Norwegian School of Economics
Bergen, Fall 2018

On Asset Pricing Models and Mutual Fund Performance

An Empirical Analysis of US Mutual Funds

Tobias Kildahl & Mari Tuhus Lunde

Supervisor: Francisco Santos

Master Thesis, Department of Finance

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgements

We would like to express our gratitude towards Francisco Santos for supervising this thesis. His valuable insight and guidance have been of great help throughout the process of writing the paper. In addition, the Department of Finance at NHH has provided fruitful discussions and inspiration. Furthermore, we would like to thank the IT-department for providing software and access to data. Lastly, the discussions and support from friends and colleagues are greatly appreciated.

Bergen, 20.12.2018

Mari Tuhus Lunde

Tobias Kildahl

Abstract

Asset pricing models introduce the challenge of testing a joint hypothesis. This thesis tests the hypothesis of model misspecification and true alpha separately, using the testing methodology of Gibbons, Ross and Shanken (1989) on US mutual fund returns. As there is extensive research on mutual fund performance, our main motivation is to analyze which asset pricing model is the most appropriate for performance evaluation. We test seven asset pricing models on 2971 US mutual funds existing in the period of January 1999 to August 2018.

First, we use the test methodology to measure mutual fund performance under the assumption of perfectly specified factor models. We find ambiguous evidence on fund mangers' ability to create abnormal return gross of fees. We conclude that the most comprehensive models argue for a small, but significant alpha. Net of fees, all models produce negative abnormal returns. This leads to a strong rejection of the null-hypothesis of alpha being equal to zero. Second, we interpret the same test statistics, but now under the assumption of zero abnormal return. This allows for using the GRS-test for testing asset pricing models. We find that the Capital Asset Pricing Model performs surprisingly well given its simplicity. Furthermore, we find that the models including the investment- and profitability factors are better at explaining mutual fund return, with our results indicating that the Fama French Five-Factor Model is the most correctly specified model. Lastly, we find the results to be sensitive to portfolio formation. We conclude that a sorting method based on two criteria is superior to one, and that portfolios sorted on characteristics lead to the strongest inference.

Keywords: asset pricing, mutual funds, factor models, GRS

List of Tables

List of Figures

Table of contents

1. Introduction

The number of mutual funds available to investors has increased rapidly over the past decades, and total assets under management have surpassed $10 trillion in the US. The literature has proposed numerous asset pricing models which can be used to explain mutual fund returns. In order to draw conclusions on the value added by fund managers, it is crucial to identify and apply the most correctly specified model. The Capital Asset Pricing Model by Sharpe (1964), Treynor (1962), Lintner (1965) and Mossin (1966) (CAPM) introduces the relationship between systematic risk exposure and expected return. They argue that an asset's covariance to the market is the only relevant risk to explain return, and that firm-specific risks can be diversified away by holding a large portfolio. However, there is weak empirical evidence of the risk-return relation being as positive as predicted by CAPM.

Motivated by observed return anomalies, additional factors have been proposed. One of the most renowned multifactor models is developed by Fama and French (1993; 1996). The model includes two factors that account for firm size and the book-to-market (B/M) ratio of the asset in addition to the market factor (henceforth FF3). The model is motivated by the observations that small-cap companies and companies with low B/M-ratio yield higher returns. Further, Jegadeesh and Titman (1993) identify that momentum stocks predict higher returns, and Carhart (1997) adds this factor to FF3 to create a four-factor model. With the Q-factor model and the Fama French five-factor model (FF5), both Hou et al. (2015) and Fama and French (2014) emphasize the role of investment- and profitability measures to explain commonality in return patterns. Furthermore, Barillas and Shanken (2018) perform a Bayesian asset pricing test in order to decide which factors to include. They combine the market factor with the size factor from FF3, the investment- and profitability factors from the Q-factor model and a new value factor HMLd based on more frequent data (Asness and Frazzini, 2013) (henceforth BS) .

Asset pricing models introduce the challenge of testing a joint hypothesis. How can we know whether an observed alpha is due to model misspecification, or if the alpha of the mutual fund market is truly different from zero? The two effects are difficult, if not impossible, to distinguish. The methodology of Gibbons, Ross and Shanken (1989) (henceforth GRS) allows for testing one of the two hypotheses at a time. The first interpretation of the test is whether or not the alphas generated by mutual funds are indistinguishable from zero, which is under the assumption of a correctly specified factor model. Secondly, model misspecification can be tested under the assumption of zero abnormal return. The current application of the GRS-test on mutual funds focuses mainly on mutual fund performance rather than asset pricing model performance. Our thesis adds to the literature on mutual funds by expanding the use of the GRS-test to include both interpretations. Most importantly, we discuss the plausibility of the underlying assumptions when applying the test to mutual funds. In addition, we include complementary summary statistics including the Sharpe ratio of the alpha (SR(a)) presented by Lewellen, Nagel and Shanken (2010).

In the first part of our analysis, we use the GRS-test to evaluate the performance of our sample of US mutual funds. If the efficient market hypothesis holds, mutual fund managers should not be able to generate abnormal returns gross of fees. However, the fact that investors are willing to invest in this asset class indicates that fund managers are able to outperform the market. Supporting this, Kosowski et al. (2006) conclude that some mutual funds are able to earn a persistent alpha even net of fees. In order to analyze the effect of fees on the GRS-test, we split the analysis into net- and gross returns. Gross of fees, the conclusion from the GRS statistics is ambiguous. For example, the results using the most comprehensive models indicate that the funds are able to create some abnormal return gross of fees. The FF5 produces an average abnormal return of 0.026% per month. On the other hand, both the FF3 and the Carhart four-factor model produce negative alphas. Furthermore, we find significant differences in the gross- and net performance of the funds. Net of fees, the returns for all models are shifted downwards resulting in the conclusion of negative abnormal return. The results are under the assumption of correct model specifications.

Patterns in mutual fund returns can differ from the identified patterns in the stock market. Wermers (1996) finds that mutual funds with a momentum strategy predict higher returns, while Carhart (1997) argues that exposure to the momentum factor does not explain persistence in US mutual fund performance. Brown and Goetzmann (1997) find significant size- and value premiums in fund return, while Huij and Verbeek (2009) find no difference in market adjusted performance for funds with different size exposures. The observed return patterns of mutual funds raise the question of which factor model that is most applicable for performance measurement. Tests of factor model performance have been conducted on mutual funds, see Huik and Verbeek (2009). However, we find that the underlying assumptions of the GRS-test on mutual funds are not sufficiently discussed. This motivates the second part of our analysis, where we perform a comprehensive test of the presented factor models on US Mutual funds. In order to analyze the implication of fees on the assumptions, we split the analysis into net- and gross returns.

The test results on gross return indicate that the FF5 model outperforms its peers. The model produces lower GRS statistics, SR(a) and average absolute alpha for the most relevant sorts. This is interesting since both the Q-factor model and the BS model prove to perform better on US stocks (Hou et al., 2016; Barillas and Shanken, 2018). Based on improved performance of FF5 relative to FF3, we conclude that also the investment- and profitability factors are important in explaining mutual fund return. Furthermore, we see a consistent reverse size pattern in the intercepts produced by the models that include this factor. This indicates that the size factor overestimates the premium of funds with the highest size exposure. We argue that mutual funds might not be able to capture the whole size premium because of trading limitations. The value pattern in the intercepts implies that the HML factor from FF3 underestimates the factor premium of the funds with the highest value exposure. However, replacing this factor with the HMLd-factor further corrects for the value exposure of the funds.

The GRS statistics are inconsistent when testing net of fees returns. We find that the CAPM produces almost exclusively negative alphas. When moving from CAPM to FF3 we identify an issue of comparing the factor models' performance using GRS when even the CAPM alphas are negative. After correcting for the size- and value exposures, the intercepts are adjusted downwards. The result is an even greater absolute mean alpha for FF3. This creates a double whammy for the GRS-test, as both the increased absolute value of the intercept and improved precision increase the test statistic. The model will be deemed to perform worse even though it might account for the portfolios' actual tilt towards these factors. This effect is also apparent when comparing FF3 to the Carhart four-factor model. In the analysis on gross returns, we observe that the momentum factor improves the model for the relevant sorts, while net of fees we identify the opposite. Based on our results, we argue that the GRS-test is inappropriate for comparing asset pricing models with test portfolios that produce negative alphas for all models. This leads us to criticize existing literature on the area, including the empirical study on mutual funds by Huij and Verbeek (2009), where the GRS-test is used and interpreted based on test portfolios generating negative alpha under CAPM.

Importantly, the second interpretation of the GRS-test is built on the assumption of mutual funds generating an alpha equal to zero. In order for this to be satisfied, we need different assumptions when analyzing gross- and net return. Testing gross returns requires the strong assumption of semi-efficient markets. However, the fact that trillions of dollars are allocated to mutual fund mangers indicates that mutual fund managers are able to generate abnormal returns. We find the assumption of a competitive mutual fund industry more reasonable. Thus, if a fund charges fees that result in negative abnormal return, the investor will choose a similar fund with better terms. This assumption implies zero alpha net of fees, and allows for testing potential model misspecification.

We sort the sample to maintain strong factor structures in the test portfolios in order to avoid dilution of factor patterns. However, Lewellen, Nagel and Shanken (2010) emphasize the sensitivity of the GRS-test towards the sorting method used for constructing test portfolios. In the spirit of their critique, we sort our sample using several different criteria. Most notably, we compare the results between funds sorted on factor exposure and on fund characteristics. Confirming the statement by Lewellen et al. (2010), we observe that a sort resulting in strong return patterns leads to improved model performance. Furthermore, we find that the sort on size leads to particularly distinct portfolios, which result in low GRS statistics for all the models. We conclude that the relative performance of the models varies when we sort on a single criterion. However, we find that sorting on more than one dimension increases the robustness of the test results. We argue that sorting on more than one criterion leads to less favoring of models including the particular factor, and is therefore preferred.

We identify differences in quality between the factor-based and characteristics-based sorting method. According to Daniel and Titman (1997), characteristics-based sorting is preferred for predicting cross-sectional stock returns. On the other hand, Fama and French (1993) argue that firm-level measures of size- and value relate sufficiently to their estimated coefficients. We find that sorting based on momentum exposure leads to results that differ from the sorting based on the 12-months trailing (TTM) return of the funds. The latter represents the momentum characteristic of the fund. In fact, we identify a negative spread between the two most extreme momentum portfolios, while we see the opposite spread pattern for the two most extreme portfolios sorted on TTM return. We argue that one reason for the weak results of the factor-based sorting on momentum is the lack of funds following a momentum strategy. Thus, we emphasize that the factor-based sorting method is most useful when there are strong factor tilts in the sample.

The thesis is structured as follows. Chapter 2 reviews literature and methodology relevant to our research. Sample selection and construction of test portfolios are presented in chapter 3. The main results of our paper are presented in chapter 4, and the robustness of our results in chapter 5. Chapter 6 concludes the analysis. Lastly, chapter 7 highlights the limitations of our work and proposes interesting topics for further research.

2. Literature Review and Methodology

The following section presents literature and methodology relevant to our research, and introduces how our results conform with the existing literature. The section begins with a brief introduction to mutual funds, followed by a review of the cross-sectional relationship between risk and return and the factor models we test in our thesis. We review some of the evidence on mutual funds' ability to capture the return from the proposed anomalies. Lastly, we introduce the GRS-test which is the methodology we use for the main analysis.

2.1 Mutual Funds and Performance

Mutual funds are investment vehicles that invest on behalf of a pool of investors, with the purpose of investing in different asset classes specified in the prospectus. Evaluation of mutual fund performance has been a longstanding debate in both academia and among investors. Fund performance is often reported over a specific benchmark, referred to as the excess return. To further examine whether the return is due to stock-picking skill or systematic exposures that can be easily mimicked, fund evaluation can be extended by regressing returns on a factor model that captures systematic factor exposures. We can interpret the coefficients and premiums from the portfolios on the right-hand-side (RHS) as the proportion of mutual fund return that is attributable to the strategies corresponding to the mimicking factor portfolio. A fund generating persistent positive alpha[1] implies a skilled fund manager who is able to generate added value. However, the theory of market efficiency argues that it is not possible to persistently outperform the market. Research by Malkiel (1995) and Fama and French (2010) support a conclusion of no abnormal returns for active funds net of management expenses, while Kosowski et al. (2006) conclude that some mutual funds are able to earn a persistent alpha net of fees. As will be shown later in our thesis, our results indicate a positive alpha for mutual funds gross of fees when using the most comprehensive factor models. On the other hand, we find strong evidence for negative abnormal return net of fees.

A following question is which factors should be included in the asset pricing model in order to correctly adjust for systematic exposure. The literature presents a broad specter of different asset pricing models that attempt to explain return anomalies. The chosen factors in an asset pricing model applied on mutual funds should have a relevant economic relationship to the strategy of the fund manger in order to obtain causality. In other words, the included factors

[1] We use alpha and abnormal return interchangeably as expressions for the intercept resulting from regressing the excess return of the mutual funds on factor models.

should be an investable strategy from the fund managers' point of view. Furthermore, the vast amount of different factors in current literature may lead to an embarrassment of riches. Including excess factors that do not causally explain return anomalies may result in overfitting and multicollinearity, hence the principle of parsimony is preferred. After considering the criteria of high explanatory power, investability and parsimony we choose to analyze CAPM, FF3, the Carhart four-factor model, the Carhart four-factor model with the more timely value factor HMLd (henceforth Carhart w/HMLd), the FF5, the Q-factor model (Q4) and the six-factor model of Barillas and Shanken[2].

2.2 Asset Pricing Models

2.2.1 The Capital Asset Pricing Model

The positive relation between risk and return is widely acknowledged. As proposed by Sharpe (1964), Treynor (1962), Lintner (1965) and Mossin (1966) with the CAPM, the return of an asset is explained by its sensitivity to the market. Based on the findings of Markowitz (1959), the CAPM captures a linear relationship between return of an asset and its exposure to the market, where β_i is the slope coefficient and a measure of the asset's sensitivity to the market, expressed by equation 2.1. R_it is the return of asset i at time t, while RF_t is the corresponding risk-free rate and R_{mt} is the market return. The authors argue that systematic market risk is the only relevant risk affecting the asset return, as idiosyncratic risk can be diversified away.

$$R_{it} = RF_t + \beta_i(R_{Mt} - RF_t) \tag{2.1}$$

Although this model is intuitively appealing, the empirical record is weak. Researchers have identified several systematic patterns in stock returns not explained by the CAPM, typically called anomalies. Confirming these findings, we identify return anomalies in our sample of mutual funds. Further, the observed anomalies have led to the development of multifactor models, expressing asset returns as a linear combination of the returns of multiple systematic risk factors in addition to the market factor from CAPM.

2.2.2 Fama-French Three-Factor Model

Based on the principles of the Arbitrage Pricing Model by Ross (1976), Fama and French (1993; 1996) argue that many of the return anomalies can be explained by a three-factor model, thus improving upon CAPM. Our analysis of asset pricing models on mutual funds indicates a slight improvement, however not as much as the authors identify on US stocks. Fama and French add a size factor based on the findings of Banz (1981), indicating higher returns for low market capitalization firms. The SMB-factor is constructed by a long-short portfolio consisting of stocks

[2]We obtain factor data from the data libraries of the respective authors.

with low market capitalization minus companies with high market capitalization. Furthermore, empirical findings on US and international stock-markets (Debondt and Thaler (1985), Fama and French (1992), Lakonisok et al. (1994), amongst others) show that assets with low market value relative to fundamentals yield higher return and vice versa. The HML-factor is constructed by a long-short portfolio of stocks with high book-to-market ratio (B/P) minus low B/P. The addition of SMB and HML to the market-factor leads to the following model:

$$R_{it} = RF_t + \beta_{market_i}(R_{mt} - RF_t) + \beta_{SMB_i}(SMB_t) + \beta_{HML_i}(HML_t) \tag{2.2}$$

2.2.3 The Carhart Four-Factor Model

DeBondt and Thaler (1985) find reversal in long-term stock returns, while Jegadeesh and Titman (1993) identify the opposite and therefore argue for a momentum effect. Based on the latter, Carhart (1997) proposes a four-factor model that adds a momentum factor to the FF3. Carhart defines the momentum factor (Mom) as a portfolio based on subtracting the equally weighted return of the lowest performing firms from the highest performing firms, lagged by one period.

$$R_{it} = RF_t + \beta_{market_i}(R_{mt} - RF_t) + \beta_{SMB_i}(SMB_t) + \beta_{HML_i}(HML_t) + \beta_{Mom_i}(Mom_t) \tag{2.3}$$

In our paper, we find that adding the momentum factor does not improve substantially upon FF3 for use on mutual funds.

2.2.4 Fama-French Five-Factor Model

Based on empirical evidence that the three-factor model partly fails to capture the variation in asset return, Fama and French (2014) extend the FF3 model by adding the investment factor (CMA) and the profitability factor (RMW). The added factors in FF5 is according to our results also explanatory for mutual fund returns, and therefore supports their argument. Novy-Marx (2013) finds that a company's gross profitability, measured as profits to assets, has equal power as the B/M-ratio on predicting average return. Furthermore, Aharoni et al. (2013) find that high expected growth in investments predicts lower return[3].

$$\mathrm{R}_{it} = RF_t + \beta_{market_i}(R_{mt} - RF_t) + \beta_{SMB_i}(SMB_t) + \beta_{HML_i}(HML_t) + \beta_{RMW_i}(RMW_t) + \beta_{CMA_i}(CMA_t) \tag{2.4}$$

Hence, RMW in equation 2.4 is the return on a portfolio of stocks with robust- minus weak operating profitability, while CMA consists of the return of a portfolio of firms with conservative minus aggressive investment style.

[3] See also Haugen and Baker (1996), Cohen et al. (2002), Fairfield et al. (2003), Titman et al. (2004) and Fama and French (2006, 2008).

2.2.5 The Q-factor Model

The Q-factor model of Hou et al. (2015) is based on neoclassical Q-theory of investment, and includes the four factors market, size, investment and profitability (equation 2.5). The market factor (ME) is measured as the difference between the arithmetic average of the returns on nine small size portfolios and nine big size portfolios. The remaining Q-factors are based on triple sorts on size, return on equity (ROE) and investment-to-assets (IA). One significant difference between the investment- and profitability factors of the FF5 and the Q-factor model is that Fama and French (2014) motivate their factor construction from a negative relation between expected investment and internal rate of return, while Hou et al. (2015) argue for a positive relation between expected investment and expected returns in a two-period investment model. Additionally, their profitability factor is calculated using monthly ROE rather than annual operating profitability. IA is measured as the annual change in total assets divided by 1-year-lagged total assets, and the investment factor is constructed of companies with low IA minus high IA.

$$R_{it} = RF_t + \beta_i(R_{Mt} - RF_t) + \beta_{ME_i}(ME_t) + \beta_{ROE_i}(ROE_t) + \beta_{IA_i}(IA_t) \tag{2.5}$$

Hou et al. (2016) compare several new asset pricing models, and find that the Q-factor model outperforms FF5. In our analysis we find the contrary, and conclude that the FF5 is more correctly specified for the use on mutual funds.

2.2.6 Barillas and Shanken Six-Factor Model

Barillas and Shanken (2018) form a six-factor model by combining the MKT and SMB-factors from FF3, the IA- and ROE-factors from the Q-factor model, the HMLd factor of Asness-Frazzini (2013) and the momentum factor (Carhart, 1997). The HMLd-factor use the book value per share divided by the current price to construct the monthly value factor, as opposed to the HML constructed by Fama and French (1996) on annual year-end data. This tweak has proven to perform worse than the original HML-factor when added to a pure value strategy, but produces better results when applied on a value and momentum strategy (Asness and Frazzini, 2013). Our results imply that the HMLd factor improves model performance when added to mutual funds sorted on their exposure to value, but performs worse in all other cases. Because of the inclusion of this factor, we find that the performance of the BS model is sensitive towards the value strategy of mutual funds.

$$\mathrm{R}_{it} = RF_t + \beta_i(R_{Mt} - RF_t) + \beta_{SMB_i}(SMB_t) + \beta_{HMLd_i}(HMLd_i) + \beta_{ROE_i}(ROE_t) + \beta_{IA_i}(IA_t) + \beta_{MOM_i}(MOM_t) \tag{2.6}$$

2.3 Patterns in Mutual Fund Return

There is mixed evidence on whether mutual funds are able to beat a passive index and capture the identified factor premiums proposed by the presented factor models. Jensen (1967) and Malkiel (1995) both indicate that mutual funds as a group do not outperform the market. Chan et al. (2002) emphasize that the funds generally do not deviate much from a wide benchmark due to managers' incentives and behavioral considerations. Consistent with this, we find that the return in our sample of US mutual funds are largely explained by the return of the market.

In their paper, Huij and Verbeek (2009) suggest that the magnitude of factor premiums harvested by mutual funds differ from the hypothetically constructed portfolios of assets. One of their arguments is that the funds have restrictions and costs that are not accounted for in the hypothetical portfolios of stocks on which the factor models are based. Carhart (1997) finds that exposure to the momentum factor does not explain return persistency of US mutual fund, contrary to the identified pattern on US stocks. He argues that one-year momentum funds underperform contrarian funds net of management fees. We find supportive evidence of this when we sort funds based on momentum exposure. These results contrast Wermers (1996) and Huij and Verbeek (2009), who find that mutual funds with a momentum strategy earn higher gross returns than the opposite. Measuring momentum as the fund's past return instead of exposure to the momentum factor leads us to the same finding. Brown and Goetzmann (1997) and Carhart (1997) argue for a significant size- and value premium for mutual funds. We find evidence of a size-effect in our sample, while we also confirm the work of Chan et al. (2002) who find that very few funds invest aggressively in value firms.

2.4 Testing Methodology

This section describes the methodology used to test performance of the funds and the factor models. We use the methodology of Gibbons, Ross and Shanken (1989) to test whether the alphas of the style portfolios are jointly equal to zero[4]. This is a pooled time-series cross-sectional methodology, and assumes independently and identically distributed regression residuals. Consider the CAPM regression:

$$\tilde{r}_{mt} = \alpha + \beta_{mt}\tilde{r}_{mt} + \epsilon_t \tag{2.7}$$

The work of Black, Jensen, and Scholes (1972) is testing whether the market portfolio is mean-variance efficient through a test of the following null hypothesis:

[4] Before we conduct our analysis we replicate the use of the GRS-test by Fama and French (2012) on the North American stock market in order to confirm that the methodology is correctly followed.

$$H_0 : \alpha_m = 0 \tag{2.8}$$

Gibbons, Ross and Shanken (1989) extend the work by testing the mean-variance efficiency on any N portfolios of assets. In addition, they include a multivariate F-test for joint significance of alpha, in contrast to Black et al. (1972) use of univariate t-tests of alpha. The GRS statistic is computed as:

$$GRS = \frac{T-N-1}{N}\hat{\alpha}'\left[(1+\frac{\bar{r}_m}{\hat{\sigma}_m})^2\hat{\Sigma}\right]^{-1}\hat{\alpha} \quad \sim F(N, T-N-1), \tag{2.9}$$

where T is the number of periods, N is the number of portfolios jointly tested, $\hat{\alpha}$ is a column vector of estimated alphas, $\bar{r}_m$ is the sample mean of excess market return, $\hat{\sigma}_m$ is the estimated standard error of the excess market return and $\hat{\Sigma}$ is the estimated residual variance-covariance matrix. The GRS-statistic can be extended to any L number of factors to allow for testing of multifactor models:

$$GRS = \left(\frac{T}{N}\right)\left(\frac{T-N-L}{T-L-1}\right)\left(\frac{\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}}{1+\bar{\mu}'\hat{\Omega}^{-1}\bar{\mu}}\right) \quad \sim F(N, T-N-L), \tag{2.10}$$

where $\bar{\mu}$ is a column vector of the factors' sample means and $\hat{\Omega}$ is the estimated covariance matrix of the factors. The GRS-test captures relative deviations for the test portfolios from the ex post efficient portfolio in terms of Sharpe ratios. Hence, a higher GRS statistic implies that the tested portfolios deviate more from the efficient portfolio than a set of equally numbered portfolios with lower GRS statistics. If the GRS statistic is higher than the critical value of the F-test at a given significance level, we reject the null-hypothesis of $\alpha = 0$.

The GRS-test is commonly used to evaluate the performance of mutual funds. Under the assumption that the model is correctly specified, the test determines whether the alphas are jointly significantly different from zero. The second application of the model is testing asset pricing models. Under the assumption of efficient markets, the alpha is equal to zero if the asset pricing model fully explains the return of the assets on the left-hand-side (LHS). However, it is important to recognize the fact that we cannot determine whether the model is rejected by the asset pricing test because it is wrongly specified or if the true alpha is different than 0. Even though the factor models might get rejected by the test, it is valuable to compare the magnitude of the GRS statistic between the chosen models.

In addition to the GRS-test, we present additional statistics in our tests. We include the mean absolute alpha of the intercepts, which is an important complementary to the GRS statistic in comparison of multifactor models. A low average absolute alpha indicates a better model. We also include the average of the standard errors of the intercepts, and the average explanatory power of the regressions. Following the recommendation of Lewellen, Nagel and Shanken (2010) we also include the Sharpe ratio of the intercepts, SR(a):

$$SR(a) = (\alpha' S^{-1} \alpha)^{1/2} \tag{2.11}$$

where α is the column vector of the regression intercepts, and S is the covariance matrix of regression residuals. Hence, SR(a) is the maximum Sharpe ratio for the excess returns on the LHS portfolios constructed to have zero slopes on the RHS returns. One advantage of this measure is that it combines the regression intercepts with the covariance matrix of the regression residuals, the latter being an important determinant of the precision of the alphas. A higher Sharpe ratio of the alphas indicates a less fitting model or an alpha more distinguishable from zero. However, this measure combines the magnitude and precision of the alpha which makes it difficult to distinguish the two effects.

3. Sample Selection and Construction of Test Portfolios

In this section we describe the sampling process for our mutual fund data, before we outline how we construct test portfolios for the GRS-test. Lastly, we take a closer look at the properties and return patterns of the constructed portfolios.

3.1 Mutual Fund Data

We obtain mutual fund return data from the Morningstar Direct global database, and follow a screening procedure inspired by Huij and Verbeek (2009) and Carhart (1997). Our sample consists of open-ended US mutual funds that existed between January 1999 and August 2018 with a minimum of 36 consecutive months of returns. Using mutual fund databases that allow for survivorship bias can lead the overall measured performance to be inflated between 40 basis points and one percent (Elton et al., 1996). Furthermore, survivorship bias affects mutual funds with particular investment styles differently, and can therefore lead to wrong inference when comparing categories of mutual funds (Elton et al., 1993). Therefore, funds are allowed to enter and exit during the period in order to avoid survivorship bias. It should be noted that by excluding funds with less than 36 consecutive months, the sample may not fully represent the ex ante investment space. However, a minimum of consecutive returns is necessary in order to run regressions with sufficient degrees of freedom when we sort mutual funds into LHS portfolios.

We are limiting our sample to funds that have a label in the Morningstar style-box,[1] and more than 90 percent of their net assets invested in US equity. Moreover, we are excluding index funds, sector funds, leveraged funds and funds-of-funds, as well as requiring the funds to be classified as true no-load. The latter restriction is imposed to avoid a sample with duplicates of funds with different share classes. After weighting the importance of large and diversified LHS portfolios against the length of the sample, these restrictions result in a sample of 2971 funds over a period from January 1999 to August 2018.

Differences in alpha between cross-categories of funds are typically very small (Elton et al., 1993), making even slight inaccuracies in the data potentially leading to incorrect inferences. Morningstar's calculation of monthly total return (TR_m) is determined each month by taking the change in the net asset value (NAV), reinvesting all income and capital-gains distributions during the selected month, and dividing by the NAV in the beginning of the month.

[1]The Morningstar style-box categorizes funds according to two dimensions of investment style: size and value.

$$TR_m = \frac{NAV_{end}(1 + \frac{Distribution}{ReinvestmentNAV})}{NAV_{start}} * 100 \tag{3.1}$$

Reinvestments are based on the actual reinvestment NAV, and daily payoffs are reinvested monthly. The monthly return accounts for management, administrative, 12b-1 fees[2] and other costs taken out of fund assets. As we only include true no-load funds, we ensure all net returns are calculated pari passu. We obtain the average monthly management, administration and 12b-1 fees for each Morningstar category, and add this back to the corresponding net return of funds in order to get an estimate of the gross return. Optimally we would use the fund specific fees in order to make the most accurate inference from our results. However, since we aggregate funds into portfolios based on investment style we accept the simplification of using fees on category level. See table A.2 in the appendix for an overview of fees.

All our returns are calculated on the basis of US dollars. Similar to Fama and French (2012), we use the one-month Treasury Bill rate from Ibbotson and Associates[3] as a proxy for the risk-free rate.

3.2 Constructing the Test Portfolios

We divide our sample of 2971 mutual funds into 25 LHS portfolios to be used as test portfolios in the GRS-test. This is a decision which serves multiple purposes. First, it reduces the variation in return due to idiosyncratic risk. Secondly, a limited number of portfolios is necessary in order to allow for sufficient degrees of freedom in the GRS-test. Furthermore, as the test is sensitive to portfolio formation (Affleck-Graves and McDonald, 1990), we test the asset pricing models with LHS portfolios constructed in several different ways to ensure that our conclusions are robust to construction methodology. We construct these portfolios on two main differentials; single- or double sort[4] and factor-based or characteristics-based sort, as presented in figure 3.1.

3.2.1 Single-sorted Portfolios

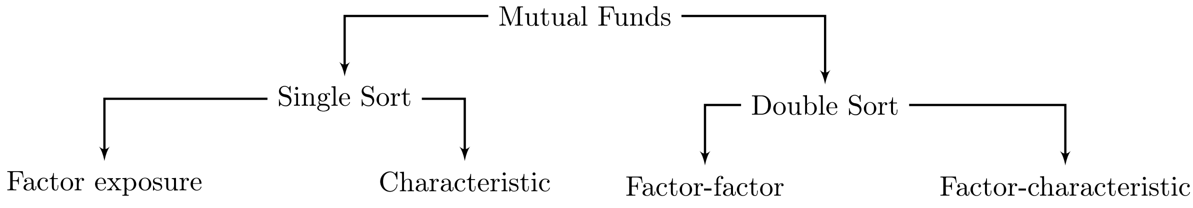
We start by dividing our sample into 25 one-dimensional portfolios in five different ways. For the first four, we sort on exposure to factors from FF3 and the Carhart four-factor model, a so-called factor-based sorting method. Lastly, we sort funds based on one fund characteristic, being the TTM return of the fund.

[2]12b-1 fees is used to pay the funds distribution and marketing cost.

[3]Now a Morningstar-owned company.

[4]As the factor models assume the presence of the included factors, we would ideally sort the portfolios according to the number of factors in the respective model. We choose the single- and double sort in line with previous research on asset pricing.

Figure 3.1 – The figure explains how we divide our data sample of 2971 US mutual funds into test portfolios for testing a null-hyothesis of alpha = 0. The first choice is between single- or double sort, and the second choice is sorting based on factor-exposure or fund characteristics.



There are several differences between a factor-based and characteristics-based sorting method. According to Daniel and Titman (1997), factor exposures provide weaker empirical predictions of cross-sectional stock returns than asset characteristics. Characteristics are directly observable, and is argued to be a better measure for the actual style of a portfolio. Furthermore, the factor-based sorting method requires long consecutive return series, and can become noisy with an insufficient number of observations. Indeed, Cremers, Petajisto, and Zitzewitz (2012) show that both the FF3 and the Carhart four-factor model can produce significant abnormal returns even for passive indices.

However, Fama and French (1993) find that the actual characteristics on size and B/M in FF3 relate sufficiently to the estimated coefficients on the factors. Using characteristics-based sorting on mutual funds requires a large amount of detailed data on holdings, which is not easily available. As a consequence, the characteristics-based method often uses hypothetical holding returns. These can deviate from the actual portfolio returns, and create agency problems such as window dressing (Meier and Schaumburg, 2006). In contrast, the factor-based method only requires monthly return data of the funds. As we have limited access to holding data on our sample of mutual funds, the factor-based sorting method is most applicable. However, in addition to exposure to the momentum factor, we also use the TTM returns of the mutual funds as a characteristics-based alternative. We compare the results of the GRS-test between the two ways of sorting. We further discuss the robustness of the factor-based portfolios in section 5.0.1.

We construct our factor-based portfolios by sorting available funds into 25 portfolios on the exposure to market (β_{MKT}), size (β_{SMB}), value (β_{HML}) and momentum (β_{Mom}). To construct the sort on exposure to the market, we first compute the monthly return over the risk-free rate of each fund for the entire sample period. Second, we obtain the monthly coefficient on the MKT-factor from CAPM using ordinary least squares (OLS) regression with a rolling window over the preceding 60 months, using the first 36 months to initialize the regression.

The window of the rolling regression and the initialization period is equal in all variations of our factor construction.

$$R_{fund_{j,t}} - RF_t = \beta_{market_{j,t}}(R_{mt} - RF_t) \tag{3.2}$$

$R_{fund_{j,t}}$ - RF_t is the excess return of fund j at time t. Every month, we sort the sample of mutual funds on exposure to the market factor, dividing them into 25 equally large groups from low to high β_{market}. This gives us dynamic portfolios based on the exposure of the fund at a given point in time. The first quantile is the portfolio with the lowest market beta and the 25th quantile is the portfolio with the highest market beta. To construct portfolios based on exposure to the size and value factors we use a similar procedure, but using coefficients deriving from OLS regressions using FF3 instead of CAPM.

$$R_{fund_{j,t}} - RF_t = \beta_{market_{j,t}}(R_{mt} - RF_t) + \beta_{SMB_{j,t}}(SMB_t) + \beta_{HML_{j,t}}(HML_t) \tag{3.3}$$

Using the coefficients resulting from equation 3.3, we sort funds into 25 portfolios from low to high β_{SMB}, and low to high β_{HML}, respectively. To construct portfolios sorted on momentum, we compute the rolling monthly coefficient on the Mom-factor resulting from the Carhart four-factor model using OLS regression.

$$\mathrm{R}_{fund_{j,t}} - RF_t = \beta_{market_{j,t}}(R_{mt} - RF_t) + \beta_{SMB_{j,t}}(SMB_t) + \beta_{HML_{j,t}}(HML_t) + \beta_{Mom_{j,t}}(Mom_t) \tag{3.4}$$

We divide the sample into 25 portfolios based on monthly exposure to the momentum factor $\beta_{MOM_{j,t}}$ resulting from equation 3.4, where the first quantile is the funds with the lowest coefficient on Mom and the 25th quantile is the funds with the highest exposure to the factor.

There is not (to our knowledge) any evidence in favor of using mutual funds' exposure to the momentum factor as a proxy for actual tilt towards momentum companies. Therefore, we continue with a sorting method based on the fund characteristic past return. We construct these portfolios by sorting available funds into 25 quantile portfolios based on their cumulative TTM[5]. We sort funds into portfolios from low- to high TTM, where the first quantile consists of the funds with the lowest past return and the 25th quantile is the funds with the highest past return. By doing this, we can compare how the results of the asset pricing tests on portfolios sorted on fund characteristics differ from the portfolios sorted on factor exposure.

3.2.2 Double-sorted Portfolios

We continue by sorting our sample into portfolios on two dimensions, in order to detect differences in asset pricing tests using single- and double sorted test portfolios. Our factor-based method of double sorting is a hybrid between the single sort on factor coefficients by Huij and

[5]We use the cumulative gross return to calculate TTM as this reflects momentum tilt better than net return.

Verbeek (2009) and the double sort on characteristics by Fama and French (2012). The former sorts mutual funds into 10, 20 and 30 portfolios based on mutual fund exposure to a single factor at a time, while Fama and French (2012) sort assets in each region based on the characteristics size (market cap) and value (B/M) and size and momentum (past return). As the mutual funds in our sample are characterized by Morningstar on tilt towards size and value, we find the double sort based on factor exposure towards these factors especially interesting to look at.

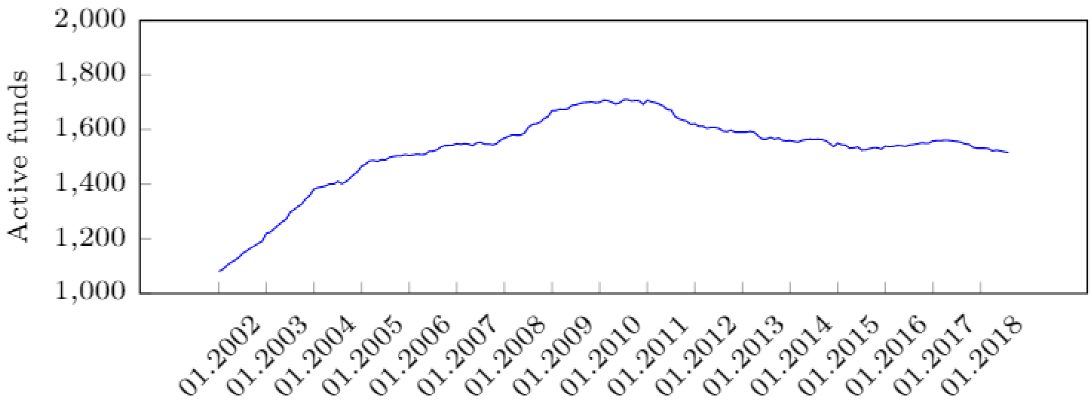
We construct our factor-based portfolios by sorting available funds into 25 quantile portfolios on the exposure to size and value and size and momentum (hereafter referred to as size-value and size-momentum). To construct the size-value sort, we compute the rolling monthly coefficient on the SMB- and HML-factors resulting from the FF3 using OLS regression (equation 3.3). Every month, we sort the sample of mutual funds on exposure to the size-factor first. The funds are divided into five equally large groups where the first quintile consists of the funds with the lowest coefficient on SMB, and the fifth quintile is the funds with the highest coefficient on SMB. We divide each of these five β_{SMB}-portfolios into five new portfolios based on monthly exposure to the value-factor, where the first quintile consists of the funds with the lowest coefficient on HML, and the fifth quintile is the funds with the highest exposure to this factor. The intersections of the 5x5 sort on the SMB- and HML-coefficient produce 25 LHS portfolios.

Equally, to construct the sort on size and momentum, we compute the rolling monthly coefficient on the SMB- and Mom-factors resulting from the Carhart four-factor model (equation 3.4). We sort on exposure to the size factor first, dividing the funds into quintile groups from low to high β_{SMB}. We divide each of these five portfolios into five new portfolios based on monthly exposure to the momentum factor. The intersections of the 5x5 sorts on SMB- and Mom-coefficient produce 25 LHS portfolios.

In order to provide a characteristics-based alternative, we continue with a sorting method partly based on fund characteristics by double-sorting on exposure to size and 12-months trailing return of the mutual funds. We construct these portfolios by sorting the funds into 25 quantile portfolios on the exposure to size (β_{SMB}) and the funds' TTM. By doing this, we can compare how the GRS-test results on portfolios sorted partly on fund characteristics differ from the portfolios sorted on factor exposure. The second sorting is now based on a fund characteristic rather than factor exposure. We divide every size quintile into five new portfolios based on TTM fund return, where the fist quintile is the funds with lowest TTM and the fifth quintile is the funds with the highest TTM. The intersections of the 5x5 sorts on the SMB-coefficient and TTM return produce 25 LHS portfolios.

After sorting the funds into quantile portfolios, we calculate the equally-weighted monthly excess return over the subsequent period for each portfolio. We do this for both net- and gross return. If a fund disappears from the sample after ranking, we re-adjust the portfolio weights correspondingly. This procedure is similar for all ways of sorting the portfolios. Due to initialization of rolling regressions, this leaves us with the monthly return for 25 LHS portfolios between January 2002 and August 2018. We end up with a varying number of funds allocated to portfolios over the sample period, as presented in figure 3.2.

Figure 3.2 – We sample 2971 US mutual funds categorized on two dimensions in the Morningstar style-box. After performing monthly rolling regressions in order to construct test portfolios for testing asset pricing models, the number of funds allocated to a portfolio varies as presented below over the period January 2002 to August 2018.



3.3 Overview of Test Portfolio Returns

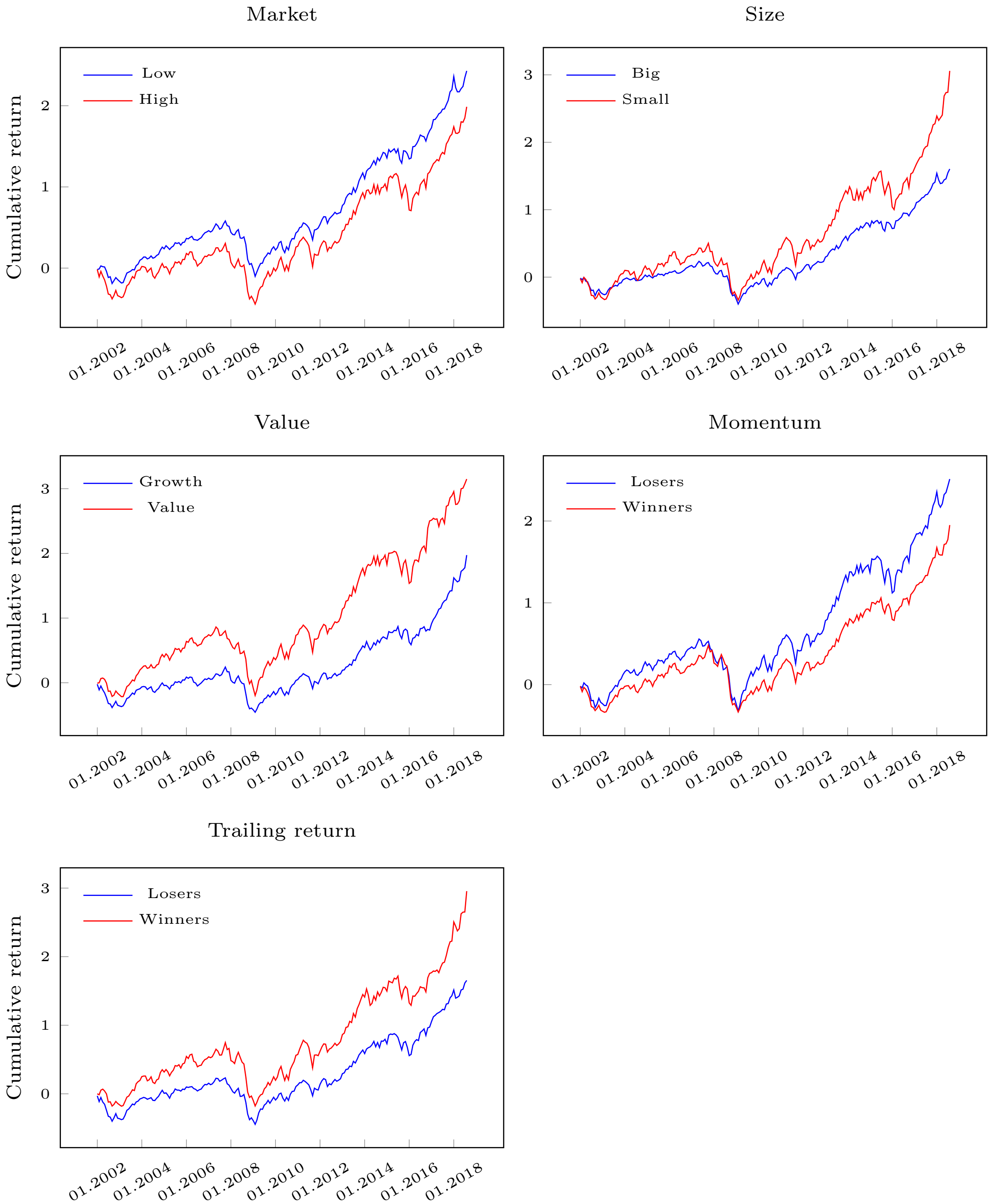
An overview of the test portfolio returns is provided to identify return patterns in our sample of mutual funds. This is a necessary supplement to explain the results from the GRS-test. We begin by looking at one anomaly at a time, before we further investigate the combined effect in the double sorted portfolios. All numbers are gross return if not otherwise stated.

3.3.1 Single Sort

In figure 3.3, we calculate the cumulative monthly excess return[6] for each portfolio over the entire sample period and draw the spread between the return series of the most and and least exposed portfolio. This provides insight to the return spread between the extreme quantiles, which can give an indication of whether there are any factor premiums in our sample of mutual funds. However, apart from the sort on market exposure, this data is not adjusted for market sensitivity. It is also important to bear in mind that the top portfolios of mutual funds constructed with the factor-based sorting method are not directly comparable with the top portfolios of assets

[6]Based on the equally-weighted post-ranking excess return of the funds in the portfolio.

Figure 3.3 – We obtain our sample data of 2971 mutual funds from the Morningstar database. We sort funds into 25 test portfolios based on on exposure to the market-, size-, value- and momentum factors from CAPM, FF3 and the Carhart four-factor model, in addition to sorts on 12-months trailing return of the mutual funds. We calculate the monthly equally-weighted excess (gross) return of each portfolio, and plot the spread between the two most extreme portfolios with regards to factor exposure for all five sorts.



sorted on characteristics. Using the size sort as an example, the exposure of the size portfolios reflect exposure to the long-short factor.

We identify a negative return spread between the funds that are most and least tilted towards the market, consistent throughout the sample period. Thus, funds that have high market exposure seem to earn slightly lower returns than the opposite. Given the CAPM relation of risk and return, funds that are more sensitive to the market should predict higher returns than the opposite. However, Baker et al. (2013)[7] find that US stocks with low market betas have surprisingly high returns between the period of 1968 and 2012, and our first impression of the market return pattern in our sample is in line with a flatter slope than the security market line (SML) predicted by CAPM.

A positive spread pattern is more evident in the portfolios sorted on size, consistent with the findings of Banz (1981). In addition, the spread increase substantially after 2015/2016. We argue that recent bull markets leads to favourable market conditions for small companies. The funds tilted towards value stocks unambiguously outperform the funds tilted towards growth stocks with an even higher spread than between the two extreme size portfolios.

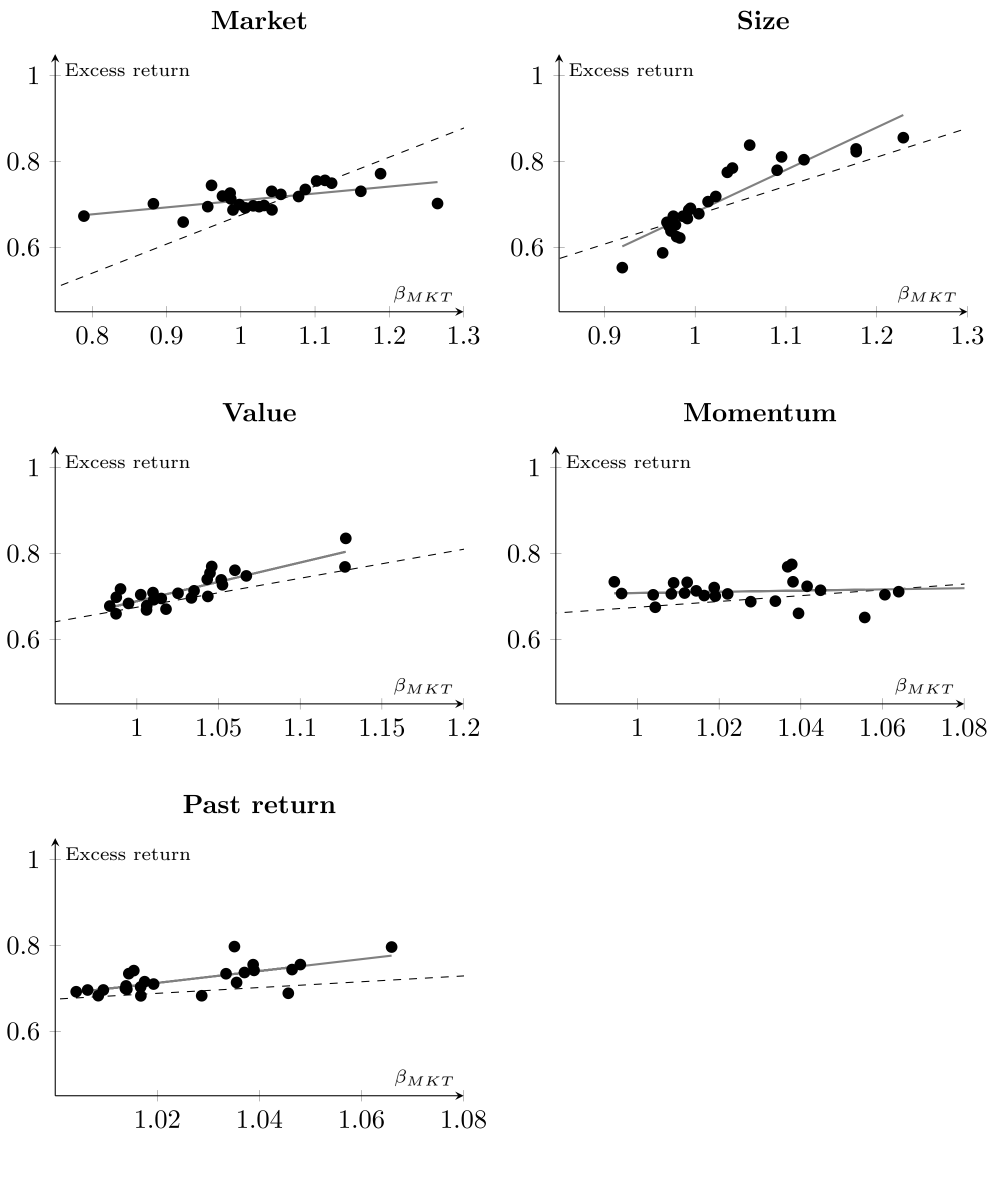
The portfolio spread for the sorting on exposure to momentum versus the sort on past return reveals a puzzling pattern. For the momentum exposure, we see a negative spread between winners and losers. On the other hand, the sort on past return indicates a persistent positive spread between the funds with highest TTM return and the funds with the lowest TTM return. This is an indication that our sorting on exposure to momentum is not capturing the true momentum tilt of the funds, as suggested by Daniel and Titman (1997). As our sample is categorized in the Morningstar style-box on the value- and size dimensions, one reason for the bad performance of the sort on momentum exposure could be that the magnitude of the momentum coefficients are small.

In addition to the monthly excess return of the portfolios we find the portfolios' coefficient on the market factor by performing time-series regressions on each portfolio using OLS and single factor CAPM (equation 2.1). Using this information, we can get an indication of whether a linear function of the market betas can explain the return of our portfolios. The SML is displayed as a dotted line in the plots in figure 3.4, while the solid line draws the relation from the empirical data on the portfolios by fitting a regression line to the 25 data points.

The empirical relation between risk and return for the portfolios sorted on market exposure is less upward sloping than the SML. This is not surprising given the negative return spread identified in figure 3.3. This further confirms our suspicion that the risk-return relation is not as strong as predicted by CAPM. Furthermore, the portfolios sorted on size yield higher return

[7] See also Haugen and Heins (1975), Chan et al. (1999), Blitz and van Vliet (2007) among several others.

Figure 3.4 – We obtain our sample data of 2971 mutual funds from the Morningstar database. We sort funds into 25 test portfolios based on on exposure to the market-, size-, value- and momentum factors from CAPM, FF3 and the Carhart four-factor model, in addition to sorts on TTM return of the mutual funds. We calculate the ex-post average monthly excess (gross) return of each portfolios and plot it against the ex-post market beta of the respective portfolio estimated using CAPM regressions, January 2002 and August 2018. The dotted line displays the SML predicted by CAPM, while the solid line is the empirical relation between the mean excess return of the portfolios and their market beta.



than predicted by CAPM, leaving some evidence of a size premium. Given the evidence on the risk-return relationship of our sample being less steep than the SML, we argue that the size effect might be even larger than the picture we get from our plot. Hence, we conclude that there is a size premium for mutual funds in our sample. The empirical line on the portfolios sorted on value is slightly higher than the SML, with the same slope. Taking into account that the CAPM relation is flatter than predicted, one could argue for a slight value premium. The sort on momentum has proven to be ambiguous, with a negative spread between losers and winners in figure 3.3. Looking at the slope of the empirically fitted line, we see that it is flatter than the SML. Hence, the scatterplot upholds the impression of a negative relation between being tilted towards momentum and return. A possible explanation for the negative correlation is trading costs of following the strategy. Sorting on past return is indicating the opposite with an upward sloping empirical line, rather parallel with the SML. Thus, using past return as a proxy for momentum gives results that are more in line with the findings of Jegadeesh and Titman (1993), while the sort on exposure to momentum equals the reversal effect proposed by Bondt and Thaler (1985).

Using net return, almost all the portfolios earn a lower return than predicted by CAPM, presented in figure A.2 in the appendix. This is in line with a large part of the literature (Malkiel (1995), Fama and French (2010) among several others), and could be due to fees imposed by mutual funds. However, the slopes of the empirical lines do not change significantly. We use fees on category level, and from table A.2 in the appendix we can see that these are quite similar for all categories. Hence, they have little impact on return patterns other than the magnitude.

3.3.2 Double Sort

Table 3.1 presents the mean return and the corresponding standard deviations of the portfolios sorted on exposure to size and value (panel A), size and momentum (panel B) and size and past return (panel C).

The returns of the test portfolios show a size pattern, with increasing returns for increasing SMB exposure for all three sorts. Furthermore, the pattern is consistent for all levels of value exposure. This is contrary to the finding of Fama and French (1993; 2012) who find that small-growth stocks perform worse than large-growth stocks. There can be several explanations for this not being an issue for mutual funds. It may indicate that our sample does not consist of a large number of funds with small-growth companies. However, looking at table A.1 in the appendix, we see that there is no indication of this issue. Another reason could be that the portfolio of funds is more diversified than a portfolio of small-growth stocks. Table A.4 in the appendix presents the coefficients on size and value of the 25 portfolios. We see that the range of coefficients on SMB is between 0.71 and 0.77 for the portfolios investing in small

Table 3.1 – We retrieve data on US mutual funds from the Morningstar database. The sample consists of 2971 funds. We form 25 test portfolios on exposure to size and value in panel A, size and momentum in panel B, and on exposure to size and the 12-months past return of the fund in panel C, and calculate the mean of monthly excess (gross) returns and the corresponding standard deviation for each portfolio, January 2002 to August 2018. Return numbers are %.

Panel A:			Mean					Standard deviation		
	Value	4	3	2	Growth	Value	4	3	2	Growth
Big	0.614	0.596	0.603	0.588	0.639	4.017	3.886	3.960	3.933	4.091
4	0.671	0.653	0.639	0.649	0.657	4.243	3.934	3.930	3.962	4.158
3	0.696	0.658	0.663	0.687	0.741	4.418	4.123	4.027	4.067	4.206
2	0.816	0.801	0.793	0.768	0.722	4.700	4.577	4.410	4.296	4.408
Small	0.845	0.802	0.814	0.854	0.831	5.356	5.122	5.068	4.918	5.206
Panel B:			Mean					Standard deviation		
	Winners	4	3	2	Losers	Winners	4	3	2	Losers
Big	0.555	0.615	0.619	0.624	0.603	3.801	3.862	3.917	3.967	4.171
4	0.658	0.661	0.667	0.644	0.670	4.014	4.001	3.972	3.947	4.206
3	0.683	0.695	0.710	0.681	0.676	4.124	4.080	4.087	4.085	4.312
2	0.717	0.752	0.795	0.825	0.761	4.492	4.520	4.493	4.386	4.461
Small	0.778	0.809	0.831	0.884	0.883	5.134	5.097	5.002	4.981	5.282
Panel C:			Mean					Standard deviation		
	Winners	4	3	2	Losers	Winners	4	3	2	Losers
Big	0.670	0.652	0.605	0.578	0.529	3.835	3.884	3.967	3.996	4.178
4	0.706	0.715	0.672	0.612	0.565	3.925	3.996	4.019	4.028	4.178
3	0.730	0.701	0.704	0.673	0.648	4.180	4.081	4.059	.4.133	4.276
2	0.869	0.843	0.797	0.721	0.658	4.452	4.452	4.390	4.427	4.593
1	0.868	0.839	0.822	0.822	0.786	5.136	5.070	5.003	5.044	5.285

stocks, while the coefficient on HML ranges between -0.17 and -0.46 for the portfolios investing in growth stocks. This indicates that the growth exposures of the funds are not as extreme as the quantile growth portfolios constructed on stocks. The funds tilted towards small-growth companies could also be distributed among the 25 portfolios, and the small-growth problem will then become diluted. Furthermore, Fama and French (2012) show that the problem is most severe for microcaps, which might be less attractive for large mutual funds to invest in. As expected, the standard deviation is increasing from big to small portfolios for all value groups.

The value effect is less evident, as only three of the columns in Panel B show slightly increasing return for increasing exposure to the value factor. Consistent with the scatterplot of the portfolios sorted on value (figure 3.4), this indicates that the funds in our sample are not fully able to capture the value premium. We argue that it is easy for mutual funds with a size strategy to mimic the factor construction of SMB, since the only input is market capitalization. However, mutual funds following a value strategy often look at several fundamental measures, potentially creating deviations from HML solely constructed on B/M.

There is no sign of a premium for being exposed to the momentum factor. Portfolio return increases slightly from left to right in table 3.1 for four of the five size groups, indicating that funds with less exposure to momentum earn higher returns than the opposite. This is consistent with our results on the single-sorted portfolios on momentum. The lack of a momentum pattern is somewhat surprising, based on the empirical finding of a momentum premium on US stocks (Fama and French, 2012). If stocks with positive momentum earn higher returns ex ante, this should also be the case for mutual funds investing in these companies. However, Carhart (1997) finds that mutual funds exposed to the momentum factor do not explain persistency in mutual fund return. We base our research on similarly calculated net returns, and our results support his finding. Fama and French (2012) argue that extreme momentum tilts in actual portfolios are rare, and the exposure to momentum could simply come from funds being accidentally exposed, as they hold last year's winners. We also observe that the coefficients on momentum are very low.

In the sort on past return, the column going from winners to losers needs to be interpreted differently than in the momentum sort. Now, we are looking at momentum effects on fund level instead of correlation with a factor. The momentum pattern for the sort on past return indicates that investing in the best mutual funds the preceding year and selling mutual funds with low TTM return generates a premium. Portfolio return increases from right to left in the table for all of the five size groups, indicating persistence in fund return. Hence, we conclude that a fund being more exposed to the momentum factor predicts lower returns than the opposite, while momentum in the fund itself predicts higher return than the opposite.

4. Main Results

This section presents the results of applying the GRS-test on the constructed test portfolios. The results have two interpretations, as the GRS-test is testing a joint hypothesis of model specification and true alpha. As one of the assumptions has to hold, the analysis is structured as follows. First, we present the evaluation of mutual performance. This is the first interpretation of the GRS-test, which is under the assumption of perfectly specified models. The second part of the analysis presents the interpretation of asset pricing model comparison. We discuss the results for both gross- and net returns in order to enlighten different perspectives. We use both single-sorted and double-sorted portfolios on the LHS, in addition to sorting criteria based on factor exposure and fund characteristics. The main results are presented in table 4.1, 4.2, 4.3 and 4.4. The intercepts of the tests using gross return are presented in table A.5 and A.6, and the intercepts for the single sort on net returns are presented in table A.7, all in the appendix. The remaining intercepts are available upon request.

4.1 Mutual Fund Performance

The most common application of the GRS-test on mutual funds is to measure fund performance. In this section we test whether our sample of mutual funds generate abnormal return, both gross- and net of fees. Since the correct model is not known ex ante, it is useful to look at all the models collectively rather than choosing one.

4.1.1 Gross Returns

In order to evaluate abnormal returns, it is necessary to assume that the models we apply are correctly specified. Thus, if we identify a significant alpha gross of fees, the conclusion is that the mutual funds are generating abnormal return. However, the plausibility of this assumption can be discussed. The factor models are built on the return of hypothetical portfolios, and not empirical portfolios like mutual funds. Trading restrictions and trading costs[1] may result in deviations between the two.

Looking at the GRS statistics in table 4.1, we cannot reject the null-hypotheses for any of the single sorts on market, size and past fund return. According to these results, there are no abnormal return gross of fees. However, the null-hypotheses for most of the models are rejected when the portfolios are sorted on value exposure. Furthermore, all of the tests on the momentum sort are rejected. The CAPM produces intercepts that are mostly positive, leading

[1]Includes bid-ask spreads and price impact.

Table 4.1 – GRS statistics for single sorted portfolios, gross return. We form our sample of 2971 US mutual funds on exposure to the market factor from CAPM, the size- and value factors from FF3, the momentum factor fro the Carhart four-factor model and on the funds TTM return. We calculate the equally weighted excess (gross) return and test seven asset pricing models; FF3, the Carhart four-factor model, Carhart with HMLd instead of HML, FF5, the Q-factor model and the six-factor model of Barillas and Shanken. The GRS statistic test whether all intercepts of the 25 test portfolios are zero. $|\alpha|$ is the average absolute alpha for the set of regressions. R^2 is the average explanatory power. S(a) is the average standard error of the alphas and SR(a) is the Sharpe ratio of the intercepts. With 25 portfolios and 200 monthly returns, critical values for the GRS test are: 90%:1.54, 95%:1.75, 99%:2.23.

Panel A:	Market					Size				
	GRS	$\|\alpha\|$	R^2	S(a)	SR(a)	GRS	$\|\alpha\|$	R^2	S(a)	SR(a)
CAPM	1.40	0.044	0.962	0.057	0.45	1.01	0.040	0.951	0.063	0.39
FF3	1.48	0.061	0.981	0.041	0.47	1.07	0.028	0.983	0.039	0.40
Carhart	1.49	0.056	0.982	0.041	0.48	0.98	0.027	0.983	0.039	0.39
Carhart w/HMLd	1.46	0.033	0.981	0.046	0.50	1.09	0.037	0.981	0.044	0.43
FF5	1.06	0.024	0.984	0.040	0.42	0.69	0.026	0.985	0.038	0.34
Q-factor	1.03	0.030	0.981	0.044	0.42	0.90	0.037	0.981	0.044	0.39
BS	1.37	0.032	0.983	0.043	0.50	0.97	0.046	0.985	0.042	0.42
Average	1.32	0.040	0.979	0.045	0.46	0.96	0.034	0.978	0.044	0.39

Panel B:	Value					Momentum				
	GRS	$\|\alpha\|$	R^2	S(a)	SR(a)	GRS	$\|\alpha\|$	R^2	S(a)	SR(a)
CAPM	1.95	0.033	0.955	0.063	0.54	2.06	0.034	0.964	0.054	0.55
FF3	1.97	0.024	0.980	0.043	0.54	2.07	0.026	0.980	0.041	0.26
Carhart	1.91	0.031	0.980	0.042	0.54	1.96	0.027	0.983	0.039	0.55
Carhart w/ HMLd	1.55	0.031	0.976	0.050	0.52	2.34	0.030	0.982	0.043	0.64
FF5	1.49	0.026	0.983	0.041	0.50	1.87	0.026	0.982	0.041	0.55
Q-factor	1.60	0.027	0.974	0.051	0.52	1.82	0.033	0.980	0.043	0.55
BS	1.31	0.057	0.980	0.047	0.49	2.05	0.037	0.984	0.041	0.62
Average	1.68	0.033	0.975	0.048	0.52	2.02	0.030	0.979	0.043	0.53

Panel C:	Past return				
	GRS	$\|\alpha\|$	R^2	S(a)	SR(a)
CAPM	1.05	0.039	0.958	0.060	0.39
FF3	1.17	0.023	0.971	0.050	0.42
Carhart	1.08	0.028	0.979	0.043	0.40
Carhart w/HMLd	1.30	0.051	0.979	0.046	0.47
FF5	1.10	0.027	0.973	0.050	0.43
Q-factor	1.20	0.053	0.976	0.049	0.45
BS	1.25	0.050	0.982	0.045	0.48
Average	1.16	0.038	0.974	0.049	0.43

to a GRS of 1.95 and a conclusion of significantly positive abnormal return. Adding the size- and value factors in FF3 yield stronger rejection, and a mean alpha of -0.003% per month. Thus, when we adjust for the funds exposure towards these factors, the performance gross of fees seems to be slightly negative. The Carhart model adjusts for the momentum factor, which yields approximately the same results as with FF3. Replacing the HML-factor with HMLd leads to an alpha indistinguishable from zero. Moving on to the less parsimonious models, the intercepts become more positive than both FF3, Carhart and Carhart with HMLd. The FF5 yields a mean alpha of 0.015%. This indicates that adjusted for exposure towards investment- and profitability, the funds generate positive abnormal return.

Table 4.2 – GRS statistics for double sorted portfolios, gross return. We double-sort our sample of 2971 US mutual funds on exposure to the the size- and value factor, the size- and momentum factor, and on exposure to size and the funds TTM return. Sorting on size- and value use the FF3 while the momentum sort uses the Carhart four factor model. We calculate the equally weighted excess (gross) return and test seven asset pricing models; FF3, the Carhart four-factor model, Carhart with HMLd instead of HML, FF5, the Q-factor model and the six-factor model of Barillas and Shanken. The GRS statistic test whether all intercepts of the 25 test portfolios are zero. $|\alpha|$ is the average absolute alpha for the set of regressions. R^2 is the average explanatory power. S(a) is the average standard error of the alphas and SR(a) is the Sharpe ratio of the intercepts. With 25 portfolios and 200 monthly returns, critical values for the GRS test are: 90%:1.54, 95%:1.75, 99%:2.23.

	Size and value					**Size and momentum**					**Size and trailing return**				
	GRS	$\lvert\alpha\rvert$	R^2	S(a)	SR(a)	GRS	$\lvert\alpha\rvert$	R^2	S(a)	SR(a)	GRS	$\lvert\alpha\rvert$	R^2	S(a)	SR(a)
CAPM	1.06	0.040	0.935	0.076	0.40	1.79	0.042	0.945	0.068	0.51	1.48	0.067	0.941	0.071	0.47
FF3	1.06	0.029	0.976	0.047	0.40	1.76	0.028	0.979	0.044	0.51	1.43	0.058	0.972	0.050	0.46
Carhart	0.99	0.033	0.977	0.046	0.39	1.68	0.034	0.981	0.042	0.51	1.33	0.040	0.978	0.045	0.45
Carhart w/HMLd	0.90	0.037	0.973	0.053	0.39	2.30	0.036	0.979	0.047	0.63	1.92	0.040	0.977	0.049	0.58
FF5	0.71	0.032	0.980	0.045	0.34	1.54	0.029	0.981	0.043	0.50	1.04	0.037	0.975	0.050	0.41
Q-factor	0.87	0.036	0.971	0.055	0.38	2.02	0.048	0.977	0.048	0.58	1.54	0.043	0.974	0.052	0.51
BS	0.89	0.065	0.978	0.050	0.41	2.09	0.045	0.982	0.044	0.62	1.80	0.051	0.980	0.048	0.58
Average	0.93	0.039	0.970	0.053	0.39	1.88	0.037	0.975	0.048	0.55	1.51	0.048	0.971	0.052	0.49

Table 4.2 presents the GRS statistics for double-sorted test portfolios. Consistent with the results on the single sorted portfolios, we see varying results for sorts and models. Looking at the size-value sort which is most in line with the investment styles of our sample, none of the models produce an alpha significantly different from zero. This indicates that the funds do not create any added value gross of fees. The sort on size and momentum leads to a rejection of the null-hypothesis in five of seven cases. The strongest rejection is given by applying Carhart with the HMLd factor, which behaves inconsistently with the momentum sort also in the test on single-sorted portfolios. Testing the size-TTM portfolios return leads to rejection for the Carhart w/HMLd and the BS model, while we cannot conclude on any abnormal alpha based on the other models.

4.1.2 Net Return

In this section we test whether our sample of mutual funds generate abnormal return net of fees. Analyzing the net performance of mutual funds is important from an investor perspective. We maintain the sorting methods used on gross returns. The results are presented in table 4.3 and 4.4.

In order to evaluate abnormal returns, it is necessary to assume that the models we apply are

Table 4.3 – GRS statistics for single sorted portfolios, net return. We form our sample of 2971 US mutual funds on exposure to the market factor from CAPM, the size- and value factors from FF3, the momentum factor fro the Carhart four-factor model and on the funds TTM return. We calculate the equally weighted excess (net) return and test seven asset pricing models; FF3, the Carhart four-factor model, Carhart with HMLd instead of HML, FF5, the Q-factor model and the six-factor model of Barillas and Shanken. The GRS statistic test whether all intercepts of the 25 test portfolios are zero. $|\alpha|$ is the average absolute alpha for the set of regressions. R^2 is the average explanatory power. S(a) is the average standard error of the alphas and SR(a) is the Sharpe ratio of the intercepts. With 25 portfolios and 200 monthly returns, critical values for the GRS test are: 90%:1.54, 95%:1.75, 99%:2.23.

Panel A:	Market					Size				
	GRS	$\|\alpha\|$	R^2	S(a)	SR(a)	GRS	$\|\alpha\|$	R^2	S(a)	SR(a)
CAPM	3.52	0.100	0.945	0.069	0.72	4.21	0.085	0.951	0.063	0.79
FF3	3.50	0.120	0.976	0.047	0.73	4.26	0.112	0.982	0.040	0.80
Carhart	3.38	0.121	0.977	0.046	0.72	4.11	0.118	0.983	0.040	0.79
Carhart w/HMLd	3.40	0.098	0.975	0.051	0.77	3.99	0.095	0.981	0.044	0.83
FF5	3.48	0.094	0.979	0.046	0.76	3.64	0.094	0.985	0.038	0.78
Q-factor	3.77	0.088	0.975	0.051	0.80	3.92	0.086	0.981	0.044	0.81
BS	3.54	0.081	0.979	0.049	0.82	3.94	0.085	0.985	0.042	0.86
Average	3.51	0.100	0.972	0.051	0.76	4.01	0.096	0.978	0.044	0.81

Panel B:	Value					Momentum				
	GRS	$\|\alpha\|$	R^2	S(a)	SR(a)	GRS	$\|\alpha\|$	R^2	S(a)	SR(a)
CAPM	2.37	0.083	0.955	0.063	0.59	2.07	0.083	0.964	0.054	0.55
FF3	2.65	0.111	0.979	0.043	0.63	2.40	0.111	0.980	0.041	0.60
Carhart	2.59	0.117	0.980	0.043	0.63	2.32	0.118	0.982	0.039	0.59
Carhart w/HMLd	2.33	0.096	0.976	0.050	0.63	2.80	0.096	0.982	0.043	0.70
FF5	2.18	0.094	0.983	0.041	0.60	1.99	0.094	0.982	0.041	0.57
Q-factor	2.17	0.086	0.974	0.051	0.61	1.87	0.087	0.980	0.043	0.56
BS	2.20	0.088	0.980	0.047	0.64	2.38	0.083	0.984	0.041	0.66
Average	2.36	0.094	0.975	0.048	0.62	2.26	0.096	0.979	0.043	0.61

Panel C:	Past return				
	GRS	$\|\alpha\|$	R^2	S(a)	SR(a)
CAPM	1.06	0.083	0.957	0.060	0.40
FF3	1.57	0.112	0.970	0.051	0.48
Carhart	1.47	0.118	0.978	0.045	0.47
Carhart w/HMLd	1.06	0.095	0.979	0.047	0.43
FF5	1.10	0.077	0.976	0.072	0.62
Q-factor	1.03	0.087	0.975	0.049	0.42
BS	1.02	0.083	0.981	0.046	0.43
Average	1.19	0.094	0.974	0.053	0.47

correctly specified. In addition to the trading restrictions and trading costs that have an effect on gross returns of mutual funds, the additional management and administrative costs must be accounted for by the factor models. We argue that this makes the assumption of perfect model specification less likely to hold net of fees.

For CAPM, the GRS statistics indicate clear rejection for all sorts except for past return. This is due to significantly negative alphas. When we add the two factors in FF3, the alpha becomes more negative and the GRS statistic increases, indicating an even stronger rejection of the null-hypothesis. The intercepts are even more negative and significant for this model. Adding more explanatory factors and arriving at the FF5 leads to relatively low GRS statistics compared with the simpler models on all sorts except the market sort. Nonetheless, we still see a clear rejection of the null hypothesis for all sorts except for past return. Overall, there is strong evidence for negative abnormal returns for the mutual funds net of fees. However, the sort on past return does not reject the null hypothesis of alpha equal to zero. As this sorting method has proven to be preferred over the factor-based method, we do not want to discard these results as an outlier. On the other hand, looking at the results of all the sorts and models jointly promotes the result of zero or negative abnormal return for the mutual funds net of fees.

Table 4.4 – GRS statistics for double sorted portfolios, net return. We double-sort our sample of 2971 US mutual funds on exposure to the the size- and value factor, the size- and momentum factor, and on exposure to size and the funds TTM return. Sorting on size- and value use FF3 while the momentum sort uses the Carhart four factor model. We calculate the equally weighted excess (net) return and test seven asset pricing models; FF3, the Carhart four factor model, Carhart with HMLd instead of HML, FF5, the Q-factor model and the six factor model of Barillas and Shanken. The GRS statistic test whether all intercepts of the 25 test portfolios are zero. $|\alpha|$ is the average absolute alpha for the set of regressions. R^2 is the average explanatory power. S(a) is the average standard error of the alphas and SR(a) is the Sharpe ratio of the intercepts. With 25 portfolios and 200 monthly returns, critical values for the GRS test are: 90%:1.54 95%:1.75 99%:2.23.

	Size and value					**Size and momentum**					**Size and trailing return**				
	GRS	$\lvert\alpha\rvert$	R^2	S(a)	SR(a)	GRS	$\lvert\alpha\rvert$	R^2	S(a)	SR(a)	GRS	$\lvert\alpha\rvert$	R^2	S(a)	SR(a)
CAPM	4.56	0.084	0.945	0.068	0.82	4.56	0.084	0.945	0.068	0.82	5.00	0.091	0.941	0.071	0.86
FF3	4.51	0.112	0.976	0.047	0.82	4.57	0.112	0.979	0.044	0.83	5.00	0.113	0.972	0.050	0.87
Carhart	4.36	0.119	0.977	0.046	0.82	4.43	0.119	0.981	0.042	0.82	4.87	0.119	0.978	0.045	0.86
Carhart w/HMLd	4.09	0.095	0.973	0.053	0.84	4.61	0.097	0.979	0.047	0.89	5.09	0.095	0.977	0.049	0.94
FF5	3.95	0.094	0.980	0.045	0.81	3.97	0.094	0.981	0.043	0.81	4.49	0.094	0.975	0.050	0.86
Q-factor	4.42	0.086	0.971	0.055	0.86	4.44	0.090	0.977	0.048	0.86	4.71	0.086	0.974	0.052	0.89
BS	4.27	0.087	0.978	0.050	0.89	4.48	0.087	0.982	0.045	0.91	5.13	0.083	0.980	0.048	0.97
Average	4.31	0.097	0.970	0.053	0.84	4.44	0.098	0.975	0.048	0.85	4.90	0.097	0.971	0.052	0.89

The results of the double sort is presented in table 4.4. Regardless of sort, the null-hypothesis of zero alpha is strongly rejected. We find that almost all of the portfolios provide excess returns that are lower than what is predicted by CAPM. The intercepts are generally negative and more significant than the intercepts using gross returns. The double sort creates stronger factor structures in the portfolios, allowing the models to more correctly adjust for factor exposure. Hence, we argue that the inference from these results are stronger than from the results using single sorted portfolios. This indicates that there is evidence in favor of a negative alpha net of fees.

4.2 Comparing Asset Pricing Models

The second interpretation of the GRS-test is not widely explored in the current literature on mutual funds. In this section we test the relative performance of the chosen asset pricing models, and discuss the assumptions required for these results to be valid.

4.2.1 Gross Return

In order for the GRS-test to allow for testing potential model misspecification, the assumption of an alpha equal to zero must hold. If we are true believers of the efficient market hypothesis, we can assume that mutual funds are not able to generate alpha gross of fees. This leads to the second interpretation of the GRS statistic. The GRS-test is now a test of model performance rather than mutual fund performance. However, it should be noted that the assumption of semi-efficient markets is strong.

Single Sort

First, we analyze the results of the GRS-tests on the 25 LHS portfolios for all single sorts, using gross return. The test statistics are presented in table 4.1.

Starting with the sort on market exposure in panel A, CAPM performs rather well with a GRS of 1.40. The average (monthly) alpha is 0.044%[2], which as described in section 2.4 is one of the main inputs in the computation of the GRS statistic. The intercepts show a slight reverse pattern for market exposure, indicating that the funds that are investing in high beta companies are not getting paid as much as predicted by CAPM. This confirms the impression from the scatter plot in figure 3.4. The GRS of CAPM among the other dimensions of sorting are ranging from 1.01 to 2.06. The CAPM seems to perform even better when we sort on size and past return than on the sort on market exposure. According to this, the sort on market exposure is not the best for reflecting the funds' market sensitivity, thus putting doubt on the factor-based sorting method. The fact that CAPM is greatly deviating between the sorts on momentum and past return puts further doubt on the factor-based sorting method.

Going from from CAPM to FF3 increases the GRS statistics when we look at the market sort. As can be seen from the table A.5 in the appendix, the intercepts of the most exposed portfolios to market are now more negative. The absolute average alpha is higher after correcting for the two added factors. There is also improved precision of the intercepts, which can be seen from both a lower average standard error of the intercepts s(a) and higher R-squared. However, with increased precision comes increased certainty that the joint alphas are different than 0, explaining the higher GRS. In the intercepts resulting from FF3 we can now see a slight reverse pattern for market exposure, indicating that the portfolios with highest market beta produce

[2]Based on all CAPM tests.

lower abnormal return compared to CAPM. One reason for this relation could be that small caps are more risky and sensitive to the market.

As expected, given the size patterns identified in the scatterplot and portfolio spread, FF3 performs best when the portfolios are sorted on size. This sort produces portfolios with distinct size tilts, thus making the return patterns co-vary with the size factor from FF3. This reveals an important feature of testing the GRS on different sorting dimensions. If the portfolios are not sorted by a factor or characteristic that is included in the model, it is difficult for the model to absorb the alpha. It will thereby perform badly in the GRS-test. Despite the good performance of FF3, we identify a slight reverse size pattern in the intercepts. This indicates that the funds might not earn sufficient returns given their size tilt. We observe a slight value premium in our sample. However, when we apply FF3 to the value sort, there is still some value pattern left in the intercepts. Thus, either the sort does not provide a sufficiently distinct pattern, or the value factor is not defined to capture all of the funds co-varying return.

Adding the momentum factor and arriving at the Carhart four-factor model increases the GRS statistic when using the market sort, but is performing better for all the other sorts. The evidence on the momentum exposure of the sample is mixed. We prefer to use the sort on past return as a proxy. The intercepts of the portfolios sorted on past return show a slight reverse momentum pattern when we apply the Carhart model. Thus, we conclude that the funds with high preceding return do not earn the full premium predicted by the Carhart four-factor model.

If we replace the original HML-factor in the Carhart four-factor model with HMLd which is constructed on more frequent data, the model performance increases when we look at the sort on value. The average absolute alpha remains unchanged. In general, this factor produces a wide range of GRS statistics from 2.34 in the momentum sort to 1.09 in the size sort. Some of this can be due to the fact that we have sorted the funds on the original HML-factor, which could automatically make it fit better to the factor models that include this factor. The Carhart four-factor model performs better than FF3 on the value sort, which suggests a momentum effect in the sample. Nevertheless, sorting on past return does not seem to provide the same fit, as the average absolute alpha increases quite substantially when replacing HML with HMLd.

Moving on to the less parsimonious models, FF5 results in lower average absolute alpha for all the single sorts. The precision of the estimate is also improving, with a higher R-squared and lower standard error of the intercepts. Even though it is easier to reject the model when the precision is high, the two effects still return a lower GRS-statistic overall. We have not sorted on investment or profitability, but since the funds are categorized according to their value- and size style we emphasize the results from these two sorting dimensions. The intercepts from FF5 applied to the size sort reveal a slight size pattern, with the funds least tilted towards size returning negative intercepts. FF5 on the value sort returns a slight reverse value pattern, with the funds most tilted towards value returning negative intercepts. Contrary to FF3, it seems like FF5 underestimates the size premium slightly and overestimates the value premium for the funds. The improved results compared to FF3 indicates that the investment- and profitability factors are important in explaining return patterns in our sample.

The Q-factor model is one of the biggest challengers of FF5. It obtains lower GRS statistics than FF5 in the sort on market exposure and momentum exposure, but performs worse on all other sorts. The intercepts produced from the Q-factor model in the size sort show a subtle size pattern, indicating that the factor is underestimating the size premium. As the size factor is slightly different in this model, and the HML factor is not present, it is not surprising that Q4 performs worse on portfolios sorted on FF3 factors.

The most comprehensive model is the model of Barillas and Shanken (2018) that includes five factors in addition to the market factor. The performance of this model is mixed across the sorting methods. For all sorts except value, the model is beaten by FF5. This strengthens the evidence of the HMLd capturing the value patterns of mutual funds better than the HML-factor. However, when looking at how the return pattern of portfolios sorted on past return is captured, the more simplistic Carhart model is deemed more fitting.

All in all, we see that the average absolute alpha is important for the performance of the model in the GRS test. However, it is also vastly effected by the residual variance-covariance matrix. The Sharpe ratio of the alpha expresses the relation between the magnitude and the precision of the alpha (equation 2.11). A large Sharpe ratio of the alpha is not optimal because this indicates a large absolute alpha relative to the precision of the estimate. Thus, if the Sharpe ratio is high we can be more certain that the alpha $\neq 0$. The Sharpe ratio of the alpha is generally higher for the sort on value and momentum, with an average of 0.52 and 0.53, and lowest for the portfolios constructed on size with an average of 0.39. This indicates that the models are relatively better at capturing return from the size factor of the funds. Between the factor models we see quite similar Sharpe ratios, however the Carhart four-factor model with HMLd seems to produce a notably higher Sharpe ratio of the intercepts in almost all sorts.

Double Sort

This section presents the results of the test on asset pricing models using gross returns and double sorted portfolios. We examine the performance of the factor models for the size-value, size-momentum and size-TTM portfolios separately. Table 4.2 presents the results of all the regressions. In table A.6 in the appendix, we display the matrices of the intercepts from all models on the size-value sort. The intercepts for the remaining sorts are available on request.

First, we perform the GRS-test on the 5x5 portfolios formed on exposure to size and value. The CAPM yields the lowest R^2 and a relatively high absolute mean alpha, which is in line with our expectations. However, the GRS statistic is low and far left of the rejection zone in the F-distribution, and CAPM is therefore accepted according to GRS. This implies a relatively imprecise alpha estimate, which is reflected in the residual variance-covariance matrix and S(a). There seems to be a size pattern in the LHS portfolios where the magnitude of the alphas increase with SMB-exposure. However, it should be noted that none of the alphas are significant at 95% level. Furthermore, running CAPM regression on a long-short portfolio of the two most extreme LHS portfolios still results in a non-significant alpha.

Performing the test on FF3 partially improves the test statistics. The R^2 increases and the mean absolute alpha decreases relative to CAPM, since the size and value factors help explain the corresponding return anomalies. The GRS statistic is unchanged from CAPM despite the lower mean absolute alpha. This indicates more precise estimates of alpha, which is also reflected by the lower S(a). The offsetting effect of lower S(a) on a lower alpha is also resulting in an equal SR(a) between CAPM and FF3. The potential size pattern in CAPM is now absent, thus FF3 seems to successfully capture mutual fund return deriving from tilts towards size. Absence of a size- or value pattern is equivalent with an improved model.

The GRS statistic of the Carhart four-factor model is slightly lower than for the three-factor model. This is indicating that there is some benefit of correcting for a momentum effect in the sample of mutual funds. The mean absolute alpha is 4bps higher, which implies that the residual variance-covariance matrix is necessarily higher. The majority of the momentum coefficients are statistically significant, but the negligible magnitude makes the factor economically insignificant. This is also reflected in the R^2, which is almost identical as in FF3. The assumption by Fama and French (2012) that the mutual funds are not that tilted towards momentum seems to be fair. Introduction of the momentum factor does not induce a new intercept pattern compared to FF3.

The Carhart model with HMLd improves the GRS statistic substantially compared to the FF3 and the vanilla Carhart variant. The lower GRS statistic is driven by a higher S(a), as the mean absolute alpha is higher than Carhart. This is strengthening the finding from the single sort on value where we found HMLd to perform better than HML on our sample of mutual funds.

The GRS statistic of FF5 tells a story of a better asset pricing model when we add the investment- and profitability factors. In their article, Fama and French (2012) emphasize that their results on the parsimonious FF3 and Carhart models might fail to explain some of the return variation of portfolios. Our results confirm their argument. The FF5 is better for asset pricing on mutual funds even though fund managers do not explicitly try to capture factor premiums from investment and profitability. A notable result is that FF5 performs better than FF3 even though the portfolios are sorted on FF3 regressions, which highlights the strength of adding the two additional factors. The R^2 is higher than all the previously discussed models, and the average absolute alpha is among the lowest.

Despite promising applications on stocks, the Q-factor model disappoints with a higher GRS statistic and lower R^2 in our analysis compared to FF5. However, an important consideration is the size-value sorting criteria which is based on Fama and French's factors. As the value factor in the previous models proves to improve model specification, we argue that the poor performance of Q4 is mainly explained by the lack of a value factor.

The BS model should produce relatively good test statistics to compensate for lacking parsimony. A GRS statistic of 0.89 is higher than both the FF5 and the Q-factor model. The mean absolute alpha is the highest among the seven models, which is a result of a distinct value pattern in the alphas. The growth portfolios generate significantly negative alphas, while the high value portfolios generate significantly positive alphas. Systematic failure to account for the value anomaly is the main driver for the weak test statistics compared to FF5.

Table 4.2 summarizes our results on the regressions of the asset pricing models to explain excess return on the 5x5 size-momentum portfolios constructed on US mutual funds. The size effect is still apparent in the intercepts from the CAPM regressions, which is not surprising given that it only controls for the market factor. Adding the size and value factors of FF3 reduce the mean absolute alpha compared to CAPM, but a more precise estimate results in a more or less equal GRS. Surprisingly, using the Carhart four-factor model seems to add a reverse momentum pattern where alpha is negatively correlated to momentum exposure. This is consistent with the spread illustrated in figure 3.3. Thus, the model overestimates the expected return for the funds' level of momentum exposure.

Using HMLd in Carhart results in the highest GRS statistic and a high SR(a). The model generates a pattern where the alphas are negative on both the highest and lowest size quintiles, indicating that the funds with positive exposure to size are not able to capture the full size premium. It seems like the Carhart model overestimates the factor premium coming from the size factor built on hypothetical portfolios.

FF5 is the best performer based on the GRS statistic, a result which is consistent with the size-value sort. FF5 outperforms both Carhart variants in the size-momentum sort, even though FF5 does not include a momentum factor. The alpha patterns are no longer apparent when using FF5. Both the Q-factor and the BS underperforms relative to the average and have the highest mean absolute alphas. This is caused by systematically negative alphas for the size-quintile with the highest exposure to small firms.

The results of the 25 portfolios double-sorted on exposure to the size factor and TTM return are presented in table 4.2. In contrast to the two previous sorting methods, trailing return is a fund characteristic rather than a factor exposure. Comparing a sorting method based on characteristics with factor-based sorting is crucial for the credibility of the previous results. Consistency between sorting on characteristics and factor exposure is important, as we cannot know if FF3 used for sorting is the correct model in the first place.

The relative performance of the models is highly consistent with the size-momentum sorting. However, it should be noted that the magnitude of the GRS statistics are lower for size-TTM sort. A possible explanation is that the funds with the highest factor exposures are expected to earn the highest past return. Thus, we expect that the highest TTM quintile has a strong factor structure. Accordingly, the mean absolute alpha is declining when adding multiple factors to CAPM, while the explanatory power is increasing. We see that opposite to the size-momentum sort, the alpha declines when adding the momentum factor. There is also lower SR(a) for the partly characteristics based sort, which indicates a better fit of the models. FF5 is once again the best performer, proving further consistency between all three sorting methods.

4.2.2 Net Returns

When performing the asset pricing test on gross of fees returns, the strong assumption of efficient markets is necessary. However, there are economic arguments for the contrary. One can argue that there should be no capital inflow to active mutual funds if fund managers are not able to generate alpha gross of fees. The rational investor will choose an index as a more efficient alternative. This is not what we observe empirically. Being a multi-trillion dollar industry indicates that fund managers are above average skilled players in the market, as investors are willing to pay for active allocation.

More reasonable, if we are true believers of a competitive mutual fund industry, the net of fee alpha will be equal to zero. If a fund charges a fee that results in negative abnormal return, the investor will choose a similar fund with better terms. This is a weaker assumption than assuming efficient markets, and it allows for testing model misspecification net of fees. This leads to the second interpretation of the GRS statistic, which is comparing asset pricing models.

A notable characteristic of the output is that the sorting criteria significantly affects the magnitude of the GRS statistic compared to the gross of fees results. All the tested models perform strictly worse when sorted on size, and strictly better when sorted on past fund return. Furthermore, going from CAPM to FF3 reveals an issue of comparing the factor models performance using GRS when even the CAPM alphas are negative. After correcting for the tilts towards size and value, the intercepts are adjusted downwards as indicated in table A.7 in the appendix. This is resulting in a greater absolute mean alpha. Combined with an increased precision of estimating the alpha in FF3, the GRS is driven upwards for all sorts except the market sort. This is the opposite of what happens when we perform the same exercise on gross return. In addition, we perform the test on the double sorted portfolios using net return of the funds. The results are equally inconsistent, presented in table A.6 in the appendix.

Lastly, we conclude that the assumption of a true positive alpha can potentially be problematic for the application of the GRS-test. The GRS-test's issues of non-zero true alphas are especially present if the fund managers are creating negative alpha e.g. due to fees. When we look at net returns of our sample, the intercepts resulting from the CAPM regression are mostly negative and therefore the mean alpha is also negative. When we correctly add more factors on the RHS, the intercepts will further decrease and the absolute average alpha increase. This creates a double whammy in the GRS-test, as both an increased absolute value of the intercept and a presumably lower standard error of the intercept increase the test statistic. A model will be deemed to perform worse even though it might just account for the portfolio's actual tilt towards this factor. Based on our results, we argue that the GRS-test is inappropriate for comparing asset pricing models with LHS portfolios that produce negative alphas for all models. This leads us to criticize existing literature on the area, including the empirical mutual fund study of Huij and Verbeek (2009), where the GRS-test is used on portfolios generating negative alphas even for CAPM.

5. Robustness

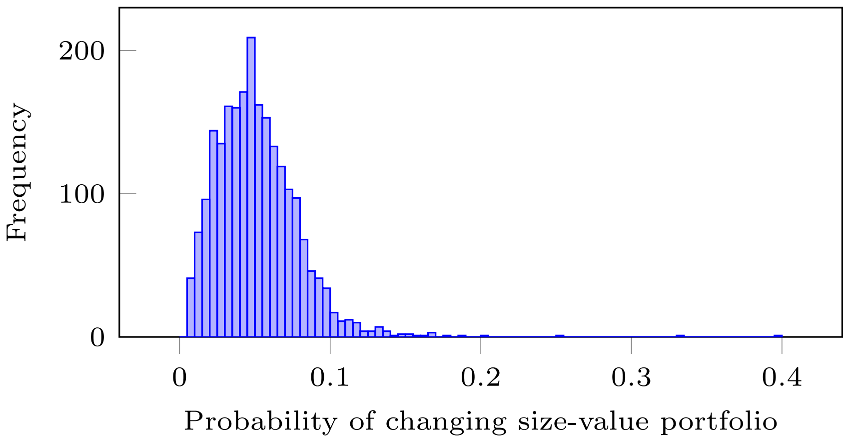
This section addresses the robustness of portfolio construction and the GRS results with different sorting methods and subsamples. First, we discuss the construction of the factor-based portfolios. Second, we divide our data set into two time periods to see if differences in market conditions have an impact on the GRS-test. Further, we discuss the robustness of our test results applying different sorting methods and number of LHS portfolios. We also apply the GRS-test on portfolios based on the Morningstar categories instead of their factor exposure, with the goal of comparing the robustness of the factor-based against the characteristics-based sorting method.

5.0.1 Robustness of The Factor-Based Portfolios

According to Chan et al. (2002), funds that have a specific tilt towards the size- and value factor are rather consistent in investment style over time. In addition, they find that funds with poor past performance shift styles relatively often. The factor-based sorting is debatable, hence it is important to ensure that the construction of portfolios using this method is robust. As a response, we analyze the funds' movement between different portfolios over the sample period. Checking for consistency helps us verify that the sorting based on factor exposure is meaningful in explaining the investment style of the portfolio.

Figure 5.1 presents how often funds change portfolio from month t to $t + 1$ when we sort on exposure to size and exposure to value. The x-axis measures $\frac{\#of times fund_i changed portfolio}{\#of periods fund_i existed}$, hereafter portfolio consistency, and the y-axis represents the frequency in which these values occur. Most of the funds have a rather low probability of changing portfolios, where the highest frequency of funds is found at a probability of around 5%. The results for the remaining double sorts are consistent with the results on the size-value portfolios (Figure A.1 in appendix). In addition, we look at the post-formation coefficients on the double-sorted portfolios (table A.4 in appendix). We see that the post-formation exposures are consistent with the factor they are sorted on, giving a clear pattern from low to high exposure. Most coefficients are also significant. When we look at the factor exposures for the funds grouped on category in the Morningstar style-box (appendix table A.3) we also see that the funds actually follow the style they are reporting.

Figure 5.1 – We sort our sample of 2971 US mutual funds into test portfolios for testing asset pricing models. The histograms display the number of US mutual funds with a certain probability of changing portfolio between January 2002 and August 2018 for the sort on exposure to size and value.



5.1 Robustness to Varying Subsamples

The GRS-test presented in table 5.1, panel A is divided into subsamples to analyze consistency throughout different time periods. The subsamples are divided by the financial crisis of 2007-2008, which marked a substantial change in the financial climate. Even though this event serves as a reasonable breaking point in the sample, it should be noted that it leads to an unequal number of observations between the two subsamples. Unequal degrees of freedom in the GRS-test affects the magnitude of the statistic, thus making the GRS not fully comparable between the subsamples. However, the models are comparable within the same subsample.

The results from the subsample 2002-2008 is consistent with few exceptions when compared to the full sample. FF5 is the best performer measured by GRS and SR(a), and it produces the lowest mean absolute alpha. Further, anomaly patterns are consistent for all models with the exception of Carhart using HMLd and BS. Both models produce a value pattern where the alphas of the LHS portfolios negatively correlate with exposure to value. It is concluded that reverse value pattern is a weakness of the two models. The mean absolute alpha is considerably higher for the pre-financial crisis sample compared to post-financial crisis and the full sample. This seems to be driven by the strong bull markets amplifying the value pattern where high growth exposure resulted in positive alphas and vice versa.

The results from the more recent subsample of 2009-2018 tell a different story. The relatively new factor models of Carhart using HMLd and BS are the best performers, while FF5 has the highest GRS. A possible explanation is that the new factor models are constructed to better explain asset returns on more recent data.

Table 5.1 – We perform the test of Gibbons, Ross and Shanken (1989) of whether the alphas of the test portfolios are jointly zero, and test on two subsamples of the total sample period from January 2002 to August 2018. The first subsample is between January 2002 and December 2008, while the second subsample is from January 2009 to August 2018. The test portfolios are constructed on a sample of US mutual funds divided into 5x5 size-value portfolios in panel A and nine portfolios in panel B, both single sort, double sort and on characteristics. The critical value for the GRS-statistics on the 95% level is the following: 1.75 for 25 portfolios and 200 observations, 1.78 for 25 portfolios with 96 observations and for 25 portfolios with 104 observations, 2.73 for nine portfolios with 200 observations.

Panel A:	Full Sample			2002-2008			2009-2018		
25 test portfolios	GRS	$\|\alpha\|$	SR(a)	GRS	$\|\alpha\|$	SR(a)	GRS	$\|\alpha\|$	SR(a)
CAPM	1.06	0.040	0.40	1.50	0.118	0.81	2.21	0.089	0.82
FF3	1.06	0.029	0.40	1.53	0.064	0.84	2.21	0.029	0.85
Carhart	0.99	0.033	0.39	1.45	0.082	0.83	2.20	0.029	0.85
Carhart w/HMLd	0.90	0.037	0.39	1.57	0.084	0.88	2.03	0.050	0.88
FF5	0.71	0.032	0.34	1.19	0.042	0.79	2.35	0.042	0.90
Q-factor	0.87	0.036	0.38	1.30	0.068	0.80	2.23	0.075	0.89
BS	0.89	0.065	0.41	1.72	0.109	0.97	2.07	0.070	0.91
Average	0.93	0.039	0.39	1.47	0.081	0.85	2.19	0.055	0.87

Panel B:	Size			Size-value			Morningstar		
9 test portfolios	GRS	$\|\alpha\|$	SR(a)	GRS	$\|\alpha\|$	SR(a)	GRS	$\|\alpha\|$	SR(a)
CAPM	0.83	0.040	0.20	0.42	0.039	0.14	1.80	0.068	0.30
FF3	1.01	0.026	0.22	0.39	0.020	0.14	1.76	0.031	0.30
Carhart	0.87	0.023	0.21	0.59	0.027	0.17	1.76	0.031	0.30
Carhart w/HMLd	0.90	0.034	0.23	0.56	0.028	0.18	1.93	0.035	0.33
FF5	0.60	0.025	0.18	0.63	0.027	0.19	2.53	0.029	0.37
Q-factor	0.64	0.036	0.19	0.92	0.037	0.23	3.30	0.028	0.43
BS	1.12	0.044	0.26	0.98	0.058	0.24	3.34	0.060	0.45
Average	0.85	0.033	0.21	0.64	0.034	0.18	2.35	0.040	0.35

5.2 Robustness to Varying Methods of Portfolio Formation

Affleck-Graves and McDonald (1990) argue that the GRS-test is sensitive to portfolio formation. Intuitively, the performance of factor models must vary when we sort on different dimensions that co-vary differently with the factors in the respective model. We confirm this with our results, and observe that there is large variation in which model that is deemed best between the different ways of sorting. FF5 performs best when looking at the size sort, while the Q-factor model has the lowest GRS statistics when we sort on exposure to momentum and market. The Carhart model is most fitting when we sort on past return. Based on the same reasoning, we also see differences between sorting on one or two dimensions.

We test if the results are consistent to the number of test portfolios in two ways. First, we perform the GRS-test on a sample of funds double-sorted on exposure towards size and value, but on a 3x3 basis instead of 5x5. The results are presented in table 5.1, Panel B. We see that the GRS statistic is slightly lower in magnitude when we sort on nine instead of 25 portfolios with both single size sort and double size-value sort. The Sharpe ratio of the alpha decreases, while the absolute average alpha is relatively unchanged. This indicates that the precision of the alphas are higher with fewer LHS portfolios, and highest when we double-sort. When we split the portfolios on size into fewer quantiles, the portfolios will show less distinct patterns as funds are forced into larger portfolios. The GRS is lower for the double sorting method than for single sorting when using nine portfolios. We argue that this is caused by more factor patterns in double sorted portfolios. Looking at the relative performance of the models, the results remain quite consistent. The FF5 produces the lowest GRS for the simple size sort, while FF3 is the best specified model for the size-value sort.

In addition, we test nine portfolios divided according to their characteristics in the Morningstar style-box. When we sort funds into portfolios based on Morningstar category, the picture changes from the nine factor-based portfolios. The results are now more reasonable when we look at the magnitude of the GRS, since we see rejection of the null-hypothesis more often. Rejection is generally what is expected in asset pricing tests, since the models are only simplifications of the return composition of the assets. While the size-value portfolios include an equal number of funds, the Morningstar portfolios are not equally divided (appendix table A.1). As a consequence, the nine equally divided portfolios might become less distinct than the portfolios sorted on Morningstar category, making it easier for the test to reject the null-hypothesis. FF5 performs worse when sorting the funds in this way as opposed to the factor-based sorting, and we now observe that FF3 and the Carhart four-factor model perform best.

5.3 Robustness to Differences in Fees

We perform the GRS-test on mutual fund returns both gross- and net of fees[1]. Since we do not have data on fees on fund level, we cannot be certain that the performance of the factor models are entirely robust to sample differences in fees. However, when we look at the univariate tests of the intercepts for the portfolios using net return, we see small differences from the portfolios using gross returns. Furthermore, the slope of the empirical relation between mean excess return and market beta is consistent between the two, see figure 3.4 and A.2.

[1]Management, administration and 12b-1 fees that are taken out of gross return when the net return is reported. Does not include load fees and other extraordinary fees

6. Conclusion

This thesis tests the joint hypothesis of model misspecification and true alpha of US mutual funds separately, using the GRS-test presented by Gibbons, Ross and Shanken (1989). We analyze the two interpretations and discuss the required assumptions in order for them to be valid. As there is extensive research on mutual fund performance, our main motivation is to use the GRS-test for determining which asset pricing model is the most appropriate for performance evaluation. We sort the sample of 2971 funds into test portfolios in the period January 2002 to August 2018. The portfolios are sorted on either a single or double criterion, using both factor exposure and fund characteristics.

In the first part of our analysis, we use the GRS-test to evaluate mutual fund performance both gross- and net of fees. We find that gross of fees, most models and sorts fail to reject the null-hypothesis, which argues for zero abnormal returns. However, the more comprehensive models which include the investment- and profitability factors lead to a significantly positive alpha. For example, the average of all FF5 tests lead to a conclusion of significant abnormal monthly returns of 0.015% for the single sort gross of fees. We argue that the positive alpha from applying the more comprehensive models supports the hypothesis of skilled portfolio managers, as the additional factors seem to explain some of the funds' return. Furthermore, we find large differences in gross- and net performance. The GRS statistics for net returns strongly reject the null-hypotheses for all single sorts except on past return. For the double-sorted portfolios, all tests indicate a significantly negative alpha. Under the assumption that any of the seven models are correctly specified, we can conclude that mutual fund managers are destroying value for the investor net of fees.

In the second part of our analysis, we use the GRS-test to evaluate the relative performance of the asset pricing models. Gross of fees, the conclusion from the GRS statistics is ambiguous. However, our first finding is that CAPM performs surprisingly well relative to the more comprehensive models when applied to our portfolios of mutual funds. For the sort on market exposure, CAPM produces low GRS statistics and performs surprisingly well. This indicates that even though the funds are actively managed, the return strongly correlates with the market. We also find a reverse market pattern, and conclude that the funds most exposed to the market yields lower return than predicted by CAPM.

Applying the models that include a size- and value factor on the double-sorted portfolios leads to slightly better performance of the models. However, we generally identify a reverse size pattern in the intercepts after testing FF3. The size pattern is the strongest factor structure in our sample, and we conclude that the funds are not earning the size premium predicted by this factor. The value pattern in our sample is weaker, and we find that there are differences between the traditional HML-factor and the more frequent HMLd-factor. The former generally underestimates the value premiums of the funds, while the latter overestimates. Testing the more comprehensive models, we conclude that FF5 overall outperforms its peers. The model produces lower GRS statistics, SR(a) and mean absolute alpha for the sort on size-value, which reflects the most relevant investment style in our sample. Based on the performance of FF5 relative to FF3, we conclude that the inclusion of the investment- and profitability factors is important for explaining mutual fund returns.

Furthermore, we find that the GRS-test is sensitive to the sorting methods and -criteria. We sort our funds with the purpose of maintaining strong factor structures, as it avoids dilution of factor patterns. However, this is a double edged sword as the test results become sensitive to factor inclusion. This is apparent in the results from the single-sorted portfolios, where the relative performance of the models varies substantially with the sorting criterion. From these sorts, we cannot make an unambiguous conclusion of which model is the best. When we sort our sample on more than one dimension, the relative performance becomes more robust towards factor inclusion. Hence, we conclude that sorting the test portfolios on two dimensions is preferred. Next, there are qualitative differences between sorting on factor exposure and fund characteristics. We find that sorting based on momentum exposure results in inferences that differ from the sorting based on past performance of the funds. We argue that the reason for the non-robust results of the factor-based momentum sort is the lack of momentum strategy in the sample. However, the size sort produces portfolios with strong factor structures, which is reflected in the low GRS statistics. Thus, we conclude that the factor-based sorting method is only appropriate when there are strong factor tilts in the sample.

Lastly, we discuss that the underlying assumptions necessary for alpha to equal zero are different when the GRS-test is applied to gross- and net returns. Accepting a model which yields an alpha insignificantly different from zero is an inaccurate conclusion if the true alpha is different from zero. We find that the GRS test issues of non-zero true alphas are especially evident if the fund managers are creating negative alpha, e.g. due to fees. Therefore, we emphasize that the model performance interpretation of the GRS-test should only be used on mutual fund return in knowledge of its limitations and necessary assumptions.

7. Limitations and Suggestions for Further Research

There are several possible limitations to our research. The first is related to the way we sort funds into test portfolios. When we use the factor-based sorting method, we get an issue of circularity because we use the same models for sorting and testing. This can create a high correlation between the LHS- and RHS portfolios. Therefore, we find a potential improvement by using fund characteristics instead of factor exposure. This would require data on the portfolio holdings for each mutual fund. Further, our results are sample-specific, thus we suggest future research to investigate samples with more aggressive- or alternative investment styles. This could potentially amplify our findings and conclusions.

Bibliography

Affleck-Graves, J. and McDonald, B. (1990). Multivariate tests of asset pricing: The comparative power of alternative statistics. *The Journal of Financial and Quantitative Analysis*, 25(2).

Aharoni, G., Grundy, B., and Zeng, Q. (2013). Stock returns and the miller modigliani valuation formula: Revisiting the fama french analysis. *Journal of Financial Economics*, 110.

Asness, C. and Frazzini, A. (2013). The devil in hml's details. *The Journal of Portfolio Management*, 39.

Baker, M. P., Bradley, B., and Taliaferro, R. (2013). The low risk anomaly: A decomposition into micro and macro effects. *Financial Analyst Journal*, 70(2).

Barillas, F. and Shanken, J. (2018). Comparing asset pricing models. *The Journal of Finance*, 73.

Blitz, D. and van Vliet, P. (2007). The volatility effect: Lower risk without lower return. *Journal of Portfolio Management*, (44).

Bondt, W. F. M. D. and Thaler, R. (1985). Does the stock market overreact? *The Journal of Finance*, 40.

Brown, S. J. and Goetzmann, W. N. (1997). Mutual fund styles. *Journal of Financial Economics*, 43.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1).

Chan, L. K., Karceski, J., and Lakonishok, J. (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. *NBER Working Paper*, (70439).

Chan, L. K. C., Chen, H.-L., and Lakonishok, J. (2002). On mutual funds investment styles. *The Review of Financial Studies*, 15(5).

Cremers, M., Petajisto, A., and Zitzewitz, E. (2012). Should benchmark indices have alpha? revisiting performance evaluation. *Critical Finance Review*, (2).

Daniel, K. and Titman, S. (1997). Evidence on the characteristics of cross sectional variation in stock returns. *The Journal of Finance*, 52(1).

Elton, E. J., Gruber, M. J., and Blake, C. R. (1996). The persistence of risk-adjusted mutual fund performance. *The Journal of Business*, 69(2).

Elton, E. J., Gruber, M. J., Das, S., and Hlavka, M. (1993). Efficiency with costly information: A reinterpretation of evidence from managed portfolios. *The Review of Financial Studies*, 6(6).

Fairfield, P. M., Whisenant, S., and Yohn, T. L. (2003). Accrued earnings and growth: Implications for future profitability and market mispricing. *The Accounting Review*, 78.

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

Fama, E. F. and French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *The Journal of Finance.*

Fama, E. F. and French, K. R. (2010). Luck versus skill in the cross-section of mutual fund returns. *The Journal of Finance*, 65.

Fama, E. F. and French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics*, 103(3).

Fama, E. F. and French, K. R. (2014). A five-factor asset pricing model. *Journal of Financial Economics.*

F.Fama, E. and R.French, K. (1992). The cross[U+2010]section of expected stock returns. *The Journal of Finance*, 47.

Gibbons, M. R., Ross, S. A., and Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, 57.

Haugen, R. A. and Baker, N. L. (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics.*

Haugen, R. A. and Heins, A. J. (1975). Risk and the rate of return on financial assets: Some old wine in new bottles. *The Journal of Financial and Quantitative Analysis*, 10(5).

Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies.*

Hou, K., Xue, C., and Zhang, L. (2016). A comparison of new factor models.

Huik, J. and Verbeek, M. (2009). On the use of multifactor models to evaluate mutual fund performance. *Financial Management*, 38(1).

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance.*

Jensen, M. C. (1967). The performance of mutual funds in the period 1945-1964. *Journal of Finance*, 23.

Jensen, M. C., Black, F., and Scholes, M. S. (1972). The capital asset pricing model: Some empirical tests. *Studies In The Theory Of Captial Markets*.

Kosowski, R., Timmermann, A., Wermers, R., and White, H. (2006). Can mutual fund "stars" really pick stocks?new evidence from a bootstrap analysis. *The Journal of Finance*, 61.

Lakonishok, J., Shleifer, A., and Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *The Journal of Finance*, 49.

Lewellen, J., Nagel, S., and Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics*, 96.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*.

Malkiel, B. G. (1995). Returns from investing in equity mutual funds 1971 to 1991. *The Journal of Finance*, 50.

Markowitz, H. M. (1959). Portfolio selection: Efficient diversification of investments. *The Journal of Finance*, 15.

Meier, I. and Schaumburg, E. (2006). Do funds window dress? evidence for u.s. domestic equity mutual funds. *Unpublished*.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, 108.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13.

Sharpe, W. F. (1964). Capital asst prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*.

Titman, S., Wei, K. J., and Xie, F. (2004). Capital investments and stock returns. *The Journal of Financial and Quantitative Analysis*, 39.

Treynor, J. (1962). Toward a theory of market value of risky assets.

W.Banz, R. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9.

Wermers, R. (1996). Momentum investment strategies of mutual funds, performance persistence, and survivorship bias.

Appendix

A. Number of companies by Morningstar category

Table A.1 – The table visualizes the two dimensions of the Morningstar style-box, and how the intersections produce nine Morningstar categories. The numbers in the table represents the number of sample funds in each category of the total sample of 2971 funds in the period January 2002 to August 2018.

	Small-cap	Mid-cap	Large-cap
Value	193	249	528
Blend	236	285	477
Growth	267	277	459

A. Tables of fees

Table A.2 – The table presents the yearly average fees of the respective nine investment styles categorized by the Morningstar style-box between January 2002 and August 2018 (%). The management and professional fee accounts for the management of the funds by the portfolio manager(s), the administrative fee is the costs related to day-to-day operations such as rental of office space, and the 12b-1 fees goes to marketing of the fund.

	Management and administrative	12b-1	Total
Large Value	1.03	0.25	1.28
Large Blend	0.85	0.25	1.10
Large Growth	1.06	0.25	1.31
Mid Value	1.37	0.25	1.62
Mid Blend	1.08	0.25	1.33
Mid Growth	1.15	0.25	1.40
Small Value	1.05	0.25	1.30
Small Blend	1.31	0.25	1.55
Small Growth	1.17	0.25	1.42

A. Factor coefficients of portfolios according to Morningstar category

Table A.3 – The table reports coefficients on the market factor (Market), size factor (SMB) and value factor (HML) deriving from OLS regressions on nine portfolios sorted on Morningstar Category using the Fama French 3 factor model, January 2002 to August 2018. * indicates significance on the 95% level.

	Markt	SMB	HML
Large Value	0.974*	−0.06 *	0.20*
Large Blend	0.977*	−0.04	0.02
Large Growth	1.02*	0.05*	−0.27 *
Mid Value	0.99*	0.29*	0.15*
Mid Blend	0.99*	0.34*	0.00
Mid Growth	1.02*	0.41*	−0.27 *
Small Value	0.95*	0.73*	0.28*
Small Blend	0.97*	0.72*	0.097*
Small Growth	1.03*	0.78*	0.24*

A. Factor coefficients of double-sorted portfolios

Table A.4 – We double-sort our sample of 2971 US mutual funds on exposure to the the size- and value factor, the size- and momentum factor, and on exposure to size and the funds 12-months trailing return. We calculate the equally weighted excess (gross) return for each portfolio. Thereafter we run regressions using FF3 to obtain the portfolios coefficient on the SMB and HML-factors, and the Carhart four-factor model to obtain the coefficient on the MOM-factor. * indicates that the coefficients are significant on a 95% level.

Size coefficient of size-value portfolios

	Value	4	3	2	Growth
Big	−0.11 *	−0.13 *	−0.13 *	−0.14 *	−0.08 *
4	−0.04 *	−0.06 *	−0.05 *	−0.04 *	−0.01
3	0.01	0.03	0.05*	0.04	0.07*
2	0.38*	0.33*	0.31*	0.31*	0.29*
Small	0.77*	0.71*	0.75*	0.75*	0.77*

Table continues on next page

Table A.4 – We double-sort our sample of 2971 US mutual funds on exposure to the the size- and value factor, the size- and momentum factor, and on exposure to size and the funds 12-months trailing return. We calculate the equally weighted excess (gross) return for each portfolio. Thereafter we run regressions using FF3 to obtain the portfolios coefficient on the SMB and HML-factors, and the Carhart four-factor model to obtain the coefficient on the MOM-factor. * indicates that the coefficients are significant on a 95% level.

Value coefficient of size-value portfolios

	Value	4	3	2	Growth
Big	0.23*	0.15*	0.08*	0.00	−0.17 *
4	0.22*	0.11*	0.00	−0.11 *	−0.32 *
3	0.17*	0.01	−0.12 *	−0.24 *	−0.39 *
2	0.09*	−0.12 *	−0.24 *	−0.3 *	−0.43 *
Small	−0.23 *	−0.07 *	−0.24 *	−0.31 *	−0.46 *

Momentum coefficient of size-momentum portfolios

	Winners	4	3	2	Losers
Big	0.17*	0.02*	0.00	−0.01 *	−0.06 *
4	0.06	0.03*	0.01	−0.01	−0.05 *
3	0.10*	0.04*	−0.02	−0.03 *	−0.06 *
2	0.11*	0.05*	0.03*	0.00	−0.04 *
Small	0.10*	0.06*	−0.03 *	0.01	−0.06 *

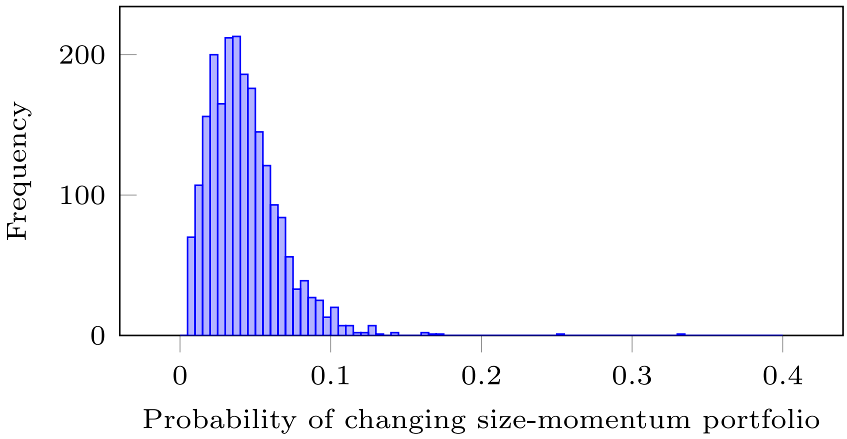
Momentum coefficient of size-TTM portfolios

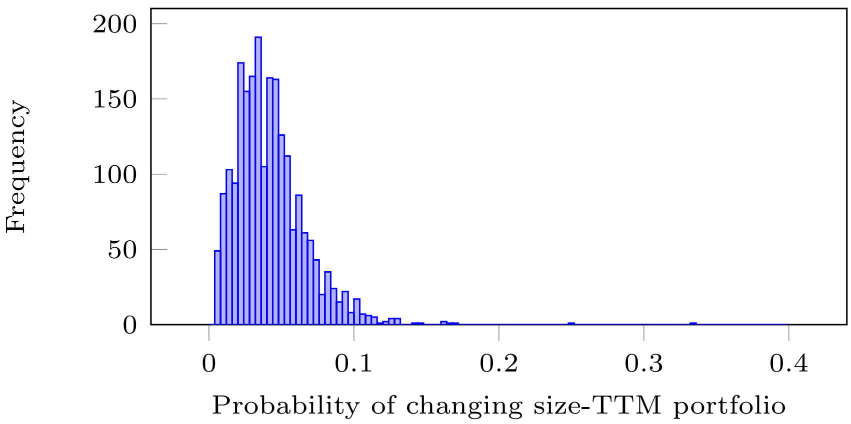
	Winners	4	3	2	Losers
Big	0.09*	0.03*	−0.01	−0.05 *	−0.11 *
4	0.11	0.05*	0.01*	−0.03 *	−0.11 *
3	0.13*	0.05*	0.01	−0.02 *	−0.11 *
2	0.14*	0.07*	0.04*	0.01	−0.07
Small	0.15	0.07	0.04	0.00	−0.10

End of Table

A. Histogram

Figure A.1 – We sort our sample of 2971 US mutual funds into test portfolios for testing asset pricing models. The histograms display the number of US mutual funds with a certain probability of changing portfolio between January 2002 and August 2018 for the sort on exposure to size and momentum and size and past return.





A. Intercepts for single sorted portfolios using gross return

Table A.5 – We form our sample of 2971 US mutual funds on exposure to the market factor from CAPM, the size- and value factors from FF3, the momentum factor from the Carhart four-factor model and on the funds 12-months trailing return. We calculate the equally weighted excess (gross) return and test seven asset pricing models; FF3, the Carhart four-factor model, Carhart with HMLd instead of HML, FF5, the Q-factor model and the six factor model of Barillas and Shanken. The GRS statistic test whether all intercepts of the 25 test portfolios are zero. The intercepts deemed most relevant from the regressions of the 25 portfolios for the five different simple sorts are included below, displayed from the portfolio with the lowest exposure towards the factor to the highest exposure towards the factor. * Indicates significance on a 95% level.

	Market				Size			Value			Momentum		Past return		
	CAPM	FF3	Carhart	FF5	FF3	FF5	Q-factor	CAPM	FF3	FF5	CAPM	Carhart	Carhart	Carhart w/HMLd	BS
Low	0.147*	0.155*	0.131*	0.081	−0.034	−0.064	−0.065	−0.083	−0.145	0.019	−0.019	−0.007	−0.090	−0.032	−0.019
2	0.113*	0.121*	0.107*	0.063	−0.036	−0.039	−0.048	0.037	−0.013	0.088	0.002	0.010	0.010	0.085	0.111*
3	0.044	0.045	0.034	0.001	−0.010	−0.013	−0.019	0.054	0.008	0.092	−0.003	−0.002	0.013	0.072	0.088
4	0.058	0.058	0.045	0.022	0.017	0.012	0.005	0.038	−0.007	0.052	−0.032	−0.031	0.011	0.067	0.075
5	0.104*	0.100*	0.083*	0.068*	0.002	−0.011	−0.013	0.073	0.025	0.074	0.000	0.001	0.018	0.064	0.077
6	0.069*	0.064*	0.053	0.048	−0.022	−0.026	−0.036	0.026	−0.020	0.035	0.003	−0.002	0.031	0.076	0.086*
7	0.069*	0.064*	0.053*	0.052	0.028	0.011	0.012	0.023	−0.021	0.020	0.043	0.028	0.030	0.080*	0.096*
8	0.034	0.030	0.022	0.015	0.016	0.015	0.008	0.024	−0.018	0.009	0.071*	0.057*	0.009	0.062	0.067*
9	0.027	0.017	0.008	0.004	0.002	0.001	−0.001	0.008	−0.032	−0.004	0.035	0.022	0.018	0.063*	0.073*
10	0.055	0.045	0.031	0.032	0.013	0.009	0.018	0.045	0.003	0.023	0.042	0.026	0.014	0.055	0.055
11	0.021	0.006	0.000	0.013	0.005	0.006	0.023	0.004	−0.032	−0.024	0.034	0.016	0.012	0.047	0.057
12	0.019	0.005	−0.002	0.014	0.022	0.021	0.043	−0.008	−0.035	−0.026	0.005	−0.016	0.006	0.027	0.026
13	0.012	−0.004	−0.015	0.014	0.015	0.022	0.048	0.019	−0.007	−0.005	0.034	0.004	0.013	0.024	0.024
14	0.010	−0.013	−0.019	0.005	−0.005	0.010	0.034	0.036	0.013	0.019	0.059	0.029	−0.003	0.012	0.016
15	−0.007	−0.032	−0.038	0.000	0.012	0.036	0.050	0.057*	0.043	0.040	0.025	−0.007	0.023	0.037	0.051
16	0.036	0.008	0.003	0.026	0.009	0.029	0.051	0.021	0.003	0.005	0.025	−0.022	−0.005	0.009	0.010
17	0.021	−0.014	−0.018	0.020	0.044	0.076	0.112	0.040	0.023	0.022	0.058	0.007	0.016	0.031	0.037
18	0.000	−0.043	−0.051	0.006	0.038	0.048	0.081	−0.002	−0.018	−0.020	0.037	−0.014	−0.014	−0.005	−0.008
19	0.010	−0.042	−0.047	0.001	0.066	0.088	0.113	0.022	0.012	−0.003	0.021	−0.031	−0.016	−0.002	0.010
20	0.020	−0.041	−0.047	−0.006	−0.022	0.021	0.051	0.008	0.003	−0.010	0.042	−0.019	−0.053	−0.044	−0.029
21	0.014	−0.056	−0.062	0.008	−0.007	0.017	0.047	0.001	−0.001	−0.032	0.078	0.003	−0.022	−0.025	−0.011
22	0.001	−0.076	−0.081	0.005	−0.037	0.012	0.027	0.036	0.031	0.012	0.018	−0.058	−0.051	−0.054	−0.038
23	−0.044	−0.129*	−0.128*	−0.027	−0.068	0.004	0.002	0.019	0.011	−0.018	0.029	−0.063	−0.068	−0.089	−0.064
24	−0.021	−0.109	−0.102	0.000	−0.071	0.015	0.010	0.058	0.038	−0.001	0.083	−0.030	−0.081	−0.113	−0.082
High	−0.141	−0.240*	−0.211*	−0.065	−0.095	0.049	0.002	0.083	0.029	−0.002	−0.053	−0.180*	−0.066	−0.098	−0.052

A. Intercepts for size-value portfolios using gross return

Table A.6 – We form our sample of 2971 US mutual funds on exposure to the size- and value factors from FF3. We calculate the equally weighted excess (gross) return and test seven asset pricing models; FF3, the Carhart four factor model, Carhart with HMLd instead of HML, FF5, the Q-factor model and the six factor model of Barillas and Shanken. The intercepts from the regressions of the 25 portfolios for the five different simple sorts are included below.

CAPM

	Value	4	3	2	Growth
Big	−0.018	−0.048	−0.040	−0.035	−0.032
4	−0.012	0.006	−0.001	0.015	0.000
3	0.019	0.001	0.013	0.029	0.052
2	0.127	0.120	0.097	0.051	0.001
Small	0.069	0.053	0.046	0.081	0.036

FF3

	Value	4	3	2	Growth
Big	0.007	−0.024	−0.019	−0.016	−0.028
4	0.004	0.019	0.006	0.015	−0.012
3	0.025	0.005	0.000	0.013	0.026
2	0.080	0.071	0.045	−0.003	−0.059
Small	−0.021	−0.044	−0.064	−0.033	−0.088

Carhart

	Value	4	3	2	Growth
Big	0.021	−0.017	−0.014	−0.016	−0.033
4	0.020	0.020	0.004	0.006	−0.028
3	0.041	0.002	−0.006	−0.006	0.002
2	0.075	0.054	0.028	−0.022	−0.086
Small	0.008	0.055	0.084	−0.057	−0.112

Carhart w/HMLd

	Value	4	3	2	Growth
Big	−0.044	−0.057*	−0.030	−0.008	0.013
4	−0.033	−0.009	0.000	0.026	0.039
3	−0.008	0.010	0.039	0.062	0.086
2	0.050	0.095	0.107	0.065	0.018
Small	−0.062	0.015	−0.012	0.006	−0.015

Table continues on next page

Table A.6 – We form our sample of 2971 US mutual funds on exposure to the size- and value factors from FF3. We calculate the equally weighted excess (gross) return and test seven asset pricing models; FF3, the Carhart four factor model, Carhart with HMLd instead of HML, FF5, the Q-factor model and the six factor model of Barillas and Shanken. The intercepts from the regressions of the 25 portfolios for the five different simple sorts are included below.

FF5

	Value	4	3	2	Growth
Big	−0.021	−0.047	−0.032	−0.022	0.002
4	−0.018	−0.015	−0.013	0.013	0.033
3	0.004	−0.004	−0.001	0.020	0.079
2	0.022	0.063	0.068	0.047	0.057
Small	−0.004	−0.006	0.019	0.062	0.087

Q4

	Value	4	3	2	Growth
Big	−0.002	−0.050	−0.041	−0.037	−0.025
4	0.010	0.007	−0.012	0.002	−0.002
3	0.047	0.034	0.026	0.027	0.054
2	0.109	0.099	0.106	0.061	0.032
Small	0.022	0.009	0.009	0.037	0.036

BS

	Value	4	3	2	Growth
Big	−0.069	−0.086	−0.056*	−0.018	0.032
4	−0.058	−0.029	−0.018	0.026	0.074
3	−0.025	0.023	0.050	0.085	0.140
2	0.020	0.091	0.140*	0.117	0.127
Small	−0.108*	−0.012	0.032	0.066	0.118

End of Table

A. Intercepts for single sorted portfolios using net return

Table A.7 – We form our sample of 2971 US mutual funds on exposure to the market factor from CAPM, the size- and value factors from FF3, the momentum factor fro the Carhart four factor model and on the funds 12-months trailing return. We calculate the equally weighted excess (net) return and test seven asset pricing models; FF3, the Carhart four factor model, Carhart with HMLd instead of HML, FF5, the Q-factor model and the six factor model of Barillas and Shanken. The GRS statistic test whether all intercepts of the 25 test portfolios are zero. The intercepts deemed most relevant from the regressions of the 25 portfolios for the five different simple sorts are included below, displayed from the portfolio with the lowest exposure towards the factor to the highest exposure towards the factor.

	Market				**Size**			**Value**			**Momentum**		**Past return**		
	CAPM	FF3	Carhart	FF5	FF3	FF5	Q-factor	CAPM	FF3	FF5	CAPM	Carhart	Carhart	Carhart w/HMLd	BS
Low	−0.090	−0.064	−0.070	−0.119*	−0.136*	−0.166*	−0.167*	−0.195	−0.257	−0.093	−0.128	−0.115	−0.195	−0.135	−0.088
2	−0.057	−0.037	−0.044	−0.088*	−0.138*	−0.141*	−0.150*	−0.076	−0.125	−0.025	−0.104*	−0.096*	−0.087	−0.015	0.014
3	−0.029	−0.016	−0.026	−0.063*	−0.111*	0.114*	−0.120*	−0.058	−0.104	−0.021	−0.110*	−0.110*	−0.088	−0.033	−0.022
4	0.010	0.009	−0.006	−0.027	−0.084*	−0.088*	−0.095*	−0.076	−0.121*	−0.061	−0.139*	−0.138*	−0.096*	−0.039	−0.030
5	0.109	0.065	0.028	0.008	−0.098*	−0.112*	−0.113*	−0.041	−0.089	−0.039	−0.107*	−0.106*	−0.061	−0.015	−0.009
6	−0.148*	−0.129*	−0.124*	−0.141*	−0.123*	−0.127	−0.137*	−0.088	−0.133*	−0.079	−0.105*	−0.109*	−0.099*	−0.054	−0.040
7	−0.094*	−0.082*	−0.087*	−0.095*	−0.073*	−0.091*	−0.090*	−0.089	−0.134*	−0.093*	−0.064*	−0.079*	−0.074	−0.022	−0.009
8	−0.051	−0.047	−0.053	−0.058*	−0.086*	−0.087*	−0.093*	−0.088	−0.130*	−0.103*	−0.035*	−0.049	−0.093*	−0.041	−0.027
9	−0.049	−0.058	−0.069	−0.064	−0.100*	−0.101*	−0.103*	−0.103	−0.143*	−0.115*	−0.072*	−0.085*	−0.091*	−0.049	−0.047
10	0.073	0.014	−0.023	−0.016	−0.090*	−0.094*	−0.084*	−0.065	−0.106*	−0.087*	−0.064	−0.080*	−0.096*	−0.055	−0.046
11	−0.179*	−0.166*	−0.156*	−0.150*	−0.098*	−0.097*	−0.080*	−0.104	−0.141*	−0.133*	−0.072*	−0.090*	−0.090*	−0.056	−0.053
12	−0.153*	−0.149*	−0.147*	−0.143*	−0.082*	−0.083*	−0.062	−0.114*	−0.141*	−0.131*	−0.101*	−0.122*	−0.095*	−0.075*	−0.072*
13	−0.108*	−0.114*	−0.118*	−0.103*	−0.090*	−0.083*	−0.057	−0.086	−0.112*	−0.109*	−0.073*	−0.102*	−0.107*	−0.085*	−0.087*
14	−0.065	−0.087*	−0.098*	−0.061	−0.111*	−0.096*	−0.071	−0.067	−0.090*	−0.084*	−0.048	−0.078*	−0.091*	−0.079*	−0.072*
15	0.021	−0.056	−0.086	−0.034	−0.096*	−0.071	−0.057	−0.045	−0.059*	−0.063*	−0.083*	−0.114*	−0.079*	−0.063*	−0.052
16	−0.208*	−0.206*	−0.186*	−0.159*	−0.099*	−0.078	−0.057	−0.082*	−0.100*	−0.098*	−0.083	−0.130*	−0.103*	−0.095	−0.098*
17	−0.162*	−0.171*	−0.168*	−0.143*	−0.068	−0.036	0.000	−0.064	−0.081*	−0.082*	−0.050	−0.102*	−0.087*	−0.074	−0.071
18	−0.085	−0.117*	−0.124*	−0.094	−0.079	−0.069	−0.036	−0.106*	−0.123*	−0.124*	−0.072	−0.123*	−0.138*	−0.124*	−0.121*
19	−0.033	−0.101	−0.122*	−0.070	−0.055	−0.033	−0.008	−0.083*	−0.093*	−0.108*	−0.088	−0.141*	−0.132*	−0.119*	−0.106*
20	0.000	−0.107*	−0.128*	−0.056	−0.141	−0.098	−0.068	−0.097	−0.102	−0.115*	−0.069	−0.129*	−0.158*	−0.156*	−0.147*
21	−0.231*	−0.252*	−0.221*	−0.151*	−0.125*	−0.101*	−0.070	−0.104*	−0.107*	−0.137*	−0.033	−0.108*	−0.152*	−0.149*	−0.133*
22	−0.184	−0.248*	−0.236*	−0.154*	−0.155*	−0.106*	−0.090	−0.071	−0.076*	−0.094*	−0.093	−0.169*	−0.175*	−0.177*	−0.165*
23	−0.129	−0.221*	−0.219*	−0.140*	−0.186*	−0.114*	−0.116	−0.088	−0.096*	−0.126*	−0.083	−0.176*	−0.182*	−0.197*	−0.166*
24	−0.111	−0.223*	−0.231*	−0.116*	−0.188*	−0.102*	−0.108	−0.050	−0.070	−0.109*	−0.031	−0.144*	−0.207*	−0.237*	−0.207*
High	−0.126	−0.254*	−0.267*	−0.106*	−0.213*	−0.069	−0.116	−0.029	−0.083	−0.114*	−0.166	−0.294*	−0.216*	−0.241*	−0.192*

A. Robustness

Table A.8 – We perform the test of Gibbons, Ross and Shanken (1989) of whether the alphas of the test portfolios are jointly zero, and test on two sub-samples of the total sample period from January 2002 to August 2018. The first sub-sample is between January 2002 and December 2008, while the second sub-sample is from January 2009 to August 2018. The test portfolios are constructed on a sample of US mutual funds divided into 25 equally weighted portfolios on exposure to size and exposure to value separately. With 25 portfolios and both 104 and 96 observations the GRS-statistic on a 95%level is 1.78.

	Size						**Value**					
	Jan.02-Dec.08			Jan.09-Aug.18			Jan.02-Dec.08			Jan.09-Aug.18		
	GRS	$\|\alpha\|$	SR(a)	GRS	$\|\alpha\|$	SR(a)	GRS	$\|\alpha\|$	SR(a)	GRS	$\|\alpha\|$	SR(a)
CAPM	1.69	0.086	0.96	0.67	0.027	0.45	1.78	0.096	0.99	1.38	0.092	0.65
FF3	1.88	0.068	1.07	0.54	0.019	0.42	1.60	0.037	0.98	1.29	0.025	0.65
Carhart	1.80	0.051	1.07	0.57	0.019	0.43	1.46	0.043	0.97	1.31	0.025	0.66
FF3 w/HMLd	1.52	0.084	1.10	0.75	0.037	0.54	1.02	0.092	0.90	1.26	0.045	0.69
FF5	1.74	0.069	1.08	0.49	0.028	0.41	1.55	0.054	1.02	1.18	0.037	0.64
Q-factor	1.52	0.053	1.01	0.67	0.060	0.49	1.45	0.053	0.99	1.31	0.070	0.69
BS	1.49	0.107	1.15	0.73	0.054	0.54	1.02	0.130	0.95	1.17	0.067	0.68
Average	1.66	0.074	1.06	0.63	0.035	0.47	1.41	0.072	0.97	1.27	0.052	0.67

A. Mean return against market beta, net return

Figure A.2 – We obtain our sample data of 2971 mutual funds from the Morningstar database. We sort funds into 25 test portfolios based on on exposure to the market-, size-, value- and momentum factors from CAPM, FF3 and the Carhart four-factor model, in addition to sorts on 12-months trailing return of the mutual funds. We calculate the ex-post average monthly excess (net) return of each portfolios and plot it against the ex-post market beta of the respective portfolio estimated using CAPM regressions, January 2002 and August 2018. The dotted line displays the SML predicted by CAPM, while the solid line is the empirical relation between the mean excess return of the portfolios and their market beta.