change drastically over time as a direct result of changes in free cash flow and the priorities as how the cash should be spent. This implies that to model dividend payouts accurately one needs to model the free cash flow accurately, in addition to predicting if and how much free cash flow the management will choose to distribute as dividends. Further on, the DDM assumes that dividends will grow at a constant rate into perpetuity which is a problematic assumption based on the levels of decision making required by management and the amount of free cash flow available year by year. All this makes it problematic for the DDM to consistently be able to predict firm value and since these problems are essentially eliminated when using the TVM, we suggest that the TVM might potentially be a more accurate method for the calculation of firm value. As support for this notion, we refer to Francis et. al. (2000) which shows that median absolute prediction error for the TVM (41%) seems to be significantly lower than for the DDM (69%). With this trail of thought in mind, we argue that the importance of the intuitive observations of a model increases as the explanatory power of the model increases. Hence, since the TVM seems to be better at predicting firm value than the DDM, we believe that derivations based on the TVM might give a clearer picture of firm value drivers than the DDM.

3.2.2 Return on Invested Capital

To deduce the implications of the TVM we must extend equation (13) further and evaluate its first derivatives. It can be shown that the TVM can be expressed as follows (Koller & Goedhart, 2015).

$$V = \frac{NOPLAT_{t=1}\left(1-\frac{g}{ROIC}\right)}{WACC-g} \tag{14}$$

Where V is the value of operations, i.e. the value of the company; $NOPLAT_{t=1}$ is the net operating profit less adjusted taxes for period 1; WACC is the weighted average cost of capital; g is growth in free cash flow; ROIC is the return on invested capital. Taking the first derivative of the function with respect to ROIC, we get the following expression.

$$\frac{\partial V}{\partial ROIC} = \frac{NOPLAT_{t=1}\left(\frac{g}{ROIC^2}\right)}{WACC-g} \tag{15}$$

This reveals that return on invested capital (used interchangeably with ROIC throughout this thesis) is a fundamental driver of firm value. We can see that an increase in the return on invested capital will always have a positive impact on the value of the firm. Taking the first derivative of the value function with respect to growth further reveals the importance of the return on invested capital.

$$\frac{\partial V}{\partial g} = \frac{NOPLAT_{t=1}(1-\frac{WACC}{ROIC})}{(WACC-g)^2} \tag{16}$$

As we can see, if the return on invested capital is not sufficiently large enough (i.e. higher than the weighted average cost of capital) then growing the operating profits or free cash flows of the firms will in fact destroy value. In summary, the implications of the TVM are that:

i. Any increase in the return on invested capital will have a positive impact on firm value
ii. The return on invested capital needs to be higher that the weighted average cost of capital for growth to have a positive effect on firm value

We have now determined that the return on invested capital indeed has a major effect on the value of a firm and now seek to understand what drives it. ROIC can be expressed as follows (Koller & Goedhart, 2015).

$$ROIC = \frac{Net\ operating\ profits\ after\ tax_t}{Total\ Invested\ Capital_{t-1}} \tag{17}$$

Recalling that the DDM simply implies that the change in invested capital from one year to the next affects firm value, the TVM takes this notion further. Not only does the effect of ROIC depend on the total invested capital, it also depends on the net operating profits for the year after the investment was undertaken. This an important aspect as the purpose of capital investments is to increase the earnings of the firm through expansion of different aspects of the firm, e.g. production facilities, development of new products etc. Hence, we observe that the size of the investment is only important when related to the extra profits it generates for the firm. Therefore, we argue that viewing the investment as an isolated variable, i.e. only at the absolute amount of the cash invested, makes little sense if you do not compare it to the profit or loss it generates. This leads us to the conclusion that if we want to capture variations in asset returns (i.e. variations in the total value of the firm divided by the total number of shares outstanding) we have to look at the return of investments undertaken by the firm rather than the absolute size of the investment. We do not challenge the idea that the total investment has an impact on asset returns, rather we argue that there is an additional aspect which might induce initial variations to asset returns that needs to be considered. Following this trail of thought, we suspect that traditional asset pricing models like the FF5-model from Fama & French (2015) fail to model the complete variation in asset returns induced by the firm's investment policy when a factor based on the absolute size of investments, CMA, is applied to the model specification. For further illustration of this point, we present an alternative presentation of the TVM formula (Koller & Goedhart, 2015). (*Full derivation of this expression can be found in appendix 9.2. Also note that the full mathematical expression of each variable used in these equations can be found in appendix 9.3.*)

$$V = IC_0 + \frac{IC_0(ROIC - WACC)}{WACC - g} \qquad (18)$$

Where IC_0 is defined as the invested capital for the current year. Indeed, we can see that if we hold all other factors constant, an increase in the invested capital will increase the value of the firm, and thus also the value of the stock, which in turn will imply higher expected returns for the investor. Hence, we do acknowledge that some of the variation in asset returns is due to the absolute size of investments.

However, we also see that an increase in invested capital will have a positive impact on value if and only if ROIC is sufficiently large enough (i.e. higher than the weighted average cost of capital). Also, we see that the higher the return on the invested capital is, the greater the impact an increase in capital invested will have on the value of the firm. This is the part of the variation induced by the firm's investment policy which we suspect is not captured by the traditional CMA-factor, leaving room for a potential increase in performance of such models by simply adjusting their specification.

3.2.3 Our Hypothesis

Based on the theory presented previously in this chapter, we suspect that traditional investment factors, e.g. CMA, could be inefficient at modelling variations in asset returns induced by a firm's investment policy. We argue that that there could be variation caused by the investment itself, but also that a potentially large portion of this variation could be left out if the return of the investment is not considered. Our hypothesis is therefore that a factor based on the return on invested capital, rather than the absolute size of the investment itself, could potentially capture the variation induced by a firm's investment policy in its entirety. Exchanging the traditional investment factor, e.g. CMA, with a factor based on the return on investment could potentially increase the explanatory power of traditional asset pricing models. Hence, this thesis will take the empirical models presented in Fama & French (2015) and Hou, Xue & Zhang (2015), exchange their original investment factors with a factor based on ROIC (referred to as EMI) and compare their performance to the original FF5-model to answer the following research question.

''Can a factor based on the return on invested capital replace the traditional investment factors and increase the performance of traditional asset pricing models?''

4.0 Methodology

4.1 Introduction

To test the previously discussed research question of this thesis, we will introduce two alternative asset pricing models. These models are inspired by the Fama French five-factor model (Fama and French, 2015), the Q-factor model (Hou, Xue and Zhang, 2015) and theory from corporate finance. In essence, we are changing the factor composition of these two well-known models, exchanging the *regular* investment factor with a factor based on the return on invested capital. Based on this, we will estimate the following three models.

(**FF5**) $R_{i,t} - R_{f,t} = \alpha_i + \beta_i(R_{m,t} - R_{f,t}) + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + \varepsilon_{i,t}$

(**FF5new**) $R_{i,t} - R_{f,t} = \alpha_i + \beta_i(R_{m,t} - R_{f,t}) + s_i SMB_t + h_i HML_t + r_i RMW_t + e_i EMI_t + \varepsilon_{i,t}$

(**Q5new**) $R_{i,t} - R_{f,t} = \alpha_i + \beta_i(R_{m,t} - R_{f,t}) + s_i SMB_t + h_i HML_t + g_i GMP_t + e_i EMI_t + \varepsilon_{i,t}$

FF5 is simply the original Fama French five-factor model, which we will use as a benchmark for the remaining two models. FF5new is a modified version of FF5 where we exchange the original investment factor with the newly constructed return on invested capital factor. We define e_i as the factor loading on the factor EMI (**e**fficient **m**inus **i**nefficient), which is the difference between the return on a portfolio of efficient return on invested capital firms and the return on a portfolio of inefficient return on invested capital firms. **Q5new** is essentially the Q-factor model as defined in Hou et al (2015) where we exchange the original investment factor with the new return on invested capital factor. We have renamed (but retained) the return on equity factor where g_i is the factor loading on the factor GMP (**g**reat **m**inus **p**oor).

Using Fama and French (2015) as a benchmark, we set out to compare the performance of the new models to the performance of the Fama French five-factor model. This translates to a comparative analysis and implies that the following two steps need to be taken.

Step one is to estimate all models and check the factors for statistical significance. This is a crucial step as it determines whether the *new* factor, EMI, incorporated in the model helps to explain asset returns or not. Further on, it will also reveal if the replacement of the old investment factor with the EMI factor will make any of the remaining factors, *size, value* and/or *market premium* redundant. An example of this happening can be found in Fama French (2015) where the inclusion of the investment and profitability factors ultimately makes the value factor redundant. For the purpose of checking for statistical significance, we will apply the cross-sectional estimation- and testing techniques discussed in Fama-MacBeth (1973), Fama (1965), Blume (1970), Officer (1971) and Cochrane (2005).

Step two is to check and compare the performance of the three models to determine whether or not the new models performed better than the model specified in Fama and French (2016), i.e. if one or both are better at excluding redundant factors and explaining the variation in asset returns. We base this part of our study on the notion that if a model is perfectly specified and explains all the variation in asset returns, the intercept of the estimated model should be equivalent to 0 (see Gibbons, Ross and Shanken (1989), Cochrane (2005) and Barillas and Shanken (2015)). Hence, we will combine several methods to study the intercepts of our estimated models to compare them and determine which one has the best performance.

First off, we will apply the method developed in Gibbons, Ross and Shanken (1989). This method is based on a classical test of the CAPM developed in Jensen (1968) and Black, Jensen and Scholes (1971) and it examines the intercepts in time-series regressions of excess test portfolio returns on market excess returns (Barillas and Shanken, 2015). It states that, given that the market is efficient and that the model is perfectly specified, the intercepts (''alphas'') of asset pricing models should be 0. By doing this, we essentially test if the intercepts for each individual model are jointly 0. In addition to the GRS-test of joint statistical significance, we will also look at the t-ratios of the intercepts from the Fama-Macbeth estimations to test if they are statistically close to 0.

Secondly, we will calculate the both the average- and average absolute values of our intercepts and compare them for each of the models and test assets (i.e. LHS portfolios). The intuition behind doing so is that the smaller the average absolute value of intercepts (i.e. closer to 0), the closer the model is to being perfectly specified. Although this specific method does not have much theoretical support, it does hold up intuitively and is used in several research papers to compare asset pricing models (see Fama and French (2015) and Hou, Xue and Zhang (2015)).

Thirdly, we apply the Sharpe ratio-method developed in Barillas and Shanken (2016). This methodology suggests that to compare to asset pricing models and determine which one is better at modelling variation in asset returns, one should compare their *maximum squared Sharpe Ratio*. Like the method described directly above, the lower the maximum squared Sharpe ratio of the model is, the better the model is at modelling variation in asset returns.

4.2 Factor definitions (RHS factors)

To estimate and examine the FF5, FF5new and Q5new we need to construct factor returns. We will follow the same methodology applied in Fama and French (1993) in their construction of factor returns for the famous three-factor model. Our versions of the five-factor model will simply be augmented versions of the three-factor model. Hence, to keep the results comparable, the new factor returns will be constructed in the same way as the operating profitability- and investment factor returns are constructed for the FF five-factor model.

The size and value factors are constructed by sorting stocks independently into two Size groups (Big or Small) and then independently to three Value groups (High, Neutral or Low). More specifically, at the end of June of each year t, stocks are first allocated into two size groups using the NYSE median market cap as breakpoint. After the stocks have been assigned to a size group they are again independently allocated to three B/M (Value) groups, and the breakpoints are the 30th and 70th percentiles of B/M for NYSE stocks. The intersection after the stocks have been allocated to two Size groups and three Value groups create six Size-B/M value weighted (VW) portfolios. By following the exact same procedure, we will in addition to these portfolios also obtain, six VW portfolios formed on Size-Inv, six

VW portfolios formed on Size-OP, six VW portfolios formed on Size-ROIC and six VW portfolios formed on Size-ROE.

The size factor, $SMB_{B/M}$ is then calculated by subtracting the average return on the three big stock portfolios (Big & Low, Big & Neutral, Big & High) from the average return on the three small stock portfolios (Small & Low, Small & Neutral, Small & High). Similarly, constructing the other 2x3 VW portfolios will also produce four additional Size factors SMB_{Inv}, SMB_{OP}, SMB_{ROIC} and SMB_{ROE}. The last step will then be to construct different SMB factor for the new models, where SMB_{FF5new} will be an equal weighted average of $SMB_{B/M}$, SMB_{OP} and SMB_{ROIC}. Equivalently, SMB_{QF5new} will be the equal weighted average of $SMB_{B/M}$, SMB_{ROIC}, and SMB_{ROE} (Table 4.1). Afterwards, the Value factor HML is constructed by subtracting the average of the two low B/M portfolio returns from the average of the two high B/M portfolio returns.

A similar approach to the creation of the HML factor is applied when constructing the profitability- (RMW), investment- (CMA), return on equity- (GMP) and return on invested capital- (EMI) factors. The only way these factors differ is that the second sort is either on operating profitability (**r**obust **m**inus **w**eak) or (**c**onservative **m**inus **w**eak) for the FF five-factor model. For our versions of the five-factor model the second sort will either be on return on equity (**g**reat **m**inus **p**oor) or return on invested capital(**e**fficient **m**inus **w**eak).

After constructing all the explanatory variables mentioned above, for each month, *t*, we will be able to evaluate the FF five-factor model and our versions of the augmented three-factor model.

Since multivariate regression slopes measure marginal effects, the regressors in our thesis will be made solely on 2x3 portfolios. Fama French (2015) also finds that the slopes for the factors in the five-factor model produced from the 2x3 and 2x2 sorts isolates exposures to the value, profitability and investment effects in returns as efficiently as the factors produced from the other sorts (e.g. 2x2x2x2). Hence, our choice of portfolio construction (2x3) will not affect the regression results. We also assume no market frictions (taxes, transaction costs, etc.) for the three models specified in the text. Lastly, we keep in mind that the regression slopes are estimated as constants which might be problematic, and leaves room for further investigation.

Table 4.1

Panel A: FF five-factor model - Construction of Size, B/M, profitability and investment factors.
Independent sorts are used to assign stocks to two size groups, and three B/M, operating profitability (OP) and investment (Inv) groups. The VW portfolios defined by the intersections of the groups are the crux for the factor construction. The portoflios are labeled with two letters. The first letter always describes the Size group, Small (S) or Big (B). The second letter will describe the B/M group, high (H), neutral (N), or low (L), the OP group, robust (R), neutral (N), or weak (W), or the Inv group, conservative (C), neutral (N), or aggressive (A). The factors are SMB (small minus big), HML (high minus low B/M), RMW (robust minus weak OP) and CMA (conservative minus aggressive Inv).

Sort	Breakpoints	Factors and their components
2x3 sorts on	Size: NYSE median	$SMB_{B/M} = (SH + SN + SL)/3 – (BH + BN + BL)/3$
		$SMB_{OP} = (SR + SN + SW)/3 – (BR + BN + BW)/3$
Size and B/M, or	B/M: 30th and 70th NYSE percentiles	$SMB_{Inv} = (SC + SN + SA)/3 – (BC + BN + BA)/3$
Size and OP, or	OP: 30th and 70th NYSE percentiles	
Size and Inv, or	Inv: 30th and 70th NYSE percentiles	$SMB_{FF5} =(SMB_{B/M} +SMB_{OP} +SMB_{Inv})/3$
		$HML=(SH + BH)/2 - (SL + BL)/2=[(SH - SL) + (BH - BL)]/2$
		$RMW=(SR + BR)/2 - (SW + BW)/2=[(SR - SW) + (BR - BW)]/2$
		$CMA=(SC + BC)/2 - (SA + BA)/2=[(SC - SA) + (BC - BA)]/2$

Panel B: Our version of the FF five-factor model and the Q-factor model - Construction of Size, B/M, ROIC and ROE factors.
Independent sorts are used to assign stocks to two size groups, and three B/M, return on invested capital (ROIC) and return on equity (ROE) groups. The VW portfolios defined by the intersections of the groups are the crux for the factor construction. The portoflios are labeled with two letters. The first letter always describes the Size group, Small (S) or Big (B). The second letter will describe the B/M group, high (H), neutral (N), or low (L), the ROIC group, efficient (E), neutral (N), or inefficient (I), or the ROE group, great (G), neutral (N), or poor (P). The factors are SMB (small minus big), HML (high minus low B/M), EMI (efficient minus inefficient ROIC) and GMP (great minus poor ROE).

Sort	Breakpoints	Factors and their components
2x3 sorts on	Size: NYSE median	$SMB_{B/M} = (SH + SN + SL)/3 – (BH + BN + BL)/3$
		$SMB_{ROIC} = (SE + SN + SI)/3 – (BE + BN + BI)/3$
Size and B/M, or	B/M: 30th and 70th NYSE percentiles	$SMB_{ROE} = (SG + SN + SP)/3 – (BG + BN + BP)/3$
Size and ROIC, or	ROIC: 30th and 70th NYSE percentiles	
Size and ROE, or	ROE: 30th and 70th NYSE percentiles	$SMB_{FF5new} =(SMB_{B/M} +SMB_{ROIC} +SMB_{OP})/3$
		$SMB_{QF5new} =(SMB_{B/M} +SMB_{ROIC} +SMB_{ROE})/3$
		$HML=(SH + BH)/2 - (SL + BL)/2=[(SH - SL) + (BH - BL)]/2$
		$EMI=(SE + BE)/2 - (SI + BI)/2=[(SE - SI) + (BE - BI)]/2$
		$GMP=(SG + BG)/2 - (SP + BP)/2=[(SG - SP) + (BG - BP)]/2$

4.3 Left Hand Side Portfolios

The returns on the LHS portfolios from January 1970 to June 2015 were extracted from the Kenneth French Data Library (French, 2018). For the purpose of comparison, we have decided to use portfolios formed on Size and B/M, Size and OP, and finally on Size and Inv since these are the LHS portfolios constructed in Fama and French (2015). By following the same methodology, we will be able to compare the performance of the FF five-factor model and our versions of the five-factor model. Fama and French (1992) argues that estimates of market betas are more precise for portfolios. By using portfolios (rather than individual securities) as test assets we also ensure to minimize the measurement error in beta, meaning that we will obtain more accurate coefficient estimates.

All portfolios are constructed at the end of each June. The stocks are first allocated independently to five Size groups (market equity) using NYSE market cap breakpoints. Thereafter the stocks are allocated independently into five Value groups (ratio of book equity to market equity, B/M-ratio), again using the NYSE breakpoints. Hence, the intersection will produce 25 VW Size-B/M portfolios. The breakpoints for Size are the NYSE market equity quintiles at the end of each June of the same year, *t*. The B/M breakpoints are also NYSE quintiles. At the time of portfolio construction in June of year *t*, market equity is market cap at the end of the previous year (December, t-1) while book equity is from the last fiscal year end, t-1.

Using a similar approach, we will also obtain 25 VW Size-OP portfolios and 25 VW Size-Inv portfolios. The procedure for creating the last two LHS portfolios will only differ in that the second sort variable will be operating profitability (OP) or Investment (Inv). OP for June of the same year *t*, is defined as revenues (annual) minus cost of goods sold, interest expense, and selling, general, and administrative expenses divided by book equity for the previous fiscal year end t-1. Investment is calculated by taking the change in total assets from the fiscal year ending in year t-2 to the fiscal year ending in t-1 and dividing by total asset for t-2. The breakpoints for both OP and investment are NYSE quintiles.

The LHS portfolios can either be equal weighted (EW) or value weighted (VW). In Fama and French (2008), we find a brief discussion about potential pitfalls applying one of the mentioned methods. One potential challenge with EW portfolios is that they may be dominated by "extremely small" stocks (stocks with market cap below the 20^{th} NYSE percentile). The main reason for the major influence of these "extremely small" stocks, even though they average only 3% of the market cap is that they account for approximately 60% of the total number of stocks included. Another challenge is that the cross-section dispersion of anomaly variables is largest among these "extremely small" stocks, meaning that they will be determinant in the extreme sort portfolios. Hence, in order to not bias the results of our analysis, we will use VW portfolios in our study. Even though we will apply VW portfolios, we are aware that also these portfolios can be dominated by a few big stocks.

4.4 Factor Significance

As mentioned in the introduction, the first step of our study is to determine the statistical significance of the factors included in each individual model. Of special importance is the EMI factor, i.e. our newly introduced factor. This part has therefore two purposes; i) to check if EMI is significant in both FF5new and Q5new, ii) to check if the inclusion of EMI makes other factors redundant. To obtain these measures, we will apply the Fama and MacBeth (1973) method of estimating and testing models.

4.4.1 Cross-section regressions - a generic description

There are two steps required when estimating regressions on a cross-sectional data set (Woolridge, 2011). *The first step* requires you to run a regular time-series regression for each point in time, *t*, for each asset, *i*. For illustrative simplicity, we present a regression equation with one factor, *F*.

$$R_{i,t} = a_i + \beta_i F_t + \varepsilon_{i,t} \tag{19}$$

We run this equation and obtain the estimates for all βs. This will be the measure of sensitivity of return, $R_{i,t}$, on the factor, F_t, and can be calculated using ordinary least squares (‘’OLS’’).

Step two is to run the actual cross-sectional regression, which will differ slightly from the original time-series regression in that we use the beta-estimates, $\hat{\beta}$, obtained from step one as the explanatory variables on the right-hand side of the equation, rather than the factor itself. Now using expected return on asset, we get the following equation.

$$E(R_{i,t}) = \gamma_i \beta_i + \alpha_i \tag{20}$$

Hence, the goal is to estimate the γ_i which can be interpreted as the ‘’risk premium’’ for exposure to that factor. This can be either positive or negative, significant or insignificant. If it turns out to be statistically significant, then we have found support that the factor possibly adds to the explanatory power of the model.

4.4.2 Fama and MacBeth (1973) approach

One major issue with cross-sectional analysis of financial data is that the error terms tend to be correlated across assets (Woolridge, 2011). For example, if oil-prices rise then the stock price of oil-producing firms will also rise. Since we do not include ''oil-prices'' as a factor in our model, this will be picked up by the error term, making the error-terms across oil-producing firms correlated.

The Fama-MacBeth (1973) provides a solution to this problem, making the methodology slightly different from the 'regular' approach. The methodology of Fama-MacBeth (1973), which is applied in this paper, is described in Cochrane (2005). *The first step* remains the same, i.e. we run a time-series regression for each point in time, *t*, for each asset, *i*, and obtain the beta-estimates. If we have a total of n assets, we have to run the following time-series regression in order to capture the exposures to each factor.

$$R_{1,t} = \propto_1 + \beta_{1,MKRF} MKRF_t + \beta_{1,SMB} SMB_t + \beta_{1,HML} HML_t + \beta_{1,GMP} GMP_t + \beta_{1,EMI} EMI_t + \varepsilon_{1,t}$$
$$R_{2,t} = \propto_2 + \beta_{2,MKRF} MKRF_t + \beta_{2,SMB} SMB_t + \beta_{2,HML} HML_t + \beta_{2,GMP} GMP_t + \beta_{2,EMI} EMI_t + \varepsilon_{2,t}$$
$$\ldots\ldots$$
$$R_{25,t} = \propto_{25} + \beta_{25,MKRF} MKRF_t + \beta_{25,SMB} SMB_t + \beta_{25,HML} HML_t + \beta_{25,GMP} GMP_t + \beta_{25,EMI} EMI_t + \varepsilon_{25,t}$$

As before, we store the 5 beta-estimates and regress these as independent variables against the return on asset. However, now we compute cross-sectional regressions at each point in time, *t*, to T (where T=546).

$$R_{1,1} = \gamma_{1,0} + \gamma_{1,1}\hat{\beta}_{1,RMRF} + \gamma_{1,2}\hat{\beta}_{1,SMB} + \gamma_{1,3}\hat{\beta}_{1,HML} + \gamma_{1,4}\hat{\beta}_{1,GMP} + \gamma_{1,5}\hat{\beta}_{1,EMI} + \varepsilon_{1,1}$$
$$R_{2,2} = \gamma_{2,0} + \gamma_{2,1}\hat{\beta}_{2,RMRF} + \gamma_{2,2}\hat{\beta}_{2,SMB} + \gamma_{2,3}\hat{\beta}_{2,HML} + \gamma_{2,4}\hat{\beta}_{2,GMP} + \gamma_{2,5}\hat{\beta}_{2,EMI} + \varepsilon_{2,2}$$
$$\ldots.$$
$$R_{25,T} = \gamma_{25,0} + \gamma_{25,1}\hat{\beta}_{25,RMRF} + \gamma_{25,2}\hat{\beta}_{25,SMB} + \gamma_{25,3}\hat{\beta}_{25,HML} + \gamma_{25,4}\hat{\beta}_{25,GMP} + \gamma_{25,5}\hat{\beta}_{25,EMI} + \varepsilon_{25,T}$$

At this point, we will have *m+1* estimates of γ, where m is the number of factors. However, in order to calculate the risk premium for each of the risk factors, we need some kind of average of $\hat{\gamma}$ for each risk factor, F. Assuming that the error terms, $\varepsilon_{?,?}$, are independent and identically distributed (''I.I.D''), we can calculate the risk premium (the average of $\hat{\gamma}$) for the *m*-th factor, F using the following equation;

$$Risk\ Premium\ for\ factor\ m,\ \hat{\gamma} = \frac{\sum_{i=1}^{T_{FMB}} \hat{\gamma}_{i,m}}{T_{FMB}} \quad (21)$$

Where T_{FMB} is the total number of cross-sectional regressions used in the second stage of our estimations.

4.4.3 Testing using the Fama-MacBeth (1973) Approach

A goal of this thesis is to identify verify if the variables, i.e. factors, in our model do in fact help to explain variations in asset returns. Essentially, this translates to statistically significant risk premiums, $\hat{\gamma}$. Hence, we want to test if the estimated risk premiums are statistically different from 0, which in turn implies that the risk factor helps to explain variations in asset returns.

Before we can start testing for significance, we need to obtain the t-statistics. Since we are working with γ-estimates that are averaged over time, we must compute the Fama and MacBeth (1973) variant of the t-statistic (See Fama (1965); Blume (1970); and Officer (1971) for justification).

$$t(\bar{\hat{\gamma}}_m) = \frac{\bar{\hat{\gamma}}}{\frac{s(\hat{\gamma}_m)}{\sqrt{n}}} \tag{22}$$

Where

$$s(\hat{\gamma}_m) = \sqrt{\frac{1}{T_{FMB} - 1} * \sum_{t=1}^{T_{FMB}} (\hat{\gamma}_{i,m} - \hat{\gamma}_i)^2} \tag{23}$$

Since we are to test the risk premiums, γ, for statistical significance, the hypothesis of main interest will be as follows.

$$H_0: \bar{\hat{\gamma}}_j = 0$$

$$H_a: \bar{\hat{\gamma}}_j \neq 0$$

If H_0 is rejected, we can conclude that the given factor helps to explain variations in asset returns. If the factors of our models prove not to be statistically significant from zero, we cannot determine that we have evidence that these factors should be included in our asset pricing models. We are especially interested in the factor created on the 'return on invested capital' as this thesis sets out to test whether or not this factor could be a better alternative for the investment factor. In other words, if the risk premium of EMI proves to be insignificant whilst CMA proves to be significant in an alternative specification, we have to discard the possibility of EMI being a potential replacement for CMA.

4.5 Step Two – Comparing Models

After checking and determining statistical significance, we must determine whether or not our *new* models outperform established models. We will use the Fama French 5-factor model (FF5) as our benchmark and compare the FF5new and Q5new to its performance. For the purpose of robustness, we will apply three well-known benchmarking techniques, namely the GRS-test, the Fama-MacBeth intercept study, a simple intercept study and a Sharpe ratio comparison.

4.5.1 Time-series regression and the Gibbons, Ross and Shankens (1989) approach

Researchers (see Roll, 1977 and 1978) have argued that since Fama and MacBeth (1973) and others use proxies for the market portfolio and not the *true* market portfolio, the regression tests are of low power and extending the model with additional factors can possibly lower it even further. The main problem is that the market portfolio is not actually mean variance efficient, i.e. that the returns are not maximized for the given level of risk (Markowitz 1959). Hence, the β-estimate representing the systematic risk will not be correct and tests on such estimates will be questionable. Gibbons, Ross and Shanken (1989) therefore suggests an alternative test for asset pricing models where the main goal is to test whether any portfolio is ex ante mean-variance efficient. If the portfolios included in a model are ex ante mean-variance efficient, then the model will explain the variations in returns perfectly. If this is the case, then $\propto_i$ (the intercept) for all assets *i* should be 0. Since this implies that the closer the intercept is to 0, the more parsimonious the model is, we can use this test to compare the performance of our estimated models. Using the GRS-statistic, as introduced in Gibbons, Ross and Shanken (1989), we can test the nullhypothesis that the intercept for our estimated models are 0 for all assets, *i*.

$$H_0 : \propto_i = 0 \ \vee i$$
$$H_1 : \propto_i \neq 0 \ \vee i$$

Hence, we are essentially testing if some linear combination of the factor portfolios included in the model is a minimum variance portfolio, or in other words on the efficient set. Before we can do such a test, we need to calculate the GRS statistic, which rather conveniently follows a regular F-distribution. Since the GRS-statistic follows an F-distribution, we have to assume that our errors and intercepts are *normal* and that they are both *uncorrelated* and *homoscedastic* (Cochrane, 2005). A full description of how this methodology was implemented, as defined in Gibbons, Ross and Shanken (1989) and illustrated in Cochrane (2005), can be found in **Appendix 9.2**. The GRS-statistic can be expressed as follows (see appendix 9.2 for details).

$$\left(\frac{T-N-1}{N}\right)\left[\frac{\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}}{1+\bar{\mu}'\hat{\Omega}^{-1}\bar{\mu}}\right] \sim F(N, T-N-K) \tag{24}$$

Using this test statistic, we can test if the intercepts in our estimated models are equivalent to 0 and compare the new models to our benchmark. If one of the models prove to have intercepts that are jointly equivalent to 0 and the others don't, we can determine that this model is better at modelling variations in asset returns than the other. However, we do not except this to be the case for any model as such a result implies that the model is perfectly specified and covers the full set of anomalies available. Either way, the GRS-statistic can be used to compare performance across several models (See Fama & French (2015)). In this case, a lower GRS-statistic will imply that the intercepts of the model are jointly closer to zero.

4.5.2 Fama-MacBeth – T-ratios of intercepts

Since the Fama-MacBeth estimation procedure grants us t-statistics for all estimated coefficients, including the intercept, we can use it to test the performance of the model. The same intuition as for the GRS-test applies, i.e. if the intercept is equivalent to 0 the model explains all variation in asset returns. Hence, we can use the t-statistic from the Fama-Macbeth estimation to test if a model's intercept is 0. Even though it is very unlikely that this will be the case, the technique adds value to the overall analysis as it shows *how* close to 0 the intercept is and the statistical significance of the value, which we can use to compare models.

4.5.3 A Simple Intercept Study

The simple intercept study is based on the intuition that the closer the intercept of the estimated model is to 0 in absolute terms, the better specified the model is. Even though this method seems simple, it is widely used and applied in well known asset pricing papers such as Fama and French (2015) and Hou, Xue and Zhang (2014). Hence, in the spirit of these famous research papers, we will also study the intercepts of our estimated models where we look at both the average absolute values of intercepts and the average value of intercepts and compare them across our three main models, for all of the test factors (left-hand side portfolios).

$$Average\ absolute\ value\ of\ intercepts = \frac{|\alpha|}{N} \tag{25}$$

$$Average\ value\ of\ intercepts = \frac{\alpha}{N} \tag{26}$$

Where N refers to the number of intercepts. Since we have 3x25 left hand portfolios for each given model, N=25.

4.5.4 The Barillas and Shanken (2015) Approach

Barillas and Shanken (2015) argues that the *regular* approaches to asset pricing comparisons *cannot serve to identify the superior model and can even be misleading in this regard*. Hence, as a safeguard and for the purpose of increasing the robustness of our conclusions, we will additionally apply the methodology as described in Barillas and Shanken (2015) to compare our models. This method is a direct extension of the GRS-approach and uses the quadratic form of the alphas, $\hat{\propto}\,'\hat{\Sigma}^{-1}\hat{\alpha}$, to compare asset pricing models. It asks us to compare two (or more) models based on their maximum squared Sharpe ratio, $Sh^2(\alpha)$, which equals the quadratic form of the alphas, $\hat{\propto}\,'\hat{\Sigma}^{-1}\hat{\alpha}$.

$$Sh^2(\alpha) = \hat{\propto}\,'\hat{\Sigma}^{-1}\hat{\alpha} \tag{27}$$

Assuming now that we have two models, each with their own maximum squared Sharpe ratio, $Sh_1^2(\alpha)$ and $Sh_2^2(\alpha)$, the one with the lowest $Sh^2(\alpha)$ will be the best at explaining variation in asset returns given the factors included in the model, no matter the test asset returns. Mathematically, this is the case when $Sh_1^2(\alpha) < Sh_2^2(\alpha)$. For this paper, we will calculate the maximum squared Sharpe ratio for each of the three estimated models and compare them to determine which one is the *best*, i.e. has the lowest $Sh^2(\alpha)$.

5.0 Data

Since the procedure of constructing factors is the crux of the methodology, we have applied a SAS sample program that extracts data from Compustat and Centre for Research in Security Prices (CRSP) to replicate all of the Fama French factors as well as our alternative factors (SAS Program, 2018). The applied procedure is how Fama and French (1993) constructed their three-factor model. To effectively evaluate the factor models, the monthly factor returns will be from January 1970 to June 2015 with a total of 546 observations. The reason for this specific start date is because when matching with the data Kenneth French publishes, we discover that during the sixties the two sets are noticeably non-comparable. These deviations might be related to the fact that book equity is not completely reported in the historical time series of Compustat. Selecting a sample that starts from 1970, we observe that both the HML and SMB factors have correlation coefficients greater than 98%, comparing with the data published by Kenneth French. Hence, we will by modifying this program construct the factors, HML, SMB, RMW, CMA, GMP and EMI. The Fama-French factor MKRF (or RM-RF) is the excess return on the market. This factor is defined as the value-weighted return on all NYSE, AMEX and NASDAQ stocks from CRSP after subtracting the one-month Treasury bill rate.

The sample, which all of the factors mentioned above are based on, includes all common stocks that are available in the CRSP stock files (stocks with share code 10 or 11). There are also some requirements that must be fulfilled for a firm to be included, such as matching entries for the company in Compustat for at least two years. Another requirement is that the firm must have CRSP stock price for December last year, t-1 and June the current year, t. Furthermore, the firm need to have book value of common equity for the previous year, t-1 based on data from Compustat.

5.1 Procedure

To be able to calculate the variables which the portfolios are constructed upon, we extract historical accounting data (annual) using Compustat. Book equity is calculated by taking the book value of stockholders` equity plus balance sheet deferred taxes and investment tax credit minus book value of preferred stocks. According to Kenneth French`s webpage it has been changes to how deferred taxes are treated (as of May 2018). However, we do not tweak the original formula since our last observation goes no longer than June 2015. Furthermore, ROE is defined as net income divided by shareholders` equity and ROIC is defined as net income minus total dividends divided by total invested capital. The operating profitability and investment factor are calculated as described in section *4.3 "LHS portfolios".*

The next step consists of merging CRSP "event" and "time-series" files. The "event" files include historical data on exchange and share codes which are crucial to identify the listed firms, common stocks as well as delisting returns. Delisting returns were added when calculating market capitalization to reduce any potential bias in the portfolio returns. Market equity at June and December was finally flagged to serve two main purposes, (1) market equity for December will be used to generate the book-to-market ratio and (2) market equity for June must be positive to be included in the portfolio.

The last step is then to merge CRSP and Compustat by matching the global company key (gvkey, year t-1) from Compustat to the unique permanent security identification number (PERMNO, year t) from CRSP. After the data was cleaned for unnecessary duplicates, portfolios and factors were constructed as explained in section *4.2 Factor definitions – RHS factors.*

5.2 Summary Statistics

Table 5.1

Panel A: Correlation between different factors for the FF5 model

	RM-RF	SMB_{FF5}	HML	RMW	CMA
RM-RF	1,00	0,24	-0,32	-0,19	-0,44
SMB_{FF5}	0,24	1,00	-0,12	-0,27	-0,43
HML	-0,32	-0,12	1,00	0,16	0,47
RMW	-0,19	-0,27	0,16	1,00	0,37
CMA	-0,44	-0,43	0,47	0,37	1,00

Panel B: Correlation between different factors for the FF5new model

	RM-RF	SMB_{FF5new}	HML	RMW	EMI
RM-RF	1,00	0,22	-0,32	-0,19	0,33
SMB_{FF5new}	0,22	1,00	-0,08	-0,21	0,07
HML	-0,32	-0,08	1,00	0,16	-0,41
RMW	-0,19	-0,21	0,16	1,00	-0,01
EMI	0,33	0,07	-0,41	-0,01	1,00

Panel C: Correlation between different factors for the QF5new model

	RM-RF	SMB_{QF5new}	HML	GMP	EMI
RM-RF	1,00	0,23	-0,32	0,04	0,33
SMB_{QF5new}	0,23	1,00	-0,12	-0,18	0,09
HML	-0,32	-0,12	1,00	-0,32	-0,41
GMP	0,04	-0,18	-0,32	1,00	0,25
EMI	0,33	0,09	-0,41	0,25	1,00

Panel D: Correlation between different versions of the SMB factor

	SMB_{FF5}	SMB_{FF5new}	SMB_{QF5new}
SMB_{FF5}	1,00	0,94	0,94
SMB_{FF5new}	0,94	1,00	1,00
SMB_{QF5new}	0,94	1,00	1,00

Panel E: Correlation between different versions of the investment factor, CMA and EMI

	CMA	EMI
CMA	1,00	-0,28
EMI	-0,28	1,00

Panel F: Correlation between different versions of the profitability factor, RMW and GMP

	RMW	GMP
RMW	1,00	0,61
GMP	0,61	1,00

Table 5.1 shows correlation coefficients between the different factors for the three models specified in the text, as well as the correlations between the different versions of the same factor (Panel D-F). The high correlations (0,94 to 1,00)

between the different versions of the SMB factors (Panel D) are not unexpected, because the Size breakpoint will always be the NYSE median market cap and the different versions of SMB use all stocks. We also observe high correlation (0,61) between the different versions of the profitability factor (Panel F). However, when comparing the traditional investment factor (CMA) with the new factor investment factor (EMI) we see that the degree of correlation (-0,28) is relatively small and negative (Panel E).

Analysing the correlation of the new added variable EMI (Panel B and C) with the other factors, we notice that the signs of the correlation coefficients are opposite of those observed for CMA with the other factors (Panel A). The only exception is that the correlation coefficients between the CMA and RMW (Panel A), and EMI and GMP (Panel C) has the same sign.

All the presented models are estimated using linear regression. Hence, in the case when multicollinearity is present, our models may potentially produce biased results. However, Table 5.1 displays that multicollinearity seems not to be an issue for any of the estimated models.

Table 5.2

Summary statistics for monthly factor percent returns: January 1970 - June 2015, 546 months.

Panel A: Averages, standard deviations, and t-statistics for monthly returns for FF5 model

	RM-RF	SMB_{FF5}	HML	CMA	RMW
Mean	0,5253	0,1414	0,3517	0,1574	0,1483
Std dev.	4,5847	3,0314	2,9873	2,7544	1,8153
t-Statistic	2,6773	1,0899	2,7510	1,3353	1,9089

Panel B: Averages, standard deviations, and t-statistics for monthly returns for FF5new model

	RM-RF	SMB_{FF5new}	HML	EMI	RMW
Mean	0,5253	0,2279	0,3517	0,3786	0,1483
Std dev.	4,5847	2,9665	2,9873	3,6967	1,8153
t-Statistic	2,6773	1,7950	2,7510	2,3931	1,9089

Panel C: Averages, standard deviations, and t-statistics for monthly returns for QF5new model

	RM-RF	SMB_{QF5new}	HML	EMI	GMP
Mean	0,5253	0,2280	0,3517	0,3786	0,0858
Std dev.	4,5847	2,9391	2,9873	3,6967	1,7142
t-Statistic	2,6773	1,8127	2,7510	2,3931	1,1696

Table 5.2 shows summary statistics for the different factor returns. Monthly average returns are in a range between 0,09% (GMP) and 0,53% (excess return on market), the former being the least volatile and the latter being the most volatile. The average return on the different versions of the SMB factor are 0,14% to 0,23%, observing a significant increase in the average monthly return comparing the FF5 model (0,14%) to the new models (0,23%), which are almost identical. The standard deviations of the different versions of SMB does not differ that much compared to the variation in average return 2,94 to 3,03. We also note that the Size factor only is significantly different from zero at the 10% level for our alternative models compared to FF5.

Comparing the different versions of the investment factor, we observe a substantial increase in average monthly returns from 0,16% (CMA) to 0,38% (EMI). The EMI factor has a t-statistic (Panel B and C) high enough to reject the hypothesis that the mean is zero at all confidence levels, but for the CMA factor the situation is opposite. With a t-statistic of 1,33 (Panel A) we cannot reject the hypothesis at any confidence level. Looking at the operating profitability and return on equity factor, we see that the average return is lower for the GMP factor (0,09%) compared to RMW (0,15%). We can also for the RMW factor, with a t-statistic of 1,91 (Panel A) reject the hypothesis that the mean is zero at 10 % level. However, we see that GMP with a t-statistic of 1,17 (Panel C) are not significantly different from zero at any confidence level.

6.0 Empirical Study

6.1 Introduction

This thesis sets out to study whether or not a factor based on the return on invested capital, EMI, can replace the traditional investment factor, CMA, and increase the performance of traditional asset pricing models. Our hypothesis is that EMI captures the full variation in asset returns caused by a firm's investment policy, whilst CMA only captures the variation caused the investment in absolute terms. Therefore, we argue that EMI could potentially be a better factor for capturing the variation caused by the firm’s investment policy. The theory behind this trail of taught was illustrated in section 3 of this paper. Section 4 outlines the methodology applied to test our hypothesis. In this section we will present the results of our empirical study.

6.2 The Fama and MacBeth Study – Verification and Factor Significance

The goal of this part is to determine if our newly included factor, EMI, is statistically significant and thus a valid factor for our newly specified asset pricing models. We will also analyse what effects the inclusion of this factor has on the other factors in the models. The results of the Fama-MacBeth procedures are reported in table 6.1**.**

Before discussing the results, we find it important to note that statistical significance, in our case, for right-hand side variables largely depends on which left-hand side test assets are used in the estimation of the models. Most notably, we observe that our models seem to have high explanatory power on left-hand side portfolios based on size-B/M. Lewellen, Nagel, and Shanken (2010) argues that left-hand side portfolios based on size-B/M suffer from high correlation between the left-hand side portfolios themselves and the factors included, and thus provide little economic meaning. We acknowledge that this might be the case for our study as well. However, we also observe that many revered research papers on asset pricing theory (see Fama and French (2015) and Hou, Xou and Zhang (2014)) still base many of their conclusions on these types of test assets. Therefore, we suggest that further research on this topic should include tests on a wider variety of left-hand side portfolios.

Table 6.1

The table reports parameter estimates (in percentages) and their corresponding t-ratios (in parantheses) obtained from the second stage Fama Macbeth tests on the VW portfolios formed on Size-BM, Size-Inv, and Size-OP for the three asset pricing models specified in the text. Column (1) displays the test asset, Cloumn (2) to Column (6) reports the slope coefficients for the different factors and Column (7) reports the intercept. ***,** and * represents the significance at 1 %, 5% and 10% significance levels respectively.

Panel A: FF5 model

	$\widehat{\gamma_{RMRF}}$	$\widehat{\gamma_{SMB}}$	$\widehat{\gamma_{HML}}$	$\widehat{\gamma_{CMA}}$	$\widehat{\gamma_{RMW}}$	$\widehat{\gamma_{FF5}}$
Size-B/M	0,3842	0,2401*	0,3655***	0,7377***	0,0281	0,8959***
	(1,0651)	(1,7592)	(2,7775)	(2,6453)	(0,1437)	(2,9211)
Size-Inv	0,40312	0,1163	0,4791***	0,5214*	-0,3408**	0,9985***
	(1,1261)	(0,8675)	(2,8944)	(1,8916)	(-2,2844)	(3,3043)
Size-OP	0,4269	0,1736	0,1563	0,4687*	0,1896**	0,1336
	(1,1243)	(1,2886)	(0,7648)	(1,6779)	(2,0209)	(0,4123)

Panel B: FF5new model

	$\widehat{\gamma_{RMRF}}$	$\widehat{\gamma_{SMB}}$	$\widehat{\gamma_{HML}}$	$\widehat{\gamma_{EMI}}$	$\widehat{\gamma_{RMW}}$	$\widehat{\gamma_{FF5new}}$
Size-B/M	0,3844	0,2169	0,3753***	1,2082**	0,4428***	0,9062***
	(1,0421)	(1,5188)	(2,8694)	(2,1149)	(2,3975)	(2,8872)
Size-Inv	0,3951	0,0000	0,6471***	1,2844	-0,1098	0,9901***
	(1,0705)	(0,0003)	(3,9427)	(1,2523)	(-0,7889)	(3,1535)
Size-OP	0,3448	0,2655*	0,2378	-0,8817	0,2766***	0,2219
	(0,9037)	(1,7608)	(1,1939)	(-0,8904)	(2,7234)	(0,6831)

Panel C: QF5new model

	$\widehat{\gamma_{RMRF}}$	$\widehat{\gamma_{SMB}}$	$\widehat{\gamma_{HML}}$	$\widehat{\gamma_{EMI}}$	$\widehat{\gamma_{GMP}}$	$\widehat{\gamma_{QF5new}}$
Size-B/M	0,4194	0,2177	0,3772***	1,1904**	0,4156**	0,9439***
	(1,1649)	(1,5351)	(2,8825)	(2,0541)	(2,1640)	(3,1157)
Size-Inv	0,3215	-0,0043	0,6334***	1,2073	-0,1766	0,9093***
	(0,8803)	(-0,0287)	(3,8563)	(1,1633)	(-1,2760)	(2,9434)
Size-OP	0,3891	0,2211	0,3285*	-0,3626	0,1976*	0,1679
	(0,9947)	(1,4624)	(1,6822)	(-0,3784)	(1,8524)	(0,4992)

6.2.1 FF5new

From table 6.1 we can see that EMI is statistically significant at a 5% confidence level with a factor loading of 1,2082% on size-B/M portfolios and only on size-B/M portfolios (Panel B). A factor loading of 1,2082%, in the Fama-Macbeth fashion, implies that a *one unit* increase in the exposure to the risk factor EMI, holding exposure to every other factor as 0, will yield 1,2082% in return as compensation. The fact that EMI is only statistically significant on Size-B/M portfolios raises concerns about our model. Lewellen, Nagel, and Shanken (2010) argues that tests on left-hand side portfolios based on size-B/M provide little economic meaning because of high correlation between the left-hand side portfolios themselves and the factors included. Hence, these results might suggest that the FF5new-model performs rather poorly at explaining variations in asset returns. However, the results also suggest that the FF5new-model performs at a level comparative to the FF5-model. On left-hand side test assets based on Size-B/M we see that RMW is statistically significant at all confidence levels in the FF5new-model, but statistically insignificant at all confidence levels in the FF5-model. For the same test assets, we also observe that SMB is statistically significant at a 10% confidence level in the FF5-model but drops completely out in the FF5new-model. Similar results between the models can be observed for left-hand side portfolios based on Size-INV and Size-OP. As support for continuation of the use of the FF5new model given that EMI only is statistically significant on Size-B/M, we lean on results from Fama & French (2015) and Hou, Xou and Zhang (2014) which also base many of their conclusions on Size-B/M test assets. In addition, since statistical significance is only found for Size-B/M portfolios, we find it unnecessary to discuss the results of the FF5new-model on Size-INV and Size-OP. However, we would like to note that these results imply that the FF5new-model is likely to underperform when used on strong- profitability and investment portfolios. We suggest that further research on this topic should be focused on testing the FF5new-model using a wider variety of left-hand side portfolios.

We observe higher slopes for the EMI-factor in the FF5new-model than for the CMA-factor in the FF5-model. Referring back to the traditional asset pricing theory of Markowitz (1952, 1959), Sharpe (1964), Lintner (1965) and Black (1972), a higher slope coefficient implies that the risk factor in question has a higher correlation with the test assets. Since the slope coefficients in asset pricing models can be interpreted as the risk premium of that particular factor and the marginal effect of EMI is greater that the marginal effect of CMA, we can deduct that investors should be compensated with higher returns for exposure to EMI than CMA. Hence, these results suggest that a portfolio of EMI stocks carries more risk than a portfolio of CMA stocks. Since higher variations in stock returns are caused by higher levels of risk, i.e. volatility, we suggest that these results show that EMI seems to capture more of the variation in asset returns caused by a firm’s investment policy than CMA does.

An observation of special interest is that the inclusion of EMI renders the Size factor (SMB) insignificant (significant at 10% in FF5). A natural question to raise is therefore whether or not EMI acts as a proxy for SMB. Banz (1981) was one of the researchers who raised the question if Size (SMB) actually is related to higher expected returns or whether it only proxies for other unidentified factors correlated with Size. In Table 5.1 (Panel B) we observe a correlation coefficient of 0,07 between SMB and EMI making it difficult to conclude that the new investment factor (EMI) could be a potential proxy for the Size factor (SMB). Horowitz, Loughran and Savin (2000) continued to study the relationship between Size and expected return by extending the data set from Fama and French (1992). Their results show that the Size effect disappears in the period 1982-1997 compared with the huge Size premiums that existed during the 1963-1981 period, concluding that the Size factor is not a systematic proxy for risk. Our results seems to corroborate these findings as SMB is only significant at a 10% confidence level in FF5 and insignificant for the two other models at all significance levels. A possible explanation could be that these results are a direct effect of the huge body of literature on this topic. With all this information available, investors started to trade on it. By investing in Small firms, the investors bid up the prices leading to a decline in average returns for the Small firms relative to the Big firms, which in turn lead to the disappearance of the Size effect.

Another interesting observation related to statistical significance is that the RMW factor becomes significant upon the inclusion EMI. A potential explanation as to why this might be the case can be found in the correlation matrix of the variables. We observe a correlation coefficient of 0,37 between CMA and RMW for the FF5F model, whilst we have a correlation coefficient close to 0 between EMI and RMW for the FF5new model. Since EMI does not pick of any of the variation caused by RMW, the entire model isolates the effect of RMW, ultimately making it significant.

To conclude our findings related to the FF5new model, we do acknowledge that FF5new seems to perform rather poorly for left-hand side test assets based on Size-OP and Size-Inv and that observations drawn from Size-B/M tests assets might potentially be questionable. However, we also argue that the estimated model is comparable to the FF5-model. We will therefore continue to study the FF5new-model but disregard estimations on Size-OP and Size-Inv test assets.

6.2.2 QF5new

Much of the story from the FF5new-model is the same for the QF5new-model. We observe statistical significance for EMI at a 5% confidence level only for left hand side portfolios formed on size-B/M. We observe a factor loading of 1,1904% for EMI, which implies that a *one unit* increase in the exposure to the risk factor EMI, holding exposure to every other factor as 0, will yield 1,1904% in return as compensation. As was the case with the FF5new-model, EMI seems only to be statistically significant for left-hand side test assets based on Size-B/M. Our concerns and arguments regarding this remain the same as for the FF5new-model since the only difference in specification between the two models is related to the construction of the profitability factor. Hence, we will also for the QF5new model only focus on results from the model estimation using Size-B/M as the left-hand side test assets. We would also like to state the the QF5new-model is likely to underperform when used on strong- profitability and investment portfolios, which was also the case with the FF5new-model. Hence, we suggest that further research on this topic should be focused on testing the QF5new-model using a wider variety of left-hand side portfolios.

Comparing the factor loadings of EMI in both models, we observe a factors loading of 1,1904 for EMI in the estimated QF5new-model and a factor loading of 0,7377 for CMA in the estimated FF5-model. This is a similar case as that found in the comparison of the FF5new-model and the FF5-model. Hence, it will also here seem that the EMI factor captures a greater portion of the variation in asset returns caused by a firm’s investment policy than the CMA factor does.

As with the FF5new-model, we observe that exchanging CMA with EMI makes the Size factor statistically insignificant and the profitability factor significant. When it comes to the Size factor, we apply the same reasoning for the QF5new-model as discussed under the FF5new-model. Our results seem to be somewhat consistent with Hou, Xue and Zhang (2015) as they also report that the factor loading on the profitability factor is significant at all confidence levels for left-hand side portfolios based on Size-BM. However, comparing the factor loading on the Size factor, we see that Hou, Xue and Zhang (2015) reports that the Size factor is significant at all confidence levels.

6.2.3 FF5new vs QF5new

The difference between the factor loading of EMI in the FF5new model and the QF5new model is very small and amounts to 0,0178%. We see that the results between the Q5new and FF5 are almost identical as the only way these two models differ is with respect to the profitability factor. The only notable difference in the results is that while the profitability factor for the Q5new model (GMP) only is significant at a 5 % confidence level, the profitability factor for the FF5new model (RMW) is statistically significant at all confidence levels.

6.3 Studying Intercepts and Comparing Models

The goal of this part is to compare the various measurements of performance, as defined in section 3 of this paper. Based on the findings in part 6.2, we will only focus on the models where the EMI-factor proved to be statistically significant, which was the case for models estimated with left-hand side test assets formed on Size-B/M.

Before we discuss the performance of our estimated models, we find it important to note that all asset pricing models are simplified explanations of the variation in asset returns. None are perfect and like in Fama and French (2015), we want to identify the best, yet imperfect model, to explain these variations. Hence, we care more about the relative performance of the models rather than if they perfectly model variations in asset returns or not. In addition, since we only are evaluation estimated models with left-hand side test assets formed on Size-B/M, our observations and conclusions will be limited to the models performance on similar Size-B/M test assets.

Table 6.2

The table tests the ability of the three models specifiied in the text to explain monthly excess returns on 25 Size-BM portfolios, 25 Size-OP portfolios, and 25 Size-Inv portfolios. For each set of 25 regressions, the table (by column) displays: (1) the LHS portfolios (2) average absolute value of the intercepts, (3) average value of the intercepts, (4) maximum squared Sharpe ratio, (5) the GRS statistic testing whether the expected values of all 25 intercepts are zero, (6) the p-value of the GRS statistics, (7) the intercepts from the second stage Fama Macbeth tests and (8) the corresponding t-statisitics in parantheses. ***,** and * represents the significance at 1 %, 5% and 10% significance levels respectively.

Panel A: FF5F

	$\lvert\bar{\alpha}\rvert$	$\bar{\alpha}$	$Sh^2(\alpha)$	GRS	p(GRS)	$\widehat{\gamma_{FF5}}$	t-statistic
Size-B/M	0,0011	0,0003	0,2033	4,0159	0,0000	0,8959***	(2,9211)
Size-Inv	0,0014	0,0008	0,2362	4,6664	0,0000	0,9985***	(3,3043)
Size-OP	0,0007	0,0003	0,0991	1,9575	0,0040	0,1336	(0,4123)

Panel B: FF5new

	$\lvert\bar{\alpha}\rvert$	$\bar{\alpha}$	$Sh^2(\alpha)$	GRS	p(GRS)	$\widehat{\gamma_{FF5new}}$	t-statistic
Size-B/M	0,0009	-0,0002	0,1757	3,4246	0,0000	0,9062***	(2,8872)
Size-Inv	0,0011	0,0004	0,2259	4,4019	0,0000	0,9901***	(3,1535)
Size-OP	0,0006	-0,0001	0,0973	1,8963	0,0058	0,2219	(0,6831)

Panel C: QF5new

	$\lvert\bar{\alpha}\rvert$	$\bar{\alpha}$	$Sh^2(\alpha)$	GRS	p(GRS)	$\widehat{\gamma_{QF5new}}$	t-statistic
Size-B/M	0,0010	-0,0003	0,1767	3,4339	0,0000	0,9439***	(3,1157)
Size-Inv	0,0011	0,0003	0,2292	4,4534	0,0000	0,9093***	(2,9434)
Size-OP	0,0006	-0,0002	0,0994	1,9311	0,0047	0,1679	(0,4992)

As illustrated by table 6.2, the GRS statistic easily rejects the null hypothesis of jointly zero intercepts for all three models. These results are as expected as this also was the case in Fama and French (2015). Hence, these results underline the fact that no asset pricing model is perfect and that there exists an unknown set of asset pricing anomalies that are yet to be discovered. Moving on, since we are interested in improvements in model performance caused by the replacement of CMA with EMI, we also benefit from looking at the GRS-statistic itself. We observe that both the FF5new-model and the QF5new-model have lower GRS-values than the FF5-model. These results suggest that both the new models should perform better at modelling variations in asset returns than the benchmark FF5-model, hence these results imply that both models are better specified than the benchmark. Isolating and comparing the two new models, we see little difference in their performance other that the FF5new having a marginally lower GRS statistic than the QF5new. In addition, and even though we are disregarding Size-OP estimation, we observe that our new models, as well as the FF5-model, fare best when using test asset portfolios based on Size-OP corroborating the results from Fama and French (2015).

Looking at the statistical significance of the intercepts from the Fama-MacBeth regressions, we see that all three are relatively close to 0 and statistically significant at all confidence levels. A *perfect* model would have statistically insignificant intercepts, implying that the intercepts themselves carry no explanatory power and are practically zero. Since no models are perfect, we have to compare the relative value of intercepts and favor the one closest to zero. The FF5-model seems to perform marginally better than the FF5new (intercept of 0,9062%) and the QF5new (intercept of 0,9439%) with its 0,8995% intercept. However, we observe that the intercept for the three models are statistically significant at all confidence levels. The fact that the intercepts are statistically significant stops us from making a firm conclusion whether the FF5-model performs better than the other two models, for the sample we examine, using this approach.

Moving over to the absolute average and average intercepts of the three models, we see little to no difference between the models. At a first glance, one observes that the FF5new and QF5new models outperform the FF5F model, but the differences are so marginal that we cannot draw sensible conclusions based on them.

The maximum squared Sharpe ratio tells a rather convincing story about the differences in model performance. We see that both the FF5new-model and the QF5new-model clearly outperform the FF5-model. Whilst the FF5-model has a maximum squared Sharpe ratio of 0,2033, the FF5new-model has a ratio of 0,1757 and the QF5new-model has a ratio of 0,1767. Hence, these results show that both the QF5new- and FF5new models are better specified than the benchmark FF5 model and should perform better on portfolios formed on size-B/M. The two models have an almost identical maximum squared Sharpe ratio, making it difficult for us to conclude one to be better than the other.

7.0 Conclusion

The purpose of this thesis was to test whether a factor based on the return on invested capital (dubbed EMI) could potentially be a better factor to model variations in asset returns created by a firm's investment policy, rather than a factor based on the absolute value of investment. Based on theoretical intuition from well-known corporate finance concepts, we deducted that this could be the case and thus formulated the null hypothesis that a model including EMI rather than CMA could potentially outperform the classical Fama-French Five Factor model.

To test our hypothesis, we applied several well-known and widely applied techniques from previous studies on asset pricing. Inspired by Fama and French (2015), much of our empirical study focuses on studying the intercepts of our estimated models to compare their performance. Several intercept studies were used to increase the robustness of our results. We used the Fama-MacBeth approach to test intercepts for statistical significance, the GRS-test to test if the intercepts were jointly zero, a simple average of the intercepts to see which one was the closest to zero and the Barillas and Shanken Sharpe Ratio approach. The GRS-test easily rejects the null hypothesis of jointly zero intercepts for all models. However, no asset pricing model is perfect, and these results are as expected. The GRS-statistic itself favours the models which include EMI rather than CMA, i.e. FF5new and QF5new over FF5. The difference between the QF5new-model and FF5new-model is marginal and we cannot determine which one is better than the other based on this test alone. Moving on to the Fama-MacBeth intercept study, we find the only indication of the FF5-model being the better. The Fama-MacBeth intercepts shows that the FF5-model outperforms the QF5new-model but is only marginally better than the FF5new-model. As for the simple intercept study, we observe that both the FF5new-model and the QF5new-model outperform the FF5-model. However, the differences are so marginal that we are careful to draw conclusions based on these results. The maximum squared Sharpe ratio strongly favours both new models over the benchmark FF5-model. The difference between the two models themselves, however, is marginal.

Our findings show that both the QF5new-model and the FF5new-model might be better specified than the benchmark FF5-model and that this increase in performance is solely due to the inclusion of the EMI factor as this is the only notable difference between the models. They also suggest that there is another dimension to investment risk than what previous theory suggests. Where previous theory only looks at the absolute investment a firm undertakes and implicitly states that investors are compensated only for the absolute amount of cash a firm invests, this thesis suggests that investors receive an additional compensation for the risk related to how management in a firm uses these funds and the returns (losses) they obtain. This is reflected by the higher slope estimates for the EMI-factor in both FF5new-model and QF5new-model than what is observed for the CMA-factor in the FF5-model, which implies higher risk premiums for exposure to EMI than for exposure to CMA. This is rather intuitive as whether a firm adds value through its investments or not, and thus obtains higher returns for their shareholders, is solely determined by the returns on these investments. However, a major drawback of these findings is that they are based on the use of Size-B/M left-hand side portfolios as we could not find statistical significance for EMI on any other of the test-assets used in the study. Hence, we cannot conclude that FF5new and QF5new outperform the FF5 on anything else than Size-B/M test-assets and our results become somewhat limited. In addition, Lewellen, Nagel, and Shanken (2010) argues that studies on left-hand side portfolios based on Size-B/M provide little economic meaning as there tends to be high correlation between the portfolios themselves and the factors included. Therefore, we strongly suggest that future research on this topic should be focused on conducting these tests using a wider variety of left-hand side portfolios.

8.0 Bibliography

Aharoni, G., Grundy, B., & Zeng, Q. (2013). Stock returns and the Miller Modigliani valuation formula: Revisiting the Fama French analysis. Journal Of Financial Economics, 110(2), 347-357. doi: 10.1016/j.jfineco.2013.08.003

Baker, M., & Wurgler, J. (2011). Behavioral Corporate Finance: An Updated Survey. SSRN Electronic Journal. doi: 10.2139/ssrn.1909013

Ball, R. (1978). Anomalies in relationships between securities' yields and yield-surrogates. Journal Of Financial Economics, 6(2-3), 103-126. doi: 10.1016/0304-405x(78)90026-0

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

Barillas, F., & Shanken, J. (2015). Comparing Asset Pricing Models. SSRN Electronic Journal. doi: 10.2139/ssrn.2676709

Barillas, F., & Shanken, J. (2018). Comparing Asset Pricing Models. The Journal Of Finance, 73(2), 715-754. doi: 10.1111/jofi.12607

Basu, S. (1983). The relationship between earnings' yield, market value and return for NYSE common stocks. Journal Of Financial Economics, 12(1), 129-156. doi: 10.1016/0304-405x(83)90031-4

Black, F. (1972). Capital Market Equilibrium with Restricted Borrowing. The Journal Of Business, 45(3), 444. doi: 10.1086/295472

Black, F., Jensen, M., & Scholes, M. (1971). The Capital Asset Pricing Model: Some Empirical Tests. Studies In The Theory Of Capital Markets, 79-12.

Blume, M. (1970). Portfolio Theory: A Step Toward Its Practical Application. The Journal Of Business, 43(2), 152. doi: 10.1086/295262

Brainard, W., & Tobin, J. (1968). Pitfalls in Financial Model Building. The American Economic Review, 58(2), 99-122. doi: 10.2307/1831802

Brav, A., Graham, J., Harvey, C., & Michaely, R. (2004). Payout Policy in the 21st Century. SSRN Electronic Journal. doi: 10.2139/ssrn.571046

Brooks, C. (2017). Introductory econometrics for finance. Cambridge: Cambridge University Press.

Chan, K., & Chen, N. (1991). Structural and Return Characteristics of Small and Large Firms. The Journal Of Finance, 46(4), 1467. doi: 10.2307/2328867

Chen, L., Novy-Marx, R., & Zhang, L. (2011). An Alternative Three-Factor Model. SSRN Electronic Journal. doi: 10.2139/ssrn.1418117

Clarke, K. (2005). The Phantom Menace: Omitted Variable Bias in Econometric Research. Conflict Management And Peace Science, 22(4), 341-352. doi: 10.1080/07388940500339183

Cochrane, J. (2008). Asset Pricing. Princeton: Princeton University Press.

De Bondt, W., & Thaler, R. (1985). Does the Stock Market Overreact?. The Journal Of Finance, 40(3), 793-805. doi: 10.1111/j.1540-6261.1985.tb05004.x

Fama, E. (1965). The Behavior of Stock-Market Prices. The Journal Of Business, 38(1), 34. doi: 10.1086/294743

Fama, E., & French, K. (1992). The Cross-Section of Expected Stock Returns. The Journal Of Finance, 47(2), 427. doi: 10.2307/2329112

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

Fama, E., & French, K. (1997). Industry costs of equity. Journal Of Financial Economics, 43(2), 153-193. doi: 10.1016/s0304-405x(96)00896-3

Fama, E., & French, K. (2015). International Tests of a Five-Factor Asset Pricing Model. SSRN Electronic Journal. doi: 10.2139/ssrn.2622782

Fama, E., & French, K. (1996). Multifactor Explanations of Asset Pricing Anomalies. The Journal Of Finance, 51(1), 55-84. doi: 10.1111/j.1540-6261.1996.tb05202.x

Fama, E., & French, K. (2003). The Capital Asset Pricing Model: Theory and Evidence. SSRN Electronic Journal. doi: 10.2139/ssrn.440920

Fama, E., & French, K. (1996). The CAPM is Wanted, Dead or Alive. The Journal Of Finance, 51(5), 1947-1958. doi: 10.1111/j.1540-6261.1996.tb05233.x

Fama, E., & French, K. (2008). Dissecting Anomalies. The Journal Of Finance, 63(4), 1653-1678. doi: 10.1111/j.1540-6261.2008.01371.x

Fama, E., & MacBeth, J. (1973). Risk, Return, and Equilibrium: Empirical Tests. Journal Of Political Economy, 81(3), 607-636. doi: 10.1086/260061

Francis, J., Olsson, P., & Oswald, D. (2000). Comparing the Accuracy and Explainability of Dividend, Free Cash Flow, and Abnormal Earnings Equity Value Estimates. Journal Of Accounting Research, 38(1), 45. doi: 10.2307/2672922

French, K. (2018). Kenneth R. French - Home Page. Retrieved from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

Gibbons, M., Ross, S., & Shanken, J. (1989). A Test of the Efficiency of a Given Portfolio. Econometrica, 57(5), 1121. doi: 10.2307/1913625

Gregory, A., Tharyan, R., & Christidis, A. (2013). Constructing and Testing Alternative Versions of the Fama-French and Carhart Models in the UK. Journal Of Business Finance & Accounting, 40(1-2), 172-214. doi: 10.1111/jbfa.12006

Hamada, R. (1972). The Effect of the Firm's Capital Structure on the Systematic Risk of Common Stocks. The Journal Of Finance, 27(2), 435. doi: 10.2307/2978486

Horowitz, J., Loughran, T., & Savin, N. (2000). The disappearing size effect. Research In Economics, 54(1), 83-100. doi: 10.1006/reec.1999.0207

Hou, K., Xue, C., & Zhang, L. (2014). A Comparison of New Factor Models. SSRN Electronic Journal. doi: 10.2139/ssrn.2520929

Hou, K., Xue, C., & Zhang, L. (2014). Digesting Anomalies: An Investment Approach. Review Of Financial Studies, 28(3), 650-705. doi: 10.1093/rfs/hhu068

Jegadeesb, N., & Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. The Journal Of Finance, 48(1), 65-91. doi: 10.1111/j.1540-6261.1993.tb04702.x

Jensen, M. (1968). The Performance of Mutual Funds in the Period 1945-1964. The Journal Of Finance, 23(2), 389. doi: 10.2307/2325404

Jensen, M. (2002). The Performance of Mutual Funds in the Period 1945-1964. SSRN Electronic Journal. doi: 10.2139/ssrn.244153

Kim, D. (1995). The Errors in the Variables Problem in the Cross-Section of Expected Stock Returns. The Journal Of Finance, 50(5), 1605. doi: 10.2307/2329328

Koller, T., & Goedhart, M. (2015). Valuation. Hoboken: Wiley.

Lakonishok, J., Shleifer, A., & Vishny, R. (1994). Contrarian Investment, Extrapolation, and Risk. The Journal Of Finance, 49(5), 1541. doi: 10.2307/2329262

Lewellen, J., Nagel, S., & Shanken, J. (2010). A skeptical appraisal of asset pricing tests☆. Journal Of Financial Economics, 96(2), 175-194. doi: 10.1016/j.jfineco.2009.09.001

Lewellen, J., Nagel, S., & Shanken, J. (2010). A skeptical appraisal of asset pricing tests. Journal Of Financial Economics, 96(2), 175-194. doi: 10.1016/j.jfineco.2009.09.001

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

Markowitz, H. (1952). Portfolio Selection. The Journal Of Finance, 7(1), 77. doi: 10.2307/2975974

Merton, R. (1973). An Intertemporal Capital Asset Pricing Model. Econometrica, 41(5), 867. doi: 10.2307/1913811

Miller, M., & Modigliani, F. (1961). Dividend Policy, Growth, and the Valuation of Shares. The Journal Of Business, 34(4), 411. doi: 10.1086/294442

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

Penman, S. (1991). An Evaluation of Accounting Rate-of-return. Journal Of Accounting, Auditing & Finance, 6(2), 233-255. doi: 10.1177/0148558x9100600204

Perold, A. (2004). The Capital Asset Pricing Model. Journal Of Economic Perspectives, 18(3), 3-24. doi: 10.1257/0895330042162340

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

Roll, R. (1978). Ambiguity when Performance is Measured by the Securities Market Line. The Journal Of Finance, 33(4), 1051. doi: 10.2307/2326939

Rosenberg, B., Reid, K., & Lanstein, R. (1985). Persuasive evidence of market inefficiency. The Journal Of Portfolio Management, 11(3), 9-16. doi: 10.3905/jpm.1985.409007

Roy, A., & Markowitz, H. (1961). Portfolio Selection. Econometrica, 29(1), 99. doi: 10.2307/1907702

SAS Program, W. (2018). WRDS SAS Program. Retrieved from https://wrds-www.wharton.upenn.edu/pages/support/applications/risk-factors-and-industry-benchmarks/fama-french-factors/

Services, W. (2018). Wharton Research Data Services. Retrieved from https://wrds-www.wharton.upenn.edu/pages/support/applications/risk-factors-and-industry-benchmarks/fama-french-factors/#fama-french-replication-sas-sample-program

Sharpe, W. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. The Journal Of Finance, 19(3), 425. doi: 10.2307/2977928

Stattman, D. (1980). Book values and stock returns. The Chicago MBA: A Journal Of Selected Papers, 4, 25–45.

Stuart, A., & Markowitz, H. (1959). Portfolio Selection: Efficient Diversification of Investments. OR, 10(4), 253. doi: 10.2307/3006625

Titman, S., Wei, K., & Xie, F. (2003). Corporate Groups, Capital Investments and Stock Returns in Japan. SSRN Electronic Journal. doi: 10.2139/ssrn.402661

Tobin, J. (1969). A General Equilibrium Approach To Monetary Theory. Journal Of Money, Credit And Banking, 1(1), 15. doi: 10.2307/1991374

Wooldridge, J. (2011). Econometric analysis of cross section and panel data. Cambridge, Mass. [u.a.]: MIT Press.

9.0 Appendix

9.1 Q-Factor Derivation

We will here illustrate the conceptual framework for the q-factor model as defined in Hou et. al. (2015) using the following assumptions:

i. We have an economy with two economical actors, households and heterogeneous firms, each of which have 1 through N participants
ii. The households are expected to maximize their utility for both periods and have a time preference, $\rho = U(C_0) + \rho E_0[U(C_1)]$.
iii. Firms produce one commodity, which can either be consumed or invested.
iv. Firm *i* produces in both periods and that it starts with productive assets A_{i0}.
v. Firms will exit at the end period with a liquidation value of 0.
vi. The operating cash flow for firm *I* is defined as $\pi_{it}A_{it}$, for t=0 and t=1. The firms operating profitability for period 1, π_{i1}, is subject to a vector of systematic factors that affect all firms simultaneously and is in addition subject to a vector of unsystematic factors affecting only firm *i*.

We can now derive the following. Denoting I_{i0} as investment for period 0, we can say that $A_{i1} = I_{i0}$ because firms exit at the end period with a liquidation value of 0, which in turn implies that A_{i0} fully depreciates at the beginning of period 1. To carry out additional investments, the firm has to bear a quadratic adjustment cost defined as follows.

$$(a/2)(I_{i0}/A_{i0})^2 A_{i0}, \text{ where a>0 is a constant parameter.}$$

Assuming that the household side is standard and defining P_{it} as the ex-dividend equity for firm *i* and D_{it} as the dividend for firm *i*, we have that the first principle of consumption states that

$$P_{i0} = E_0[M_1(P_{i1} + D_{i1})] \text{ or } E_0[M_1 r_{i1}^s] = 1$$

Where $r_{i1}^{s} \equiv \frac{P_{i1}+D_{i1}}{P_{i0}}$ is the stock return and $M_1 \equiv \frac{\rho U\prime(C_1)}{U\prime(C_o)}$ is the stochastic discount factor.

On the production side, firm *i* uses the period 0 operating cash flows to pay for the investment and adjustment costs. We can then define the free cash flow as

$$D_{i0} \equiv \pi_{i0}A_{i0} - I_{i0} - (^{a}/_{2})(^{I_{i0}}/_{A_{i0}})^2 A_{i0}$$

Now note that if the free cash flow in period 0 is positive, the firms will distribute it back to the households. If it turns out to be negative, the firm will have to resort to external equity. Further on, at period 1, the firm will use assets, A_{i1}, to obtain the operating cash flow, $\pi_{i1}A_{i1}$, which in turn gets redistributed as dividends, D_{i1}.

Since we only account for two points in time, the firm will not invest at time 1, resulting in an ex-dividend equity value, P_{i1}, equal to zero. Given that the ex-dividend equity value in period 1 is 0, firm *i* will choose I_{i0} to maximize the cumulative dividend equity value at the beginning of date 0:

$$P_{i0} + D_{i0} \equiv Max\ \{I_{i0}\}$$
$$\pi_{i0}A_{i0} - I_{i0} - (^{a}/_{2})\left(^{I_{i0}}/_{A_{i0}}\right)^2 A_{i0} + E_0[M_1\pi_{i1}A_{i1}]$$

Where the first order condition will give us the first principle of investment:

$$1 + a\frac{I_{i0}}{A_{i0}} = E_o[M_1\pi_{i1}]$$

The first order condition (Euler Equation) illustrates Tobin's Q (Tobin, 1969) and states that firms will continue to invest until the marginal cost of investment is equal to the marginal benefit of investment. Rearranging the equation even further we get the following equation for the expected return on stock *i*.

$$E_0[r_{i1}^{s}] = \frac{E_0[\pi_{i1}]}{1 + a(I_{i0}/A_{i0})}$$

9.2 Terminal Value of Free Cash Flow – Derivations

We below provide a derivation of equation (18). A similar illustration can be found in Appendix A of *Valuation* (Koller & Goedhart, 2015).

The value of a firm's operations can be expressed as follows.

$$V = \frac{FCF_{t=1}}{WACC-g} \tag{9.2.1}$$

Where V is the value of operations; $FCF_{t=1}$ is the free cash flow in year 1; WACC is the weighted average cost of capital; g is the growth in NOPLAT and free cash flow; NOPLAT is the net operation profit less adjusted taxes. $FCF_{t=1}$ can be expressed as follows.

$$FCF_{t=1} = NOPLAT_{t=1}\left(1 - \frac{g}{RONIC}\right)$$

Where RONIC is the return on new invested capital. Adding this formulation of $FCF_{t=1}$ into equation (9.2.1), we get the *key value driver formula*.

$$V = \frac{NOPLAT_{t=1}\left(1-\frac{g}{RONIC}\right)}{WACC-g} \tag{9.2.2}$$

Since the discounted cash flow is equal to the current book value of invested capital plus the present value of future economic profit, it follows, by definition, that NOPLAT at time one can be expressed as follows.

$$NOPLAT_{t=1} = Invested\ capital_0 * ROIC$$

Adding this formulation of $NOPLAT_{t=1}$ into (9.2.2.) we get the following expression of V.

$$V = \frac{Invested\ capital_0 * ROIC\left(1-\frac{g}{RONIC}\right)}{WACC-g} \tag{9.2.3}$$

Assuming that the return of invested capital (ROIC) equals the return in new invested capital (ROIC), we can simplify equation (9.2.3) to the following equation.

$$V = Invested\ capital_0 * \left(\frac{ROIC - g}{WACC - g}\right) \tag{9.2.4}$$

The next step involves adding and subtracting WACC in the numerator.

$$V = Invested\ capital_0 \left(\frac{ROIC - WACC + WACC - g}{WACC - g}\right) \tag{9.2.5}$$

Separating (9.2.5) into two components and simplifying it leads to the following and final specification of the value function.

$$V = Invested\ capital_0 \left(\frac{ROIC - WACC}{WACC - g}\right) + Invested\ capital_0 \left(\frac{WACC - g}{WACC - g}\right) \tag{9.2.6}$$

Simplification finally leads to equation (18).

$$\boldsymbol{V = Invested\ capital_0 + \left(\frac{Invested\ capital_0 (ROIC - WACC)}{WACC - g}\right)} \tag{18}$$

9.3 Mathematical expressions of variables used in Terminal Value calculations

9.3.1 Free Cash Flow

Free cash flow can be expressed as follows.

FCF = Cash flow from operating activities – Capital expenditures

9.3.2 NOPLAT

NOPLAT is the abbreviation for *net operating profit less adjusted taxes* and can be expressed as follows.

*NOPLAT = Earnings before interest and tax * (1 – Tax rate)*

9.3.2 g (Growth)

(g) is defined as the growth in NOPLAT and free cash flow since the growth in NOPLAT should also equal the growth in free cash flow. Hence, g can be expressed as either

$$g = \frac{NOPLAT_t - NOPLAT_{t-1}}{NOPLAT_{t-1}}$$

Or

$$g = \frac{FCF_t - FCF_{t-1}}{FCF_{t-1}}$$

9.3.3 WACC

WACC is the abbreviation for *weighted average cost of capital* and can be expressed as follows.

$$WACC = \left(\frac{Equity}{Debt + Equity}\right) * Cost\ of\ equity + \left(\frac{Debt}{Debt + Equity}\right) * Cost\ of\ debt * (1 - Tax\ rate)$$

9.3.4 ROIC/RONIC

ROIC is the abbreviation for *return on invested capital*. Similarly, RONIC is the abbreviation for *return on new invested capital.* In our derivation of the terminal value of free cash flow formula we view these as being equal. ROIC can be expressed as follows.

$$ROIC = \frac{Net\ income - total\ dividends}{Total\ invested\ capital}$$

9.4 Gibbons, Ross and Shanken Methodology

We will below provide a description of how this methodology was implemented, as defined in Gibbons, Ross and Shanken (1989) and illustrated in Cochrane (2005).

The first step of the GRS-approach is to estimate an ordinary time-series regression of the model in question. As we did with the Fama-MacBeth part of this chapter, we will illustrate the methodology applied using model (3) as an example. We run the following time-series regression for each of the 3x25 left hand side portfolios. That is, for each of the models (1), (2) and (3), we run three tests where we test for each of the three left hand side portfolios individually. In sum, this translates to a grand total of 9 GRS-statistics (3x3). If we for example want to test model (3) using the left-hand side portfolios based on *size-inv*, we run the following regression 25 times since we have 25 left hand side size-inv portfolios.

$$R_{i,t} - R_{f,t} = \propto_i + \beta_i\left(R_{M,t} - R_{f,t}\right) + s_i SMB_t + h_i HML_t + g_i GMP_t + e_i EMI_t + \varepsilon_{i,t}$$

The second step is to collect all the estimated intercepts, $\hat{\alpha}_i$, from the time-series regression in step one and generate an *intercept vector,* which we refer to as $\hat{\alpha}$. Since we use 25 left hand side portfolios and thus estimate the model 25 times, we will get a 25x1 vector of intercepts. In general, this is a Nx1 vector where N refers to the number of left-hand side portfolios.

$$\hat{\alpha} = \begin{bmatrix} \widehat{\alpha_1} \\ \widehat{\alpha_2} \\ . \\ . \\ . \\ \widehat{\alpha_N} \end{bmatrix} \rightarrow \quad \hat{\alpha} = \begin{bmatrix} \widehat{\alpha_1} \\ \widehat{\alpha_2} \\ . \\ . \\ . \\ \widehat{\alpha_{25}} \end{bmatrix}$$

The third step is to calculate a TxN matrix using the estimated residuals for each of the 25 time-series regressions calculated in step one, which is termed as the *residual matrix*. Since the residual matrix is a TxN matrix, where T refers to the number of time periods in our data set, of which we have 546, and N refers to the number of estimated equations from step one, we need to estimate 546 residuals, $\hat{\varepsilon}$, for each of the 25 estimated regressions from step one.

$$\hat{\varepsilon}_{i,t} = \left(R_{i,t} - R_{f,t}\right) - \widehat{\alpha_i} - \hat{\beta}_i\left(R_{M,t} - R_{f,t}\right) - \hat{s}_i SMB_t - \hat{h}_i HML_t - \hat{g}_i GMP_t - \hat{e}_i EMI_t$$

After estimating all the residuals, $\hat{\varepsilon}_{i,t}$, for all T time periods and all 25 test portfolios, we form them into a 546x25 matrix.

$$\hat{\varepsilon} = \begin{bmatrix} \hat{\varepsilon}_{1,1} & \hat{\varepsilon}_{1,2} & \cdots & \hat{\varepsilon}_{1,N} \\ \hat{\varepsilon}_{2,1} & \hat{\varepsilon}_{2,2} & \cdots & \hat{\varepsilon}_{2,N} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{\varepsilon}_{T,1} & \hat{\varepsilon}_{T,2} & \cdots & \hat{\varepsilon}_{T,N} \end{bmatrix} \rightarrow \hat{\varepsilon} = \begin{bmatrix} \hat{\varepsilon}_{1,1} & \hat{\varepsilon}_{1,2} & \cdots & \hat{\varepsilon}_{1,25} \\ \hat{\varepsilon}_{2,1} & \hat{\varepsilon}_{2,2} & \cdots & \hat{\varepsilon}_{2,25} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{\varepsilon}_{546,1} & \hat{\varepsilon}_{T,2} & \cdots & \hat{\varepsilon}_{546,25} \end{bmatrix}$$

The fourth step is to calculate an unbiased estimate of the covariance matrix of the estimated residuals, $\hat{\Sigma}$, which is done by multiplying the transposed of the covariance matrix with the covariance matrix itself and dividing it by the total number of time periods, T.

$$\hat{\Sigma} = \frac{1}{T}\hat{\varepsilon}'\hat{\varepsilon}$$

Since we are multiplying the inverse of $\hat{\varepsilon}$ (25x546) with $\hat{\varepsilon}$ (546x25) itself, we will obtain a *covariance matrix of residuals* with a 25x25 (NxN) specification.

The fifth step is to calculate factor mean vector, $\bar{\mu}$, which is a Kx1 vector of the sample means of the factor portfolios, i.e. the sample means of all the individual factors in model. Note that K refers to the number of factors included in the model, hence in our case K = 5 for all models and we get a 5x1 mean vector for each individual model.

$$\bar{\mu} = \begin{bmatrix} \bar{F}_1 \\ \bar{F}_2 \\ . \\ . \\ . \\ \bar{F}_K \end{bmatrix} \rightarrow \bar{\mu} = \begin{bmatrix} \bar{F}_1 \\ \bar{F}_2 \\ . \\ . \\ . \\ \bar{F}_5 \end{bmatrix}$$

The sixth step is to construct a TxK (546x5) matrix, F, using the factor portfolio returns.

$$F = \begin{bmatrix} F_{1,1} & F_{1,2} & \dots & F_{1,L} \\ F_{2,1} & F_{2,2} & \dots & F_{2,L} \\ \dots & \dots & \dots & \dots \\ F_{T,1} & F_{T,2} & \dots & F_{T,L} \end{bmatrix} \rightarrow F = \begin{bmatrix} F_{1,1} & F_{1,2} & \dots & F_{1,5} \\ F_{2,1} & F_{2,2} & \dots & F_{2,5} \\ \dots & \dots & \dots & \dots \\ F_{546,1} & F_{546,2} & \dots & F_{546,5} \end{bmatrix}$$

We use this factor matrix to create an unbiased estimate of the KxK (5x5) covariance matrix of the factors.

$$\widehat{\Omega} = \frac{1}{T}(F - \bar{F})'(F - \bar{F})$$

Where $\bar{F}_K$ refers to the average factor return for factor K and $\bar{F}$ (546x5) is defined as follows.

$$\bar{F} = \begin{bmatrix} \bar{F}_1 & \bar{F}_2 & \dots & \bar{F}_K \\ \bar{F}_1 & \bar{F}_2 & \dots & \bar{F}_K \\ \dots & \dots & \dots & \dots \\ \bar{F}_1 & \bar{F}_2 & \dots & \bar{F}_K \end{bmatrix} \rightarrow \bar{F} = \begin{bmatrix} \bar{F}_1 & \bar{F}_2 & \dots & \bar{F}_5 \\ \bar{F}_1 & \bar{F}_2 & \dots & \bar{F}_5 \\ \dots & \dots & \dots & \dots \\ \bar{F}_1 & \bar{F}_2 & \dots & \bar{F}_5 \end{bmatrix}$$

The final step is to compute the GRS statistic, which is also the test-statistic used to for the null hypothesis. Assuming i.i.d. errors, we get that the quadratic form, $\widehat{\propto}\,'\widehat{\Sigma}^{-1}\hat{\alpha}$, has the following distribution.

$$\left(\frac{T - N - 1}{N}\right)\left[\frac{\widehat{\propto}\,'\widehat{\Sigma}^{-1}\hat{\alpha}}{1 + \bar{\mu}'\widehat{\Omega}^{-1}\bar{\mu}}\right] \sim F(N, T - N - K)$$

Finally, using this test statistic, we can test if the intercepts in our estimated models are equivalent to 0 and compare the new models to our benchmark. If one of the models prove to have intercepts that are jointly equivalent to 0 and the others don't, we can determine that this model is better at modelling variations in asset returns than the other.

9.5 Basic Code for Data Collection and Factor Creation (SAS)

This is a sample code describing how Fama and French (1993) methodology can be implemented to construct the factors for the three-factor model. The code itself and its justification can be found on Wharton Research Data Services website (Services, 2018). We have by modifying this code extracted factor returns for the period, January 1970 to June 2015. The code below merges Compustat XpressFeed (annual data) and CRSP in order to be able to extract data and construct the factor returns.

```
libname comp '/wrds/comp/sasdata/naa';
libname crsp ('/wrds/crsp/sasdata/a_stock'
'/wrds/crsp/sasdata/a_ccm');
libname myh  '~';

/************************ Part 1: Compustat
****************************/
/* Compustat XpressFeed Variables:
*/
/* AT      = Total Assets
*/
/* PSTKL   = Preferred Stock Liquidating Value
*/
/* TXDITC  = Deferred Taxes and Investment Tax Credit
*/
/* PSTKRV  = Preferred Stock Redemption Value
*/
/* SEQ     = Total Parent Stockholders' Equity
*/
/* PSTK    = Preferred/Preference Stock (Capital) - Total
*/

/* In calculating Book Equity, incorporate Preferred Stock
(PS) values */
/*  use the redemption value of PS, or the liquidation value
*/
/*    or the par value (in that order) (FF,JFE, 1993, p. 8)
*/
/* USe Balance Sheet Deferred Taxes TXDITC if available
*/
```

```
/* Flag for number of years in Compustat (<2 likely
backfilled data)   */

%let vars = AT PSTKL TXDITC PSTKRV SEQ PSTK ;
data comp;
  set comp.funda
  (keep= gvkey datadate &vars indfmt datafmt popsrc consol);
  by gvkey datadate;
  where indfmt='INDL' and datafmt='STD' and popsrc='D' and
consol='C'
  and datadate >='01Jan1959'd;
 /* Two years of accounting data before 1962 */
  PS = coalesce(PSTKRV,PSTKL,PSTK,0);
  if missing(TXDITC) then TXDITC = 0 ;
  BE = SEQ + TXDITC - PS ;
  if BE<0 then BE=.;
  year = year(datadate);
  label BE='Book Value of Equity FYear t-1' ;
  drop indfmt datafmt popsrc consol ps &vars;
  retain count;
  if first.gvkey then count=1;
  else count = count+1;
run;

/************************ Part 2: CRSP
**********************************/
/* Create a CRSP Subsample with Monthly Stock and Event
Variables       */
/* This procedure creates a SAS dataset named "CRSP_M"
*/
/* Restrictions will be applied later
*/
/* Select variables from the CRSP monthly stock and event
datasets      */
%let msevars=ticker ncusip shrcd exchcd;
%let msfvars =  prc ret retx shrout cfacpr cfacshr;

%include '/wrds/crsp/samples/crspmerge.sas';

%crspmerge(s=m,start=01jan1959,end=30jun2011,
sfvars=&msfvars,sevars=&msevars,filters=exchcd in (1,2,3));

/* CRSP_M is sorted by date and permno and has historical
returns     */
/* as well as historical share codes and exchange codes
*/
/* Add CRSP delisting returns */
proc sql; create table crspm2
 as select a.*, b.dlret,
```

```
  sum(1,ret)*sum(1,dlret)-1 as retadj "Return adjusted for
delisting",
  abs(a.prc)*a.shrout as MEq 'Market Value of Equity'
 from Crsp_m a left join
crsp.msedelist(where=(missing(dlret)=0)) b
 on a.permno=b.permno and

intnx('month',a.date,0,'E')=intnx('month',b.DLSTDT,0,'E')
 order by a.date, a.permco, MEq;
quit;

/* There are cases when the same firm (permco) has two or
more          */
/* securities (permno) at same date. For the purpose of ME
for          */
/* the firm, we aggregated all ME for a given permco, date.
This        */
/* aggregated ME will be assigned to the Permno with the
largest ME     */
data crspm2a (drop = Meq); set crspm2;
  by date permco Meq;
  retain ME;
  if first.permco and last.permco then do;
    ME=meq;
  output; /* most common case where a firm has a unique
permno*/
  end;
  else do ;
    if  first.permco then ME=meq;
    else ME=sum(meq,ME);
    if last.permco then output;
  end;
run;

/* There should be no duplicates*/
proc sort data=crspm2a nodupkey; by permno date;run;

/* The next step does 2 things:
*/
/* - Create weights for later calculation of VW returns.
*/
/*   Each firm's monthly return RET t willl be weighted by
*/
/*   ME(t-1) = ME(t-2) * (1 + RETX (t-1))
*/
/*     where RETX is the without-dividend return.
*/
/* - Create a File with December t-1 Market Equity (ME)
*/
```

```
data crspm3 (keep=permno date retadj weight_port ME exchcd
shrcd cumretx)
decme (keep = permno date ME rename=(me=DEC_ME) )  ;
     set crspm2a;
 by permno date;
 retain weight_port cumretx me_base;
 Lpermno=lag(permno);
 LME=lag(me);
     if first.permno then do;
     LME=me/(1+retx); cumretx=sum(1,retx);
me_base=LME;weight_port=.;end;
     else do;
     if month(date)=7 then do;
        weight_port= LME;
        me_base=LME; /* lag ME also at the end of June */
        cumretx=sum(1,retx);
     end;
     else do;
        if LME>0 then weight_port=cumretx*me_base;
        else weight_port=.;
        cumretx=cumretx*sum(1,retx);
     end; end;
output crspm3;
if month(date)=12 and ME>0 then output decme;
run;

/* Create a file with data for each June with ME from
previous December */
proc sql;
  create table crspjune as
  select a.*, b.DEC_ME
  from crspm3 (where=(month(date)=6)) as a, decme as b
  where a.permno=b.permno and
  intck('month',b.date,a.date)=6;
quit;

/***************   Part 3: Merging CRSP and Compustat
***********/
/* Add Permno to Compustat sample */
proc sql;
  create table ccm1 as
  select a.*, b.lpermno as permno, b.linkprim
  from comp as a, crsp.ccmxpf_linktable as b
  where a.gvkey=b.gvkey
  and substr(b.linktype,1,1)='L' and linkprim in ('P','C')
  and (intnx('month',intnx('year',a.datadate,0,'E'),6,'E')
>= b.linkdt)
  and (b.linkenddt >=
intnx('month',intnx('year',a.datadate,0,'E'),6,'E')
  or missing(b.linkenddt))
```

```
  order by a.datadate, permno, b.linkprim desc;
quit;

/*  Cleaning Compustat Data for no relevant duplicates
*/
/*  Eliminating overlapping matching : few cases where
different gvkeys     */
/*  for same permno-date --- some of them are not 'primary'
matches in CCM  */
/*  Use linkprim='P' for selecting just one gvkey-permno-
date combination   */
data ccm1a; set ccm1;
  by datadate permno descending linkprim;
  if first.permno;
run;

/* Sanity Check -- No Duplicates */
proc sort data=ccm1a nodupkey; by permno year datadate; run;

/* 2. However, there other type of duplicates within the
year                */
/* Some companiess change fiscal year end in the middle of
the calendar year */
/* In these cases, there are more than one annual record for
accounting data */
/* We will be selecting the last annual record in a given
calendar year      */
data ccm2a ; set ccm1a;
  by permno year datadate;
  if last.year;
run;

/* Sanity Check -- No Duplicates */
proc sort data=ccm2a nodupkey; by permno datadate; run;

/* Finalize Compustat Sample                              */
/* Merge CRSP with Compustat data, at June of every year  */
/* Match fiscal year ending calendar year t-1 with June t */
proc sql; create table ccm2_june as
  select a.*, b.BE, (1000*b.BE)/a.DEC_ME as BEME, b.count,
  b.datadate,
  intck('month',b.datadate, a.date) as dist
  from crspjune a, ccm2a b
  where a.permno=b.permno and intnx('month',a.date,0,'E')=
  intnx('month',intnx('year',b.datadate,0,'E'),6,'E')
  order by a.date;
quit;

/************************ Part 4: Size and Book to Market
Portfolios ***/
```

```
/* Forming Portolio by ME and BEME as of each June t
*/
/* Calculate NYSE Breakpoints for Market Equity (ME) and
*/
/* Book-to-Market (BEME)
*/
proc univariate data=ccm2_june noprint;
  where exchcd=1 and beme>0 and shrcd in (10,11) and me>0
and count>=2;
  var ME BEME; * ME is Market Equity at the end of June;
  by date; /*at june;*/
  output out=nyse_breaks median = SIZEMEDN pctlpre=ME BEME
pctlpts=30 70;
run;

/* Use Breakpoints to classify stock only at end of all
June's */
proc sql;
  create table ccm3_june as
  select a.*, b.sizemedn, b.beme30, b.beme70
  from ccm2_june as a, nyse_breaks as b
  where a.date=b.date;
quit;

/* Create portfolios as of June                        */
/* SIZE Portfolios          : S[mall] or B[ig]         */
/* Book-to-market Portfolios: L[ow], M[edium], H[igh] */
data june ; set ccm3_june;
 If beme>0 and me>0 and count>=2 then do;
 positivebeme=1;
 * beme>0 includes the restrictioncs that ME at Dec(t-1)>0
 * and BE (t-1) >0 and more than two years in Compustat;
 if 0 <= ME <= sizemedn     then sizeport = 'S';
 else if ME > sizemedn      then sizeport = 'B';
 else sizeport='';
 if 0 < beme <= beme30 then           btmport = 'L';
 else if beme30 < beme <= beme70 then btmport = 'M' ;
 else if beme  > beme70 then          btmport = 'H';
 else btmport='';
end;
else positivebeme=0;
if cmiss(sizeport,btmport)=0 then nonmissport=1; else
nonmissport=0;
keep permno date sizeport btmport positivebeme exchcd shrcd
nonmissport;
run;

/* Identifying each month the securities of              */
/* Buy and hold June portfolios from July t to June t+1  */
proc sql; create table ccm4 as
```

```
 select a.*, b.sizeport, b.btmport, b.date as portdate
format date9.,
        b.positivebeme , b.nonmissport
 from crspm3 as a, june as b
 where a.permno=b.permno and  1 <=
intck('month',b.date,a.date) <= 12
 order by date, sizeport, btmport;
quit;

/*************** Part 5: Calculating Fama-French Factors
**************/
/* Calculate monthly time series of weighted average
portfolio returns */
proc means data=ccm4 noprint;
 where weight_port>0 and positivebeme=1 and exchcd in
(1,2,3)
      and shrcd in (10,11) and nonmissport=1;
 by date sizeport btmport;
 var retadj;
 weight weight_port;
 output out=vwret (drop= _type_ _freq_ ) mean=vwret
n=n_firms;
run;

/* Monthly Factor Returns: SMB and HML */
proc transpose data=vwret(keep=date sizeport btmport vwret)
 out=vwret2 (drop=_name_ _label_);
 by date ;
 ID sizeport btmport;
 Var vwret;
run;

/************************ Part 6: Saving Output
************************/
data myh.ff_factors;
set vwret2;
 WH = (bh + sh)/2  ;
 WL = (sl + bl)/2 ;
 WHML = WH - WL;
 WB = (bl + bm + bh)/3 ;
 WS = (sl + sm + sh)/3 ;
 WSMB = WS - WB;
 label WH   = 'WRDS High'
       WL   = 'WRDS Low'
       WHML = 'WRDS HML'
       WS   = 'WRDS Small'
       WB   = 'WRDS Big'
       WSMB = 'WRDS SMB';
run;
```

```
/* Number of Firms */
proc transpose data=vwret(keep=date sizeport btmport
n_firms)
               out=vwret3 (drop=_name_ _label_) prefix=n_;
by date ;
ID sizeport btmport;
Var n_firms;
run;

data myh.ff_nfirms;
set vwret3;
 N_H = n_sh + n_bh;
 N_L = n_sl + n_bl;
 N_HML = N_H + N_L;
 N_B =  n_bl + n_bm + n_bh;
 N_S =  n_sl + n_sm + n_sh ;
 N_SMB = N_S + N_B;
 Total1= N_SMB;
 label N_H   = 'N_firms High'
       N_L   = 'N_firms Low'
       N_HML = 'N_firms HML'
       N_S   = 'N_firms Small'
       N_B   = 'N_firms Big'
       N_SMB = 'N_firms SMB';
run;

/* Clean the house*/
proc sql;
   drop table ccm1, ccm1a,ccm2a,ccm2_june,
              ccm3_june, ccm4, comp,
              crspm2, crspm2a, crspm3, crsp_m,
              decme, june, nyse_breaks;
quit;
```

9.6 Basic Code for Fama MacBeth Procedure (Eviews)

This is a sample code describing how the Fama and Macbeth (1973) methodology can be implemented to run the second stage Fama-Macbeth tests on the VW portfolios formed on Size-BM for the FF5 model. The code can be found in the book Introductory Econometrics for Finance (Brooks, 2017).

LHS: Size_BM, RHS: FF5 factors

```
'TRANSFORM ACTUAL RETURN INTO EXCESS RETURNS
me1_bm1=me1_bm1-rf
me1_bm2=me1_bm2-rf
me1_bm3=me1_bm3-rf
me1_bm4=me1_bm4-rf
me1_bm5=me1_bm5-rf
me2_bm1=me2_bm1-rf
me2_bm2=me2_bm2-rf
me2_bm3=me2_bm3-rf
me2_bm4=me2_bm4-rf
me2_bm5=me2_bm5-rf
me3_bm1=me3_bm1-rf
me3_bm2=me3_bm2-rf
me3_bm3=me3_bm3-rf
me3_bm4=me3_bm4-rf
me3_bm5=me3_bm5-rf
me4_bm1=me4_bm1-rf
me4_bm2=me4_bm2-rf
me4_bm3=me4_bm3-rf
me4_bm4=me4_bm4-rf
me4_bm5=me4_bm5-rf
me5_bm1=me5_bm1-rf
me5_bm2=me5_bm2-rf
me5_bm3=me5_bm3-rf
me5_bm4=me5_bm4-rf
me5_bm5=me5_bm5-rf

'DEFINE THE NUMBER OF TIME SERIES OBSERVATIONS
!NOBS=546

'CREATE SERIES TO PUT BETAS FROM STAGE 1
'AND LAMBDAS FROM STAGE 2 INTO
SERIES BETA_C
SERIES BETA_RMRF
SERIES BETA_RMW
SERIES BETA_CMA
SERIES BETA_HML
SERIES BETA_SMB
```

```
SERIES LAMBDA_C
SERIES LAMBDA_RMRF
SERIES LAMBDA_RMW
SERIES LAMBDA_CMA
SERIES LAMBDA_HML
SERIES LAMBDA_SMB
SERIES LAMBDA_R2
SCALAR LAMBDA_C_MEAN
SCALAR LAMBDA_C_TRATIO
SCALAR LAMBDA_RMRF_MEAN
SCALAR LAMBDA_RMRF_TRATIO
SCALAR LAMBDA_RMW_MEAN
SCALAR LAMBDA_RMW_TRATIO
SCALAR LAMBDA_CMA_MEAN
SCALAR LAMBDA_CMA_TRATIO
SCALAR LAMBDA_HML_MEAN
SCALAR LAMBDA_HML_TRATIO
SCALAR LAMBDA_SMB_MEAN
SCALAR LAMBDA_SMB_TRATIO
SCALAR LAMBDA_R2_MEAN

'THIS LOOP CREATES THE SERIES TO PUT THE
'CROSS-SECTIONAL DATA IN
FOR !M = 1 TO 546
SERIES TIME {!M}
NEXT

'NOW RUN THE FIRST STAGE TIME-SERIES REGRESSIONS
'SEPARATELY FOR EACH POTFOLIO AND
'PUT THE BETAS INTO APPROPRIATE SERIES
SMPL 1970:01 2015:06
!J=1
FOR %Y me1_bm1 me1_bm2 me1_bm3 me1_bm4 me1_bm5 me2_bm1
me2_bm2 me2_bm3 me2_bm4 me2_bm5 me3_bm1 me3_bm2 me3_bm3
me3_bm4 me3_bm5 me4_bm1 me4_bm2 me4_bm3 me4_bm4 me4_bm5
me5_bm1 me5_bm2 me5_bm3 me5_bm4 me5_bm5
'THE PREVIOUS COMMAND (ABOVE) WITH VARIABLE NAMES

'NEEDS TO ALL GO ON ONE LINE

EQUATION EQ1.LS {%Y} C RMRF RMW CMA HML SMB
BETA_C(!J)=@COEFS(1)
BETA_RMRF(!J)=@COEFS(2)
BETA_RMW(!J)=@COEFS(3)
BETA_CMA(!J)=@COEFS(4)
BETA_HML(!J)=@COEFS(5)
BETA_SMB(!J)=@COEFS(6)
!J=!J+1
NEXT

'NOW RESORT THE DATA SO EACH COLUMN IS A
```

```
'MONTH AND EACH ROW IS RETURNS ON PORTFOLIOS
FOR !K=1 TO 546
TIME!K(1)=me1_bm1(!K)
TIME!K(2)=me1_bm2(!K)
TIME!K(3)=me1_bm3(!K)
TIME!K(4)=me1_bm4(!K)
TIME!K(5)=me1_bm5(!K)
TIME!K(6)=me2_bm1(!K)
TIME!K(7)=me2_bm2(!K)
TIME!K(8)=me2_bm3(!K)
TIME!K(9)=me2_bm4(!K)
TIME!K(10)=me2_bm5(!K)
TIME!K(11)=me3_bm1(!K)
TIME!K(12)=me3_bm2(!K)
TIME!K(13)=me3_bm3(!K)
TIME!K(14)=me3_bm4(!K)
TIME!K(15)=me3_bm5(!K)
TIME!K(16)=me4_bm1(!K)
TIME!K(17)=me4_bm2(!K)
TIME!K(18)=me4_bm3(!K)
TIME!K(19)=me4_bm4(!K)
TIME!K(20)=me4_bm5(!K)
TIME!K(21)=me5_bm1(!K)
TIME!K(22)=me5_bm2(!K)
TIME!K(23)=me5_bm3(!K)
TIME!K(24)=me5_bm4(!K)
TIME!K(25)=me5_bm5(!K)
NEXT

'RUN 2ND STAGE CROSS-SECTIONAL REGRESSIONS
FOR !Z = 1 TO !NOBS
EQUATION EQ1.LS TIME!Z C BETA_RMRF BETA_RMW BETA_CMA
BETA_HML BETA_SMB
LAMBDA_C(!Z)=@COEFS(1)
LAMBDA_RMRF(!Z)=@COEFS(2)
LAMBDA_RMW(!Z)=@COEFS(3)
LAMBDA_CMA(!Z)=@COEFS(4)
LAMBDA_HML(!Z)=@COEFS(5)
LAMBDA_SMB(!Z)=@COEFS(6)
LAMBDA_R2(!Z)=@R2
NEXT


'FINALLY ESTIMATE THE MEANS AND THE T-RATIOS
'FOR THE LAMBDA ESTIMATES IN THE SECOND STAGE
LAMBDA_C_MEAN=@MEAN(LAMBDA_C)

LAMDA_C_TRATIO=@SQRT(!NOBS)*@MEAN(LAMBDA_C)/@STDEV(LAMBDA_C)
```

```
LAMBDA_RMRF_MEAN=@MEAN(LAMBDA_RMRF)

LAMDA_RMRF_TRATIO=@SQRT(!NOBS)*@MEAN(LAMBDA_RMRF)/@STDEV(LAM
BDA_RMRF)

LAMBDA_RMW_MEAN=@MEAN(LAMBDA_RMW)

LAMDA_RMW_TRATIO=@SQRT(!NOBS)*@MEAN(LAMBDA_RMW)/@STDEV(LAMBD
A_RMW)

LAMBDA_CMA_MEAN=@MEAN(LAMBDA_CMA)

LAMDA_CMA_TRATIO=@SQRT(!NOBS)*@MEAN(LAMBDA_CMA)/@STDEV(LAMBD
A_CMA)

LAMBDA_HML_MEAN=@MEAN(LAMBDA_HML)

LAMDA_HML_TRATIO=@SQRT(!NOBS)*@MEAN(LAMBDA_HML)/@STDEV(LAMBD
A_HML)

LAMBDA_SMB_MEAN=@MEAN(LAMBDA_SMB)

LAMDA_SMB_TRATIO=@SQRT(!NOBS)*@MEAN(LAMBDA_SMB)/@STDEV(LAMBD
A_SMB)

LAMBDA_R2_MEAN=@MEAN(LAMBDA_R2)
```