BI Norwegian Business School - campus Oslo

GRA 19502

Master Thesis

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

U.S. Mutual Fund Performance: Skill or Luck?

Navn:	Kenneth Nordlund Hoem, Erik Rosmer Aasland

Start:	02.03.2018 09.00
Finish:	03.09.2018 12.00

Kenneth Nordlund Hoem
Erik Rosmer Aasland

BI Norwegian Business School
Master Thesis

U.S. Mutual Fund Performance: Skill or Luck?

Supervisor:
Paul Ehling

Hand-In Date:
03.09.2018

Examination Code and Name:
GRA 19502 – Master Thesis

Campus:
BI Norwegian Business School, Oslo

Programme:
Master of Science in Business – Major in Finance

Abstract

This thesis examines the performance of 1704 actively managed U.S. open-end, domestic equity mutual funds in the period of January 1995 to December 2017. Regression results for an equal-weighted portfolio suggest that fund managers in aggregate do not possess sufficient skill to cover their costs. We use a bootstrap procedure to distinguish skill from luck in the cross-section of three-factor $t(\alpha)$ estimates for net and gross fund returns. The bootstrap results show that a sizeable minority of fund managers do have sufficient skill to cover their costs. The evidence of skill is stronger when examining performance gross of management fees. Under the assumption that the cross-section of true α has a normal distribution with mean zero and standard deviation σ, we inject α into fund returns in the bootstrap simulations. We find that the σ for the left tail is about 0.75% a year, while the right tail is about 1.25%.

Acknowledgements

We would like to thank our supervisor Paul Ehling for suggestions on our topic and valuable guidance throughout the process of this thesis. We would also like to thank BI Norwegian Business School for providing us with access to the Bloomberg database.

Table of contents

1. Introduction

The U.S. mutual fund industry has experienced a remarkable growth over the last decades, especially due to increased incentives of individuals to engage in financial markets without having to make individual investment decisions. Total net assets in the U.S. mutual fund industry have grown from $2.81 trillion to $18.75 trillion from 1995 to 2017 (illustrated in Figure A1 in the appendix), which corresponds to an annual growth rate of 8.6% (Investment Company Fact Book, 2018). Throughout this period, some fund managers have been able to achieve returns above their respective benchmarks, while others have struggled to justify fees imposed on investors.

There are two main issues that previous literature on mutual fund performance have tried to provide answers to. The first is whether average risk adjusted abnormal fund performance (net of expenses) is positive, negative or zero. The second is whether abnormal performance can be identified ex-ante and for how long it persists. The empirical findings in the U.S. are somewhat mixed. However, the lion's share of U.S. studies of mutual funds suggests little or no superior performance, but stronger evidence of underperformance (e.g Jensen 1968, Fama 1970, Malkiel 1995, Carhart 1997). This suggests that active fund managers in general do not possess skill to produce risk adjusted expected returns that cover their costs. On the other hand, there are studies that find evidence of superior stock-picking skills of active fund managers (e.g. Wermers 2000, Kosowski et al. 2006, Berk and Van Binsbergen 2015). This thesis contributes to the literature by examining the performance of actively managed U.S. mutual funds over the 1995 to 2017 period, as it includes the post-financial crisis period.

According to the efficient market hypothesis (EMH), introduced by Malkiel and Fama (1970), fund managers are not able to outperform a benchmark index, because stock prices have already incorporated all available information. Additionally, there is a constraint to active management that Fama and French (2010) refer to as equilibrium accounting. Passive investors earn returns equal to the market portfolio (before costs), which implies that they have α equal to zero relative to passive benchmarks. This means that the aggregate α is also equal to zero before costs for active investment. As stated by Fama and French (2010), active investors who obtain positive excess returns does so at the expense of other active investors, and consequently, active investment must be a “negative sum game”. However, this does not rule out the possibility that some active managers consistently outperform the market, but they do so at the expense of other active investors. Based on these theories and previous literature, we proceed with the following hypothesis:

> *H_o: Mutual fund managers do not possess skill (good or bad) and performance is only due to luck*
> *H_1: Mutual fund managers do possess skill (good or bad) and performance is not only due to luck, but some skill as well*

The purpose of this thesis is to examine the performance of 1704 actively managed U.S. mutual funds during the period from January 1995 to December 2017. We use four different factor models as performance measurement: The capital asset pricing model (CAPM), the Fama-French three-factor model, the Carhart four-factor model with Fama and French’s version of the momentum factor and the Fama-French five-factor model. We use bootstrap simulations to distinguish skill from luck in the cross-section of three-factor t(α) estimates for mutual fund returns. Kosowski et al. (2006) were first to apply a bootstrap procedure to examine the performance of actively managed U.S. mutual funds. We use the bootstrap methodology of Fama and French (2010), where we jointly sample fund returns rather than simulating for individual funds.

In order to obtain an initial overview of the aggregate performance of U.S. mutual funds, we construct an equally weighted portfolio consisting of all the funds in our sample. We estimate regressions on monthly excess returns from January 1995 to December 2017 using the above-mentioned factor models. We find that alpha (α) is close to zero for all performance models on both equally weighted net and gross fund returns. Additionally, the exposure to the market portfolio is close to 1. These results are in accordance with the observations of Fama and French (2010) and suggest that if there are active fund managers who produce positive true α, they are balanced by active managers with negative α.

When we turn to individual fund performance, we aim to distinguish skill from luck to draw inferences about the existence of superior and inferior managers. We compare the actual cross section of $t(\alpha)$ estimates to simulated values from 10,000 bootstrap simulations. The simulated fund returns have the same characteristics as the actual fund returns, but true α is set to zero. Thus, the distribution of simulated $t(\alpha)$ estimates represents what we can expect if abnormal returns are only due to luck. Then we examine whether fund managers are skilled/unskilled by comparing the simulated $t(\alpha)$ estimates to the actual values.

When $t(\alpha)$ is estimated on net fund returns, the empirical results show evidence of both inferior and superior fund managers. The distribution of $t(\alpha)$ estimates display a majority of inferior managers. This suggests that most active fund managers do not possess sufficient skill to cover their costs. However, when we add back the management fees and compare the $t(\alpha)$ estimates from gross fund returns to the simulated averages, we find much stronger evidence of skill.

We find stronger evidence of skill than Fama and French (2010), which is rather interesting. Since our sample period covers the post-financial crisis period, our results suggest that active fund managers are more skilled today than they were in previous decades. Fama and French (2010) found stronger evidence of skill to cover fees examining the sample period used by Kosowski et al. (2006), which is an earlier sample period from 1975 to 2002. For the most part, they rationalized this by saying that there was a higher percentage of skilled managers in olden times due to fewer funds. Our findings question the validity of this argument.

Under the assumption that the cross-section of true α has a normal distribution with mean zero and standard deviation σ, then σ around 1.25% per year seems to capture the right tail of the cross-section of α estimates, while σ around 0.75% per year captures the left tail. The σ estimates do not suggest much superior or inferior performance in producing returns gross of fees. Fama and French (2010) found that σ around 1.25% captured both tails of the cross-section of α estimates for their sample period. We find a lower σ for the left tail, which indicates a lower level of inferior performance for the funds in our sample.

This thesis is structured as follows: Section 2 presents a review of existing literature, Section 3 explains the methodology used, Section 4 covers how the data (fund returns and Fama-French factors) was collected and handled, Section 5 provides empirical results from regressions and bootstrap simulations and Section 6 concludes.

2. Literature review

Most of the empirical literature on mutual fund performance concentrates on whether mutual funds outperform or underperform the market and the degree to which performance persists. The empirical findings based on U.S. mutual funds are somewhat mixed. However, the vast majority of the studies finds little or no superior performance and argues that active managers are not able to beat the market. Most of these studies stem from the early work of Jensen (1968), who examined the performance of U.S. mutual funds in the 1945-1964 period. Jensen discovered that mutual funds were not able to outperform the market and cover their costs, which is in line with the efficient market hypothesis of Malkiel & Fama (1970).

Malkiel (1995) examined the performance of actively managed equity U.S. mutual funds from 1971 to 1991 and found that funds generally underperformed their benchmark portfolios even gross of expenses. Carhart (1997) used net α to investors to assess whether U.S. fund managers are skilled or not. He found that common factors in stock returns and investment expenses almost completely explain persistence in fund returns, which did not support the existence of skilled or informed fund managers. Fama and French (2010) examined the performance of actively managed U.S. mutual funds from 1984 to 2006 using a bootstrap procedure to distinguish skill from luck. The simulation tests on net returns showed little evidence of managers with sufficient skill to cover their costs. When costs were added back, they found evidence of both superior managers with positive true α and inferior managers with negative true α.

On the contrary, there are studies that find evidence of skilled active managers, despite the general consensus that it is not possible to beat the market. Grinblatt and Titman (1989) examined the performance of U.S. mutual funds in the 1975-1984 period and found evidence of fund managers that were able to persistently earn abnormal returns. The value of active management is further encouraged by Wermers (2000), who found that mutual funds on average hold stocks that outperform the market. Kosowski et al. (2006) were the first to use a bootstrap methodology to distinguish skill from luck in mutual fund performance. They found that a sizeable minority of active fund managers persistently outperform the market. Cuthbertson et al. (2008) applied the bootstrap methodology to U.K. mutual funds and observed similar results. Finally, Berk and Van Binsbergen (2015) investigated the performance of U.S. mutual funds from 1962 to 2011 using the value a mutual fund manager adds as the measure of skill. Their findings revealed that skilled managers exist and that this skill is persistent up to ten years. They also found that there is a strong positive correlation between current compensation and future performance, and that better funds earn higher aggregate fees.

3. Methodology

3.1 Factor models

In this thesis, we apply four different factor models of performance evaluation: The Capital Asset Pricing Model (CAPM) of Sharpe (1964), Lintner (1965) and Mossin (1966), the three-factor model of Fama and French (1993), Carhart's four-factor model (1997) and the five-factor model of Fama and French (2015). Fama and French (1993) argued that the three-factor model can explain over 90% of the variability in returns, whereas the CAPM is only able to explain around 70%. The CAPM (1), the three-factor model (2), the four-factor model (3) and the five-factor model (4) can be specified as follows:

$$R_{i,t} = \alpha_{i,t} + \beta_i R_{m,t} + \varepsilon_{i,t} \tag{1}$$

$$R_{i,t} = \alpha_{i,t} + \beta_i R_{m,t} + s_i SMB_t + h_i HML_t + \varepsilon_{i,t} \tag{2}$$

$$R_{i,t} = \alpha_{i,t} + \beta_i R_{m,t} + s_i SMB_t + h_i HML_t + m_i MOM_t + \varepsilon_{i,t} \tag{3}$$

$$R_{i,t} = \alpha_{i,t} + \beta_i R_{m,t} + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + \varepsilon_{i,t} \qquad (4)$$

where $R_{i,t}$ is the excess gross or net fund return, $R_{m,t}$ is the excess return on the market portfolio, SMB_t and HML_t are the size and value factors of Fama and French (1993), MOM_t is Fama and French's version of Carhart's (1997) momentum factor, RMW_t and CMA_t are the profitability and investment factors of Fama and French (2015), $\alpha_{i,t}$ is the abnormal return and $\varepsilon_{i,t}$ is the regression residual. The abnormal return (the intercept in the models above) refers to the average return left unexplained by the benchmark models and is commonly known as Jensen's alpha (1968). This parameter lets us know whether a fund is able to generate significant abnormal returns. If the intercept is positive and statistically significant (true α), the fund manager has outperformed a passive benchmark portfolio. Conversely, if the intercept is negative and statistically significant, the fund manager has underperformed a passive benchmark portfolio.

3.2 The bootstrap procedure

To distinguish skill from luck, we employ the bootstrap procedure introduced by Kosowski et al. (2006) with the modifications proposed by Fama and French (2010). The modified version involves simulating residuals and factor returns for all funds jointly, rather than only residuals for one fund at the time, which maintains the cross-correlation of returns across funds (Fama and French, 2010). The bootstrap procedure is outlined in the following paragraphs. Technical details of the procedure can be found in Appendix B.

The bootstrap procedure starts by estimating one benchmark regression model for each fund in the dataset, using the three-factor model. We obtain historical returns on each fund and regress them against the risk factors in (2). Estimates of actual alpha ($\alpha_{i,t}$) and its corresponding t-statistic, t(α), as well as coefficient estimates for the risk factors ($R_{m,t}$, SMB_t, HML_t) and residuals ($\varepsilon_{i,t}$) are collected for each fund.

We construct a set of simulation runs S, which are identical for all funds. For each simulation run, we draw a random vector T_s from the uniform distribution $\{U_t(0,1)\}$, where T denotes the number of periods in the dataset and s is a unique simulation run. We then round up to the nearest integer, which results in the following vector of time indices, randomly drawn with equal probability and replacement:

$$T_s = roundup(T * \{U_t(0,1)\}_{t-1}^{T})$$

where $s = 1, \ldots, S$. The next step is to construct new time series of zero-α fund returns and risk factor returns utilizing the simulated time indices. The α-free fund returns have the same properties of actual fund returns (except that true α is equal to zero) and are now computed as follows:

$$r_{i,T_s}^{e} = \sum_{j=1}^{K} \beta_{i,j} * f_{j,T_s} + \varepsilon_{i,T_s}$$

where K denotes the number of factors used in the regression and f_{j,T_s} is the new series of risk factor returns, while ε_{i,T_s} consists of drawn residuals from the original regression model.

Furthermore, we run regressions on the constructed zero-α fund returns from the simulation runs as dependent variables and the new risk factor returns as explanatory variables. We do this S times to obtain S simulated estimates of α with corresponding t-statistics, t(α). The estimates of α and t(α) obtained from using historical returns are then compared to the α and t(α) estimates from the bootstrap simulations. In line with previous research, we focus on estimates of t(α), rather than the α estimates. This is because the t-statistics control for differences in precision due to differences in residual variance (Fama and French, 2010).

3.3 Historical versus simulated returns

In order to compare the historical returns to the simulated returns, we sort the estimates of t(α) for both the benchmark regressions and the bootstrap simulations. The actual cross-section of t(α) estimates are compared to the t(α) estimates obtained from 10,000 bootstrap simulations on selected percentiles. We compute the average over all the simulated values at the different percentiles to compare them against the actual t(α) estimates. Thus, we can observe how well the funds in our sample perform at different percentiles compared to how they should perform if there is no presence of skill and returns are only due to luck. For instance, the t(α) of the funds that did best in our sample (the 99th percentile) is compared to the best average t(α) estimate from the bootstrap simulations.

Similar to previous studies, we also measure the fractions of simulation runs that produce lower t(α) values than the actual t(α) at the selected percentiles. These fractions allow us to judge more formally whether the tails of actual t(α) estimates are extreme relative to the simulated values. For example, if low fractions of simulation runs are below the actual t(α) estimates in the left tail, we infer that there is presence of bad skill and active managers cannot cover their costs. Likewise, if large fractions of the simulation runs below the actual t(α) estimates in the right tail, we infer that there is presence of good skill and active managers can more than cover their fees.

3.4 Estimating the distribution of true α

To examine the likely size of the skill effects in fund returns, we continue to follow the methodology of Fama and French (2010) and repeat the bootstrap procedure with α injected into the fund returns. The standard deviation of α is altered from 0.0% to 2.0% (in steps of 0.25%) for each simulation run. We then compare the results from these simulation runs to the actual values of α. As stated by Fama and French (2010), we are then able to investigate (i) how much α is necessary to reproduce the distribution of t(α) estimates for actual gross fund returns and (ii) which levels of α that are too extreme to be consistent with the t(α) estimates for actual gross fund returns.

The bootstrap procedure with injected α is similar to the previous simulation runs. However, instead of leaving out α and create a "luck distribution", we now inject different values of α into the fund returns. The fund returns are now computed as follows:

$$r_{i,T_s}^{e} = \frac{\alpha_{i,S}}{12} * S_i \sum\nolimits_{j=1}^{K} \beta_{i,j} * f_{j,T_s} + \varepsilon_{i,T_s}$$

where $\alpha_{i,S}$ denotes the annual α. This number is drawn from a normal distribution with mean equal to zero and standard deviation σ per year. S_i is a scalar that adjusts for different levels of diversification that funds may pursue due to different strategies, which is defined as follows:

$$S_i = \frac{SE(\varepsilon_i)}{\left(\frac{\sum_{i=1}^{N} SE(\varepsilon_i)}{N}\right)}$$

where $SE(\varepsilon_i)$ is the standard error of the residuals from the initial benchmark regression for fund *i* and *N* is the total number of funds included in our sample. Hence, the denominator is the average standard error of the residuals and the scalar S_i is lower for more diversified funds. As stated by Fama and French (2010), the scalar is included because funds that hold a highly diversified portfolio are less likely to generate true α.

There are two ways to figure out (i) how much α is necessary to reproduce the distribution of t(α) estimates for actual gross fund returns. We can find the value of σ that generates simulation t(α) estimates below those from actual returns in approximately 50% of the simulation runs, or we can look for the value of σ that produces average percentiles of simulations equal to the actual fund returns.

To find (ii) which levels of α that are too extreme to be consistent with the t(α) estimates for actual gross fund returns, we apply what Fama and French (2010) refers to as 'the 20% rule'. Under this rule, we accept a 20% probability of setting an upper bound that is too low and a 20% probability of setting a lower bound that is too high. The upper bound for the left tail estimate of σ is the value that yields simulated percentiles below corresponding actual percentiles in about 80% of the simulation runs, while the lower bound for the left tail estimate of σ is the value that yields simulated percentiles below corresponding actual percentiles in about 20% of the simulation runs. On the contrary, the upper bound for the right tail estimate of σ is the value that yields simulated percentiles below corresponding actual percentiles in about 20% of the simulation runs. The lower bound for the right tail estimate of σ is the value that yields simulated percentiles below corresponding actual percentiles in about 80% of the simulation runs.

4. Data

4.1 Mutual fund returns

Monthly mutual fund net returns are retrieved from the Bloomberg database. The data consists of 1704 actively managed U.S. open-end, domestic equity mutual funds in the period of January 1995 to December 2017. All of the funds in our sample were still active as of December 2017. The total number of active funds in our dataset throughout the sample period is illustrated in Figure A2 in the appendix.

In line with Fama and French (2010), we exclude funds that do not reach $5 million in asset under management (AUM) during their lifetime. The reason for this is to limit the effects of incubation bias, which occurs when funds include pre-release returns in mutual fund databases when they open to the public. These returns are typically positive and lead to an upward bias in return histories. Similar to Fama and French (2010), we only include funds that have at least 60 monthly returns (active for at least five years) to avoid having a lot of funds with short return histories. To avoid including passive funds that are labelled as actively managed, we remove every fund with a variation of the word "index" in its name. We manually check every questionable fund for its investment strategy. Finally, funds with irregularities in returns are removed from the sample.

Gross fund returns are constructed by adding back the management fees (as of December 2017) to the net returns. Management fees are retrieved from the Bloomberg database. These fees are stated annually and not available in time series. Thus, we assume that management fees have been constant over time. This is a caveat, because we know that the fees have varied over time. In sum, the gross fund returns are estimates and not completely accurate.

4.2 Survivorship bias

Since our sample is limited to funds that are active as of December 2017, we have to be aware of the issue of survivorship bias. Excluding defunct funds from the sample is likely to result in an overestimation of historical performance. This is because mutual funds that perform poorly is more likely to be liquidated than funds with good performance (Carhart, 1997). Brown et al. (1992) and Malkiel (1995) find that excluding defunct funds from the sample significantly biases empirical results. We are aware of the potential effect survivorship bias has on our results.

4.3 Risk factors

The risk factors used in this thesis are constructed by Fama and French and retrieved from Kenneth R. French Data Library. These factors are the excess return on the market (Rm-Rf), small minus big (SMB), high minus low (HML), momentum (MOM), profitability (RMW) and investment (CMA). The excess return on the market is the return on a value-weight market portfolio of NYSE, Amex, and NASDAQ stocks minus the 1-month Treasury bill rate (risk-free rate).

4.4 Summary statistics

Table 1 shows descriptive statistics for the excess net (Rn) and gross return (Rg) of an equal-weighted portfolio, excess market return (Rm-Rf), risk-free return (Rf) and returns of five additional risk factors. The equal-weighted portfolio consists of all the funds in our sample. The excess net and gross return display the highest average monthly returns of 0.99% (t = 3.74) and 0.74% (t = 2.79), which are both higher than the excess market return of 0.71% (t = 2.74). For the five additional risk factors, MOM yields the highest average return of 0.42% (t = 1.36), though insignificant. RMW is the only risk factor significant at the 5% level with an average return of 0.35% (t = 2.00). The SMB, HML and CMA display average returns of 0.14% (t = 0.68), 0.20% (t = 1.09) and 0.23% (t = 1.79). Table 1 also displays the correlation-matrix between all the variables mentioned above. As expected, the excess net and gross returns of the equal-weighted portfolio and the excess market return (Rm-Rf) yield the highest correlations of 0.98. The second highest correlation is between the CMA and HML (0.65), while the highest negative correlation is between the SMB and RMW (-0.58).

Table 1: Descriptive statistics

This table shows descriptive statistics for the risk-free rate, excess net and gross returns of an equal-weighted portfolio and all the risk factors we have used in this thesis on a monthly basis. The average return, median, standard deviation, max and min are reported as percentages. Average returns are calculated as the arithmetic mean. Rm is the return on a value-weight market portfolio of NYSE, Amex, and NASDAQ stocks and Rf is the 1-month Treasury bill rate. SMB (Small Minus Big) is the average return on three small portfolios minus the average return on three big portfolios. HML (High Minus Low) is the average return on two value portfolios minus the average return on two growth portfolios. MOM (Momentum) is the average return on two high prior return portfolios minus the average return on two low prior return portfolios. RMW (Robust Minus Weak) is the average return on two robust operating profitability portfolios minus the average return on two weak operating profitability portfolios. CMA (Conservative Minus Aggressive) is the average return on two conservative investment portfolios minus the average return on two aggressive investment portfolios.

199501-201712	Rf	Rn	Rg	Rm-Rf	SMB	HML	MOM	RMW	CMA
Summary statistics									
Average return	0.19	0.74	0.99	0.71	0.14	0.20	0.42	0.35	0.23
Median	0.13	1.34	1.53	1.33	0.03	-0.06	0.56	0.38	-0.03
Standard deviation	0.18	4.39	4.39	4.33	3.35	3.11	5.10	2.89	2.15
t-statistics	17.39	2.79	3.74	2.74	0.68	1.09	1.36	2.00	1.79
Max	0.56	12.79	12.86	11.35	21.71	12.90	18.36	13.51	9.58
Min	0.00	-19.11	-18.97	-17.23	-16.88	-11.10	-34.39	-18.72	-6.87
Skewness	0.34	-0.80	-0.81	-0.77	0.79	0.16	-1.50	-0.41	0.66
Kurtosis	-1.55	2.00	1.98	1.39	8.30	2.54	10.39	9.65	2.38
Cross-correlation									
Rf	1								
Rn	-0.02	1							
Rg	0.02	1.00	1						
Rm-Rf	-0.03	0.98	0.98	1					
SMB	-0.07	0.35	0.34	0.22	1				
HML	0.08	-0.10	-0.10	-0.15	-0.28	1			
MOM	0.10	-0.29	-0.28	-0.28	0.08	-0.20	1		
RMW	0.06	-0.50	-0.49	-0.49	-0.58	0.44	0.09	1.00	
CMA	0.06	-0.30	-0.29	-0.34	-0.12	0.65	0.02	0.30	1.00

5. Empirical results

5.1 Regression results

We examine the performance of active fund managers in aggregate by constructing an equal-weighted (EW) portfolio of all the funds in our sample. The funds are weighted equally each month. Table 2 shows regression results where the dependent variables are excess net and gross return on the EW portfolio. The intercept is the average annualized α, which reveals whether funds on average generate returns that deviate from those implied by their exposure to different risk factors. A positive, statistically significant α coefficient reveals that active fund managers on average possess sufficient skill to cover their costs for EW net returns and whether they are able to beat the market for EW gross returns.

For EW net returns, we observe that the estimated α coefficients are close to zero for all factor models, but only statistically significant for the five-factor model. The average annualized α of the five-factor model is -0.50% (t = -3.34). This is close to what we expect, because the returns are weighed down by fees. Actually,

Table 2: Regression results for different factor models for an EW portfolio of actively managed mutual funds

This table provides regression results for an EW portfolio of actively managed U.S. mutual funds. The dependent variables are the net and gross returns of the portfolio. The sample period is 276 months and number of funds used to compute the EW mean returns varies from 417 to 1704. Regression results are reported with coefficients and corresponding t-statistics. Alphas are annualized (α*12) and their t-statistics are calculated with annualized standard errors. The table also shows regression slopes for the explanatory variables (Rm-Rf, SMB, HML, MOM, RMW, CMA) with corresponding t-statistics and the regression R^2. The t-statistics test whether the coefficient is different from 1 for the market slope and different from 0 for the remaining factors. We use OLS estimation and the standard errors corrected for heteroscedasticity and autocorrelation with the Newey and West (1986) method.

		α*12	Rm-Rf	SMB	HML	MOM	RMW	CMA	R^2
	Net returns								
CAPM	Coef	0.0036	0.99						0.96
	t-stat	1.38	-0.45						
FF3	Coef	-0.0007	0.97	0.20	0.12				0.98
	t-stat	-0.44	-2.73	6.77	3.90				
FF4	Coef	0.0004	0.96	0.20	0.11	-0.01			0.98
	t-stat	0.22	-2.56	7.41	3.50	-0.65			
FF5	Coef	-0.0050	0.98	0.24	0.06		0.06	0.01	0.98
	t-stat	-3.34	-1.26	11.15	2.14		3.39	0.20	
	Gross returns								
CAPM	Coef	0.0106	0.99						0.96
	t-stat	4.10	-0.45						
FF3	Coef	0.0063	0.97	0.20	0.12				0.98
	t-stat	3.89	-2.73	6.77	3.90				
FF4	Coef	0.0074	0.96	0.20	0.11	-0.01			0.98
	t-stat	4.21	-2.57	7.41	3.50	-0.65			
FF5	Coef	0.0020	0.98	0.24	0.06		0.06	0.01	0.98
	t-stat	1.37	-1.27	11.14	2.14		3.39	0.20	

we expect a value closer to the negative of the average management fee in our sample (0.73%).

The α coefficient is positive and statistically significant for the CAPM, three-factor and four-factor models for EW gross fund returns. This suggests that active managers on average are to beat the market. The CAPM produces an average annualized α of 1.06% (t = 4.10), which corresponds to 0,09% per month. However, the annualized α is lower when we control for more risk factors. The three-factor and four-factor models produce an annualized α of 0.63% (t = 3.89) and 0.74% (t = 4.21), respectively.

For both EW net and gross returns, the exposure to the excess return on the market portfolio is close to 1, and statistically significantly different from 1 for the three-factor and four-factor models. The three-factor model for the EW net fund returns provides a market coefficient equal to 0.97. This means that when the market portfolio increases by 1%, the EW net fund return portfolio increases by 0.97%. Conversely, if the market decreases by 1%, the EW net fund return portfolio decreases by 0.97%.

Both the EW net and gross returns show a positive, significant exposure to the size portfolio (SMB), value portfolio (HML) and profitability portfolio (RMW), while the coefficients on the momentum portfolio (MOM) and investment portfolio (CMA) are close to zero and insignificant.

In sum, the regression results suggest that active fund managers on average do not possess sufficient skill to cover their costs, which is consistent with Carhart (1997) and Fama and French (2010). This indicates that if there exist active fund managers with skill to cover costs, they are balanced by active fund managers who lack skill. However, we find evidence that suggests that active managers on average are able to beat the market. This is consistent with the predictions of Berk and Green (2004).

5.2 Bootstrap results – net returns

Table 3 reports t(α) estimates for actual and simulated net fund returns at selected percentiles, sorted from lowest to highest. We observe that the left tail percentiles of t(α) estimates from actual net fund returns are very low compared to the corresponding percentile of simulated t(α) estimates. For example, the 1^{st} and 5^{th} percentile of actual t(α) estimates are -3.62 and -2.38, which means that 1% and 5% of funds have t(α) below -3.62 and -2.38, respectively. The average value of the corresponding percentiles from the bootstrap simulations are -2.40 and -1.67. The t(α) estimates from the actual net fund returns are below the average values from the simulation runs for all percentiles below the 96^{th}.

The last column in Table 3 shows the percentage of simulation runs that produce lower values of t(α) at a given percentile than those from the actual fund returns. For example, 0.12% of the simulation runs produce t(α) values below the actual value for the 1^{st} percentile, which corresponds to only 12 out of 10,000 bootstrap simulations. We observe that up to and above the 30^{th} percentile, the simulated values are above the actual value in more than 90% of the simulation runs. Additionally, up and until the 90^{th} percentile, the simulation runs generate average simulated values above the actual values. In short, the bootstrap simulations suggest that the vast majority of fund managers do not possess sufficient skill to cover their costs and fund managers that perform poorly lack skill (negative true

Table 3: Percentiles of t(α) estimates for actual and simulated net fund returns from January 1995 to December 2017

The table shows t(α) values at different percentiles of the distribution of t(α) estimates for actual net fund returns (Actual). The simulated average (Simulated) is the average of the distribution of t(α) estimates at the selected percentiles produced by 10,000 bootstrap simulations. The last column (% < Actual) shows the percentage of simulation runs that produce lower values of t(α) at a given percentile than those from the actual fund returns. The results are based on the Fama-French 3-factor model, where Rm-Rf, SMB and HML are the explanatory variables. We use OLS estimation and the standard errors corrected for heteroscedasticity and autocorrelation with the Newey and West (1986) method.

	T-statistics of alpha		
Percentile	Actual	Simulated	% < Actual
1%	-3.62	-2.40	0.12
2%	-3.13	-2.10	0.28
3%	-2.84	-1.92	0.46
4%	-2.53	-1.78	1.25
5%	-2.38	-1.67	1.51
10%	-1.84	-1.30	3.18
20%	-1.25	-0.85	6.51
30%	-0.88	-0.53	8.62
40%	-0.56	-0.25	11.32
50%	-0.27	0.00	13.48
60%	0.03	0.26	18.01
70%	0.39	0.53	30.27
80%	0.73	0.85	33.61
90%	1.28	1.30	50.00
95%	1.69	1.68	55.01
96%	1.85	1.79	62.16
97%	2.04	1.92	67.50
98%	2.32	2.11	76.79
99%	2.67	2.40	80.24

α). Thus, we reject the null hypothesis that poor performance is only due to bad luck.

The right tail of the distribution suggests that a sizeable minority of fund managers do have sufficient skill to produce risk-adjusted expected returns that cover their costs (positive true α). The t(α) estimates from the actual net fund returns are above the average values from the simulation runs from the 96th percentile and up. We observe that the actual t(α) estimates are higher than the simulated average in more than half of the simulation runs above the 90th percentile. For the top 1% performing funds, the simulated t(α) estimates are smaller than the actual values in almost 20% of the simulation runs. Hence, we can reject the null hypothesis that good performance is only due to luck.

Figure 1 provides a graphical illustration of the results in Table 3. The solid curve shows the percentage of actual t(α) estimates below each value, while the dashed line shows the equivalent for the simulated averages. We observe that the distribution of t(α) from actual fund returns is situated to the left of the simulated averages, except from somewhere between the 90th and 95th percentile, where the actual t(α) estimates are higher than the simulated averages. In other words, the top 5-10% performing funds have an actual t(α) that is higher than the simulated averages, which is consistent with the results from Table 1.

Figure 1: Cumulative distribution function of t(α) estimates for actual and simulated net fund returns

The figure shows empirical cumulative distribution functions of actual and simulated t(α). The CDFs are based on the values of actual and simulated t(α)s from Table 3.

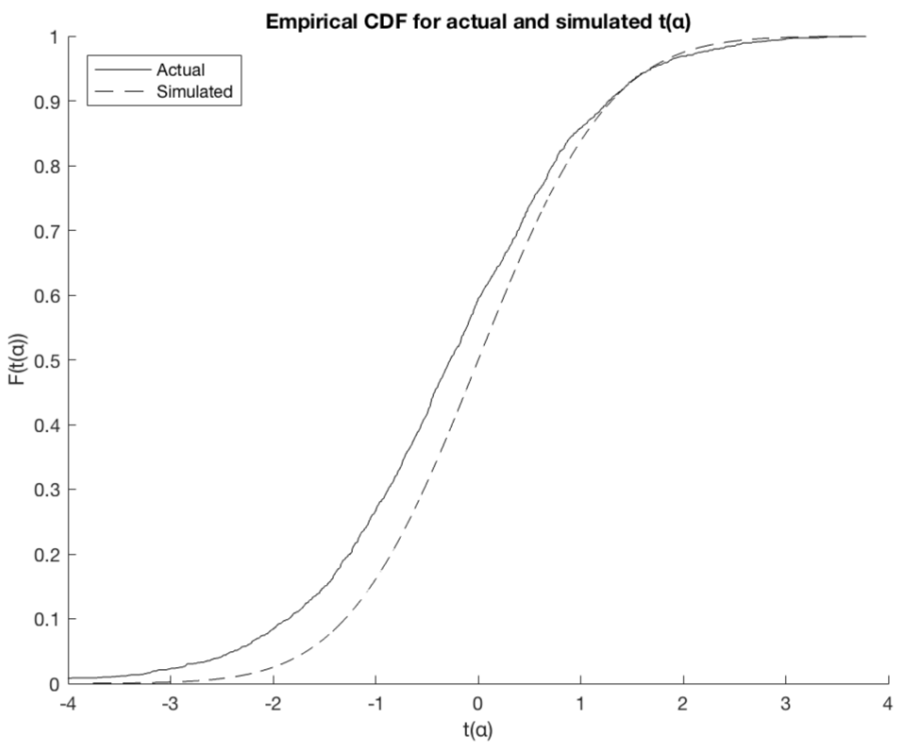


Figure 2 shows Kernel density estimates of actual and bootstrap distributions of t(α), inspired by the methodology of Cuthbertson et al. (2008). The dotted line shows the distribution of simulated t(α) if performance is only due to luck, while the solid line shows the actual t(α) distribution. The results show that the left tail of the actual t(α) distribution lies largely to the left of the luck distribution and reinforces the evidence that bad performance cannot be explained by bad luck alone. Additionally, we observe that the extreme right tail of the actual t(α) distribution lies outside the luck distribution, which reinforces the evidence of fund managers with skill to cover their fees in the top performing funds.

Figure 2: Kernel density estimates of actual and simulated t_{α} for net fund returns

The figure shows Kernel density estimates of actual and simulated t(α) values for net fund returns. The Kernel density estimates are based on the actual and simulated t(α) values from Table 3.

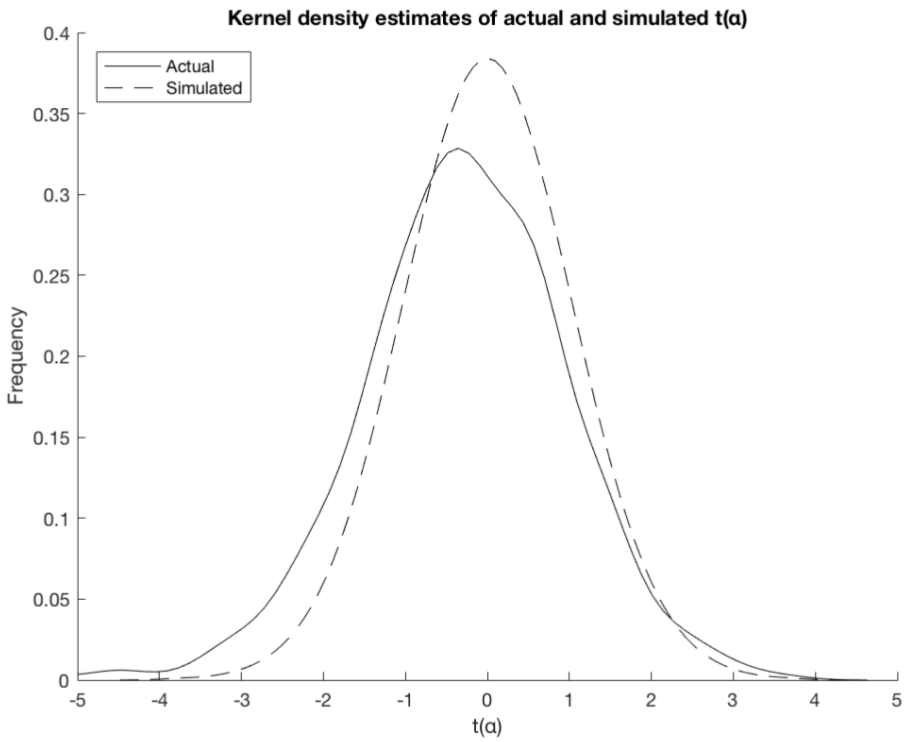
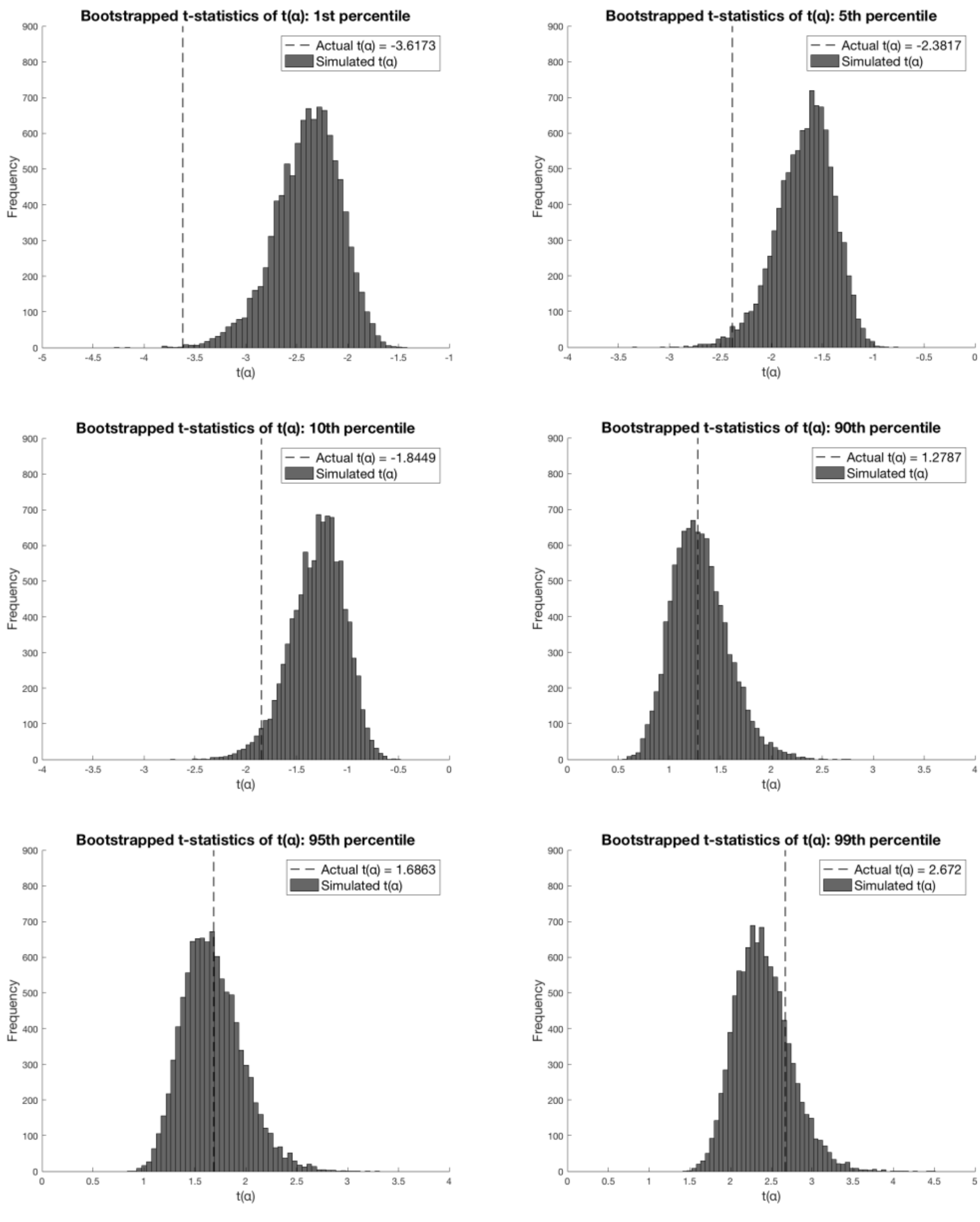


Figure 3 shows histograms for actual and simulated values of t(α) at different percentiles. The dotted lines constitute the actual values of t(α), while the histograms present the distribution of simulated (α) values obtained from 10,000 bootstrap simulations. For example, the bottom right figure represents the results for the top 1% performing funds (99^{th} percentile), where the actual value of t(α) is 2.672. We observe from the histogram that most of the simulated values of t(α) situated below the actual value. This is consistent with the findings from Table 2, which shows that 80.24% of the simulated values of t(α) are below the actual value.

Figure 3: Histograms of bootstrapped net return t(α) estimates at different percentiles

The figure shows histograms of bootstrapped t(α) estimates at different percentiles. The dashed lines present actual t(α) estimates at the corresponding percentile.



When t(α) is estimated on net returns, we find that a sizeable minority of active fund managers have sufficient skill to cover their costs. This is consistent with Kosowski et al. (2006) and Cuthbertson et al. (2008) and contrary to Fama and French (2010).

5.3 Bootstrap results – gross returns

Table 4 displays t(α) estimates for actual and simulated gross fund returns using the same percentiles as for the net returns. We observe that the distribution of actual t(α) estimates is now shifted to the right, because the fees are added back to the returns. For example, the 1st and 5th percentile of actual t(α) estimates for gross returns are -2.90 and -1.82, while the corresponding percentiles for actual t(α) estimates for net returns are -3.62 and -2.38. The actual t(α) estimates are still below the average values from the simulation runs up to the 10th percentile. This suggests that there is still evidence of bad skill even though fees are added back and not all fund managers are able to beat the market.

Table 4: Percentiles of t(α) estimates for actual and simulated gross fund returns from January 1995 to December 2017

The table shows t(α) values at different percentiles of the distribution of t(α) estimates for actual gross fund returns (Actual). The simulated average (Simulated) is the average of the distribution of t(α) estimates at the selected percentiles produced by 10,000 bootstrap simulations. The last column (% < Actual) shows the percentage of simulation runs that produce lower values of t(α) at a given percentile than those from the actual fund returns. The results are based on the Fama-French 3-factor model, where Rm-Rf, SMB and HML are the explanatory variables. We use OLS estimation and the standard errors corrected for heteroscedasticity and autocorrelation with the Newey and West (1986) method.

	T-statistics of alpha		
Percentile	Actual	Simulated	% < Actual
1%	-2.90	-2.40	6.92
2%	-2.46	-2.10	11.88
3%	-2.15	-1.92	20.57
4%	-1.97	-1.78	24.07
5%	-1.82	-1.67	27.84
10%	-1.30	-1.30	46.38
20%	-0.73	-0.85	65.53
30%	-0.39	-0.53	69.47
40%	-0.05	-0.25	79.07
50%	0.22	0.00	81.07
60%	0.55	0.26	87.54
70%	0.86	0.53	89.40
80%	1.19	0.85	89.73
90%	1.80	1.30	95.30
95%	2.19	1.68	94.88
96%	2.30	1.79	94.50
97%	2.51	1.92	95.98
98%	2.84	2.10	97.76
99%	3.24	2.40	98.52

However, the simulation runs generate average simulated values above the actual values only up until somewhere between the 10th and 20th percentile. Conversely, we find strong evidence of superior management in the right tail t(α) estimates. For example, only 4.7% of the simulation runs yield a higher t(α) estimate than the actual value at the 90th percentile, in which the actual value is substantially higher than the simulated value (1.80 to 1.30). In fact, the majority of the right side of the distribution of actual t(α) estimates lies to the right of the simulated values.

Figure 4 provides the graphical illustration of the results in Table 4. We observe that the distribution of t(α) from actual gross fund returns is situated to the right of the simulated averages from just above the 10th percentile, where the actual t(α) estimates are higher than the simulated averages. In other words, only the bottom 10% performing funds have an actual t(α) that is lower than the simulated averages. This is consistent with the results from Table 4.

Figure 4: Cumulative distribution function of t(α) estimates for actual and simulated gross fund returns

The figure shows empirical cumulative distribution functions of actual and simulated t(α). The CDFs are based on the values of actual and simulated t(α)s from Table 4.

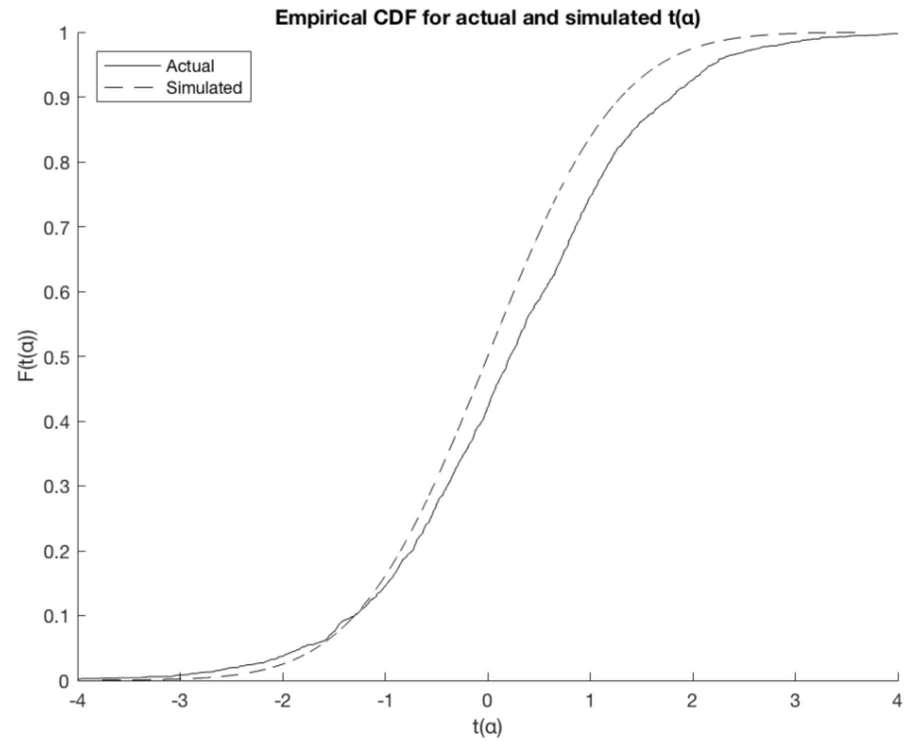


Figure 5 shows Kernel density estimates of actual and bootstrap distributions of t(α) for gross returns. The results show that the right tail of the actual t(α) distribution lies largely to the right of the luck distribution and reinforces the evidence that good performance cannot be explained only by chance. Additionally, we observe that the extreme left tail of the actual t(α) distribution lies outside the luck distribution, which reinforces the evidence that some managers are not able to beat the market.

Figure 5: Kernel density estimates of actual and simulated t_α for gross fund returns

The figure shows Kernel density estimates of actual and simulated t(α) values for net fund returns. The Kernel density estimates are based on the actual and simulated t(α) values from Table 4.

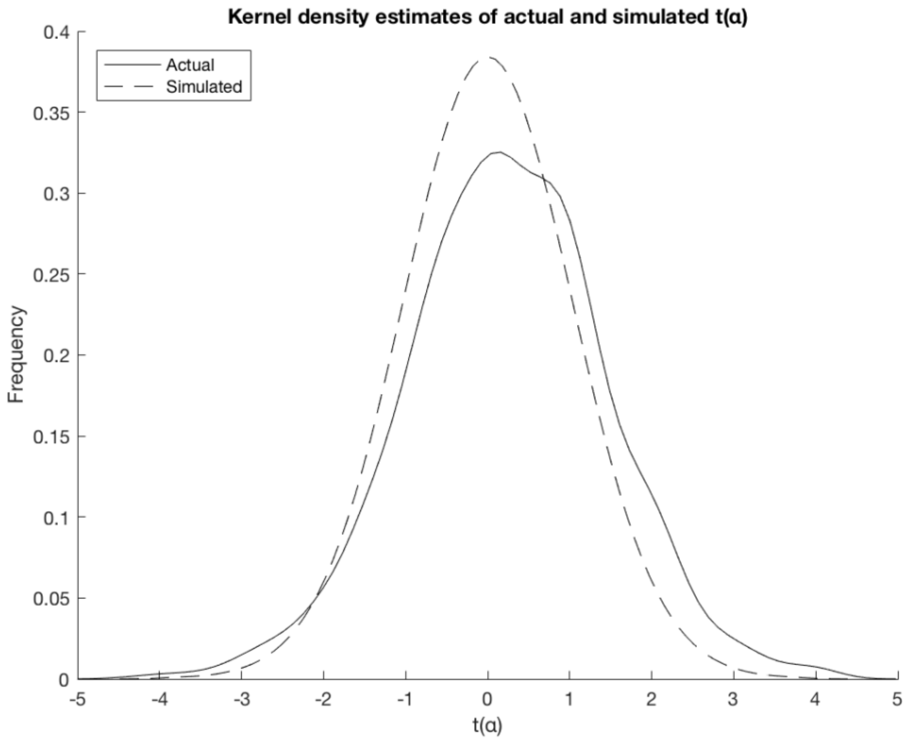
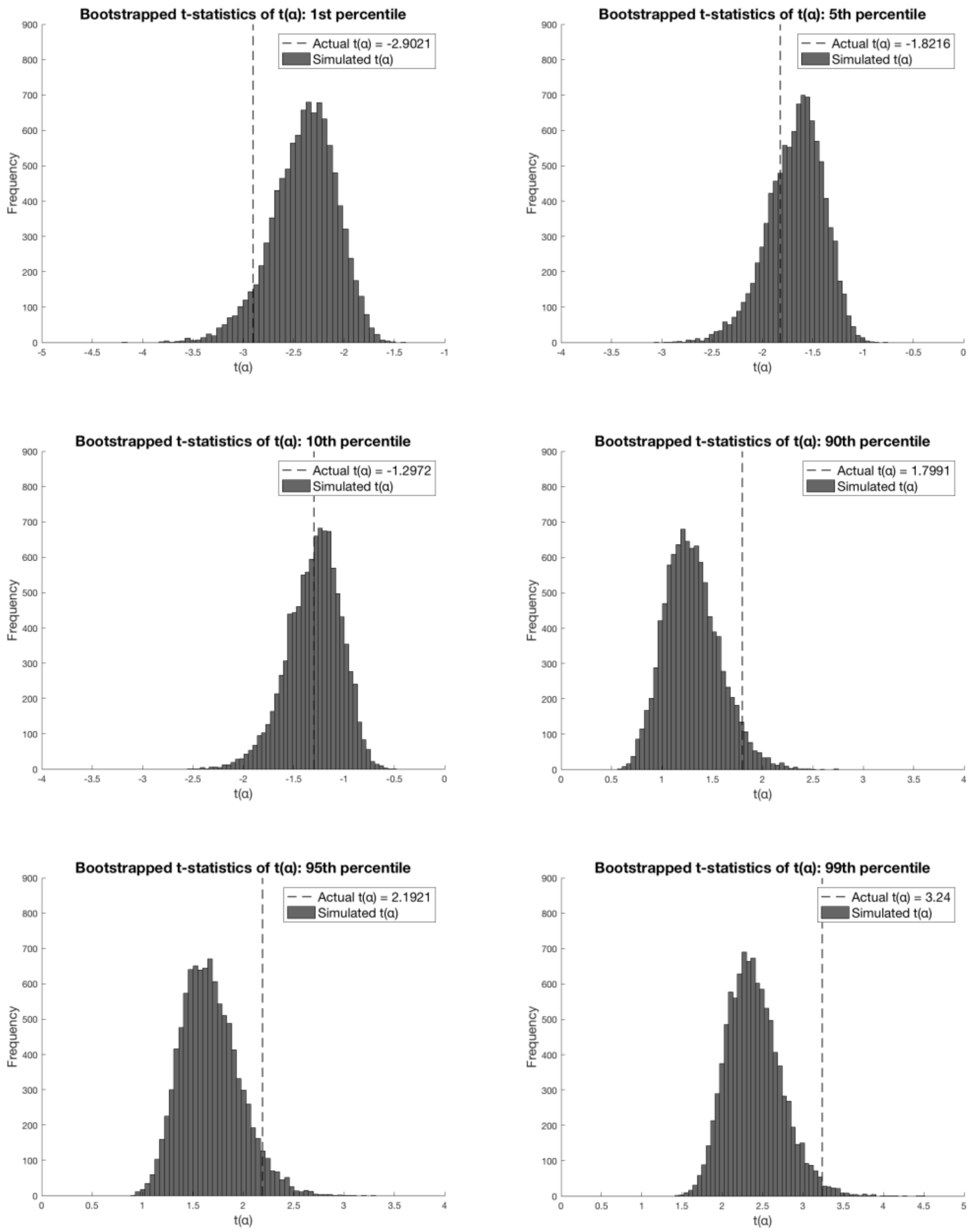


Figure 6 shows histograms for actual and simulated values of t(α) at different percentiles based on gross returns. The bottom right figure represents the results for the top 1% performing funds, where the actual value of t(α) is 3.24. Here as well as for net returns, we observe from the histogram that most of the simulated values of t(α) are situated below the actual value. This is consistent with the findings from Table 4, which shows that 98.52% of the simulated values of t(α) are below the actual value.

Figure 6: Histograms of bootstrapped gross return t(α) estimates at different percentiles

The figure shows histograms of bootstrapped t(α) estimates at different percentiles. The dashed lines present actual t(α) estimates at the corresponding percentile. The histograms on are based on the actual and simulated t(α) values from Table 4.



In sum, these results suggest that mutual fund managers for the most part are able to produce positive excess return relative to passive benchmarks before taking fees into account. This is consistent with Berk and Green (2004), who predict that most fund managers have sufficient skill to cover their costs.

5.4 Injected alpha

The likely levels of performance are measured by applying a normality assumption for true α (symmetric around zero with standard deviation σ). Even if we allow for different levels of σ for the left and right tail, we do not expect that a single value of σ will perfectly capture the tails of the t(α) estimates for actual fund returns (Fama and French, 2010). Table 5 shows the results from the bootstrap procedure with α injected into the fund returns (altered from 0.0% to 2.0% in steps of 0.25%). The simulations suggest that σ around 1.25% to 1.50% per year captures the extreme right tail of the t(α) estimates for gross fund returns, with 1.25% as the best estimate. However, the σ that captures the extreme left tail is lower, about 0.75% to 1.00% per year, with 0.75% as the best estimate. For perspective, the average of the standard errors from the individual fund α estimates is 0.145% per month (1.75% per year), slightly higher than the injected α estimates.

The σ estimates do not suggest much superior or inferior performance in producing returns gross of fees. For the right tail, σ = 1.25% indicates that about 16% of funds have true annual gross return α above 1.25% (0.10% per month) and roughly 2.4% of funds have true annual gross return α above 2.5% (0.21% per month). The left tail estimate is somewhat lower (σ = 0.75%), which suggests that about 16% of funds have true annual gross return α below -0.75% and roughly 2.4% of funds have true annual gross return α below -1.50%. Thus, the σ estimates indicate a higher level of superior performance than inferior performance.

When analyzing the unlikely levels of performance, we use the simulation results from Table 5. Similar to Fama and French (2010), we accept a 20% chance of setting a lower bound too low and a 20% chance of setting an upper bound too high. This leads to intervals for σ from the estimates for likely levels of performance plus minus 0.5%. The best right tail estimate is σ = 1.25% and leads to an interval ranging from 0.75% to 1.75%, whereas the interval for the best left tail estimate of σ = 0.75% ranges from 0.25% to 1.25%. Hence, we find it rather unlikely that σ lies below 0.25% or above 1.75%.

Table 5: Percentiles of t(α) estimates for actual and simulated gross fund returns from January 1995 to December 2017 with injected α

The table shows t(α) estimates at different percentiles of the distribution of t(α) estimates for actual gross fund returns (Actual). The simulated values are the average of the distribution of t(α) estimates at the selected percentiles produced by 10,000 bootstrap simulations, for eight values of annual standard deviation of injected α. See '*3.4 Estimating the distribution of true α*' for a description of this procedure. The table also shows the percentage of simulation runs that produce lower t(α) estimates at a given percentile than those from the actual fund returns. The results are based on the Fama-French three-factor model, where Rm-Rf, SMB and HML are the explanatory variables. We use OLS estimation and the standard errors corrected for heteroscedasticity and autocorrelation with the Newey and West (1986) method.

		Average t(α) from simulations								
Percentile	Actual	0%	0.25%	0.50%	0.75%	1.00%	1.25%	1.50%	1.75%	2.00%
1%	-2.90	-2.40	-2.43	-2.52	-2.67	-2.86	-3.10	-3.36	-3.66	-3.97
2%	-2.46	-2.10	-2.13	-2.21	-2.34	-2.51	-2.71	-2.94	-3.19	-3.46
3%	-2.15	-1.92	-1.95	-2.02	-2.14	-2.29	-2.47	-2.68	-2.90	-3.14
4%	-1.97	-1.78	-1.81	-1.88	-1.98	-2.12	-2.29	-2.48	-2.69	-2.91
5%	-1.82	-1.67	-1.70	-1.76	-1.86	-1.99	-2.15	-2.32	-2.52	-2.72
10%	-1.30	-1.30	-1.32	-1.37	-1.44	-1.54	-1.66	-1.79	-1.94	-2.09
20%	-0.73	-0.85	-0.86	-0.89	-0.94	-1.01	-1.08	-1.17	-1.26	-1.36
30%	-0.39	-0.53	-0.54	-0.56	-0.59	-0.63	-0.67	-0.72	-0.78	-0.84
40%	-0.05	-0.25	-0.26	-0.27	-0.28	-0.30	-0.32	-0.35	-0.37	-0.40
50%	0.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
60%	0.55	0.26	0.26	0.27	0.28	0.30	0.33	0.35	0.38	0.41
70%	0.86	0.53	0.54	0.56	0.59	0.63	0.67	0.73	0.78	0.84
80%	1.19	0.85	0.86	0.90	0.94	1.01	1.08	1.17	1.26	1.36
90%	1.80	1.30	1.32	1.37	1.44	1.54	1.66	1.80	1.94	2.09
95%	2.19	1.68	1.70	1.76	1.86	1.99	2.15	2.33	2.52	2.72
96%	2.30	1.79	1.81	1.88	1.99	2.13	2.29	2.48	2.69	2.91
97%	2.51	1.92	1.95	2.02	2.14	2.29	2.47	2.68	2.90	3.14
98%	2.84	2.10	2.14	2.22	2.34	2.51	2.71	2.94	3.19	3.46
99%	3.24	2.40	2.43	2.52	2.67	2.86	3.10	3.36	3.65	3.97

		Percent of simulations below Actual								
Percentile	Actual	0%	0.25%	0.50%	0.75%	1.00%	1.25%	1.50%	1.75%	2.00%
1%	-2.90	6.92	7.87	11.97	23.10	42.70	70.02	91.67	99.25	99.97
2%	-2.46	11.88	14.25	20.20	32.83	52.83	78.00	94.65	99.45	99.97
3%	-2.15	20.57	23.75	31.45	44.50	65.78	86.35	97.27	99.83	99.98
4%	-1.97	24.07	27.17	35.25	48.00	67.70	87.13	97.27	99.82	99.98
5%	-1.82	27.84	31.10	38.60	51.35	70.22	87.83	97.32	99.63	99.97
10%	-1.30	46.38	49.27	56.90	67.88	82.00	92.10	97.83	99.68	99.95
20%	-0.73	65.53	67.10	72.33	79.13	86.57	92.17	96.68	98.77	99.68
30%	-0.39	69.47	70.17	73.35	78.13	83.30	88.00	91.78	95.00	97.18
40%	-0.05	79.07	79.47	81.17	82.73	84.98	87.45	89.87	91.57	93.43
50%	0.22	81.07	80.92	81.37	81.45	81.97	82.30	82.62	82.43	82.68
60%	0.55	87.54	87.18	86.62	85.68	84.22	82.60	79.90	76.65	73.25
70%	0.86	89.40	88.70	87.42	85.65	82.37	77.55	71.10	63.02	53.42
80%	1.19	89.73	88.93	87.05	83.15	77.37	68.48	55.78	41.23	27.22
90%	1.80	95.30	94.50	92.98	89.13	82.48	70.85	52.98	32.43	14.07
95%	2.19	94.88	93.95	91.63	86.63	76.67	58.63	35.07	13.30	2.75
96%	2.30	94.50	93.68	90.77	85.23	73.95	54.35	29.90	9.43	1.58
97%	2.51	95.98	95.27	92.98	87.58	77.58	57.98	32.43	9.98	1.37
98%	2.84	97.76	97.22	95.97	92.65	84.72	67.63	40.38	14.23	2.12
99%	3.24	98.52	98.20	96.93	94.10	86.17	69.02	39.33	11.92	1.43

In sum, the right tail estimate of σ of 1.25% is consistent with Fama and French (2010), while the left tail estimate of σ of 0.75% is in contrast to their study. Our left tail estimate suggests less inferior performance.

6. Conclusion

Throughout this thesis, we examine the performance of 1704 actively managed U.S. open-end, domestic equity mutual funds in the period of January 1995 to December 2017. In line with previous research, we do not find evidence that fund managers in general have sufficient skill to cover their costs. However, when we add back the management fee of each fund, the results reveal that fund managers on average are able to beat a passive benchmark portfolio.

We use a bootstrap procedure to distinguish skill from luck. When $t(\alpha)$ is estimated on net returns, the bootstrap simulations suggest that a few active fund managers have sufficient skill to cover their costs, while the majority of active fund managers lack skill. This is consistent with Kosowski et al. (2006) and Cuthbertson et al. (2008), Berk and Van Binsbergen (2015) and contrary to Fama and French (2010). When $t(\alpha)$ is estimated on gross returns, the bootstrap simulations suggest that the majority of active fund managers have sufficient skill to beat a passive benchmark portfolio, while a few active fund managers lack skill. This is consistent with Berk and Green (2004).

Under the normality assumption that true alpha is symmetric around 0 with standard deviation σ, we find that the annual standard deviation of true α is about 1.25% per year for the right tail and about 0.75% for the left tail. In addition, σ is unlikely to be less than 0.25% or more than 1.75%. The right tail estimate is consistent with Fama and French (2010), while the left tail estimate is in contrast to their study.

6.1 Limitations

We have made our best efforts to minimize limitations, but there are certain aspects of the thesis that needs to be addressed. First, our sample of fund returns is subject to survivorship and incubation bias, which is likely to result in an overestimation of historical performance. The second concern is that we were unable to obtain consistent historical data on assets under management (AUM). Thus, we were not able to construct a value-weighted portfolio, which is regarded as more accurate than an equal-weighted portfolio. The third and final limitation is that management fees were not available to us in time series in the Bloomberg database, which prevented us from constructing accurate gross returns.

Bibliography

Andrews, D. W. K., & Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica: Journal of the Econometric Society*, 60(4), 953-966. Doi: 10.2307/2951574.

Berk, J. B., & Green, R. C. "Mutual Fund Flows and Performance in Rational Markets." Journal of Political Economy 112 (2004), 1269-1295. Doi: 10.3386/w9275.

Berk, J. B., & Van Binsbergen, J. H. (2015). Measuring skill in the mutual fund industry. *Journal of Financial Economics*, 118(1), 1-20. Doi: 10.1016/j.jfineco.2015.05.002.

Blake, D., & Timmermann, A. (1998). Mutual fund performance: evidence from the UK. *Review of Finance*, 2(1), 57-77. Doi: 10.1023/A:1009729630606.

Brown, S. J., Goetzmann, W., Ibbotson, R. G., & Ross, S. A. (1992). Survivorship bias in performance studies. *The Review of Financial Studies*, 5(4), 553-580. Doi: 10.1093/rfs/5.4.553.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The journal of Finance*, 52(1), 57-82. Doi: 10.1111/j.1540-6261.1997.tb03808.x.

Cuthbertson, K., Nitzsche, D., & O'Sullivan, N. (2008). UK mutual fund performance: Skill or luck? *Journal of Empirical Finance*, 15(4), 613-634. Doi: 10.1016/j.jempfin.2007.09.005.

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

Fama, E. F., & French, K. R. (2010). Luck versus skill in the cross-section of mutual fund returns. *The Journal of Finance*, 65(5), 1915-1947. Doi: 10.1111/j.1540-6261.2010.01598.x.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1-22. Doi: 10.1016/j.jfineco.2014.10.010.

Grinblatt, M., & Titman, S. (1989). Mutual fund performance: An analysis of quarterly portfolio holdings. *Journal of business*, 393-416. Doi: 10.1086/296468.

Investment Company Institute. (2018). 2018 Investment Company Fact Book: A Review of Trends and Activities in the Investment Company Industry, 58th edition. https://www.ici.org/pdf/2018_factbook.pdf

Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *The Journal of Finance*, 23(2), 389-416. Doi: 10.1111/j.1540-6261.1968.tb00815.x.

Kosowski, R., Timmermann, A., Wermers, R., & White, H. (2006). Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. *The Journal of Finance*, 61(6), 2551-2595. Doi: 10.1111/j.1540-6261.2006.01015.x.

Lintner, J. (1965). Security prices, risk, and maximal gains from diversification. *The Journal of Finance,* 20(4), 587-615. Doi: 10.1111/j.1540-6261.1965.tb02930.x.

Malkiel, B. G. (1995). Returns from investing in equity mutual funds 1971 to 1991. *The Journal of Finance,* 50(2), 549-572. Doi: 10.1111/j.1540-6261.1995.tb04795.x.

Malkiel, B. G., & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417. Doi: 10.1111/j.1540-6261.1970.tb00518.x.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica: Journal of the Econometric Society*, 768-783. Doi: 10.2307/1910098.

Newey, W., & West, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica: Journal of the Econometric Society*, 55(3), 703-08. Doi: 10.2307/1913610.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3), 425-442. Doi: 10.1111/j.1540-6261.1964.tb02865.x.

Wermers, R. (2000). Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. *The Journal of Finance, 55(*4), 1655-1695. Doi: 10.1111/0022-1082.00263.

Appendix A - Figures

Figure A1: Total net assets in the U.S. mutual fund industry

This figure shows the growth of total net assets in the U.S. mutual fund industry from 1995 to 2017, which corresponds to our sample period. This data is retrieved from the Investment Company Fact Book (2018).

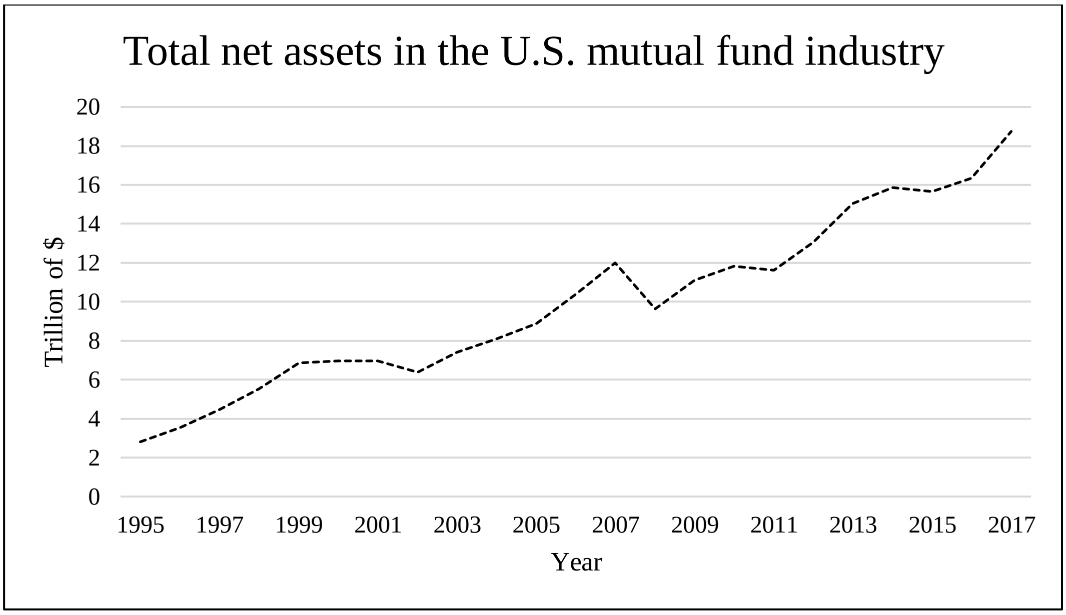
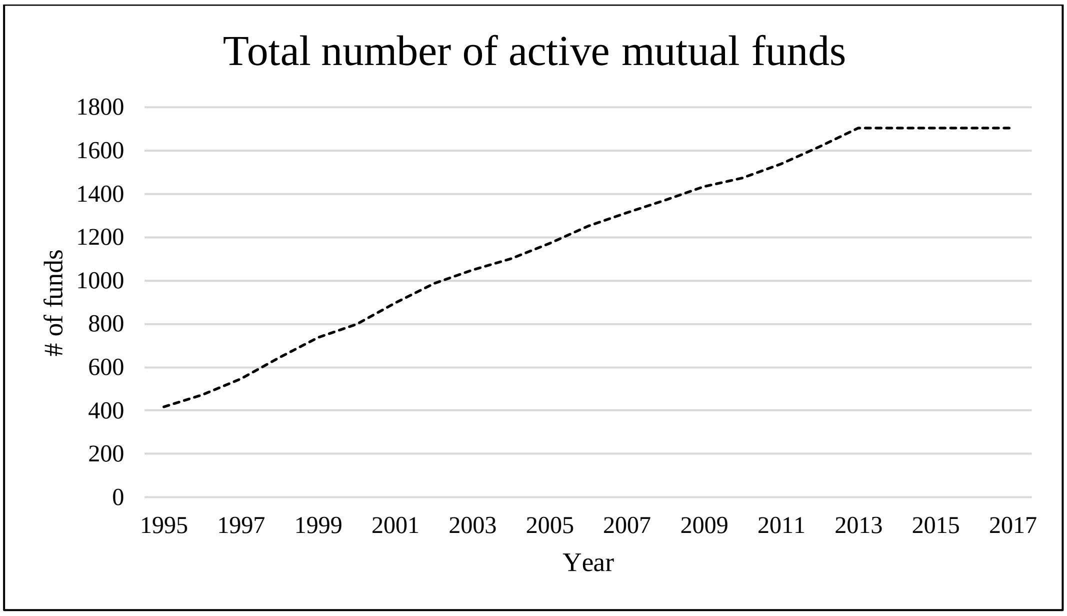


Figure A2: Total number of active mutual funds

This figure shows the total number of active mutual funds included in our dataset.



Appendix B - Technical details of the bootstrap procedure

In the following paragraphs, we provide technical details of the bootstrap procedure used in this thesis. This procedure follows the methodology of Fama and French (2010). We have chosen MATLAB as our programming tool.

The first step is to create matrices of dependent and independent variables in MATLAB. The benchmark model for the bootstrap procedure is the Fama-French three-factor model. We import monthly excess net and gross fund returns, excess market returns, size returns and value-growth returns from Excel to MATLAB. The time period is from January 1995 to December 2017. We create a matrix of excess net fund returns and a matrix of excess gross fund returns, which are the dependent variables. We create two matrices of dependent variables, because we run bootstrap simulations on both net and gross returns. We also create a matrix of independent variables, consisting of the excess market returns, size returns and value-growth returns.

The next step is to estimate benchmark regressions for each fund. We calculate the lag selection parameter for the standard Newey-West HAC estimate (Andrews and Monahan, 1992) as follows:

$$maxlag = floor(4 * \left(\frac{T}{100}\right)^{\frac{2}{9}})$$

where T is the number of observations for each fund. We estimate the standard Newey-West OLS coefficient covariance using the command ‘hac’ by setting the bandwidth to ‘maxLag+1’. We save coefficient estimates and corresponding standard errors. We compute t-statistics for each coefficient estimate. We sort the t(α) estimates and convert them into selected percentiles. In addition, standard error of residuals for each fund are calculated.

To setup the simulation runs, we create a matrix of uniformly distributed pseudorandom integers using the ‘randi’ command. From this uniform distribution, we draw random vectors of time indices. To perform 10,000 bootstrap simulations, we need 10,000 vectors of random integers. The simulation process requires a lot of memory and we were unable to perform all the 10,000 bootstrap simulations at once. A solution to this problem is to run 2,000 simulation runs five times rather than 10,000 at once. To control the random number generation, we alter ‘rng’ from 0 to 4. We concatenate the five simulation runs along the third dimension using the ‘cat’ command.

Furthermore, we construct new time series of simulated factor returns and fund returns with α set to zero. We create a new matrix of original coefficients, leaving out the intercept (α). The vectors of time indices pick corresponding numbers of original factor returns and residuals, so that the order of the time series is changed for each simulation run. Simulated fund returns are then calculated by multiplying the simulated factor returns with the new coefficient matrix and adding residuals. Thus, we have now created a matrix of simulated fund returns and a matrix of simulated factor returns.

We estimate regressions on the simulated time series in the same manner as for the actual returns. The $t(\alpha)$ estimates are sorted for each simulation run and converted into selected percentiles. Then we calculate the average of the $t(\alpha)$ estimates over all the simulation runs. Thus, we can compare the simulated $t(\alpha)$ estimates to the actual $t(\alpha)$ estimates. The results are shown in a table, where we display actual and simulated $t(\alpha)$ estimates at different percentiles. In addition, the table shows the percentage of simulation runs that produce lower values of $t(\alpha)$ at a given percentile than those from the actual fund returns. We also display the results in cumulative distribution functions (CDF), Kernel density functions and histograms.

Furthermore, we perform the exact same process with the only modification being that alpha is injected into the gross fund returns. The series of alpha is scaled for each fund using the original standard error of residuals to adjust for different levels of diversification. Similar to Fama and French (2010), we apply nine different values of annual standard deviation of alpha, ranging from 0.0% to 2.0% in steps of 0.25%. We display actual and simulated $t(\alpha)$ estimates at different percentiles in a table, as well as the percentage of simulation runs that produce lower values of $t(\alpha)$ at a given percentile than those from the actual fund returns.