Erasmus University Rotterdam
Erasmus School of Economics
Master Thesis Econometrics & Management Science, Quantitative Finance Specialization

On the robustness of factor p-values in multi-factor asset pricing models

Tijs Bronnenberg (479124)

Supervisor:	dr. (Mikhail) M. Zhelonkin
Second assessor:	dr. (Erik) H.J.W.G. Kole
Date final version:	27th January 2024

Abstract

This thesis paper researches the robustness of factor sensitivity p-values in multi-factor asset pricing models using OLS and robust MM estimation. To do so, factors are added to the Fama-French three factor model resulting in four factor models. The significance of the newly added fourth factor is studied extensively in a case study using the Carhart four factor model. Building upon this case study, results are aggregated for over 140 factors. These aggregations lead to the conclusion that outliers rendering a factor significant is not an uncommon phenomenon in asset pricing models.

Contents

1 Introduction **2**

2 Relevant Literature **4**

3 Methodology **8**
3.1 Factor Models 8
3.2 Robustness 9
3.3 Robustness Procedure 16

4 Results **19**
4.1 The Data 19
4.2 Fama & French three factor Model 20
4.3 Carhart four factor Model 23
4.4 Aggregation of four factor models 35

5 Conclusion **40**

References **42**

A Factor abbreviations **45**

B Programming code **52**

Chapter 1

Introduction

One of the most well-studied problems in Financial Econometrics is explaining the cross-variation in stock returns. This seemingly complex problem is studied through the use of multi-factor asset pricing models. These asset pricing models use factors which have been empirically tested to explain the variation in returns. A widely known example of such a model is the capital asset pricing model, created by Sharpe (1964) and Litner (1965). The most adopted approach to finding factors is to create factors which capture certain firm characteristics. Fama and French (1993), for example, develop a three factor model in which they capture the effects of firm size (market equity) and firm value (book-to-market). Since the introduction of this model, the search for factors explaining the cross-variation in stock returns has exploded (Cochrane, 2011).

To determine the relationship between a set of factors and returns, a linear model is usually used. This model has a specific return series as the dependent variable, and the set of factors as explanatory variables. The model is then estimated using ordinary least squares (OLS). A drawback to OLS, however, is that it is not robust to outliers. This could lead to possibly wrong conclusions, given that financial return series are well known to contain outliers (Mandelbrot, 1963).

Tying into this, the world of asset pricing models has faced two larger problems. The first problem is that of replicability. Multiple studies have tried to replicate multi-factor models, albeit with slightly different data (e.g. Hou, Xue and Zhang (2018), Jensen, Kelly and Pedersen (2023)). They find mixed evidence in favor of the significance of well researched factors. The second problem is that of an overwhelmingly large number of factors, otherwise known as the factor zoo. Both of these problems either could be related to outliers, or could be better understood by accounting for outliers. Replicability issues could imply that a factor is rendered significant by a handful of outliers. Similarly, the large number of factors might be reduced when accounting for outliers. As the relevance of a factor is determined by its statistical significance, the influence of outliers on the significance of factors will be researched. The main research question of this thesis is: To what extent are p-values of estimated factor sensitivities affected by outliers in financial return data? In other words, how robust are p-values of factor sensitivities in asset pricing models?

To research this, 200 dependent variables and a set of 149 risk factors is used. The Fama and French (1993) three factor model is used as a base model. The robustness of the estimated coefficients and p-values of this model is examined through robust MM-regression estimates.

Comparisons are made between the OLS estimates and the MM estimates. If the model assumptions hold and there are no outliers present, there should be no large differences between the methods.

Using the base model, four factor models are created by adding an additional factor to the Fama-French three factor model. The focal point of this research is the robustness of the estimated p-value for the fourth factor. For the fourth factor, the estimated p-values and the respective statistical conclusions based on OLS and MM estimation are compared. Clearly, if a factor sensitivity is robust, both methods should have a concurring statistical conclusion. Using the aforementioned dependent variables and set of factors, the robustness of p-values of factor sensitivities is extensively researched.

Before aggregating results for all factors, however, the Carhart (1997) four factor model is used as a detailed case study. Using this model, differences in statistical conclusions can be clearly highlighted. Using a single difference and MM estimation, it can be studied if this difference is in fact the result of outliers. Finally, the results from this case study can be aggregated for all factors. This should provide a comprehensive overview of the robustness of the p-values of factors, and simultaneously answers if outliers drive differences in p-values of factors.

The set up of this paper is as follows. Firstly, Section 2 provides a broad review of the existing literature and how this research positions itself between it. Section 3 formalizes the models, estimators, and procedures as well as the concepts of robustness used in this research. Section 4 contains empirical results for the Fama and French (1993) three factor model, the Carhart (1997) four factor model, and the other factors in the aforementioned set of factors. Finally, Section 5 contains concluding remarks.

Chapter 2

Relevant Literature

The foundation upon which this research is built is multi factor asset pricing models. These linear models relate multiple firm characteristics to returns. Perhaps the most widely known multi-factor model is the three factor model presented by Fama-French (Fama & French, 1993). In their study, they examine to what extent firm size (market-equity) and firm value (book-to-market) are related to returns. They find strong evidence in favor of the three factor model explaining cross sectional variation in portfolio returns sorted on size and value.

After the creation of the three factor model, many other factor models building upon their findings and methodology followed. Carhart (1997) created a four factor model which uses the findings of Jegadeesh and Titman (1993) in which past winners (high return stocks) seem to outperform losers (low return stocks). This effect is also known as momentum. Additionally, Pastor and Stambaugh (2003) find that adding a liquidity factor improves the Fama-French three factor model. As a final example, Fama and French (2015) improve on their three factor model by adding two factors capturing profitability and investment.

As mentioned in Chapter 1, the number of factors related to returns has exploded. This development has resulted in the so-called factor zoo. Multiple researchers have been involved in creating an overview of the factor zoo. Firstly, Harvey and Liu (2019) document more than 400 published factors. They provide an open source Google spreadsheet, in which information as well as citation links on each factor are provided. Although there is value to be found in the census provided, a shortcoming of this paper is that it provides no factor data itself. Fortunately, there have been researchers who also recognized the need for an open factor data source. Chen and Zimmermann (2022) provide code and data which can successfully reproduce more than 300 factors related to stock returns. Hou et al. (2018) compile a large data library containing more than 440 factors. Feng, Giglio and Xiu (2020) create an open factor data set, consisting of 150 factors. Additionally, the Fama & French three factors are provided via an online data library, courtesy of Professor French. On this website, multiple return series for different portfolio sorts are presented.

The need for open factor data can be attributed to the two aforementioned problems within the realm of multi-factor asset pricing models. The first problem is that of replication. Hou et al. (2018) try to replicate multi-factor studies using slightly different methodology or data. They find that roughly 64% of the factors tested turn out to be insignificant at the 5% level, indicating replicability issues. Jensen et al. (2023) construct and estimate a Bayesian model used to replicate factors. They find evidence, however, in favor of the replicability of factor models. The second problem faced in factor models, arises from the overwhelming number of factors. Given such a large amount of factors, surely not all of them should be included in an asset pricing model. Feng et al. (2020), for example, create a method to examine whether or not a factor should be added to an already existing factor model.

These two problems reflect the need of open-source factor data. Yet, more importantly for this paper, is that both of these problems also touch upon the main question of this research: To what extent are p-values of estimated factor sensitivities affected by outliers in financial return data? If p-values of factor sensitivities are influenced by a single, or handful of, outliers, replicability issues could arise when using different data. The influence of outliers on p-values is also relevant for the second problem. If a factor is rendered significant due to a handful of outliers, questions should be raised whether to include it an already existing factor model.

Researching the robustness of sensitivities of factors is not a new concept, and many different approaches have been used. Fama and French (1993) themselves examine the robustness of their inference. As a robustness check, they test their model using different test assets. In the original estimations, the dependent variables are returns on portfolios sorted by firm value and size. Lo and MacKinlay (1990) show that using portfolios sorted on the characteristics which are also captured in the factors can give rise to misleading statistical conclusions. Hence, Fama & French test their model on portfolios sorted by earnings-to-price as well as dividend-to-price. They find that their model captures the cross section of average stock returns. This approach does raise the question if using merely two different portfolio sorts is enough to deem their findings robust? Feng et al. (2020), for example, use 8 different portfolio sorts, totalling 202 different portfolio returns as test assets for their own robustness checks.

A similar, yet slightly different approach to determine the robustness of factor models is through the use of equities from different markets. The Fama and French (1993) three factor model was tested using equities listed in US Markets. Malin and Veeraraghavan (2004) find evidence of a size effect in Germany, France and the United Kingdom, yet they find no evidence of a value effect. Contrarily, Boamah (2015) finds that there is in fact evidence for a size and value effect, as well as a momentum effect on the South-African stock market. Although estimating the models using equities from different markets makes sense, such an approach does not take the findings of Lo and MacKinlay (1990) into account.

Fortunately, other methods can be employed to research how robust asset pricing models are. McLean and Pontiff (2016) study the pre-publication out-of-sample and post-publication return predictability of 97 factors. In essence, they study whether or not the factors are related to returns outside of the study's respective sample period. If the factor's predictability disappears out of sample, there is some sort of bias in the factor. They this bias to be about 26% in the pre-publication out-of-sample period. Additionally, they find returns to be 58% lower after the findings regarding a factor have been published. This makes sense given the efficient market hypothesis. It is interesting, however, that there seems to be a 26% bias in their sample of factors. As this could also be the result of in-sample outliers, it provides another reason to research the robustness of inference for factors.

A different robustness method is provided by Harvey, Liu and Zhu (2015). They argue that the standard procedure for determining significance does not make sense, given the large amount of available factors. They introduce a multiple testing framework in which they argue that a new factor needs to pass a much higher threshold to be deemed significant, with a t-statistic greater than three.

Although the aforementioned approaches yield interesting results, this paper will employ robust estimation techniques. In the creation of factor models, OLS is usually used to determine and test the relation between a certain predictive factor and returns. It is well researched, however, that financial return series contain outliers and are seen as heavy tailed (Mandelbrot, 1963) & (Guillaume et al., 1997). Rousseeuw and Leroy (1987) show multiple examples of how OLS estimates lead to incorrect conclusions when outliers are present in data. Therefore, estimation methods which are robust to outliers must be taken into account.

The robustness of an estimator is measured in two ways. Firstly, it can be quantified through the (asymptotic) breakdown point of an estimator (Hampel, 1971). The basic idea behind the breakdown point is to measure what proportion of the data is needed for an estimator to become biased. As an example, the sample mean has a breakdown point of $1/n$, indicating that it only takes one observation within the sample to bias the estimator. Similarly, Rousseeuw and Yohai (1984) show that the OLS estimator has a breakdown point of $1/n$, and hence an asymptotic breakdown point of 0%.

The second measure of an estimator's robustness is its influence function (IF). The influence function describes the effect of infinitesimal contamination on an estimator. A more formal definition for the influence function is provided in Section 3.2. A measure which is derived from the influence function, is the gross error sensitivity (GES). This is defined as the maximum effect of an observation on the estimate (Rousseeuw & Yohai, 1984). Although one would assume that for an estimator to be robust its IF or GES must be bounded, this turns out to not be the case. Maronna, Martin and Yohai (2006) point out that although the GES may be unbounded, this does not imply that the maximum bias introduced by the contamination is infinite.

Clearly, a high asymptotic breakdown point is needed for an estimator to be robust. Similarly, its influence function must be carefully taken into account. The first estimator to satisfy a high asymptotic breakdown point is the least trimmed squares (LTS) estimator. The idea behind this estimator is to minimize the sum of the h smallest squared residuals. In other words, to use a concentrated subset of data points. Contrary to the OLS estimator, this estimator has a breakdown point of 50% (Mount, Netanyahu, Piatko, Silverman & Wu, 2014). Rousseeuw and Yohai (1984) show that the influence function (IF) of the LTS estimator is unbounded. This would seem to be a problem, yet they also show that the IF is unaffected by data outside the middle half of the data. Combining this with the high asymptotic breakdown point, the estimator is robust. A drawback to the LTS estimator, however, is its efficiency. Given that a subset of h data points are selected, it will inevitably result in a loss of information.

A more efficient method is that of M-estimation. M-estimators are a generalization of maximum likelihood estimators. The idea behind making M-estimators robust lies within the loss function. Recall that OLS minimizes the sum of squared residuals. Clearly, in the case of a large outlier, the squared residual will be large. The OLS estimate will therefore be more influenced by a large residual (outlier) and compensate for it.

In a regression setting, M-estimators were first developed by Huber (1973). An advantage of M-estimators is their efficiency. However, Maronna et al. (2006) also show that M-estimators are not robust against so-called bad leverage points. They also show that M-estimators have unbounded influence, however, this is not a problem as the maximum bias introduced by the contamination is bounded. Although M-estimators are not robust, they will be useful for the development of robust estimators.

The next class of estimators is an S-estimator, which was developed by Rousseeuw and Yohai (1984) This estimator uses the Huber (1964) scale M-estimator, and a robust loss function. The S-estimator has an asymptotic breakdown point of 50%, but again an unbounded IF. Note, however, that an S-estimator is essentially an M-estimator. Thus, the same argument regarding the maximum bias introduced by the contamination can be made. Hence, the S-estimator is in fact robust. Unfortunately, a drawback of the S-estimator is that it suffers from the trade-off between robustness and efficiency, as described by Albuquerque and Biegler (1996). So although the S-estimator is robust, it is not efficient.

This is where so called MM-estimators provide a solution. MM-estimators were first proposed by Yohai (1987) as high breakdown point and high efficiency estimators for regressions. It builds upon the S-estimator by (Rousseeuw & Yohai, 1984), where the residual scale estimate is computed. Keeping this residual scale estimate fixed, an M-estimator is computed. Yohai (1987) proves that the breakdown point of the MM estimator is inherited from the S-estimator in the first step. Additionally, it is shown that the asymptotic efficiency is inherited from the M-estimator in the second step. Finally, MM-estimators have an unbounded IF, similar to M- and S-estimators. However, the maximum bias introduced by potential outliers is bounded.

To conclude, the literature on factors related to returns is vast, and suffers from replicability issues and an overwhelmingly large number of factors. Although the robustness of factor models has been researched using various techniques, robust estimation is not one of them. This paper will use robust MM-regression estimates to research the robustness of the significance of factors in four factor asset pricing models. Hopefully, this will shed some light on the effect of outliers in asset pricing models, and provide new insights into the existing replicability issues and the factor zoo.

Chapter 3

Methodology

3.1 Factor Models

The workhorse of this paper are the previously mentioned multi factor asset pricing models. These models relate some set of factors, usually based on certain characteristics of a firm, to an expected excess return. More formally, a k (return) factor model is given by

$$E[R] = R_f + \boldsymbol{\beta}'(E[\mathbf{F}] - R_f \boldsymbol{\iota}_k), \tag{3.1}$$

where $E[R]$ is the expected returns, R_f is the risk free rate, $\boldsymbol{\beta}$ is a vector containing the loadings onto the k factors, $\mathbf{F}$ are the k return factors and $\boldsymbol{\iota}_k$ is a $k \times 1$ vector of ones.

Note that $\mathbf{F}$ contains return factors. In other words, these factor realizations are the realizations of the returns of a portfolio of securities. This proves to be a problem, as firm characteristics themselves cannot be traded and thus are not return factors. Hence, factors which are returns and capture a firm specific effect must be created. This can be done through mimicking portfolios.

Mimicking portfolios are created by sorting a large number of stocks on certain characteristics. Based on specific break-off points, or threshold values, these stocks are grouped into sorted portfolios. This results in portfolios with a larger exposure to the characteristic and portfolios with a smaller exposure. Ranking these portfolios on the exposure to the characteristic gives rise to a top portfolio and a bottom portfolio. The strategy between going long on a top portfolio, and shorting a bottom portfolio is known as the so-called top-minus-bottom or mimicking portfolio. Taking the average returns of this strategy, this portfolio captures the returns which are related to the previously sorted on characteristic. Hence, the top-minus-bottom portfolio is a return factor which captures the effect of a characteristic.

Fama and French (1993), for example, use a 2×3 sort on size and book-to-market, as they study the relation between returns and size & value. The stocks they sort are listed on NYSE, Amex and NASDAQ (after 1972). Using the 2×3 sort gives 6 portfolios, which are then held for a year. The Small-minus-Big (SMB) factor is then created using the average return of the three small minus the average return of the three big portfolios. Similarly, the High-minus-Low (HML) factor is created by the average return of the two value portfolios minus the average return of the two growth portfolios. Combining this procedure with (3.1) and their empirical research gives rise to the Fama and French (1993) model:

$$E[R_t] - R_{f,t} = \alpha_t + \beta_{MKT}(R_{M,t} - R_{f,t}) + \beta_{SMB}(SMB_t) + \beta_{HML}(HML_t) + \epsilon_t, \tag{3.2}$$

where $E[R_t]$ is the expected return at time t, and SMB_t, HML_t are the factors which were created according to the outlined procedure above. $R_{M,t} - R_{f,t}$ is the excess market return factor, where $R_{f,t}$ is the risk-free rate and $R_{M,t}$ is the return on the market.

The next step in the development of factor models is to estimate and test the model. Firstly, a dependent variable must be selected. As mentioned previously in Section 2, portfolio returns are used to reduce noise. The standard procedure for this is to form portfolios based on the characteristics which are meant to explain the variation in stock returns. Fama & French use 25 portfolios formed on size and book-to-market. They do so as they seek to determine the whether the factors they create capture common factors in returns related to those characteristics. However, Lo and MacKinlay (1990) point out that this procedure might lead to misleading conclusions.

The final step is to estimate the factor sensitivities and test statistics. Given the dependent variables, a linear model is fit to the data. The significance of a factor can then be assessed using the estimated test statistics. These are usually estimated using OLS. OLS minimizes the sum of squared residuals, which is prone to influence of outliers.

3.2 Robustness

A quantitative method to determine the robustness of an estimator is to look at its breakdown point and its influence function (IF). The concept of the breakdown point is the proportion of observations required to make an estimator arbitrarily biased. As an intuitive example, one can look at a location estimator. For a location estimator T_l, the formal definition of a breakdown point is as follows.

Let ϵ_n be the breakdown point of the estimator. Then ϵ_n is defined as:

$$\epsilon_n(T_l, \mathbf{x}_n) = \min_{1 \leq m \leq n} \left\{ \frac{m}{n} : \sup_{x_{n,m}} |T_l(\mathbf{x}_n) - T_l(\mathbf{x}_{n,m})| = \infty \right\}, \tag{3.3}$$

where $\mathbf{x}_n$ is the original data and $\mathbf{x}_{n,m}$ is the data where m points are replaced with arbitrary values. Then, the breakdown point ϵ_n is the proportion of data points required to make T_l arbitrarily biased.

A perhaps more intuitive way of interpreting the breakdown point is through its asymptotic behavior. The asymptotic breakdown point of any estimator T is defined as

$$\epsilon(T) = \lim_{n \to \infty} \epsilon_n(T, \mathbf{x}_n). \tag{3.4}$$

For the sample mean, one would get a breakdown point equal to

$$\epsilon_n(\overline{x}, \mathbf{x_n}) = \frac{1}{n}, \tag{3.5}$$

which intuitively makes sense, given that an extremely large outlier can heavily influence the sample mean. This also implies that the sample mean then has an asymptotic breakdown point of 0%.

The interpretation behind the influence function (IF) is to quantify the change in an estimator, given an infinitely small amount of contamination. The definition of an influence function for an estimator T is given by:

$$IF(x, T, F) = \lim_{\epsilon \to 0+} \frac{T\{(1-\epsilon)\, F + \epsilon \Delta_x\} - T(F)}{\epsilon}, \tag{3.6}$$

where F is the assumed distribution of the data, ϵ is the contamination level and Δ_x is the point mass distribution at x. The gross error sensitivity (GES) γ^* is then defined as:

$$\gamma^*(T, F) = \sup_x |IF(x, T, F)|, \tag{3.7}$$

which is the largest change in an estimator if one observation is added. As an example, the sample mean has an IF given by $IF(x; \bar{x}, F) = x - \mu$, where μ is the expectation of the data under F. Clearly, this has unbounded influence.

As stated in Section 2, factor sensitivities and their respective test statistics are usually estimated using ordinary least squares (OLS). OLS estimates the parameters $\boldsymbol{\beta}$ which minimize the sum of squared residuals:

$$\widehat{\boldsymbol{\beta}}_{OLS} = \arg\min_{\boldsymbol{\beta}} \sum_{t=1}^{T} (y_t - \mathbf{x_t}^\top \boldsymbol{\beta})^2, \tag{3.8}$$

where y_t is the return of a sorted portfolio at time t and $\mathbf{x_t}$ is the set of factors at time t, including an intercept.

Clearly, the squared residual loss function is prone to influence from outliers. Given the large residual of an outlier, this loss function will increase the influence of that outlier in the minimization problem. Hence, the OLS estimator will overcompensate for the data point with the large residual. Clearly, this provides a qualitative argument against the robustness of the OLS estimator. This argument is formalized by Rousseeuw and Yohai (1984) in which they show that the breakdown point of the OLS estimator is equal to $1/n$, implying an asymptotic breakdown point of 0%. The influence function of the OLS estimator is given by:

$$IF(\boldsymbol{z}, \boldsymbol{\beta}_{OLS}, F) = \boldsymbol{M}_{OLS}^{-1} (y_t - \mathbf{x_t}^\top \boldsymbol{\beta}_{OLS}) \mathbf{x_t}, \tag{3.9}$$

where $\boldsymbol{M}_{OLS} = \int \mathbf{x_t} \mathbf{x_t}^\top dF$ and $\boldsymbol{z}$ represents the statistical unit $(y, \mathbf{x_t}^\top)$. Note that this is proportional to the residual, and unbounded. The combination of the IF and the asymptotic breakdown point shows that the OLS estimator is not robust.

An intuitive approach to solving the lack of robustness of the OLS estimator, is to change its loss function. Huber (1973) uses this approach in the development of M-estimators. The

M-estimate of β is given by

$$\widehat{\boldsymbol{\beta}}_M = \arg\min_{\boldsymbol{\beta}} \sum_{t=1}^{T} \rho\left(y_t - \mathbf{x_t}^\intercal \boldsymbol{\beta}\right), \tag{3.10}$$

where $\rho(\cdot)$ is a loss function. The loss function $\rho(\cdot)$ determines the solution to (3.10). If $\rho(z) = \frac{1}{2}z^2$, for example, then the M-estimator simply becomes the OLS estimator. Now, let $\psi(\cdot)$ be the derivative of $\rho(\cdot)$. A robust ψ-function proposed by Huber is as follows:

$$\psi(z) = \begin{cases} z, & \text{if } |z| \leq c, \\ c \cdot \text{sign}(z), & \text{if } |z| > c, \end{cases} \tag{3.11}$$

where c is a tuning constant which determines the asymptotic efficiency. A well researched different influence function is the Tukey bisquare loss function, which is given by

$$\psi(z) = \begin{cases} z\left(\left(\frac{z}{c}\right)^2 - 1\right)^2, & \text{if } |z| \leq c, \\ 0, & \text{if } |z| > c, \end{cases} \tag{3.12}$$

where again, c is a tuning constant which determines the asymptotic efficiency. Both of these loss functions will be used in the upcoming estimators.

These functions also give a more intuitive insight into M-estimation. Given the minimization problem in (3.10), the first order condition (FOC) becomes:

$$\sum_{t=1}^{T} \psi\left(y_t - \mathbf{x_t}^\intercal \boldsymbol{\beta}\right) \mathbf{x_t} = 0. \tag{3.13}$$

This equation can be rewritten to

$$\sum_{t=1}^{T} w_t \left(y_t - \mathbf{x_t}^\intercal \boldsymbol{\beta}\right) \mathbf{x_t} = 0, \tag{3.14}$$

where

$$w_t = w\left(\left(y_t - \mathbf{x_t}^\intercal \boldsymbol{\beta}\right)/\widehat{\sigma}\right) = \frac{\psi\left(\left(y_t - \mathbf{x_t}^\intercal \boldsymbol{\beta}\right)/\widehat{\sigma}\right)}{\left(y_t - \mathbf{x_t}^\intercal \boldsymbol{\beta}\right)/\widehat{\sigma}}, \tag{3.15}$$

and $\widehat{\sigma}$ is a residual scale estimate. In other words, the M-estimator can be interpreted as a weighted least squares estimator, where the weights depend on the data (Maronna et al., 2006). For the Tukey bisquare loss function a weight of zero will be assigned once the standardized residual becomes larger than the tuning constant c in absolute value. For the Huber loss function, the numerator in (3.15) becomes equal to the tuning constant multiplied by the sign of the standardized residual, once the standardized residual becomes larger than the tuning constant in absolute value. The denominator is equal to the standardized residual.

The computation of M-estimators is not straightforward. Maronna et al. (2006) note that in the case that w_t would be known in advance, OLS could be used. However, the weights are unknown. This can be solved through the use of iterative least squares:

For iteration k, with current estimates $\widehat{\boldsymbol{\beta}}_k$ and scale estimate $\widehat{\sigma}_k$:

1. Update weights:

$$w_t\left(\widehat{\boldsymbol{\beta}}_k, \widehat{\sigma}_k\right) = \frac{\psi\left(\left(y_t - \mathbf{x_t}^\mathsf{T}\widehat{\boldsymbol{\beta}}_k\right)/\widehat{\sigma}_k\right)}{\left(y_t - \mathbf{x_t}^\mathsf{T}\widehat{\boldsymbol{\beta}}_k\right)/\widehat{\sigma}_k}.$$

2. Obtain new estimate $\widehat{\boldsymbol{\beta}}_{k+1}$ by solving:

$$\widehat{\boldsymbol{\beta}}_{k+1} = \arg\min_{\boldsymbol{\beta}} \frac{1}{T}\sum_{t=1}^{T} w_t\left(\widehat{\boldsymbol{\beta}}_k, \widehat{\sigma}_k\right)\left(y_t - \mathbf{x_t}^\mathsf{T}\boldsymbol{\beta}\right)^2.$$

3. Obtain new residual scale estimate

$$\widehat{\sigma}_{k+1} = \text{MAD}\left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{k+1}\right).$$

where MAD is the median absolute deviation, which for a vector of data points $\mathbf{x_t}$ equals MAD $= a \cdot \text{med}\left|\mathbf{x_t} - \text{med}\left(\mathbf{x_t}\right)\boldsymbol{i}\right|$ and a is a consistency correction. Repeating this procedure until convergence then gives the desired estimator $\widehat{\boldsymbol{\beta}}^M$.

Unfortunately, Maronna et al. (2006) show that $\widehat{\boldsymbol{\beta}}^M$ is not robust against bad leverage points due to the residual scale estimate $\widehat{\sigma}$ breaking down. More formally, the M-estimator has an asymptotic breakdown point of 0%. For its unbounded influence function, they show that this need not be problem, as the maximum bias of the estimator as a result of contamination is in fact bounded.

To achieve a high breakdown point estimator, two conditions must be satisfied. Firstly, a robust scale estimate is needed. Secondly, a redescending ψ-function is needed, as it shrinks the influence of outliers down to zero. The latter condition proves to be a problem, as redescending ψ-function will lead to a non-convex optimization problem. This could in turn lead to a local minimum instead of a global minimum. Thus, the starting values used in the iterative least squares algorithm must be chosen carefully.

An estimator with a high breakdown point is the so-called S-estimator. The derivation of the S-estimator follows from the OLS estimator, and was first proposed by Rousseeuw and Yohai (1984). Recall (3.8), which can be rewritten to:

$$\widehat{\boldsymbol{\beta}}_{OLS} = \arg\min_{\boldsymbol{\beta}} \frac{1}{T}\sum_{t=1}^{T}(y_t - \mathbf{x_t}^\mathsf{T}\boldsymbol{\beta})^2. \tag{3.16}$$

As the first building block of a robust estimator, a robust scale estimate must be used to replace the residual variance given in (3.16). This is where the M-estimator of scale (Huber, 1964) is used, which is given by the following equation.

$$\frac{1}{T}\sum_{t=1}^{T}\rho\left(\frac{x_t}{\widehat{\sigma}_M}\right) = \delta, \tag{3.17}$$

where $x_t \sim F$ with mean 0 and variance σ^2 and $\delta = E\left(\rho\left(\frac{X}{\sigma}\right)\right)$ for consistency at model distribution F.

Using this M-estimator of scale, the S-estimator is then defined as:

$$\widehat{\boldsymbol{\beta}}_S = \arg\min_{\boldsymbol{\beta}} \widehat{\sigma}^2_M(\boldsymbol{\beta}). \tag{3.18}$$

Similar to (3.17), $\widehat{\sigma}_M$ is obtained by solving:

$$\frac{1}{T}\sum_{t=1}^{T} \rho\left(\frac{y_t - \mathbf{x}_t^{\mathsf{T}}\boldsymbol{\beta}}{\widehat{\sigma}_M(\boldsymbol{\beta})}\right) = \delta, \tag{3.19}$$

where δ is some constant which ensures consistency. For the loss function, typically the Tukey bisquare loss function (3.12) is used.

The computation of S-estimators can be seen as an M-estimator with fixed preliminary scale $\widehat{\sigma}_S = \widehat{\sigma}_M\left(\widehat{\boldsymbol{\beta}}_S\right)$:

$$\widehat{\boldsymbol{\beta}}_S = \arg\min_{\boldsymbol{\beta}} \sum_{t=1}^{T} \rho\left(\frac{y_t - \mathbf{x_t}^{\mathsf{T}}\boldsymbol{\beta}}{\widehat{\sigma}_S}\right). \tag{3.20}$$

Then also similar to the M-estimator, the S-estimator can also be seen as a weighted least squares estimator with data-dependents weights. To compute the S-estimator, the FAST-S algorithm (Salibian-Barrera & Yohai, 2006) is used. The first step in the algorithm is the I-step, which builds upon the iterative least squares algorithm proposed in this section.

For a current iteration k with estimate $\widehat{\boldsymbol{\beta}}_k$:

1. Compute M-estimate of scale $\widehat{\sigma}_M\left(\widehat{\boldsymbol{\beta}}_k\right)$.

2. Update weights:

$$w_t\left(\widehat{\boldsymbol{\beta}}_k\right) = \frac{\psi\left(\left(y_t - \mathbf{x_t}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}_k\right)/\widehat{\sigma}_M\left(\widehat{\boldsymbol{\beta}}_k\right)\right)}{\left(y_t - \mathbf{x_t}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}_k\right)/\widehat{\sigma}_M\left(\widehat{\boldsymbol{\beta}_k}\right)}.$$

3. Solve weighted least squares problem

$$\widehat{\boldsymbol{\beta}}_{k+1} = \arg\min_{\boldsymbol{\beta}} \frac{1}{T}\sum_{t=1}^{T} w_t\left(\widehat{\boldsymbol{\beta}}_k\right)\left(y_t - \mathbf{x_t}^{\mathsf{T}}\boldsymbol{\beta}\right)^2.$$

As a result of a redescending loss function, multiple initial values must be used to guarantee a global minimum. The entire procedure then becomes:

1. Start with m sets of initial values.

2. For $i = 1, ..., m$ use the iterated least squares algorithm to obtain coefficient estimates.

3. $\widehat{\boldsymbol{\beta}}_S$ is then obtained as the solution with the smallest M-estimate of scale.

To formalize the robustness of the S-estimator, the breakdown point is discussed. The developers of the S-estimator show that the estimator has an asymptotic breakdown point of 50%, given that $F \sim N(0, \sigma^2)$ and $\frac{\delta}{\rho(c)} = \frac{1}{2}$. Clearly, this implies that the asymptotic breakdown point depends on the tuning parameter c. For the IF of the S-estimator, Maronna et al. (2006) show that it is unbounded. However, since S-estimation is essentially M-estimation, the same argument as for M-estimation can be applied.

In spite of their robustness, an improvement on S-estimators can be made. Hössjer (1992) shows that S-estimators with a breakdown point of 50% have a maximum efficiency of 33%. M-estimators, however, are the exact opposite: highly efficient, yet not robust. Yohai (1987) combines both S-estimation and M-estimation to create the highly robust and efficient MM-estimator. This estimator uses the preliminary scale estimate $\widehat{\sigma}_S = \widehat{\sigma}_M\left(\widehat{\boldsymbol{\beta}}_S\right)$ and estimated coefficients $\widehat{\boldsymbol{\beta}}_S$ obtained from the S estimation in (3.18) as initial values for an M-estimator. Using the scale estimate, the M-estimator is computed, as per (3.10).

The MM-estimator is given by:

$$\widehat{\boldsymbol{\beta}}_{MM} = \arg\min_{\boldsymbol{\beta}} \sum_{t=1}^{T} \rho_2\left(\frac{y_t - \mathbf{x_t}^{\mathsf{T}}\boldsymbol{\beta}}{\widehat{\sigma}_S}\right), \tag{3.21}$$

with preliminary scale estimate $\widehat{\sigma}_S$ obtained from (3.18) and (3.19). As for this procedure an S-estimator is needed, two loss functions are required. The loss function used to find the S-estimator of scale is denoted as ρ_1. The loss function used to find the MM-estimator is denoted as ρ_2.

The computation of the MM-estimator follows from that of the M and S-estimator, and is done in two steps:

1. Compute the S-estimator and save residual scale estimate $\widehat{\sigma}_S$.

2. Compute M-estimator keeping residual scale $\widehat{\sigma}_S$ fixed.

To compute each step, the algorithms previously described can be used. Given that the MM-estimator uses both M- and S-estimation, the two tuning constants associated with each estimation step can be tuned accordingly. Typically, both steps use the Tukey bisquare loss function. For the M-estimator, a tuning constant must be chosen such that the efficiency is guaranteed. Maronna et al. (2006) show that a tuning constant of $c_2 = 4.685$ leads to an efficiency of 95%. For the S-estimator a high breakdown point is desired. It can also be shown that this results in a tuning constant $c_1 = 1.547$ which leads to an asymptotic breakdown point of 50%. Similar to the S-estimator, the IF of the MM-estimator is unbounded. As mentioned in previous sections, Maronna et al. (2006) point out that although there may be an infinite gross error sensitivity, this does not imply that the resulting maximum bias is infinite. They show that although the IF is indeed unbounded, the maximum bias resulting from an additional observation is bounded.

Finally, a highly efficient and highly robust estimator is obtained. The focal point of this research is to examine the robustness of factor p-values. This can be researched using OLS and MM estimation. To estimate a p-value, firstly the test statistics must be estimated. For an MM-estimator, this is given by:

$$\widehat{t}_{MM} = \frac{\widehat{\beta}_{MM}}{SE(\widehat{\beta}_{MM})}, \tag{3.22}$$

where $\widehat{\beta}_{MM}$ is the MM estimator and $SE(\widehat{\beta}_{MM})$ is the standard error of the estimated coefficient. The effect of outliers on the t-statistic is two fold. Firstly, the estimated coefficient $\widehat{\beta}_{MM}$ is affected by outliers. Secondly, the standard error of the coefficient is affected by outliers. Note that the t-statistic can also be calculated for the OLS coefficient using the same equation but with OLS estimates.

Using the t-statistic from Equation (3.22), the p-value can be calculated. For the MM estimator, this is given by:

$$\widehat{p}_{MM} = 2 \cdot \mathbb{P}\left(t_{n-p-1,1-\frac{\alpha}{2}} > |\widehat{t}_{MM}|\right), \tag{3.23}$$

where $t_{n-p-1,1-\frac{\alpha}{2}}$ represents the respective quantile of the t-distribution, α is the significance-level and $\widehat{t}_{MM}$ is the aforementioned t-statistic. Note that the p-value can straightforwardly also be calculated for the OLS coefficient. The estimated p-values of a factor using both estimation methods allows for the examination of the robustness of the factor's p-value.

To provide an intuition behind the robustness of p-values, a simulation is used. In this simulation, there is one dependent variable (y_t) and one explanatory variable (x_t). The values of y_t and x_t are drawn from a standard normal distribution. As both x_t and y_t are randomly drawn, x_t should have no explanatory power. The entire length of the series of y_t and x_t equals 100 observations.

To illustrate the influence of an outlier on the p-value, the final observation in the time series is set to a relatively large value. This observation mimics an outlier. x_{100} is arbitrarily fixed at 4. However, y_{100} takes on values $3, 4, ..., 10$. Using the outlier, 1000 time series are simulated. Based on a simulation, the p-value of the coefficient is estimated using OLS and MM estimation. This results in 1000 combinations $(\hat{p}^i_{OLS}, \hat{p}^i_{MM})$, where i denotes the simulation. For each outlier and estimation method, a box-plot is made of the estimated p-values. This can be found in Figure 3.1.

Box-plot of estimated P-values using OLS and MM estimation for different Outlier Values

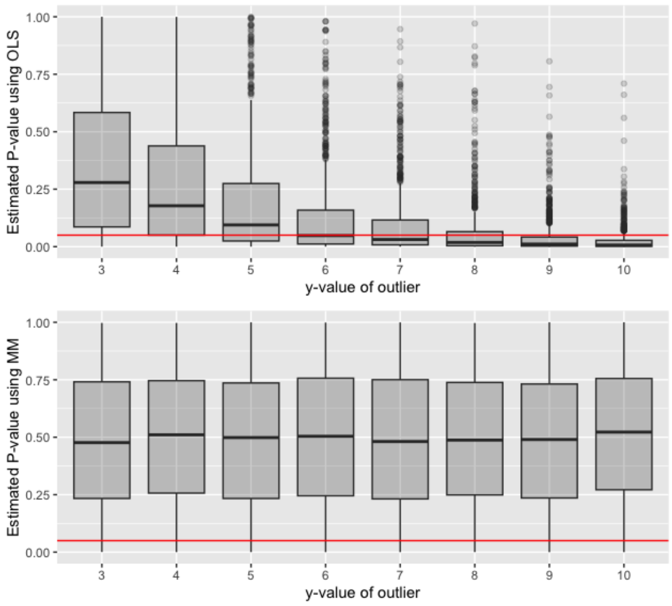

Figure 3.1: Box-plots for estimated p-values. The red line depicts the significance level of 0.05

Figure 3.1 shows that for OLS, x_t is rendered significant as a result of the outlier. The larger the outlier, the more often x_t is deemed significant. For MM estimation, however, x_t is consistently estimated to be insignificant. Given the randomness of the data, this plot illustrates the need for a robust p-value estimator if outliers are present. Although this example is a simplified version of the models studied in this paper, the intuition and concepts behind the influence of outliers on p-values remains the same.

3.3 Robustness Procedure

The robustness of factor sensitivity p-values are studied by estimating the p-values using both OLS and MM estimation. As can be seen in Figure 3.1, if models assumptions hold and outliers are not present, there should be no large differences in estimation methods. To examine the robustness of a whole host of factor p-values, first two models are studied in more detail.

The first model to be studied will be the Fama and French (1993) three factor model. This model is one of the most influential factor models and functions as a foundation for multiple models (e.g. Carhart (1997) four factor model, Pastor and Stambaugh (2003) liquidity model). Additionally, this model will function as a base model in this paper.

Firstly, the model will be estimated using both methods. The dependent variables in these estimation methods are portfolio returns sorted on size and book-to-market. These are the same dependent variables as used by the authors in their respective paper. An addition to this will be the use of multiple dependent variables sorted on different characteristics. This is done to account for the findings of Lo and MacKinlay (1990), in which they find that portfolios sorted on the same characteristic of a stock could potentially create biases in the test statistics. For this model, both estimated coefficients and estimated p-values will be presented even though p-values are the main quantity of interest in this research. This should give the reader a more intuitive, and step-by-step understanding of the estimation methods and their respective results. However, no explanations or detailed analysis for potential differences will be given for this model, as the focal point of this thesis will be on four factor models.

The motivation behind the focus on four factor models is primarily illustrative. However, the advantage of studying four factor models over the three factor models lies in the large number of additional factors which can be added to the Fama and French (1993) three factor model. This gives an opportunity to study the behavior of the estimators of an additional fourth factor in the presence of three well known factors on a large scale. Additionally, adding only one factor has a higher interpretability than adding say two or three factors. The general four factor model is given by:

$$E[R_t] - R_{f,t} = \alpha_t + \beta_{MKT}(R_{M,t} - R_{f,t}) + \beta_{SMB}(SMB_t) + \beta_{HML}(HML_t) + \beta_G(G_t) + \epsilon_t, \quad (3.24)$$

where G_t represents the additional fourth factor. This fourth factor is the factor of interest for this study.

To achieve an in-depth understanding of the robustness of p-values of the factor G_t, the Carhart four factor model is selected to function as in depth case study. That is, $G_t = UMD_t$, where UMD_t is the Up-minus-Down factor capturing momentum in stocks. The case study is conducted as follows.

Firstly, multiple sorted portfolio returns will be used as dependent variables. Using these returns, a model will be fit multiple times. For example, if there are 10 dependent variables, this results in the Carhart four factor model being fit 10 times for each method (i.e. 10 times for OLS estimation, 10 times for MM estimation). Each time the parameters and test statistics are estimated. The dependent variable resulting in the largest difference in estimated p-value of the UMD factor will be selected to function as a case study.

As mentioned in Section 3.2, outliers can lead to different statistical conclusions. In this case study, it will be researched if the differences in conclusions are the result of outliers in the underlying data. To do so, firstly regression diagnostics will be run. This will be used to visualize potential outliers. For a formal outlier identification method, the weights resulting from the robust MM regression can be used. Recall from Section 3.2 that MM estimation can be seen as weighted least squares. An outlier can then be defined as an observation with a weight smaller than some threshold. Potential thresholds are discussed in Section 4.3.

Once outliers have been identified, a so-called skipped estimator can be calculated ((Johansen & Nielsen, 2013) & (Bickel, 1975)). To do so, the outliers must be removed from the data resulting in a new data set. Using the new data set, the model is fitted using OLS. If outliers are indeed the cause of differences between MM and OLS estimation, it is expected that the statistical conclusions based on skipped estimators and MM estimators should overlap. If this is in fact the case, the argument can be made that outliers lead to differences in statistical conclusions.

Note that although the skipped estimator is a more robust estimator than simply using the OLS estimated p-value, it is not as robust as the MM estimated p-value. Also note that MM estimation assigns different weights to different observations. The skipped estimator assigns equal weight to different observations. It could therefore well be the case that certain differences between estimation methods remain.

Once the case study has been completed other dependent variables can be considered. The skipped p-value estimator can be computed for all the dependent variables which lead to differences in statistical conclusions. Aggregating these results should then provide the reader with a comprehensive and in-depth understanding of the robustness of factor sensitivities for the UMD factor. For this, two aggregation metrics are calculated. First, the percentage of regressions leading to different statistical conclusions is calculated. For example, suppose that there are ten dependent variables. This will lead to the model being fit ten times for both OLS and MM estimation. Suppose then that out of the ten dependent variables, two lead to different statistical conclusions. The percentage of regressions leading to different conclusions is 20%. The higher this percentage is, the less robust an estimated p-value is and the more skepsis on the relevance of a factor there should be.

The second metric is the percentage of previously established differences which can be explained by outliers. This is calculated using the skipped estimator. For each previously established difference, the skipped-estimator is calculated. If the then newly estimated p-value is larger than 0.05, the conclusion based on MM estimation coincides with the conclusion based on the skipped estimator. This then also means that the difference between OLS and MM estimation can be attributed to the presence of outliers. Referring to the previous example, there are two dependent variables leading to different statistical conclusions. Then, for both of these differences the skipped p-value is calculated. If out of these two skipped p-values one is larger than 0.05, the percentage of differences which can be explained by outliers is 50%. Using these metrics, a coherent picture can be formed on the robustness of the UMD sensitivity.

In addition to the UMD factor, there is an entire factor zoo. The respective factor p-values are estimated for all these factors, using OLS estimation, MM estimation and the available set of dependent variables. Using the estimated p-values, the same procedure can be done for the other factors as for the UMD factor. That is, dependent variables leading to an estimated p-value larger than 0.10 using MM estimation and smaller than 0.05 using OLS estimation are selected. Using these p-values, the percentage of cases with different conclusions is computed. For all the cases where differences are present, the skipped p-value is estimated. Finally, the percentage of the differences explained by outliers is computed for each factor.

Chapter 4

Results

In this section the results for each factor model are presented. As a first step, the Fama and French (1993) three factor model is examined as it is used as a base model. Additionally, the results of this model function as a stepping stone towards the results of the cornerstone of this research, four factor models. The analysis starts with the most basic dependent variables, which replicates the original study conducted by Fama & French (albeit with data with a different time span). Building upon these results other dependent variables are also used as test assets. Using both sets of dependent variables, estimated coefficients and respective p-values are reported, but not explained.

Next, the results for the Carhart (1997) four factor model are presented. Similar to the three factor model, the estimated coefficients and p-values are presented. As mentioned, this model will act as a case study for the robustness procedure in which differences between conclusions based on OLS and MM estimation are explained. To do so, a dependent variable where differences in statistical conclusions occur are selected. Based on this dependent variable, the results of an in depth study will be presented as described in Section 3.3. Using this case study, results for multiple dependent variables as well as multiple factors are aggregated and presented.

4.1 The Data

As dependent variables, sorted portfolio returns are used. Given the previously presented results by Lo and MacKinlay (1990), using portfolios sorted on merely two characteristics could potentially lead to incorrect conclusions. Thus, similar to Feng et al. (2020), this paper uses multiple portfolio sorts. Similar to the Fama and French (1993) study, 5×5 bi-variate portfolio sorts on size and book-to-market are used. Additionally, 5×5 portfolio sorts on size & momentum, size & long term reversal, size & operating profitability, size & investment, book-to-market & operating profitability, book-to-market & investment and operating profitability & investment are used. This gives a total of 200 test assets, with a wide range of firm specific characteristics. Note that for each of these test assets, the risk free rate is subtracted to obtain excess returns. The data can be downloaded from Prof. French's website (French, 2023).

For the factors, an open data library is used. This open data library is created by Feng et al. (2020) and contains 150 factors. The data spans from July, 1976 to December, 2017 with a monthly frequency leading to 498 observations per factor. However, these 150 risk factors also contain the risk-free rate which is not used as a factor. Additionally, some of the factors have missing data values. These factors are then dropped. This leads to 146 factors, of which the three Fama-French factors. In total, this leaves 143 possible four factor models.

4.2 Fama & French three factor Model

The first model to be examined is the Fama-French three factor model, as in (3.2). First, a comparison between OLS and MM estimates is made using only the returns of the 5×5 portfolio sort on size and book-to-market. This leads to 25 different OLS estimates and 25 MM estimates for factor coefficients and p-values. If estimators for factor sensitivities are robust, the differences in coefficients should not be large. This translates into the estimates forming a linear relationship with coefficient 1 and intercept 0. In more simpler terms, a "$y = x$" relationship. In Figure 4.1 the estimated coefficients for different estimation methods are plotted against one another. A red line is added which depicts this relationship.

Fama-French three factor Model: Coefficients under OLS and MM estimation

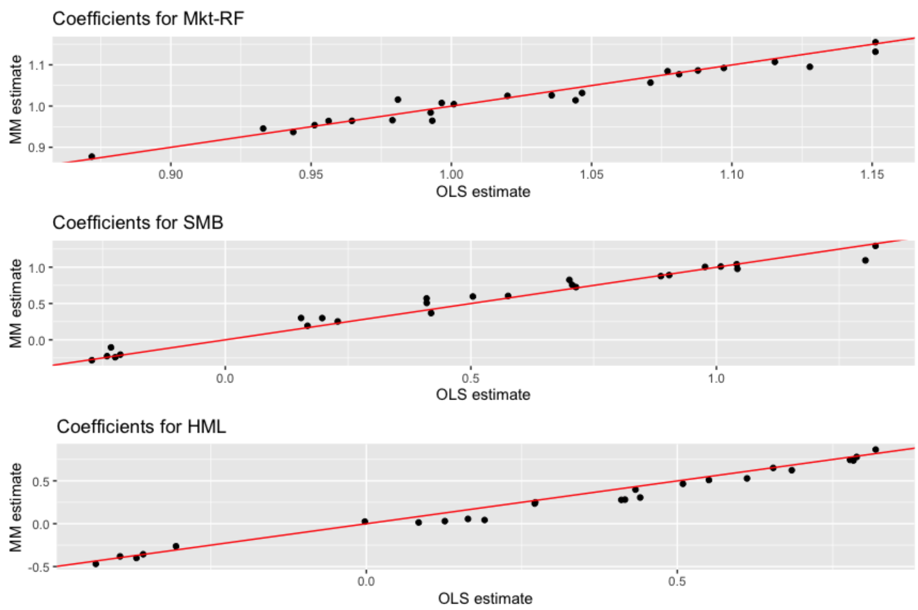

Figure 4.1: These plots show the different estimates for factor coefficients for 25 different portfolio returns sorted on size and book-to-market. The red line depicts a linear relationship with slope 1 and intercept 0.

The above plot provides a first insight into the robustness of sensitivity estimators of the three factor model, using the original dependent variables. Given the plot, OLS and MM estimation seem to roughly coincide. However, estimated p-values must also be taken into account before conclusions on the robustness of the estimators in this model can be drawn.

As with the coefficients, it is expected that the estimated p-values using both estimation techniques should be relatively similar if robust. In Figure 4.2, the p-value pairs for each factor and dependent variable are plotted.

Fama-French three factor Model: Estimated p-values using OLS and MM

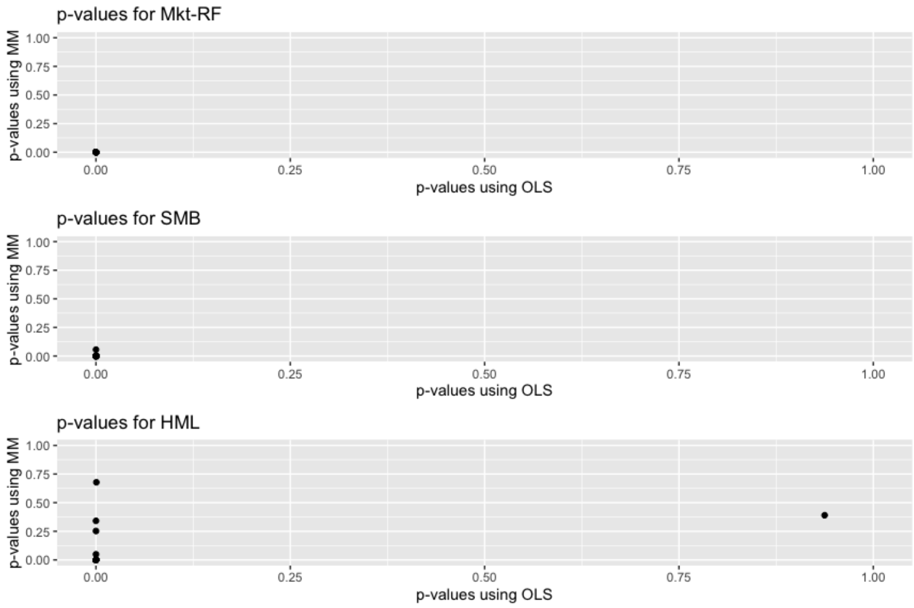

Figure 4.2: These plots show estimated p-value pairs using OLS and MM estimation and 25 portfolios sorted on size and book-to-market.

Nearly all estimated coefficients are significant, and OLS and MM estimation are in agreement. Combining these results together with Figure 4.1 provide evidence in favor of the robustness of the Fama and French (1993) findings.

A problem with this approach however, has been clearly described in Section 3.3 and mentioned multiple times. By using portfolio returns which are created by sorting on the same characteristics as captured in the factors, Lo and MacKinlay (1990) show that this can lead to biases in test statistics. As a remedy against this problem, 200 portfolio returns will be used as dependent variables. The portfolio's are sorted on multiple characteristics which have been described in Section 4.1. Repeating the same procedure as previously described yields 200 estimates for coefficients and p-values for both methods. In Figure 4.3 the estimated coefficients for both OLS and MM are plotted against one another.

Fama-French three factor Model: Estimated Coefficients using 200 test portfolios

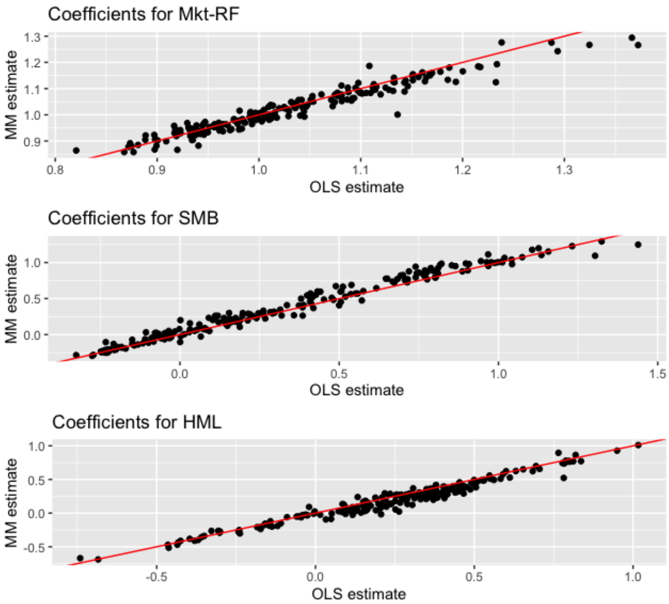

Figure 4.3: The estimated coefficients using OLS & MM estimation combined with 200 test portfolios. A linear relationship with slope 1 and intercept 0 is added as a benchmark for robustness.

In accordance with the findings in Figure 4.1, the coefficients are estimated similarly. To quantify if this is in fact the case, a simple linear regression can be fit. As mentioned previously, for a robust estimator an intercept of zero and a coefficient of one is expected. The results for this regression are presented in Table 4.1.

Fama-French coefficient regressions

Estimates	$\widehat{\alpha}_{(R_M-R_f)}$	$\widehat{\beta}_{(R_M-R_f)}$	$\widehat{\alpha}_{SMB}$	$\widehat{\beta}_{SMB}$	$\widehat{\alpha}_{HML}$	$\widehat{\beta}_{HML}$
Values	0.1126 (<0.001)	0.8817 (<0.001)	0.0345 (<0.001)	1.007 (<0.001)	-0.0369 ((<0.001)	0.9522 (<0.001)

Table 4.1: Regression estimates for a simple linear regression of MM and OLS estimates of coefficients in the Fama-French three factor model. Note that $\widehat{\alpha}$ represents the estimate of the intercept, and $\widehat{\beta}$ represents the estimate of the coefficient. Also, p-values are in brackets.

Clearly, the estimated coefficients and intercepts provide evidence in favor of a linear relationship between the OLS and MM estimates. This in turn, provides a first piece of evidence that the Fama-French three factor sensitivities are robust. It becomes more interesting, however, when taking inference into account. A similar plot of MM and OLS p-values yields noteworthy results.

Fama-French three factor Model: Estimated p-values using 200 test portfolios

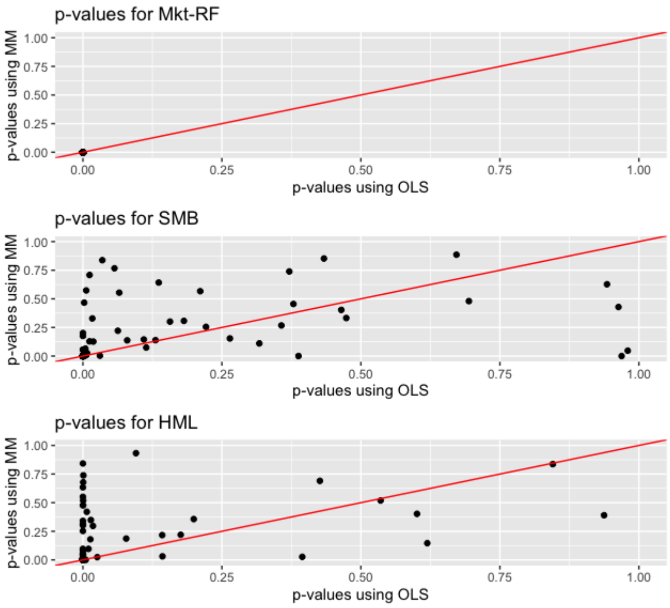

Figure 4.4: The estimated p-values using OLS & MM estimation combined with 200 test portfolios. A linear relationship with slope 1 and intercept 0 is added as a benchmark for robustness.

Firstly, note the concentration of p-values for the first factor around 0. This makes sense as one would expect portfolio returns to be correlated with the market. For the other two factors the results seem a bit more dispersed, although also largely in line with previous findings. For the SMB factor 92% of the regressions yield similar statistical conclusions using a significance threshold of 5%. Similarly, for the HML factor this turns out to be 88%. Although this is still quite high, there are some clear differences in conclusions based on OLS and MM estimation for the HML and SMB factors.

4.3 Carhart four factor Model

After examining the Fama-French three factor model, the next step is to research models that built upon the findings of the three factor model. As described in Section 3.3 factors can be added to the existing three factors, resulting in a four factor model. Also described in Section 3.3 is that the Carhart (1997) four factor model will function as a detailed case study. More specifically, it will be studied why the UMD factor is rendered significant using OLS, but insignificant using MM for a specific dependent variable. This case study will lead to a deeper understanding of underlying differences in statistical conclusions resulting from different estimation methods. First and foremost, the Carhart four factor model is given by:

$$E[R_t] - R_{f,t} = \alpha_t + \beta_{MKT}(R_{M,t} - R_{f,t}) + \beta_{SMB}(SMB_t) + \beta_{HML}(HML_t) + \beta_{UMD}(UMD_t) + \epsilon_t,$$

where UMD_t is the Up-minus-Down factor capturing momentum and the other factors are from the Fama and French (1993) model. In this Section, the results of estimations using both OLS and MM-estimation are presented for the previously described 200 portfolio returns.

Although coefficients are not the focal point of this research, it is informative to take them into account. Figure 4.5 displays the results.

Carhart four factor Model: Estimated coefficients using 200 test portfolios

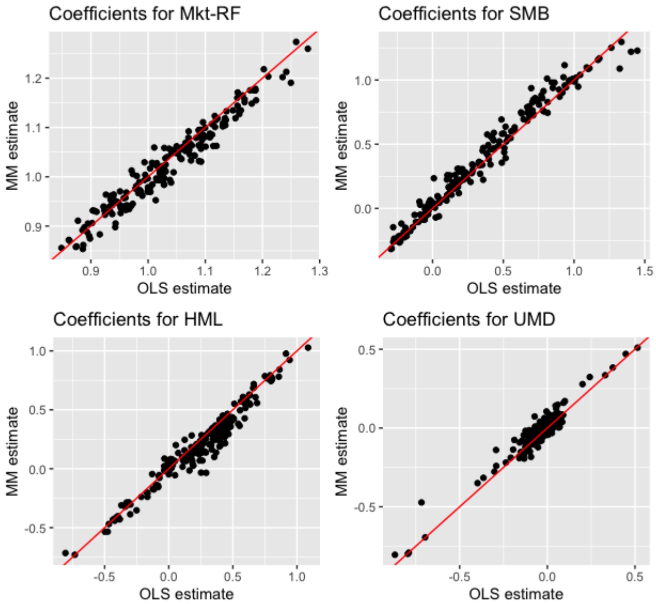

Figure 4.5: The estimated coefficients using OLS & MM estimation combined with 200 test portfolios. A linear relationship with slope 1 and intercept 0 is added for easy comparison between methods.

Similar to the Fama-French three factor model, both estimation methods roughly concur on coefficient sign and magnitude. A simple linear regression of the MM estimates on the OLS estimates gives an intercept equal to 0.0114 ($\widehat{p} < 0.001$) and a slope of 0.9306 ($\widehat{p} < 0.001$) for the sensitivities of the UMD factor. Interestingly though, is the high concentration of UMD estimates around zero.

As mentioned in previous sections, the focal point of this research lies on the estimated p-values. In this case, that is the estimated p-values for the UMD factor. The OLS and MM estimates for p-values are plotted against one another in Figure 4.6.

Carhart four factor Model: Estimated p-values using 200 test portfolios

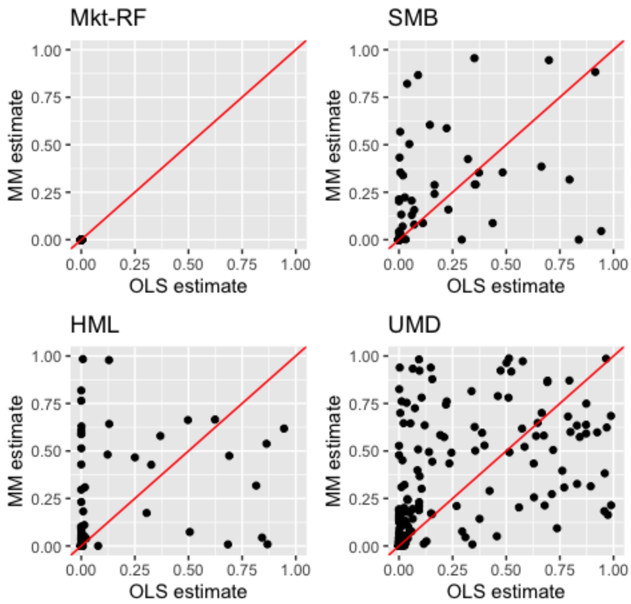

Figure 4.6: The estimated p-values using OLS & MM estimation combined with 200 test portfolios. A linear relationship with slope 1 and intercept 0 is added for easy comparison between methods.

A first, yet again unsurprising result is the highly significant excess market return factor. For the SMB factor, the estimation methods yield concurring conclusions in 94.5% of the regressions. Similarly, for the HML factor, the conclusions based on OLS and MM estimation are identical in 90% of the regressions. Additionally, both factors are significant in 87.5% and 91% respectively using OLS. For MM estimates, the factors are significant in 83.5% and 83% of the cases respectively. For the Fama-French three factors, both methods seem in agreement.

For the UMD factor, however, the data tells a different story. This factor is rendered insignificant in 43.5% of regressions using OLS. For MM estimation, the factor is rendered insignificant in 60.5% of the regressions. Already, there seems to be a difference in estimation methods larger than previously seen. In fact, in 19% of the regressions MM-estimates are rendered insignificant, while OLS estimates are significant. For the SMB and HML factors, this is 5% and 10% respectively. If an estimator is robust against outliers, such differences should not occur.

To examine these differences more clearly, a distinction will be made on estimates having a p-value larger than 0.10 using MM estimation, while simultaneously having a p-value smaller than 0.05 using OLS estimation. This encompasses 14.5% of all 200 regressions. The region is marked with a red rectangle in Figure 4.7.

UMD factor: Estimated p-values using 200 test portfolios

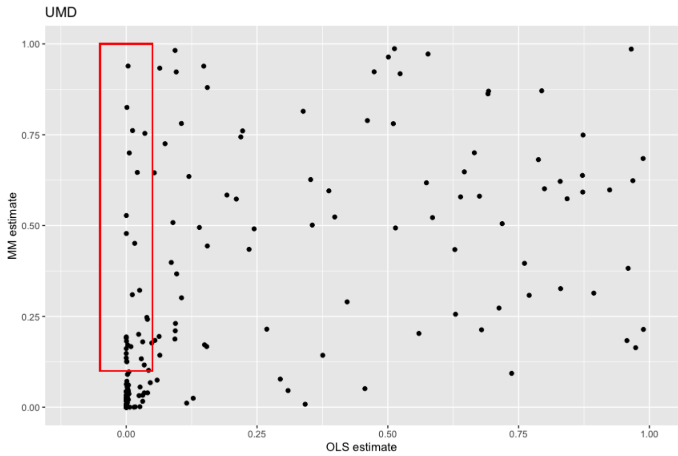

Figure 4.7: The estimated p-values using OLS & MM estimation combined with 200 test portfolios for the UMD factor. A red rectangle defines the region of interest.

To research what might cause such differences, a single dependent variable is selected. Using this single dependent variable, potential outliers in the underlying data are studied. For this research, the dependent variable resulting in the largest difference between methods for the UMD factor is selected. This dependent variable is a portfolio return of the sorted portfolio ranked second largest on size (market equity) and highest in operating profitability (ME4 OP5). This leads to the UMD factor having an estimated coefficient of 0.053 and respective p-value of 0.0034 using OLS. Hence, the factor is significant when estimating the sensitivity with OLS. For MM-estimation, however, the factor has an estimated coefficient of 0.0014 and respective p-value of 0.9390. The conclusion based on this p-value leads to an insignificant effect of the UMD factor. Based on this dependent variable, a detailed case study is conducted as to why these differences occur.

The goal of this case study is to determine if the differences in estimated p-values are the result of outliers. To be able to do so, outliers must first be identified. As a starting point the model estimated by OLS is used, on which regression diagnostics can be applied. The regression diagnostics are used as a stepping stone towards a formal identification of outliers. Firstly, the standardized residuals are plotted against the fitted values in Figure 4.8.

ME4 OP5: Standardized Residuals vs Fitted Values

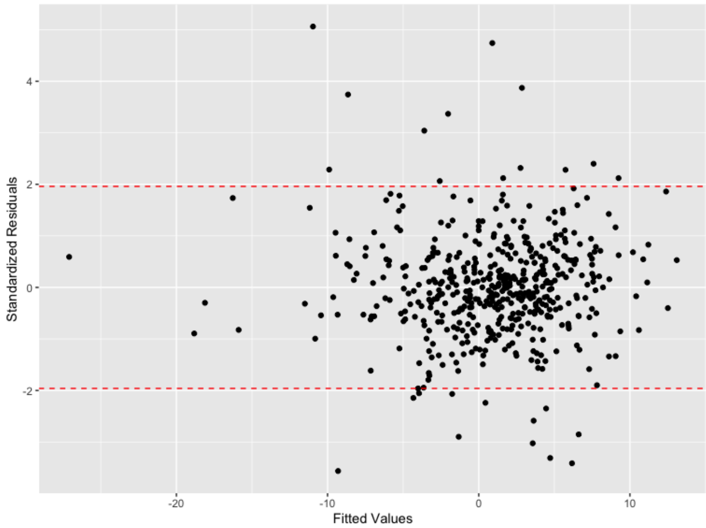

Figure 4.8: Standardized residuals resulting from OLS estimation are plotted against the fitted values. The two horizontal, dashed lines are at 1.96 and −1.96, respectively.

The dashed lines correspond to the 95% region of the standard normal distribution. Residuals outside of these confidence bands are suspicious. However, this plot is not conclusive. It could be the case, for example, that certain outliers have such a large effect on the estimated coefficients that its residual is not large (Heritier, Cantoni, Copt & Victoria-Feser, 2009). A different plot which can be used to assess the normality of the residuals is a Quantile-Quantile plot. In this plot the residual quantiles are plotted against their theoretical normal counter parts. If outliers are present, which in turn lead to larger residuals, the residual distribution is expected to be heavy tailed. This can also be seen in Figure 4.9.

ME4 OP5: Residual Quantiles vs Normal Quantiles

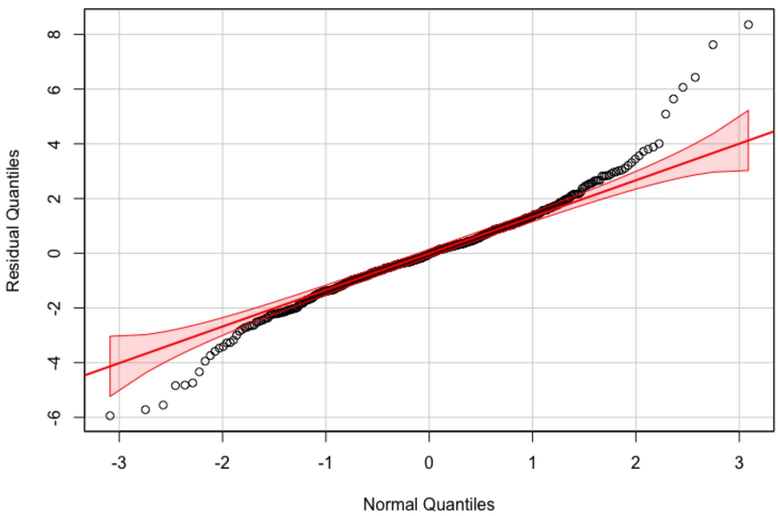

Figure 4.9: Observed residual quantiles using OLS versus theoretical normal quantiles. A red confidence envelope at a 95% confidence level is plotted to accurately assess the normality of the residuals.

The next step is to run regression diagnostics for the model estimated by MM. If it is the case that outliers are present, MM estimation should assign a lower weight to the outliers, as per Section 3.2. As a result, the residuals for the outliers should increase. The observation having the largest standardized residual in Figure 4.8 is identified as observation 293 out of 498. Using OLS this observation has a standardized residual equal to 5.0608, rounded to 4 decimals. Similarly, the largest standardized residual for MM estimation is identified as observation 293 out of 498. This standardized residual has a value of 5.6025, rounded to 4 decimals. This can also been seen in Figure 4.10.

ME4 OP5: Standardized Residuals vs Fitted Values using MM estimation

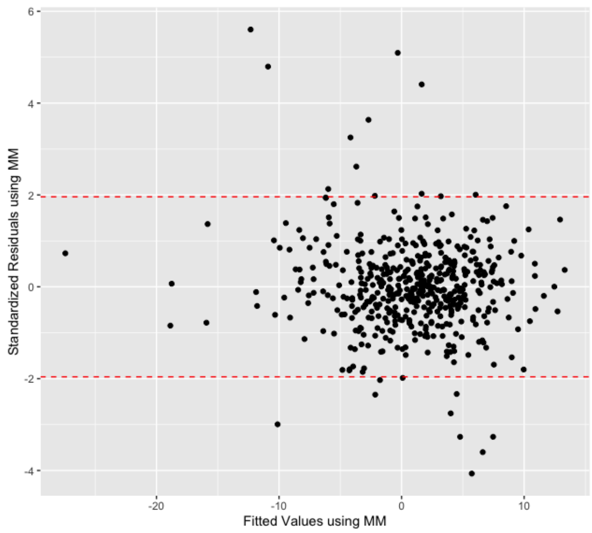

Figure 4.10: Standardized residuals resulting from MM estimation are plotted against the fitted values. The two horizontal, dashed lines are at 1.96 and −1.96, respectively.

In addition to the largest standardized residual, it also seems as if the residuals have a higher density around 0. The mean of the standardized OLS residuals is 0.0012, rounded to 4 decimals. The mean of the standardized MM residuals is 0.000, rounded to 4 decimals. Additionally, the median absolute deviation (MAD) of the standardized OLS residuals is 0.8067, rounded to 4 decimals. For the standardized MM residuals, the MAD equals 0.7297. These results are in line with expectation. As MM estimation places a lower weight on outliers, the model should provide a better fit for non-outliers than with OLS estimation. As a result the residuals for non-outliers should be smaller.

In addition to 4.10, a Quantile-Quantile (QQ) plot is presented in Figure 4.11. In this plot, the effect of larger residuals for outliers using MM estimation can clearly be seen in the heavy tails.

ME4 OP5: Residual Quantiles vs Normal Quantiles using MM estimation

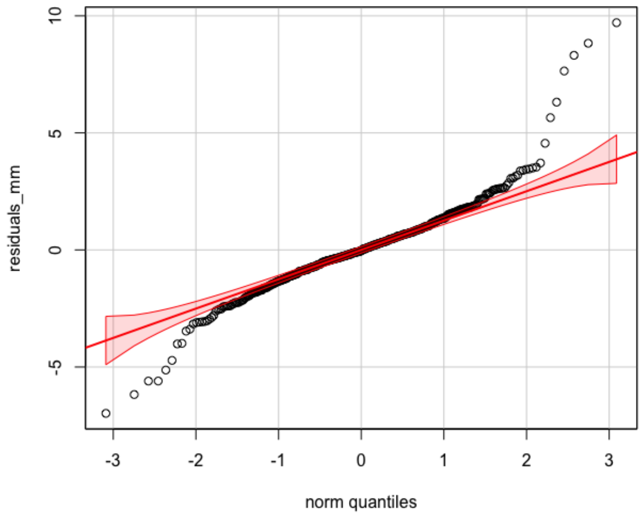

Figure 4.11: Observed residual quantiles using MM versus theoretical normal quantiles. A red confidence envelope at a 95% confidence level is plotted to accurately assess the normality of the residuals.

Although these plots provide a strong intuition that outliers are in fact present, a more formal approach is needed. To identify outliers in a structured manner, MM-regression can be used. Recall from Section 3.2 that MM-estimation can essentially be seen as weighted least squares. Outliers can then be identified using the data-dependent weights assigned in the estimation. The smaller the weight of an observation, the larger the outlier. To provide an intuition behind the weights and their respective values, a density plot of the weights is presented in Figure 4.12.

ME4 OP5: Density Plot of Weights in MM Estimation

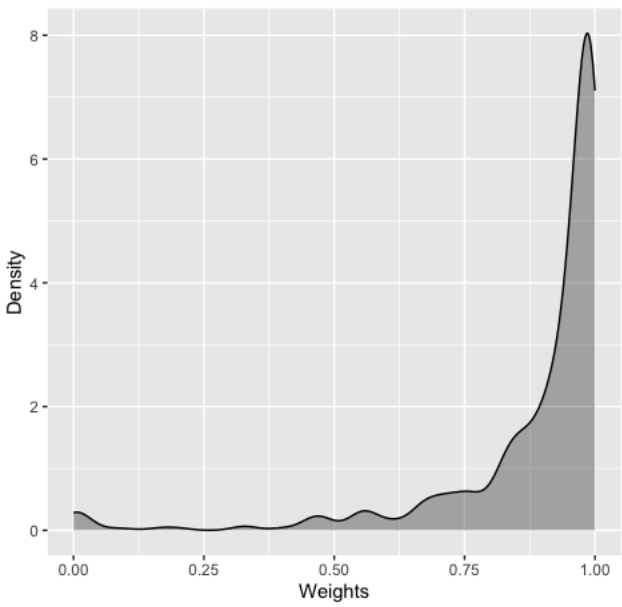

Figure 4.12: A density plot for the weights of the ME4 OP5 dependent variable. Note that the weights have a value between 0 and 1 and that the smaller the weight, the large the outlier.

As expected, the majority of the data points are concentrated in the right tail and thus not seen as outliers. Also in line with expectations, is the fact that the density of the weights decreases with the value of the weight. However, in the left tail an increase can be seen in the density. This implies that a relatively large number of observations is given a low weight, and provides further evidence of the presence of outliers.

To formally define outliers, an arbitrary threshold value is selected where observations having a weight smaller than the threshold value are seen as outliers. For example, if the threshold value is 0.10 all observations having a weight smaller than 0.10 are identified as outliers. It is important to consider the total number of observations, as well as the value of the weight itself as the threshold value is arbitrary. To illustrate this, a table containing different threshold values and the consequential percentage of observations identified as outlier is presented in Table 4.2.

Threshold Values and the Percentage of Data Deemed an Outlier

Threshold	**Percentage Deemed Outlier**
0.01	1.41%
0.05	2.01%
0.10	2.21%
0.25	2.61%
0.50	4.82%

Table 4.2: Different threshold values and the percentage of observations deemed outliers. Note the non-linear relationship between threshold values and percentage deemed outlier. Additionally, the effect of the increase in the left-tail of the empirical distribution of the weights can clearly be seen.

Based on Table 4.2 the decision is made to take 0.01 and 0.50 as threshold values for outliers. The reason for this is that a threshold equal to 0.01 straightforwardly deems the smallest percentage of data points to be outliers. However, the argument can also be made that observations having a weight smaller than 0.50 are significantly down weighted. Thus, both thresholds will be used. Consequentially, results for both thresholds will be reported. Using the thresholds, the outliers can be identified. In Figure 4.13 the residuals resulting from outliers are marked red, using a threshold of 0.01.

ME4 OP5: Residuals Resulting from Outliers vs Fitted Values using OLS, Threshold = 0.01

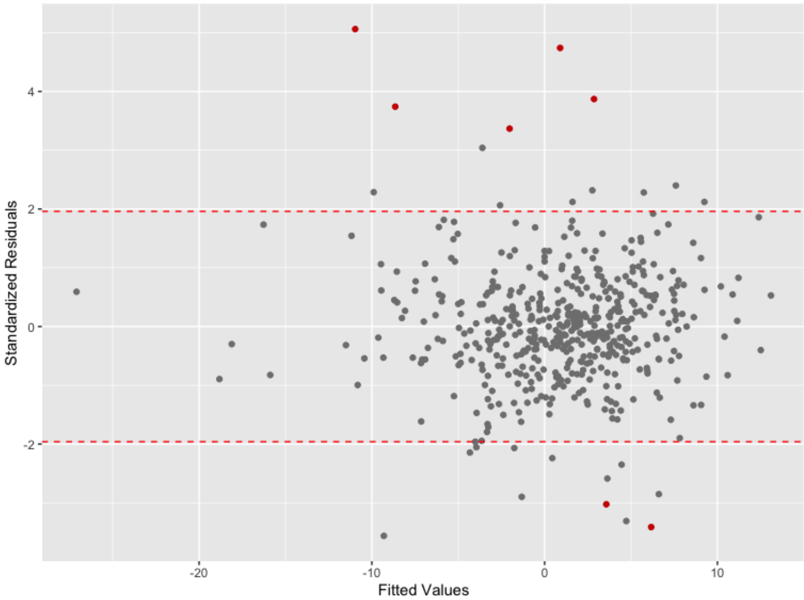

Figure 4.13: The outliers correspond to the red points in the scatter plot. The dashed red lines are at 1.96 and −1.96 respectively, indicating the bounds for the 95% interval of a standard normal distribution.

Similarly, in Figure 4.14 outliers are identified using a threshold of 0.50. The residuals resulting from outliers are marked red. Comparing Figure 4.13 with Figure 4.14, logically a larger number of observations are deemed outliers. Additionally, some of the residuals are inside of the 95% confidence bounds, albeit near the edges. This reflects the need for a formal identification method, as merely using visual tools would not have identified these residuals as outliers.

ME4 OP5: Residuals Resulting from Outliers vs Fitted Values using OLS, Threshold = 0.50

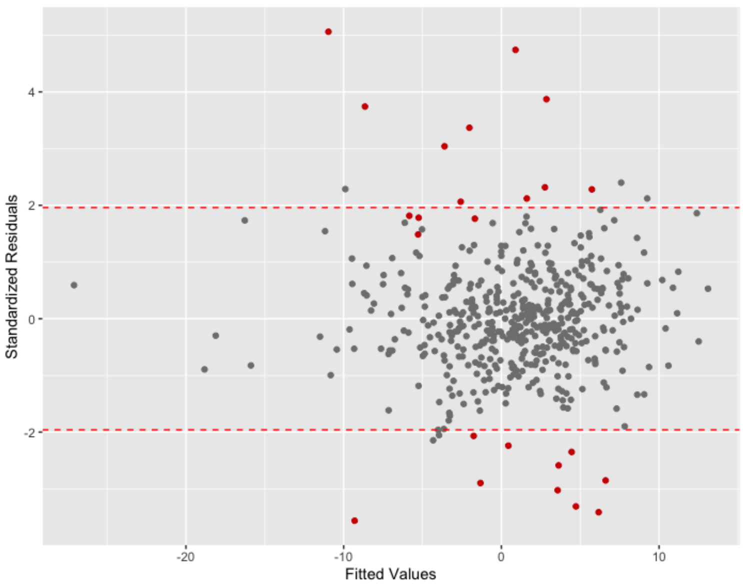

Figure 4.14: The outliers correspond to the red points in the scatter plot. The dashed red lines are at 1.96 and −1.96 respectively, indicating the bounds for the 95% interval of a standard normal distribution.

Now that the outliers have been identified, their influence on the previously estimated p-value can be determined. This is done according to the procedure described in Section 3.3, where the outliers are removed from the data and the model is estimated again using OLS. The newly estimated p-value of the UMD factor is known as the skipped p-value. This skipped p-value is more robust than the OLS p-value, but less robust than the MM p-value. Additionally, the MM estimator uses different weights for different observations. The skipped p-value, however, is estimated using a different data set but assigns equal weight to all data points. Therefore, it could well be the case that differences between MM and skipped estimation are present.

However, since the largest difference between the OLS p-value and MM p-value is selected, it is expected that removing the outliers from the data should result in the same statistical conclusion. In other words, the factor is expected to have a skipped p-value of at least larger than 0.05 and thus be insignificant. The estimated p-values for all three methods can be seen in Table 4.3.

ME4 OP5: Three estimated p-values

$\widehat{\mathbf{p}}_{\mathbf{OLS}}$	$\widehat{\mathbf{p}}_{\mathbf{SKIP}}$	$\widehat{\mathbf{p}}_{\mathbf{MM}}$
0.0034	0.6503	0.9390

Table 4.3: Three different estimations of the the p-value for the UMD factor using ME4 OP5 as dependent variable. The skipped estimator is denoted by $\widehat{p}_{SKIP}$. Note that in the calculation of the skipped estimator a threshold of 0.01 is used to identify outliers.

In spite of the difference between $\widehat{p}_{SKIP}$ and $\widehat{p}_{MM}$, this result is in line with expectations. As can be seen, the UMD factor is rendered insignificant when the previously identified outliers are dropped from the data. This implies that outliers in the data lead to the significance of the UMD factor, and hence the estimator is not robust. Clearly, this could lead to potentially misleading conclusions and the incorrect inclusion of the UMD factor in an asset pricing model.

The results from this case study can be extrapolated to all of the dependent variables in the region of interest in Figure 4.7. That is, the 'skipped' p-value estimator can be estimated for all the dependent variables leading to differences in conclusions for the UMD factor. A histogram containing the skipped p-values for all 29 cases can be seen in Figure 4.15.

Estimated p-values After Removing Outliers

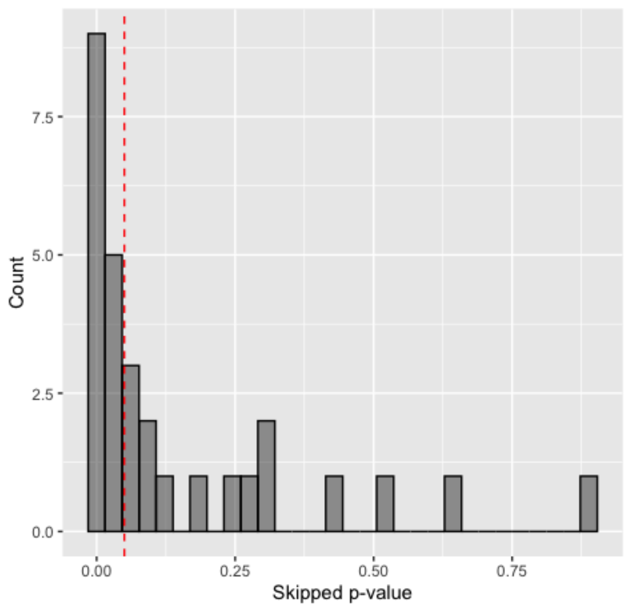

Figure 4.15: The histogram contains 29 estimated p-values for each of the dependent variables under which differences between OLS and MM estimation arose. The p-values shown in this plot are estimated through the aforementioned outlier removal procedure (skipped estimation) with a threshold of 0.01. A dashed red line is added to indicate the significance level of 0.05.

Interestingly enough, in the majority of the cases the UMD factor is in fact rendered significant even when outliers are removed based on a threshold equal to 0.01 (51.72%). However, as also mentioned there are multiple reasons for this. Additionally, the threshold value of 0.01 is chosen arbitrarily. It is interesting to study what happens once this value is increased.

For the previously selected dependent variable (ME4 OP5) the threshold value is varied according to the values in Table 4.2 , and the same procedure is conducted. That is, for each different threshold value, the skipped p-value is calculated. The results can be seen in Table 4.4.

ME4 OP5: Skipped p-value for Different Thresholds

Threshold	$\widehat{\mathbf{p}}_{\mathbf{SKIP}}$
0.01	0.6503
0.05	0.7281
0.10	0.6893
0.25	0.8766
0.50	0.9166

Table 4.4: The estimated p-value based on removing outliers identified by the respective thresholds for ME4 OP5, rounded to 4 decimals.

In accordance with previous expectations, the skipped p-value increases towards the MM estimated p-value of 0.9390. As per Figure 4.15, this does not seem to be the case for all dependent variables. As an example, the BM4 OP2 dependent variable can be used. This dependent variable is created by sorting on firm value (book-to-market) and Operating Profitability, and then selecting the fourth highest value (BM4) and the second lowest Operating Profitability (OP2). The skipped p-values for this dependent variable can be seen in Table 4.5.

BM4 INV2: Skipped p-value for Different Thresholds

Threshold	$\widehat{\mathbf{p}}_{\mathbf{SKIP}}$
0.01	0.034
0.05	0.053
0.10	0.001
0.25	0.003
0.50	< 0.001

Table 4.5: The estimated p-value based on removing outliers identified by the respective thresholds for BM4 OP2, rounded to 4 decimals. The p-value of < 0.001 indicates a p-value smaller than 0.001

Contrary to the results for the ME4 OP5 return series, there seems to be no pattern in the estimated p-values. For a threshold of 0.01 the skipped p-value is estimated at 0.034. For the next threshold it increases to 0.053. In this it case it is marginally significant, but given the p-value the evidence is weak. For the next thresholds the p-value decreases.

Given that these are merely two examples, the results can be aggregated for all dependent variables leading to initial differences in methods. To provide an insightful aggregation, a new metric is introduced. This metric is the percentage of differences which can be explained by outliers. Outliers are said to explain differences if the skipped p-value and the MM p-value lead to the same conclusions.

As an example, take the 0.01 threshold, the ME4 OP5 return series and the BM4 OP2 returns series. From Tables 4.3 and 4.5 it can be seen that the skipped p-values are estimated at 0.6503 and 0.034, respectively. The MM p-values are estimated at 0.9390 and 0.1162, respectively. This means that for the ME4 OP5 return series, the skipped p-value and MM p-value lead to the same conclusion that the UMD factor is insignificant. For the BM4 OP2 return series, the skipped p-value and MM p-value lead to different conclusions. Therefore, it is said that the percentage of differences which can be explained by outliers is 50% us a threshold of 0.01.

This metric can be computed for each different threshold and all of the dependent variables leading to differences. That is, all dependent variables having $\widehat{p}_{MM} > 0.10$ and $\widehat{p}_{OLS} < 0.05$. The results for each threshold can be found in Table 4.6.

Percentage Explained by Outliers for Different Thresholds

Threshold	**Percentage explained by Outliers**
0.01	48.28%
0.05	55.17%
0.10	58.62%
0.25	51.72%
0.50	58.62%

Table 4.6: The Percentage explained by outliers for each respective threshold for the UMD factor.

As expected the percentage explained by outliers increases for the first three threshold values. However, for the threshold of 0.25, the percentage explained by outliers interestingly decreases and then bounces up again for the threshold of 0.50. With exception of the 0.01 threshold, for all threshold the majority of the differences can be explained by removing the outliers from the data. As stated previously, it makes sense that there still remain differences even when outliers are removed. This is due to the lower robustness of the skipped estimator, and the fact that equal weights for data points are used in skipped estimation.

To conclude on the Carhart (1997) four factor model, it has been established that differences in statistical conclusions arise based on the estimation method and different dependent variables. The research then focused on the cases in which OLS rendered the UMD factor significant (i.e. a p-value smaller than 0.05) and MM estimation estimated the p-value to be larger than 0.10. A specific dependent variable was selected to function as a case study. Using this dependent variable, outliers were identified through the weights used in MM regression, as per Section 3.2. By removing these outliers, skipped estimators were used to re-estimate the p-values. Using the skipped p-values, differences between OLS-estimated and MM-estimated p-values could be explained.

4.4 Aggregation of four factor models

Given the large number of factors related to returns, it makes sense to extrapolate the methodology and procedures used in the previous Section for the UMD factor to other factors. As described in Section 2, this research uses an open source data library provided by Feng et al. (2020). Excluding the risk-free rate, $Mkt - R_f$, SMB, HML and UMD this leaves 142 risk factors which can be examined. The models for each of these factors are as follows:

$$E[R_t] - R_{f,t} = \beta_{MKT}(R_{M,t} - R_{f,t}) + \beta_{SMB}(SMB_t) + \beta_{HML}(HML_t) + \beta_G(G_t) + \epsilon_t, \quad (4.1)$$

where G_t represents the additional fourth factor. Similar to the Carhart four factor model, the final factor is the factor of interest. And also similar to the Carhart four factor model, the focus will be on the fourth factor's p-value.

The goal by studying these other factors is to find evidence in favor of, or against, the influence of outliers on the significance of factors. To do so, aggregation metrics are computed for each factor. This is done as follows. Using the previously described 200 test assets, the models are fit using both OLS estimation and MM estimation. This results in two estimated p-values for the combination of a factor and a dependent variable. The combination of a factor and a dependent variable is denoted as a case. Note that as there are 200 different dependent variables, each factor has 200 different cases.

Then, all the cases in which the estimated p-value using OLS is smaller than 0.05, and using MM estimation is larger than 0.10 are selected. This is where the first metric, reflecting the robustness of the estimator of a factor is computed. Clearly, the number of cases where differences arise is informative about the robustness. The first metric is the percentage of all 200 cases having a MM-estimated p-value larger than 0.10 and OLS-estimated p-value smaller than 0.05. This is also denoted as the percentage of regressions (cases) leading to different conclusions.

As an example, let's assume the UMD factor has 20 cases for which $p_{MM} > 0.10$ and $p_{OLS} < 0.05$. Also described is that there are 200 different test assets. Then the percentage of regressions which lead to differences is 10%.

The second metric which is computed has already been mentioned in Section 4.3: the percentage of differences which can be explained by outliers. This percentage is calculated through the use of skipped estimators and a threshold value for the weights assigned to observations in MM estimation. Using the threshold value, outliers are identified.

For the factors, two different thresholds are used. The first threshold is set at 0.01. This threshold selects the smallest number of observations. The second threshold is 0.50. The argument for using this threshold is that an observation having a weight smaller than 0.50 is significantly down-weighted. Finally, it is beneficial to use two thresholds as the behavior of the skipped estimator can be studied.

Using the threshold value, outliers are identified. These outliers are then removed from the data. Using the new data-set, the model is refit using OLS. The resulting p-value is known as the skipped p-value. Recall that using the entire data-set and OLS resulted in an estimated p-value smaller than 0.05.

Outliers are then said to explain the difference in p-values between OLS and MM estimation if the skipped p-value and MM p-value lead to the same conclusion. That is if $\widehat{p}_{SKIP} > 0.05$. Note that a skipped p-value of 0.05 is used instead of 0.10. This is done to account for the fact that the skipped estimator is less robust than the MM-estimator.

Returning to our previous example of the UMD factor, 20 out of 200 cases lead to differences in conclusions. Suppose that out of those 20 cases, 10 of the newly estimated skipped p-values are larger than 0.05. This means that in 10 out of the 20 cases, the conclusions based on MM estimation and skipped estimation overlap. It is then said that the percentage of differences explained by outliers is 50%.

Table 4.7 shows these two metrics for a number of factors, using a threshold of 0.01. Note that the table is incomplete as there are 142 factors for which these metrics are computed.

Aggregated Differences and Explanations for Other Factors, Threshold = 0.01

Factor Abbrevation	Percentage Differences	Percentage of Differences Explained by Outliers
beta	11.50%	52.17%
ep	15.50%	41.94%
dy	6.500%	46.15%
sue	13.00%	53.85%
pps	8.000%	43.75%
...	...	...

Table 4.7: Differences in statistical conclusions aggregated for each factor. Additionally, the percentage of the differences explained by outliers is added. Note that the table is incomplete as there are 142 factors, results for all factors will be presented. A list containing the full names of each of the factors and their respective abbreviations can be found in Appendix A.

To summarize the results for all 142 factors, Figure 4.16 shows a density plot for the percentage of regressions with different conclusions. In other words, the percentage of regressions where the p-value of the factor is estimated below 0.05 using OLS, and above 0.10 using MM estimation.

Density plot of The Percentage of Regressions with Different Conclusions for All Factors

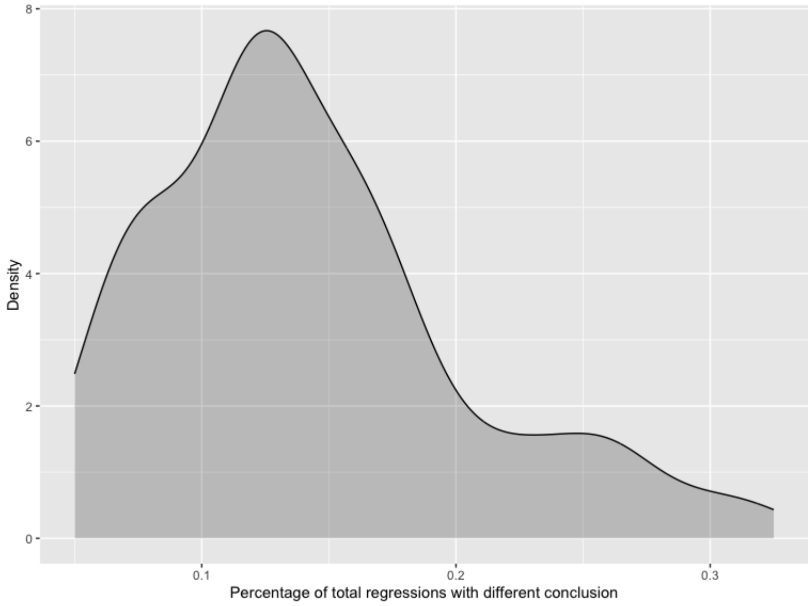

Figure 4.16: The density plot shows the approximate distribution of the percentage of regressions with different conclusions.

In addition to Table 4.7 and Figure 4.16 the descriptive statistics of what percentage of regressions have different statistical conclusions for each factor are presented in Table 4.8.

Descriptive statistics of the Percentage of Regressions with Different Conclusions

Min.	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max**	**Std.**
5.00%	10.00%	13.00%	14.24%	17.00%	32.50%	6.10%

Table 4.8: This table contains the descriptive statistics of the percentage of regressions with different conclusions. Note that Qu. is an abbreviation for quantile.

.

It can be seen that the mean of the percentage of regressions with different conclusions for each factor lies around 14%, with a minimum of 5% and a maximum of 32.5%. The results are based on the aggregations of all factors. This means that on average, 14% of the estimations for a factor will result in a p-value larger than 0.10 using MM estimation, and a p-value smaller than 0.05 using OLS. This percentage provides a first piece of evidence that, on average, factor p-value estimations are influenced by outliers.

To come to a full conclusion, the percentage explained by these outliers is taken into account. A difference can be explained by outliers if the skipped p-value and MM p-value give rise to the same conclusion. Figure 4.17 shows the density of the percentage of differences explained by outliers, using a threshold of 0.01.

Density plot of The Percentage of Differences which can be Explained by Outliers, Threshold = 0.01

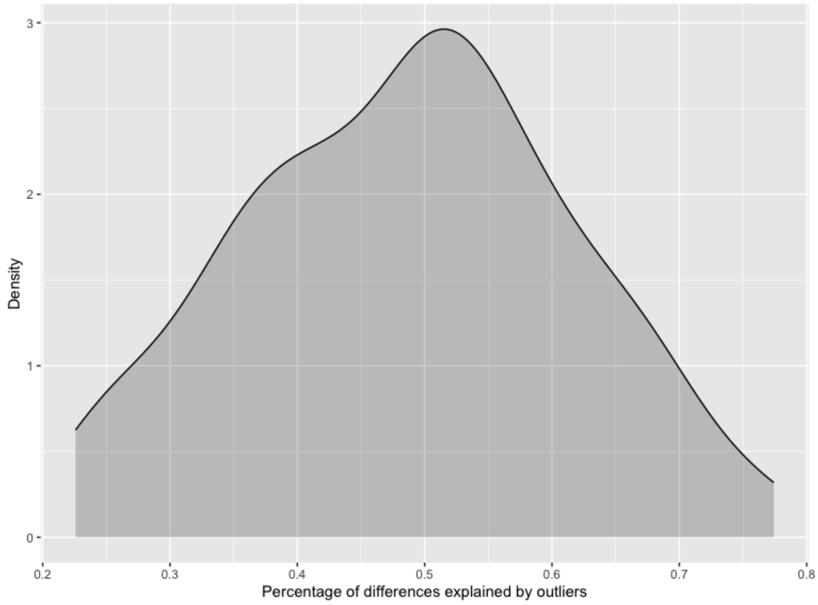

Figure 4.17: The density plot shows the approximate distribution of the percentage of differences explained by outliers.

Additionally, the descriptive statistics for the percentage of the differences which can be explained by outliers are presented in Table 4.9.

Descriptive Statistics of the Percentage of Differences Explained by Outliers, Threshold = 0.01.

Min.	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max**	**Std.**
22.58%	39.13%	50.00%	48.68%	57.69%	77.42%	12.81%

Table 4.9: . This table contains the descriptive statistics of the percentage of differences which can be explained by outliers. A threshold of 0.01 is used. Note that Qu. is an abbreviation for Quantile.

On average, 48.68% of the differences for a factor disappear when outliers are removed from the data and a threshold of 0.01 is used. Additionally, the lowest percentage of differences explained by outliers for a factor was 22.48%. Contrary, the highest percentage for a factor was 77.42%. At a first sight, the average of 48.68% seems to lead to the conclusion that outliers cannot explain the differences in estimation methods. However, the results using a threshold of 0.50 should also be taken into account.

Consequentially, the final step in this study is to increase the value of the threshold. The new threshold is set to 0.50. As the threshold increases, more observations are identified to be an outlier. This is in turn should yield more concurring conclusions between skipped estimation and MM estimation.

The reason for this is as follows. Suppose that an observation has a weight of 0.10. Given this weight, it should be seen as suspicious. Using a threshold equal to 0.01, this observation will not be identified as an outlier. Consequentially, this observation will be included in skipped estimation. Its influence in MM estimation, however, is relatively small. Clearly, if the threshold value is increased the effect of suspicious observations on the skipped estimator will disappear. This should then lead to a higher percentage of overlapping conclusions.

In Figure 4.18 a density plot can be seen reflecting the approximate density of the percentage of differences which can be explained by outliers using a threshold of 0.50.

Density plot of The Percentage of Differences which can be Explained by Outliers, Threshold = 0.50

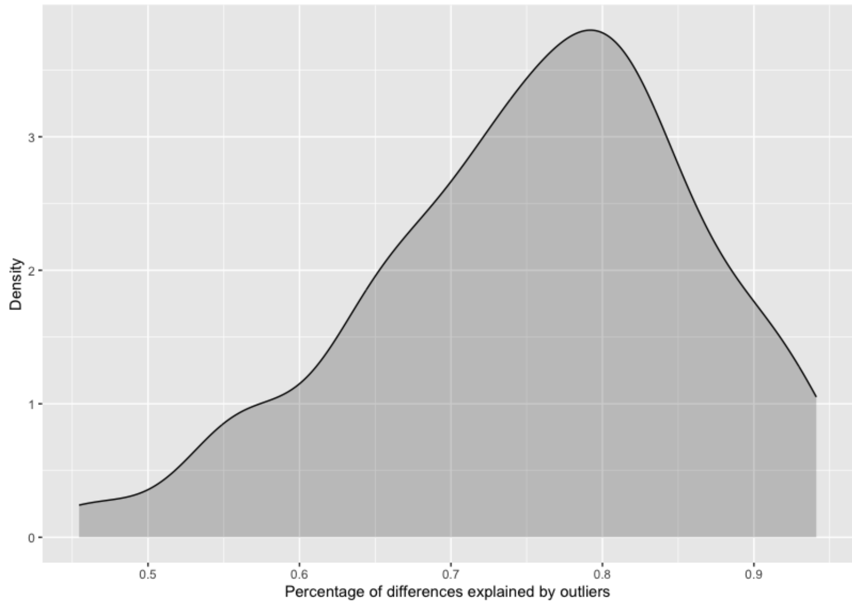

Figure 4.18: The density plot shows the approximate distribution of the percentage of differences explained by outliers.

Already it can be seen that a higher percentage of the differences is explained by outliers using a threshold of 0.50 compared to a threshold of 0.01. Furthermore, the descriptive statistics for the percentage of differences explained by outliers can be seen in Table 4.10.

Descriptive Statistics of the Percentage of Differences Explained by Outliers, Threshold = 0.50.

Min.	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max**	**Std.**
45.45%	68.81%	76.92%	75.44%	82.55%	94.12%	10.77%

Table 4.10: This table contains the descriptive statistics of the percentage of differences which can be explained by outliers. A threshold of 0.50 is used. Note that Qu. is an abbreviation for Quantile.

Using a threshold of 0.50, on average, 75.44% of the differences in statistical conclusions based on OLS and MM estimation are the result of outliers. Clearly, this average percentage provides strong evidence in favor of outliers leading to differences in conclusions between OLS and MM estimation. It also concludes the study, and shows that outliers significantly influence estimated factor p-values in multi factor asset pricing models.

Chapter 5

Conclusion

In this paper, the robustness of estimated p-values for factor sensitivities was researched. For this purpose an open source factor library containing 149 risk factors was used. Additionally, 200 different test assets were used as dependent variables. Using the Fama and French (1993) three factor model as a base model, the estimated p-values were studied for 143 factors functioning as an additional fourth factor. On average, 14.24% of the regressions for each factor have an MM estimated p-value larger than 0.10, yet an OLS estimated p-value smaller than 0.05. So, on average, the estimated p-value of a factor was not robust in 14.24% of the cases.

Using the robustness weights from the MM-regression and two different threshold values of 0.01 and 0.50 outliers were identified. It was then researched if these outliers are at the root of these differences. On average, 48.68% of the differences in conclusions for a factor could be explained by the presence of outliers when a threshold of 0.01 is used. For the threshold of 0.50, 75.44% of the differences in conclusions for a factor could be explained by the presence of outliers.

These findings provide strong evidence in favor of outliers leading to misleading conclusions in multi-factor asset pricing models. It is therefore advised to take these findings into account when researching potential new factors. Once a factor has been developed, it should be tested using both OLS and MM estimation. Additionally, the findings of this study can be used to assess whether a factor should be included into an already existing asset pricing model.

The results of this study are, to the best of the author's knowledge, the first of its kind. However, the fact that misleading conclusions in multi factor asset pricing models are present is not new. Hou et al. (2018), for example, find replicability issues for factor significance. Similarly, McLean and Pontiff (2016) find that factors do not perform as well out-of-sample. An interesting topic for further research would be to examine if the results in the two previously mentioned papers are the result of outliers.

A potential shortcoming of this study is the dependent variables that are used. There are 200 different test assets, resulting from 8 different portfolio sorts. However, 7 out of the 8 portfolio sorts are either sorted on size, value or both. It could be interesting to include other dependent variables that are sorted on neither size nor value, to even further account for the findings of Lo and MacKinlay (1990). Additionally, it would be interesting to study the observations identified as outliers in Section 4.3 in more detail. It could be the case that these are the result of noteworthy economic events, such as large recessions.

On top of this, this study decides to focus on the estimated p-value of factors which are added to the Fama and French (1993) three factor model. The coefficient sign and magnitudes, however, are only taken into account for the Carhart (1997) four factor model. The effect outliers have on the estimated coefficients could also be studied. Additionally, only the fourth factor is taken into account in this thesis. It might also be interesting to examine the three Fama and French (1993) factors in combination with the fourth factor.

Finally, it could be interesting to include results for the intercept of the asset pricing models used in this thesis. The intercept, also known as alpha, tells us more about the model fit. Including results for both MM and OLS estimation for alpha allows for an additional evaluation of the estimators' robustness. Using the Gibbons-Ross-Shanken test (Gibbons, Ross & Shanken, 1989), alpha's resulting from each estimation method can be evaluated on their joint significance. For example, if the alpha's resulting from OLS estimation turn out to not be significantly different from zero, this would indicate the asset pricing model fits the returns well. If it then turns out to be the case that the alpha's are significantly different from zero using MM estimation, one should be concerned that the model's estimators are not robust.

To conclude, this research showed that, on average, the OLS estimated and MM estimated p-value for a factor lead to different conclusions in 14.24% of the cases. This implies that, on average, 14.24% of the cases the statistical conclusion of the significance of a factor is not robust. Out of these cases, on average, 48.68% of the differences could be explained by outliers using a threshold of 0.01. Using a threshold of 0.50, 75.44% of the differences could be explained by outliers on average. Tying all of these findings together leads to the conclusion that in some cases, outliers lead to misleading statistical conclusions on the significance of a factor. It would be interesting to research how these findings tie into the replicability issues and factor zoo within the realm of asset pricing.

References

Albuquerque, J. S. & Biegler, L. T. (1996, 10). Data Reconciliation and Gross-Error Detection for Dynamic Systems. *Aiche Journal*, *42*(10), 2841–2856. doi: 10.1002/aic.690421014

Bickel, P. J. (1975, 6). One-Step Huber estimates in the linear model. *Journal of the American Statistical Association*, *70*(350), 428–434. doi: 10.1080/01621459.1975.10479884

Boamah, N. A. (2015, 11). Robustness of the Carhart four-factor and the Fama-French Three-Factor Models on the South African Stock Market. *Review of Accounting and Finance*, *14*(4), 413–430. doi: 10.1108/raf-01-2015-0009

Carhart, M. M. (1997, 3). On Persistence in Mutual Fund Performance. *Journal of Finance*, *52*(1), 57–82. doi: 10.1111/j.1540-6261.1997.tb03808.x

Chen, A. C. N. & Zimmermann, T. (2022, 1). Open Source Cross-sectional Asset Pricing. *Critical Finance Review*, *11*(2), 207–264. doi: 10.1561/104.00000112

Cochrane, J. H. (2011, 8). Presidential Address: Discount Rates. *Journal of Finance*, *66*(4), 1047–1108. doi: 10.1111/j.1540-6261.2011.01671.x

Fama, E. F. & French, K. R. (1993, 2). 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. F. & French, K. R. (2015, 4). A Five-Factor Asset Pricing Model. *Journal of Financial Economics*, *116*(1), 1–22. doi: 10.1016/j.jfineco.2014.10.010

Feng, G., Giglio, S. & Xiu, D. (2020, 6). Taming the Factor Zoo: A Test of New Factors. *Journal of Finance*, *75*(3), 1327–1370. doi: 10.1111/jofi.12883

French, K. R. (2023, 6). *U.S. Research Returns Data.* Retrieved from `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

Gibbons, M. R., Ross, S. A. & Shanken, J. (1989, 9). A test of the efficiency of a given portfolio. *Econometrica*, *57*(5), 1121–1152. doi: 10.2307/1913625

Guillaume, D., Dacorogna, M. M., Dave, R., Müller, U., Olsen, R. B. & Pictet, O. V. (1997, 4). From the Bird's Eye to the Microscope: A Survey of New Stylized Facts of the Intra-Daily Foreign Exchange Markets. *Finance and Stochastics*, *1*(2), 95–129. doi: 10.1007/s007800050018

Hampel, F. (1971, 12). A General Qualitative Definition of Robustness. *Annals of Mathematical Statistics*, *42*(6), 1887–1896. doi: 10.1214/aoms/1177693054

Harvey, C. R. & Liu, Y. (2019, 2). A Census of the Factor Zoo. *Social Science Research Network*. doi: 10.2139/ssrn.3341728

Harvey, C. R., Liu, Y. & Zhu, H. (2015, 10). ... and the Cross-Section of Expected Returns. *Review of Financial Studies*, *29*(1), 5–68. doi: 10.1093/rfs/hhv059

Heritier, S., Cantoni, E., Copt, S. & Victoria-Feser, M.-P. (2009). *Robust Methods in Biostatistics.* John Wiley & Sons. doi: 10.1002/9780470740538

Hou, K., Xue, C. & Zhang, L. (2018, 12). Replicating Anomalies. *Review of Financial Studies*, *33*(5), 2019–2133. doi: 10.1093/rfs/hhy131

Huber, P. J. (1964, 3). Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics*, *35*(1), 73–101. doi: 10.1214/aoms/1177703732

Huber, P. J. (1973, 9). Robust Regression: Asymptotics, Conjectures and Monte Carlo. *Annals of Statistics*, *1*(5). doi: 10.1214/aos/1176342503

Hössjer, O. (1992, 7). On the optimality of S-estimators. *Statistics Probability Letters*, *14*(5), 413–419. doi: 10.1016/0167-7152(92)90103-c

Jegadeesh, N. & Titman, S. (1993, 3). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance*, *48*(1), 65–91. doi: 10.1111/j.1540-6261.1993.tb04702.x

Jensen, T. Z. T., Kelly, B. T. & Pedersen, L. H. (2023, 10). Is There a Replication Crisis in Finance? *Journal of Finance*, *78*(5), 2465–2518. doi: 10.3386/w28432

Johansen, S. & Nielsen, B. (2013, 5). Outlier detection in regression using an iterated One-Step approximation to the Huber-Skip estimator. *Econometrics*, *1*(1), 53–70. doi: 10.3390/econometrics1010053

Litner, J. (1965, 2). The valuation of risk assets and the selection of risky investment in stock portfolios and capital budgets. *The Review of Economics and Statistics*, *47*, 13–37. doi: 10.2307/1924119

Lo, A. W. & MacKinlay, A. C. (1990, 7). Data-Snooping Biases in Tests of Financial Asset Pricing Models. *Review of Financial Studies*, *3*(3), 431–467. doi: 10.1093/rfs/3.3.431

Malin, M. & Veeraraghavan, M. (2004, 1). On the robustness of the Fama and French multifactor model: Evidence from France, Germany and United Kingdom. *International Journal of Business and Economics*, *3*(2), 155–176.

Mandelbrot, B. B. (1963, 1). The Variation of Certain Speculative Prices. *The Journal of Business*, *36*(4), 394 – 419. doi: 10.1086/294632

Maronna, R. A., Martin, R. D. & Yohai, V. J. (2006). *Robust Statistics: Theory and Methods (2nd ed.).* Wiley. doi: 10.1002/9781119214656

McLean, R. D. & Pontiff, J. (2016, 1). Does Academic Research Destroy Stock Return Predictability? *Journal of Finance*, *71*(1), 5–32. doi: 10.1111/jofi.12365

Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R. & Wu, A. Y. (2014, 5). On the Least Trimmed Squares Estimator. *Algorithmica*, *69*(1), 148–183. doi: 10.1007/s00453-012-9721-8

Pastor, L. & Stambaugh, R. F. (2003, 6). Liquidity Risk and Expected Stock Returns. *Journal of Political Economy*, *111*(3), 642–685. doi: 10.1086/374184

Rousseeuw, P. J. & Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* John Wiley & Sons. doi: 10.1002/0471725382

Rousseeuw, P. J. & Yohai, V. J. (1984). *Robust Regression by Means of S-Estimators.* Springer. doi: 10.1007/978-1-4615-7821-5\{_}15

Salibian-Barrera, M. & Yohai, V. J. (2006, 6). A Fast Algorithm for S-Regression Estimates. *Journal of Computational and Graphical Statistics*, *15*(2), 414–427. doi: 10.1198/106186006x113629

Sharpe, W. F. (1964, 9). Capital Asset Prices: a Theory of Market Equilibrium. *Journal of Finance*, *19*(3), 425–442. doi: 10.1111/j.1540-6261.1964.tb02865.x

Yohai, V. J. (1987, 6). High Breakdown-Point and High Efficiency Robust Estimates for Regression. *Annals of Statistics*, *15*(2), 642–656. doi: 10.1214/aos/1176350366

Appendix A

Factor abbreviations

The tables below contain the factor abbreviations, factors and their respective authors which have been used in this study. The factor library used was provided by Feng et al. (2020).

Factor Abbreviation	**Factor**	**Authors**
MktRf	Excess Market Return	Black and Jensen and Scholes
beta	Market Beta	Fama and Macbeth
ep	Earnings to price	Basu
dy	Dividend to price	Litzenberger and Ramaswamy
sue	Unexpected quarterly earnings	Rendelman and Jones and Latane
pps	Share price	Miller and Scholes
LTR	Long-Term Reversal	De Bondt and Thaler
lev	Leverage	Bhandari
cashdebt	Cash flow to debt	Ou and Penman
currat	Current ratio	Ou and Penman
pchcurrat	% change in current ratio	Ou and Penman
pchquick	% change in quick ratio	Ou and Penman
pchsaleinv	% change sales-to-inventory	Ou and Penman
quick	Quick ratio	Ou and Penman
salecash	Sales to cash	Ou and Penman
saleinv	Sales to inventory	Ou and Penman
salerec	Sales to receivables	Ou and Penman
baspread	Bid-ask spread	Amihud and Mendelson
depr	Depreciation / PP&E	Holthausen and Larcker
pchdepr	% change in depreciation	Holthausen and Larcker
SMB	Small Minus Big	Fama and French
HML	High Minus Low	Fama and French
STR	1-month momentum	Jegadeesh and Titman
mom6m	6-month momentum	Jegadeesh and Titman
mom36m	36-month momentum	Jegadeesh and Titman
sgr	Sales growth	Lakonishok and Shleifer and Vishny
cp	Cash flow-to-price	Lakonishok and Shleifer and Vishny
IPO	New equity issue	Loughran and Ritter

Factor Abbreviation	Factor	Authors
divi	Dividend initiation	Michaely and Thaler and Womack
divo	Dividend omission	Michaely and Thaler and Womack
acc	Working capital accruals	Sloan
sp	Sales to price	Barbee and Mukherji and Raines
cto	Capital turnover	Haugen and Baker
UMD	Momentum	Carhart
turn	Share turnover	Datar and Naik and Radcliffe
pchgm_pchsale	% change in gross margin - % change in sales	Abarbanell and Bushee
pchsale_pchinvt	% change in sales - % change in inventory	Abarbanell and Bushee
pchsale_pchrect	% change in sales - % change in A/R	Abarbanell and Bushee
pchsale_pchxsga	% change in sales - % change in SG&A	Abarbanell and Bushee
etr	Effective Tax Rate	Abarbanell and Bushee
lfe	Labor Force Efficiency	Abarbanell and Bushee
os	Ohlson's O-score	Dichev
zs	Altman's Z-score	Dichev
pchcapx_ia	Industry adjusted % change in capital expenditures	Abarbanell and Bushee
nincr	Number of earnings increases	Barth and Elliott and Finn
indmom	Industry momentum	Moskowitz and Grinblatt
ps	Financial statements score	Piotroski
bm_ia	Industry-adjusted book to market	Asness and Porter and Stevens
cfp_ia	Industry-adjusted cash flow to price ratio	Asness and Porter and Stevens
chempia	Industry-adjusted change in employees	Asness and Porter and Stevens
mve_ia	Industry-adjusted size	Asness and Porter and Stevens
dolvol	Dollar trading volume	Chordia and Subrahmanyam and Anshuman
std_dolvol	Volatility of liquidity (dollar trading volume)	Chordia and Subrahmanyam and Anshuman
std_turn	Volatility of liquidity (share turnover)	Chordia and Subrahmanyam an-dAnshuman
adm	Advertising Expense-to-market	Chan and Lakonishok and Sougiannis
rdm	R&D Expense-to-market	Chan and Lakonishok and Sougiannis

Factor Abbreviation	Factor	Authors
rds	R&D-to-sales	Chan and Lakonishok and Sougiannis
kz	Kaplan-Zingales Index	Lamont and Polk and Saa-Requejo
chinv	Change in inventory	Thomas and Zhang
chtx	Change in tax expense	Thomas and Zhang
ill	Illiquidity	Amihud
LIQ_PS	Liquidity	Pastor and Stambaugh
idiovol	Idiosyncratic return volatility	Ali and Hwang and Trombley
grltnoa	Growth in long term net operating assets	Fairfield and Whisenant and Yohn
ob_a	Order backlog	Rajgopal and Shevlin and Venkatachalam
grltnoa_hxz	Changes in Long-term Net Operating Assets	Fairfield and Whisenant and Yohn
cfp	Cash flow to price ratio	Desai and Rajgopal and Venkatachalam
rd	R&D increase	Eberhart and Maxwell and Siddique
cinvest	Corporate investment	Titman and Wei and Xie
roavol	Earnings volatility	Francis and LaFond and Olsson and Schipper
cinvest_a	Abnormal Corporate Investment	Titman and Wei and Xie
noa	Net Operating Assets	Hirshleifer and Hou and Teoh and Zhang
dnoa	Changes in Net Operating Assets	Hirshleifer and Hou and Teoh and Zhang
tb	Tax income to book income	Lev and Nissim
pricedelay	Price delay	Hou and Moskowitz
age	years since first Compustat coverage	Jiang and Lee and Zhang
egr	Growth in common shareholder equity	Richardson and Sloan and Soliman and Tuna
lgr	Growth in long-term debt	Richardson and Sloan and Soliman and Tuna
dcoa	Change in Current Operating Assets	Richardson and Sloan and Soliman and Tuna
dcol	Change in Current Operating Liabilities	Richardson and Sloan and Soliman and Tuna
dwc	Changes in Net Non-cash Working Capital	Richardson and Sloan and Soliman and Tuna

Factor Abbreviation	Factor	Authors
dnca	Change in Non-current Operating Assets	Richardson and Sloan and Soliman and Tuna
dncl	Change in Non-current Operating Liabilities	Richardson and Sloan and Soliman and Tuna
dnco	Change in Net Non-current Operating Assets	Richardson and Sloan and Soliman and Tuna
dfin	Change in Net Financial Assets	Richardson and Sloan and Soliman and Tuna
ta	Total accruals	Richardson and Sloan and Soliman and Tuna
dsti	Change in Short- term Investments	Richardson and Sloan and Soliman and Tuna
dfnl	Change in Financial Liabilities	Richardson and Sloan and Soliman and Tuna
egr_hxz	Change in Book Equity	Richardson and Sloan and Soliman and Tuna
ms	Financial statements score	Mohanram
chmom	Change in 6-month momentum	Gettleman and Marks
grcapx	Growth in capital expenditures	Anderson and Garcia-Feijoo
retvol	Return volatility	Ang and Hodrick and Xing and Zhang
zerotrade	Zero trading days	Liu
pchcapx3	Three-year Investment Growth	Anderson and Garcia-Feijoo
cei	Composite Equity Issuance	Daniel and Titman
nef	Net equity finance	Bradshaw and Richardson and Sloan
ndf	Net debt finance	Bradshaw and Richardson and Sloan
nxf	Net external finance	Bradshaw and Richardson and Sloan
rs	Revenue Surprises	Jegadeesh and Livnat
herf	Industry Concentration	Hou and Robinson
ww	Whited-Wu Index	Whited and Wu
roic	Return on invested capital	Brown and Rowe
tang	Debt capacity/firm tangibility	Almeida and Campello
op	Payout yield	Boudoukh and Michaely and Richardson and Roberts

Factor Abbreviation	Factor	Authors
nop	Net payout yield	Boudoukh and Michaely and Richardson and Roberts
ndp	Net debt-to-price	Penman and Richardson and Tuna
ebp	Enterprise book-to-price	Penman and Richardson and Tuna
chcsho	Change in shares outstanding	Pontiff and Woodgate
aeavol	Abnormal earnings announcement volume	Lerman and Livnat and Mendenhall
ear	Earnings announcement return	Kishore and Brandt and Santa-Clara and Venkatachalam
moms12m	Seasonality	Heston and Sadka
dpia	Changes in PPE and Inventory-to-assets	Lyandres and Sun and Zhang
pchcapx	Investment Growth	Xing
cdi	Composite Debt Issuance	Lyandres and Sun and Zhang
rna	Return on net operating assets	Soliman
pm	Profit margin	Soliman
ato	Asset turnover	Soliman
chatoia	Industry-adjusted change in asset turnover	Soliman
chpmia	Industry-adjusted change in profit margin	Soliman
cashpr	Cash productivity	Chandrashekar and Rao
sin	Sin stocks	Hong and Kacperczyk
rsup	Revenue surprise	Kama
stdcf	Cash flow volatility	Huang
absacc	Absolute accruals	Bandyopadhyay and Huang and Wirjanto
invest	Capital expenditures and inventory	Chen and Zhang
roaq	Return on assets	Balakrishnan and Bartov and Faurel
stdacc	Accrual volatility	Bandyopadhyay and Huang and Wirjanto
realestate_hxz	Industry-adjusted Real Estate Ratio	Tuzel
pctacc	Percent accruals	Hafzalla and Lundholm and Van Winkle
maxret	Maximum daily return	Bali and Cakici and Whitelaw
ol	Operating Leverage	Novy-Marx

Factor Abbreviation	**Factor**	**Authors**
ivg	Inventory Growth	Belo and Lin
poa	Percent Operating Accruals	Hafzalla and Lundholm and Van Winkle
em	Enterprise multiple	Loughran and Wellman
cash	Cash holdings	Palazzo
HML_Devil	HML Devil	Asness and Frazzini
gma	Gross profitability	Novy-Marx
orgcap	Organizational Capital	Eisfeldt and Papanikolaou
BAB	Betting Against Beta	Frazzini and Pedersen
QMJ	Quality Minus Junk	Asness and Frazzini and Pedersen
hire	Employee growth rate	Bazdresch and Belo and Lin
gad	Growth in advertising expense	Lou
ala	Book Asset Liquidity	Ortiz-Molina and Phillips
RMW	Robust Minus Weak	Fama and French
CMA	Conservative Minus Aggressive	Fama and French
HXZ_IA	HXZ Investment	Hou and Xue and Zhang
HXZ_ROE	HXZ Profitability	Hou and Xue and Zhang
Intermediary	Intermediary Risk Factor	He and Kelly and Manela
convind	Convertible debt indicator	Valta

Appendix B

Programming code

This section contains a brief description of the programming files which have been submitted as a ZIP-file in Sin-Online, as per the thesis manual available on canvas. The ZIP-file contains six files of code, all written in R. The decision was made to use R as MM-regression is easily done using the "robustbase" package. Below a description can be found of all the programming files.

func.R: A file containing helper functions to improve code readability in other sections. Includes functions to estimate a model via OLS and MM, and store relevant results in a data-frame. Includes functions to estimate a p-value via skipped estimation (get_pwols and wols_4factor), but requires relevant packages to be loaded before calling functions. Additionally, a function to compute the percentage of regressions leading to different conclusions and percentage of differences explained by outliers is in the file (run_wols4factor). Please refer to Section 4.4 for a more detailed explanation on these metrics.

ff93_sizebtm.R : This file replicates the Fama and French (1993) study while simultaneously also estimating the model using MM estimation. It uses 25 portfolio returns sorted on size and book-to-market as dependent variables, and the three Fama and French (1993) three factors as explanatory variables.

ff93_multisort.R: This file is similar to ff93_sizebtm.R, however instead of using merely 25 dependent variables, the estimations here use 200 dependent variables according to Section 4.1.

deep_dive_c4.R: This file estimates the Carhart (1997) four factor model using 200 dependent variables. Additionally, all code needed to replicate the extensive case study in Section 4.3 is included in this file.

agg_calc.R: This file computes the percentage of regressions leading to different conclusions and percentage of differences explained by outliers for each factor in the open-source data library provided by Feng et al. (2020). Computing the metrics for all factors is computationally heavy. Therefore it is also possible to read from a CSV-file containing the metrics and factors in the case that the computations have already been done.

p_val_sensitivity_example.R: This file is used to simulate p-values using MM and OLS estimation in the presence of outliers. It contains the code to reproduce Figure 3.1, in Section 3.