Norwegian School of Economics
Bergen, Spring 2023

Should Investors Trust Morningstar Ratings?

An empirical study on Morningstar ratings ability to predict future performance using evidence from mutual funds in the U.S.

Diego Gerardo Molina Herbas & Anna Oline Bruun Setsaas

Supervisor: Petter Bjerksund

Master thesis, Economics and Business Administration

Major: Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgements

This thesis completes our Master of Science in Economics and Business Administration with a specialization in Financial Economics, at the Norwegian School of Economics (NHH). The thesis investigates if investors should trust Morningstar Ratings when investing in mutual funds in the U.S. large-cap market.

The thesis has been very challenging, but educational in many different ways. We have spent endless hours developing a code from scratch to our dataset and the purpose of our research. Working independently to apply the concepts we have studied in class for the past 5 years has given us a whole new perspective and understanding of the field.

We want to thank our supervisor, Petter Bjerksund, for his guidance throughout the process, and for challenging our abilities to work independently. We also want to thank the Norwegian School of Economics for providing access to literature and several databases that have been crucial for our thesis, such as Morningstar Direct and WRDS. Finally we want to thank our families and friends for their moral support through the semester.

Norwegian School of Economics

Bergen, May 2023

Diego Gerardo Molina Herbas

Anna Oline Bruun Setsaas

Abstract

This thesis studies the ability Morningstar ratings have to predict future performance, particularly if funds that are top rated tend to outperform the worst rated funds. Our study is focused on the U.S large-cap mutual fund market, and is based on two panel data regression models that use different performance metrics. Specifically, we use three factor models to measure performance; CAPM, Fama-French 3-factor and Carhart 4-factor, to calculate alphas for all funds in the sample over different time horizons. S&P500 is used as benchmark index and we use factor returns developed by Kenneth French. We then use two different panel data regressions, where we regress the alphas against dummies for each rating. The first regression utilizes simple monthly alphas and lagged dummies, whereas the second regression is based on alphas over different time horizons.

We find that low Morningstar ratings generally indicate poor future performance. This holds consistently for both analyses we conduct. For investment periods of 6, 12 and 36 months, low rated funds generally underperformed the higher rated funds. We also find that 5-star rated funds fail to consistently outperform the 4- and 3-star rated funds.

Based on our findings, it can be concluded that the ratings are more effective in identifying potentially poor investments rather than potentially good, or at least, great ones. Therefore, while the ratings do possess some informational value, they should not be the sole determinant when making investment decisions in mutual funds. The findings are consistent with previous research conducted by Blake and Morey (2000) and Morey and Gottesman (2006), although with stronger evidence of the underperformance of low rated funds.

Keywords – Morningstar Ratings, Mutual Funds

Contents

List of Figures

List of Tables

1 Introduction

According to ICI, mutual funds play a crucial role in building wealth for the majority of households in the United States. Almost 70 million American households across all income groups, owned mutual funds in 2022 (Investment Company Institute, 2023). The U.S. mutual fund market stands out as one of the world's largest and most diverse, offering investors with a vast array of investment opportunities. Extensive research conducted by Ippolito (1992) and Del Guercio and Tkac (2002) has documented the remarkable expansion of the market, characterized by a significant increase in the number of funds tailored to diverse investment objectives and asset classes. The mutual fund market experiences substantial investor participation from both individual and institutional investors. Studies conducted by Sirri and Tufano (1998) and Poterba and Samwick (1997) have explored the behavior of individuals in the mutual fund market, and highlighted factors such as fund flows, investor sentiment, and the impact of market conditions on investment decisions.

A recent report from ICI reveals that 62% of U.S. households find mutual fund rating services very or somewhat important when selecting what mutual funds to invest in (Investment Company Institute, 2022). Morningstar Inc. holds a prominent position in the mutual fund marketplace, providing ratings for thousands of funds in markets worldwide. The company's one- to five-star rating system has become a widely recognized benchmark for evaluating fund performance. In fact, Morningstar's star ratings are so influential that they are often the primary factor considered by retail investors when making investment decisions.

The impact of Morningstar's ratings on investor behavior is significant. The ratings have been shown to influence the flow of funds into and out of specific funds, which is a key measure of the aggregate impact of individual investors' allocation decisions. While many studies have established a correlation between investor fund flow and various measures of fund performance, including Morningstar ratings, a clear causal relationship between any given measure of fund quality and investor behavior has proven elusive.

Despite the lack of a direct causal link, Morningstar's ratings continue to hold a crucial position in the mutual fund marketplace. The company's reputation for providing reliable and independent ratings has made it a trusted source of information for both individual and institutional investors. As a result, Morningstar's ratings are likely to remain a primary factor in investment decision-making for the foreseeable future. The Morningstar ratings are also broadly used in marketing of funds. This suggests that mutual fund companies and financial service firms recognize the significance investors place on these ratings.

Because of the relationship between high ratings and inflows, it is important to assess whether these inflows are justified or reasonable. This thesis addresses the Morningstar ratings' ability to predict future performance of mutual funds. The purpose of this analysis is not to establish a causal relationship between morningstar ratings and future performance, but rather to examine potential differences in future performance based on the construction of the ratings themselves. Our empirical analysis investigates investigates the following research question:

Should the Morningstar ratings be used by investors to choose what mutual funds to invest in?

We will investigate this by using a data set consisting of 272 mutual funds across a 12 year period from January 2011 to December 2022. We conduct two analyses where we utilize fixed effects panel data regressions. Specifically, we regress different performance measures against the funds' ratings, represented by dummy variables. The performance measures are alphas estimated with the help of three factor models: CAPM, Fama-French 3-factor and Carhart 4-factor. The two analyses differ in the way that they calculate the alpha, where the first analysis is based on the simple monthly alpha, while the second analysis will look at cumulative alphas achieved over different investment horizons. S&P500 is used as the benchmark index. The two analyses are both also split into two sample periods, where the first one ranges from 2011 to 2016, and the second one from 2017 to 2022.

To be consistent with Morningstar's rating methodology, where they rank funds relatively compared to similar funds, we only compare the rankings and performance of funds within the same category. That is, all the funds in our sample are categorized in the peer group U.S. Large-Cap Blend.

The research question and methods in this paper are inspired by several previous studies, but our study differs essentially by expanding the quantitative analysis in the form of a complex panel data regression. In addition, we only compare funds within the same peer category group. To assess potential differences across categories, we also conduct the second analysis on an additional data set that consists of U.S. Small-Cap Growth funds.

The thesis is structured in the following way: the second section goes deeper into the background and reviews previous relevant literature. Section three describes the methodology behind the Morningstar rating system. In section four we explain how the necessary data was collected before we in section five present the methods used in our analysis, which is explained and discussed in section six. Finally, we present our conclusion in section seven.

2 Background

2.1 Mutual Funds Related Literature

Mutual funds and what characteristics drive their performance have been studied extensively in academic research, especially in the U.S. However, the results are often mixed and inconclusive. The efficient market hypothesis, developed by Fama et al. (1969), has had a significant influence on academic research. This hypothesis suggests that asset prices accurately incorporate all available information at any given time. When assuming semi-strong market efficiency, it is theoretically impossible to consistently outperform the market, unless one has access to private information, since asset prices immediately react to new public information.

Shiller (2003) introduces the concept of inefficient markets. Shiller argues that market prices can deviate from their fundamental values due to psychological biases and investor sentiment. He emphasizes the role of irrational behavior and the potential for mispricing in financial markets, creating opportunities for profit.

In contrast to the extremes of efficient and inefficient markets, the Grossman and Stiglitz (1980) framework presents the notion of efficiently inefficient markets. They suggest that while markets are generally efficient, there are instances where inefficiencies exist due to factors such as information asymmetry or transaction costs. In such cases, informed investors have the potential to exploit these inefficiencies and profit from them, resulting in temporarily inefficient markets.

Actively managed funds remain a popular investment tool, which suggests that many investors hold the belief that actively managed funds offer the potential to achieve risk-adjusted returns higher than the overall market. Additionally, throughout history, certain funds have consistently provided investors with abnormal positive or negative alphas. This leads to the question of whether the outperformance or underperformance of these funds persists over time.

Jensen (1968) concludes that when evaluating historical performance using his own measure called Jensen's alpha, there is no evidence to support the notion that past high performance indicates superiority over a randomly constructed portfolio. This perspective aligns with the commonly accepted statement that past performance is no guarantee of future performance.

Grinblatt and Titman (1992) present evidence supporting positive persistence in mutual fund performance, suggesting that past performance can be informative for investors considering investments in mutual funds. This finding is reinforced by Brown and Goetzmann (1995), who find evidence of relative performance persistence, indicating that historical information can be utilized by investors to outperform other funds. Previous research, such as Carhart (1997) found that management fees are inversely related to future fund performance, meaning that low (high) cost funds tend to perform better (worse) in the future. This finding can be attributed to the lack of manager skills and the inability of funds with high management fees to generate returns that justify the associated costs.

The contrasting views lead to an important question regarding the usefulness of ratings, such as those provided by Morningstar, in providing valuable information about the quality of funds.

2.2 Morningstar Ratings and Fund Performance

The impact of Morningstar's ratings on investor behavior is significant. The ratings have been shown to influence the flow of funds into and out of specific funds, which is a key measure of the aggregate impact of individual investors' allocation decisions. Many studies have established a correlation between investor fund flow and various measures of fund performance, including Morningstar ratings. Del Guercio and Tkac (2008) conducted an event study on more than 10,000 instances of Morningstar rating changes. Their findings revealed that Morningstar ratings have a significant impact on the investment decisions made by individual investors in mutual funds. Specifically, they discovered that the discrete change in the star rating itself influenced fund flows, rather than changes in the actual performance of the funds. They find significant evidence of positive abnormal flows from star upgrades, and negative abnormal flows from star downgrades.

A key question is whether this behavior is rational. It is very possible, even common, for a fund to perform well for a few years, receive large inflows, and then fail to live up to expectations. There are a few studies that analyze the relationship between Morningstar ratings and future performance. Blake and Morey (2000) conducted a study on the Morningstar rating as a predictor of mutual fund performance for U.S. domestic equity funds. They find that low ratings from morningstar generally indicate relatively poor future performance. While there is little evidence that Morningstar‘s highest-rated funds manage to outperform the next-to-highest and medium rated funds.

A newer study from Morey and Gottesman (2006) find that higher rated funds, for the most part, significantly outperform lower rated funds. As the present study was conducted in 2006, the available data and analysis period were relatively limited. However, given the current year of 2023, there exists a much larger period for conducting a more comprehensive analysis.

Recent master theses, such as Kirkeby (2020) and Røed and Høiden (2022), have also investigated the predictive ability of Morningstar ratings on mutual funds future performance. Kirkeby (2020) specifically examines the ability of Morningstar ratings to forecast future performance for mutual funds primarily investing in Norwegian equity. On the other hand, Røed and Høiden (2022) focuses on global funds accessible to Norwegian investors. Kirkeby (2020) analyzes two time periods, one before the financial crises and one after. For the first period, she finds that low ratings from Morningstar indicated relatively poor future performance. For the second period, she found the opposite, low ratings indicated relatively high future performance. Finally, she found little statistical evidence that Morningstar‘s highest rated funds managed to outperform the four-, three-, and two-star rated funds. Røed and Høiden (2022) found some, although not very significant, evidence that the highest rated funds managed to outperform the lowest rated funds, but with varying results across different time periods and different investment categories. One notable challenge in conducting research on the relationship between Morningstar ratings and future performance in the Norwegian market arises from the limited availability of mutual funds and their corresponding Morningstar ratings over an extended time period. Consequently, this scarcity of data poses limitations for these studies. Given that Morey and Gottesman (2006) research dates back 17 years, we find it appropriate to undergo a new analysis on the U.S. market. This is particularly warranted due to the availability of more recent data and Morningstar ratings assigned to a larger number of funds, in addition to several updates in the methodology behind the well-known rating system.

Both theses build upon the work of Blake and Morey (2000), with Røed and Høiden (2022) employing a similar methodology. They employ a cross-sectional regression analysis based on two representative time periods. Kirkeby (2020) introduces another approach, that consists of a buying winners vs losers strategy. We build further upon Blake and Morey (2000), Morey and Gottesman (2006) and Røed and Høiden (2022), and we expand their approach by utilizing panel data analysis. This way we consider each individual month as a potential start for each investment horizon, providing a more comprehensive examination. The methodology section of the thesis will delve deeper into this.

3 Morningstar Rating Methodology

The Morningstar Rating was initially introduced in 1985 with the aim of aiding investors and advisors in choosing a few funds from a wide range of options within different asset classes. However, as the investment industry has evolved, the focus has shifted towards viewing funds as part of a portfolio rather than standalone investments. This has highlighted the importance of having suitable alternatives within a specific rating group to construct diversified portfolios. As a result, Morningstar now assigns ratings by comparing funds within a specific Morningstar Category, rather than comparing them across all funds in a broader asset class (Morningstar, 2021).

3.1 Morningstar Rating System

Morningstar is recognized for its rating system, which employs a one-to-five-star scale to evaluate funds. The assigned stars indicate how funds have historically performed on a risk-adjusted basis, in comparison to other funds belonging to the same investment category, based on Morningstar's own *Morningstar Category™*. In this context, one-star funds are categorized as the worst performers, whereas five-star funds are regarded as the best performers. Following the 2016 revised methodology for Morningstar's rating system, the rating calculation involves a three-step process (Morningstar, 2021).

3.1.1 Category Peer Groups

Firstly, Morningstar arranges the funds it evaluates into peer groups, which are typically determined by the fund's investment focus, including the countries where it invests and the types of securities it primarily invests in. Additionally, for certain geographic regions, a fund's classification may also depend on its long-term or "normal" investment style profile, based on a minimum of three years of portfolio statistics (Morningstar, 2021).

Morningstar's reasoning behind this is that the relative star ratings of two funds should be affected more by potential manager skill differences, than by market circumstances or events out of the funds managers' control. Another principle emphasizes the importance of peer groups that align with the investment opportunities available to investors. The purpose of fund ratings is thus to assist investors in determining the relative quality of a fund compared to other funds in the same category. These ratings provide a measure of whether a fund is considered favorable or unfavorable within its peer group. In general, funds in the same category should be considered reasonable substitutes for the purposes of portfolio construction (Morningstar, 2021).

With this in mind, this study will be based on Morningstar's category groups when determining if there are future performance differences across different rating groups. For the main sample analyzed, funds in the Morningstar category "US large-cap blend" will be compared against each other. For the additional analysis, funds in the Morningstar category "US small-cap growth" will be compared against each other. This distances us from previous research, which mostly analyzes broader fund categories.

3.1.2 Morningstar Risk-Adjusted Return (MRAR)

The second step in the rating calculation is for Morningstar to calculate a fund's MRAR. The calculations for MRAR are done on a monthly basis first and then the results are annualized.

Firstly, Morningstar calculates a funds monthly total returns using the following formula (Morningstar, 2021, p. 9):

$$TR_t = \left[\frac{P_e}{P_b}\prod_{i=1}^{n}\left(1+\frac{D_i}{P_i}\right)\right]-1 \tag{3.1}$$

where:

TR_t = total return for the fund for month t
P_e = end of month net asset value (NAV) per share
P_b = beginning of month NAV per share
D_i = per share distribution at time i
P_i = reinvestment NAV per share at time i
n = number of distributions during the month

Morningstar calculates the total return of a fund on a monthly basis, expressed as a percentage. This calculation involves measuring the change in the fund's net asset value (NAV) over the month, including the reinvestment of all income and capital-gains distributions. The resulting value is then divided by the starting NAV. Reinvestments are made using the actual reinvestment NAV, and any daily payoffs are reinvested on a monthly basis. It's important to note that Morningstar's total return calculation does not adjust for sales charges like front-end loads, deferred loads, or redemption fees. This approach aims to provide a clear and unobscured view of a fund's performance. However, the total returns do take into account management fees, administrative fees, 12b-1 fees, and other costs that are deducted from the fund's assets (Morningstar, 2021).

The cumulative total return is then calculated by:

$$TR_c = \prod_{t=1}^{T} (1 + TR_t) - 1 \tag{3.2}$$

where:

TR_c = cumulative return for the fund
TR_t = total return for the fund for month t
T = number of months in the period

Morningstar then calculates the Morningstar return for each fund, which adjusts the total return for an appropriate risk-free rate (Morningstar, 2021, p. 11):

$$ER_t = \frac{1 + TR_t}{1 + RF_t} - 1 \tag{3.3}$$

where:

ER_t = the geometric excess return for the fund for month t
TR_t = the total return for the fund for month t
RF_t = the total return for the risk-free rate for month t

Lastly, Morningstar adjusts for risk to get to the funds MRAR, calculated by the following formula:

$$MRAR(\gamma) = \frac{1 + TR_t}{1 + RF_t} - 1 \tag{3.4}$$

where:

γ = parameter that describes the degree of risk aversion

For the calculation of MRAR, Morningstar uses expected utility theory to model how investors trade off return and risk. Morningstar's choice of the utility function relies on a few assumptions. Firstly, investors generally prefer higher expected returns than lower expected returns. Secondly, the general investor is risk-averse, and so the utility function penalizes risk. This means an investor will prefer a less volatile fund with the same expected return as a more volatile fund. Thirdly, an investor demands a larger risk premium for choosing a riskier portfolio, than a less risky one, which holds true when gamma is bigger than 0. When gamma is 0, the investor is indifferent between a riskless choice and a risky choice as long as the geometric average expected return is the same. When gamma is greater than 0, the risk premium must be larger than the difference between the arithmetic and geometric average returns (Morningstar, 2021).

By applying expected utility theory to risk-adjusted returns, Morningstar quantifies how investors perceive different distributions. The underlying idea is that investors prefer a distribution with high expected return and low risk compared to one with low expected return and high risk. As risk increases or expected return decreases, the trade-off between risk and return changes. If the level of risk becomes too high, investors are willing to accept a lower return to reduce the risk. For instance, an investor might be indifferent between a moderately risky fund generating a 12% return (as observed) and a risk-less fund generating an 8% return (as determined by the utility function). In this case, the investor is willing to sacrifice a 4% return to eliminate the risk. Morningstar analysts have determined that using a gamma value of 2 yields fund rankings aligned with the risk preferences of typical retail investors. Therefore, Morningstar applies $\gamma = 2$ in the calculation of its star ratings (Morningstar, 2021).

3.1.3 Fund Ranking

The last step is to rank all funds within a category, based on the risk adjusted return the fund has achieved. A higher MRAR corresponds to a higher rating. Generally, the top 10% of funds receive a 5-star rating, the subsequent 22.5% receive a 4-star rating, and the following 35%, 22.5%, and 10% are awarded 3, 2, and 1 star, respectively. Funds must have a minimum 3-year return history to be eligible for a rating (Morningstar, 2021).

Table 3.1: Morningstar Fund Rating

Star	Percent	Word Label
5	Top 10%	High
4	Next 22.5%	Above Average
3	Next 35%	Average
2	Next 22.5%	Below Average
1	Next 10%	Low

Morningstar also provides an overall rating for funds. They do this by first calculating ratings for three, five, and 10-year periods, depending on the available return history of a fund. The overall Morningstar Rating is then determined by a weighted average of the available time-period ratings as shown in Table 3.2 (Morningstar, 2021):

Table 3.2: Morningstar Fund Rating

Months of Total Returns	Overall (Weighted) Morningstar Rating
36-59	100% three-year rating
60-119	60% five-year and 40% three-year rating
120 or more	50% ten-year, 30% five-year and 20% three-year rating

3.2 Morningstar Rating System Critics

To effectively evaluate the future performance of different ratings, it is important to comprehend how the ratings are constructed. Morningstar categorizes funds into peer groups before assigning rankings, which means the system is designed to compare funds within specific markets, industries, or styles rather than identify good or bad funds across different categories. The rating becomes useful when comparing the past performance of similar funds. Therefore, neglecting to consider this distinction may lead to incomplete or misleading assessments of funds performance.

The Morningstar rating methodology has several limitations. The rating system is mostly dependent on past performance, which rarely persists over longer periods of time. The system does not account for outliers, such as abnormally good or bad years, consistent leadership, or other qualitative aspects of fund management.

The system does account for management fees, as the risk adjusted returns are after management fees. Considering Carhart (1997), this may play a role in spotting potential differences in future performance across ratings, as a 5-star fund is more likely to have low fees.

4 Data

This section provides an overview of the dataset utilized in the thesis, and covers the collection, description, and critics of the dataset. Firstly, it provides information on the data source for the mutual funds and the selection process. Secondly, it elaborates on the choice of the reference index, risk-free rate, and the various factor returns used as explanatory variables in the analysis.

4.1 Mutual Funds Data

The dataset on U.S. mutual funds used in this thesis is obtained through Morningstar Direct. As earlier mentioned Morningstar is most known for its independent analysis and ratings of stocks, mutual funds and other investments, but also offers extensive research resources, such as Morningstar Direct which is a platform that provides data on different types of investments worldwide.

The sample used contains data on 272 mutual funds in the period January 2011 - December 2022. In determining the starting period for our analysis, several factors were considered, including the number of funds, the availability of Morningstar ratings, and market conditions. It was crucial to avoid commencing the analysis immediately before or during the financial crisis characterized by extreme market volatility, as these factors could significantly influence the outcomes. After careful consideration, we selected 2011 as the starting point since the worst phase of the crisis had passed, and market conditions had stabilized to a more typical state. Furthermore, our choice of 2011 also takes into account the fact that Morey and Gottesman (2006) had already examined the time period from 2002 to 2005. Hence, we sought to investigate a new period that had not been extensively explored before. Additionally, starting from 2011 onwards provided us with a substantial number of rated funds, which is advantageous for our analysis.

By taking these factors into account, we aim to ensure a more robust and comprehensive analysis of the relationship between Morningstar ratings and fund performance during a period that offers meaningful insights and sufficient data availability. We thus ended with a period of 12 years, which we considered sufficient enough for this analysis. As the sample consists of time series data of multiple funds, it is considered what is called panel data. It includes monthly data for mutual fund returns and their respective Morningstar ratings.

Morningstar calculates the monthly returns based on the following formula:

$$R_t = \left[\frac{P_e}{P_b}\left(1 + \frac{D_i}{P_i}\right)\right] - 1 \tag{4.1}$$

where:

R_t = return for the fund for month t

P_e = end of month net asset value (NAV) per share

P_b = beginning of month NAV per share

D_i = per share distribution at time i

P_i = reinvestment NAV per share at time i

n = number of distributions during the month

4.1.1 Selection Criteria

It is essential to achieve consistency between the data sample and the research objectives of the thesis. Consequently, Morningstar's mutual fund database is utilized to apply multiple filters to the mutual fund sample. We started out with Morningstar's database of 347 294 global mutual funds, which includes a substantial number of mutual funds from all around the world. We thus present 7 criterias that need to be met for the inclusion of a fund in this analysis:

1. The fund has to be an open-end fund. Open-end funds means that the shares in the funds can be issued and redeemed at any time.

2. The fund must be an actively managed fund. The fund cannot be an index fund which tries to replicate an index, and these are thus removed from the sample. Comparing the performance of actively managed funds against index funds can lead to biased conclusions. Index funds are designed to closely track the performance of a benchmark, and their returns are expected to align closely with the index. In contrast, actively managed funds seek to outperform the benchmark through active investment decisions. Therefore, excluding index funds allows for a more focused

analysis of the performance of actively managed funds in relation to their investment objectives.

3. As this thesis focuses on mutual funds that invest primarily in U.S. equity, we screened our dataset to consist only of funds investing in the Morningstar group of "US Equity".
4. We only include funds in the Morningstar category U.S. Large-Cap Blend.
5. This thesis focuses on mutual funds that are offered for investors in the U.S. We thus only include funds that are available for sale in the U.S.
6. Some of the funds are listed multiple times in different classes which tend to have very similar returns. To avoid potential skewness in the results, we only include one share class for each mutual fund. Only the oldest share class is included.
7. In order to investigate whether there are differences in future performance across Morningstar ratings, the funds included in this thesis must be rated by Morningstar. For this, the fund must have at least a 3 year return history.
8. If the fund is terminated during the analysis period, it will still be included in the study if it meets the remaining criteria. We include these funds to avoid survivorship bias. Further discussion on this topic is provided in section 4.1.2.

After all adjustments are made, we are left with a dataset containing 272 mutual funds.

For the additional data set, we do the exact same adjustments, except from point 4, where the Morningstar category is "U.S. Small-Cap Growth".

4.1.2 Survivorship Bias

Survivorship bias occurs when we only consider the success stories or the winners in a given situation, neglecting the failures or non-successful outcomes. For example, in the business world, survivorship bias may occur when we study successful companies and try to identify common characteristics or strategies that led to their success.

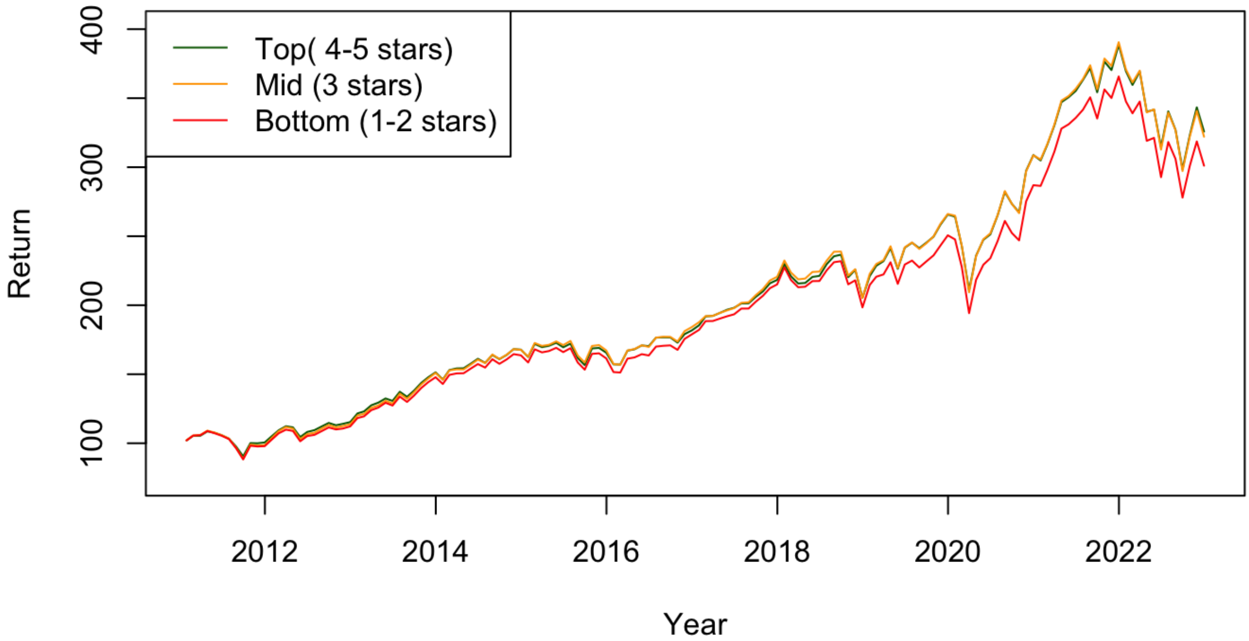
In a dataset of mutual funds, like ours, the main reasons for a fund to terminate is that it performs very poorly over time and is liquidated or that it is merged with another fund. citeelton1996survivor examines how survivorship bias has impacted prior studies on mutual funds. They find that failure to eliminate this type of bias can result in spurious conclusions when it comes to fund characteristics' effect on return.

To mitigate this bias, it's important to consider the complete data set, including both the successes and failures, to gain a more accurate understanding of the factors that contribute to outcomes in a given context. Hence, we also include the funds that are either liquidated or merged with another fund during the sample period.

4.1.3 Visualization of the Sample

To get a better look at the sample period, we divide the funds into groups of high rated (4-5 stars), medium rated (3 stars), and low rated funds (1-2 stars). The groups are divided based on which Morningstar rating the fund had at the start of the period. Figure 4.1 shows the cumulative return on $100 invested in three equal-weighted portfolios of each group. At first glance there does not seem to be a significant difference between high and medium rated funds in terms of raw cumulative returns. There is however a clear gap between these groups and the low rated funds, indicating that there may be differences between the higher/medium and lower rated funds. These plots are however based on raw returns, that do not take into consideration the risk taking of the funds. The same graph divided into the two time periods can be found in section A1 in the appendix.

Figure 4.1: Return on Benchmark Index



4.2 Benchmark Index

To assess the performance of a mutual fund properly, it is necessary to have a suitable benchmark index. There are multiple perspectives in the literature regarding this topic. One of the primary issues is whether to use the benchmark index fund stated in the fund's prospectus or to use an objectively selected benchmark index with the same risk profile as the funds.

This thesis will employ one single benchmark index as the market return for all the funds in the data sample. This is done to achieve comparability of the funds‘ performance, which is crucial for the purpose of this analysis. The chosen index is the S&P 500, which is also the most common by the different funds. The S&P 500, also known as the Standard & Poor's 500 Index, is a widely followed stock market index in the United States. It represents the performance of 500 largest publicly traded companies listed on stock exchanges in the U.S. It is also the benchmark that best suits the chosen sample, as we are investigating mutual funds in the Morningstar category, "U.S Large-Cap Blend" (S&P Dow Jones Indices, 2023).

Data for the monthly historical returns on the S&P 500, are downloaded from the Wharton Research Data Service database. WRDS is widely used by researchers, academics, and professionals in the fields of finance, economics, and business to analyze and extract insights from large-scale data sets (Wharton Research Data Services, 2023).

Figure 4.2: Return on Benchmark Index

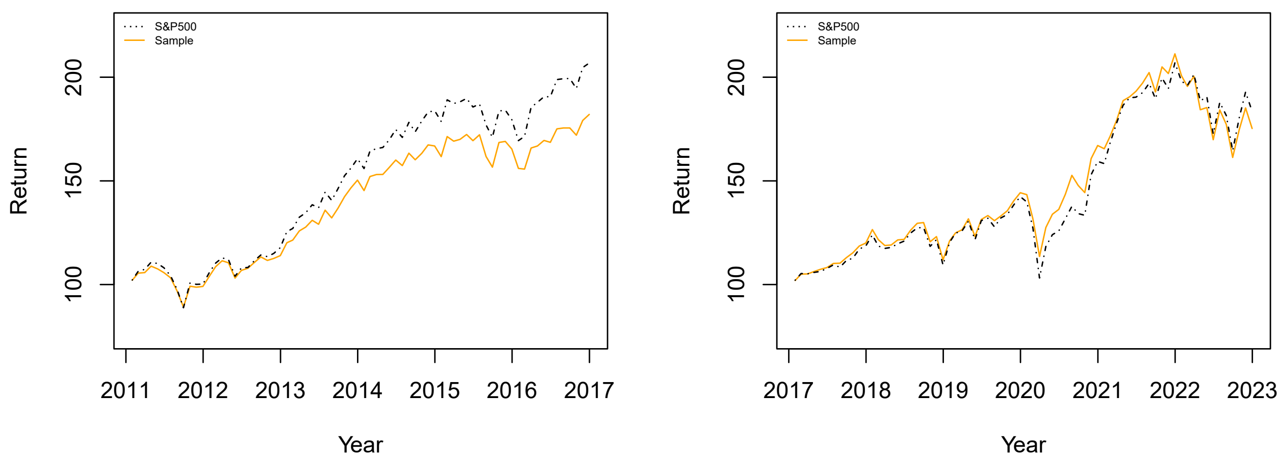


Figure 4.2 shows the cumulative average returns of our sample alongside the benchmark index (S&P 500).

The first period is characterized by relatively stable growth in the SP 500 and global markets, despite occasional negative events like the European debt crisis. The second period, covering the years 2017 to 2022, was characterized by heightened volatility in the financial markets. This was largely influenced by the COVID-19 pandemic in 2020, which had a significant impact on the global economy. Following the initial shock of the pandemic, there was a period of rapid growth fueled by monetary and fiscal stimulus measures, along with positive market sentiment surrounding the reopening of economies and vaccine developments. However, in 2022, concerns about rising inflation, fears of recession, and Russia-Ukraine conflict led to aggressive increases in interest rate and further volatility in the markets, which continues to be a concern until today.

We observe that the samples cumulative returns are below those of the S&P 500 for the first period (January 2011 - December 2016). This indicates that, on average, the returns of our sample have been lower than the market average represented by the SP 500. This suggests that the portfolio or strategy associated with the mutual funds selected, may not have been able to generate returns at the same level as the overall market. It could indicate factors such as suboptimal asset allocation, poor stock selection, or an inability to capture the full benefits of the market's upward trends. For the second period, the cumulative returns of the sample seem to align well with the the market index.

For the additional analysis done on the data set of "U.S. Small-Cap Growth" funds, we use the Russell 2000 Growth Index, which is appropriate for small-cap growth funds.

4.3 Risk-free Rate

The risk-free rate is a theoretical term which refers to the rate of return on an investment with no risk. It is the rate of return an investor would expect to receive from an investment with zero risk, typically a government-issued bond or treasury bill. It represents the minimum rate of return an investor should expect to earn when taking on additional risk (Ang, 2014). The performance of the funds is thus evaluated in excess of the risk-free rate.

The chosen risk free rate for this analysis is the one-month treasury bill rate, retrieved along with the factor returns from the Kenneth French Data Library (French, 2023).

4.4 Factor Returns

To calculate the different funds' risk adjusted performance, we compare their returns to those implied by the different factor models (further explained in section 4). The monthly factor returns are retrieved from the Kenneth French Data Library (French, 2023).

5 Methodology

This section provides the methodology used to evaluate differences in future performance across the different Morningstar-ratings in the U.S. market, measured by risk-adjusted returns.

Firstly, we present the different performance metrics used in the analysis. These are: CAPM, Fama-French 3-factor model, and Carhart 4-factor model. Secondly, we present the two main analyses conducted, which are largely based on the methodology used by (Blake and Morey, 2000) and (Morey and Gottesman, 2006), who conducted similar studies on mutual funds in the American market.

The first part of the analysis evaluates whether there are significant performance differences across funds with different Morningstar ratings. Additionally, we add lags for the returns of 1 and 2 months, to examine the Morningstar ratings predictive ability in the very short run. The second part of the analysis differs from the first by the way the performance metrics are calculated. The first part looks at monthly performance, while this part deals with performance over short (6 months), medium (1 year), and long term (3 years) investment horizons. Both methods are conducted with a fixed effects panel data regression. To strengthen the robustness of the analysis, the methods are divided in two time periods, and also conducted for a different Morningstar category.

5.1 Performance Metrics

We use three metrics for the evaluation of the funds performance. Primarily to test Morningstar ratings ability to predict future performance, and to observe if there is consistency across the different results.

Extensive literature that analyzes and tries to explain what drives asset prices exists. Various factor models have been developed to evaluate risk-adjusted performance. It can be used to explain either an individual security or a portfolio of securities. For fund performance evaluation, the fund's actual excess return is compared to the excess return implied by various factor models, which is often referred to as the expected rate of return (Ang, 2014). A factor model follows the following formula:

$$r_i = \beta_1 F1 + \beta_2 F2 + ... + \beta_N FN + \alpha_i \tag{5.1}$$

The performance is measured by the alpha given by the different factor models. The alpha is calculated as the excess actual rate of return of the portfolio for a given period, over the expected rate of return given by the model.

In this thesis we use different measures of alpha. For the first analysis, we use monthly alphas. For the second analysis, we use cumulative alphas for different investment periods. We also employ estimates of alpha from three different factor models. We use multiple models to be consistent with previous fund performance literature, as well as to increase robustness due to different models giving different results. The three models will be presented and explained in the following sub-sections.

The different funds exposure to the different factors is measured by the beta. In this analysis the beta is calculated for the different models over the whole sample period.

The three factor models employed in this thesis are the CAPM, Fama-French 3-factor model, and Carhart 4-factor model. These models are widely recognized and commonly used in the analysis of mutual fund performance.

5.1.1 Capital Asset Pricing Model (CAPM)

In 1968 Michael Jensen introduced one of the most used performance indexes in finance today. Jensen's alpha measures the excess return of a portfolio compared to its expected return, based on CAPM. The CAPM assumes that the expected return of an asset or portfolio is a function of its systematic risk, measured by its beta coefficient. Systematic risk is defined by risk that cannot be diversified and we typically measure this by the asset's correlation to the market portfolio, i.e. the asset's beta coefficient (Ang, 2014).

$$r_{i,t} = r_{f,t} + \beta_i(r_{m,t} - r_{f,t}) + \alpha_{i,t} \tag{5.2}$$

Where, r_i, is the monthly return achieved by the mutual fund. β represents the funds exposure to the market factor. $r_m - r_f$ is the markets return in excess of the risk free rate.

Table 5.1. shows the top 10 best and worst performers in our sample based on Jensen's

Alpha. The top performers achieved a monthly average alpha of 0.262% and the worst performers an average alpha of -0.340%. This indicates there are funds that outperformed the market over the sample period. The table also reports what rating the fund started out with. Among the top performers, 9 funds initially had ratings of 5-, 4-, or 3-stars. However, it is interesting to note that the best-performing fund in our sample had an initial rating of only 1-star, which is surprising. On the other hand, the worst performers were predominantly rated as 3-stars, with some observations of 2-stars and even 4-stars. This suggests that the ratings do not always accurately indicate overperformance or underperformance.

There are several weaknesses associated with the CAPM model. Firstly, it assumes a linear relationship between the portfolio's return and the market return, which may not always hold true in real-world scenarios. Secondly CAPM does not consider other risk factors that could impact the portfolio's performance.

5.1.2 Fama-French 3-factor

The Fama-French 3-factor model was developed in the early 1990s by Eugene Fama and Kenneth French (Fama and French, 1992). This model is an extension of the CAPM by incorporating additional factors that capture the systematic risk associated with specific characteristics of a stock or a portfolio. The model explains returns by considering three different factors: market risk, size and value. Excess returns are given by combining these factors in a linear model as follows:

$$r_{i,t} = r_{f,t} + \beta_{1,i}(r_{m,t} - r_{f,t}) + \beta_{2,i}SMB_t + \beta_{3,i}HML_t + \alpha_{i,t} \tag{5.3}$$

SMB: This is the size factor, and is measured by the difference between the returns in small and big companies (returns on small companies minus returns on big companies).

HML: This is the value factor, and we measure this by calculating the difference between returns on value and growth companies. Value companies are defined by high book-to-market and growth companies have low book-to-market ratio.

Table 5.1: Best and worst Performers 201101-202212

(a) Top 10 funds

Fund	Avg. Alpha	Avg. Beta	Avg. Rating	Rating Jan 2011	Rating Dec 2022
Pear Tree Quality Ordinary	0.738	0.263	3.056	1	5
GMO Quality IV	0.305	0.739	4.611	5	5
Jensen Quality Growth J	0.218	0.774	3.764	5	5
T. Rowe Price Dividend Growth	0.213	0.778	4.347	4	5
Hartford Stock HLS IA	0.212	0.755	3.375	3	5
Parnassus Core Equity Investor	0.200	0.769	4.736	5	5
Hartford Core Equity Y	0.195	0.826	4.465	3	4
Vanguard PRIMECAP Inv	0.185	0.780	4.306	5	4
Hartford Disciplined Equity HLS IA	0.185	0.838	4.417	3	4
American Funds Washington Mutual A	0.164	0.865	3.528	3	4
Average	0.262	0.739	4.061	3.6	4.6

(b) Worst 10 funds

Fund	Avg. Alpha	Avg. Beta	Avg. Rating	Rating Jan 2011	Rating Dec 2022
Selected American Shares S	−0.240	0.964	2.243	3	1
DWS Equity Sector Str Fund Class S	−0.257	0.682	2.292	2	1
Weitz Partners Value Investor	−0.274	0.899	2.389	4	1
Victory Special Value A	−0.284	0.967	1.639	3	2
Permanent Portfolio Aggr Growth I	−0.286	1.171	1.757	3	1
Fidelity Leveraged Company Stock	−0.320	1.158	2.326	3	1
Fidelity Advisor Leveraged Co StkI	−0.351	1.196	2.542	3	1
First Eagle Rising Dividend Fund C	−0.425	0.927	2.639	4	1
Bridgeway Aggressive Investors 1	−0.437	1.130	1.819	2	1
IMS Capital Value	−0.524	1.026	1.257	3	1
Average	−0.340	1.012	2.100	3	1.1

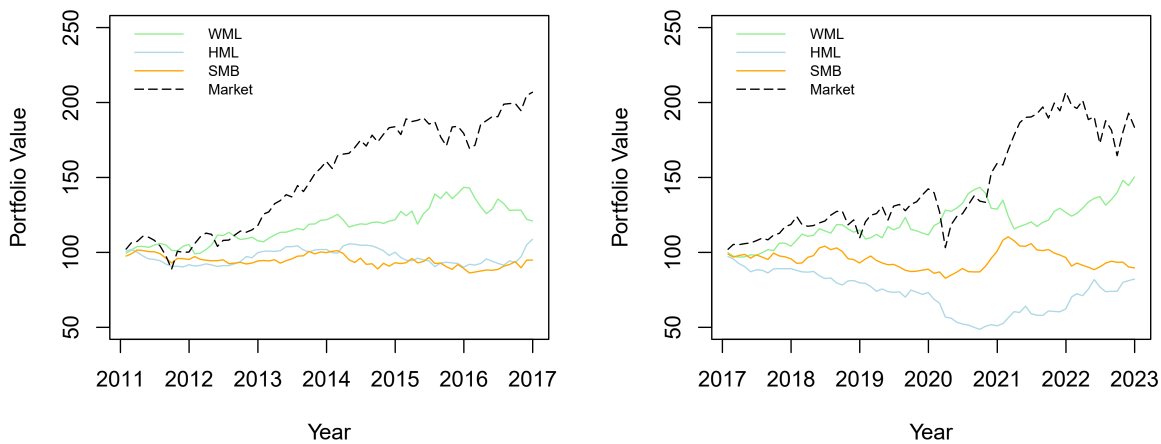
5.1.3 Carhart 4-factor

The Carhart 4-factor model is an expansion of the Fama-French 3-factor model, and includes one more factor, momentum. The momentum factor captures the tendency that stocks that have previously performed well, will also outperform in the future, and stocks that have previously performed poorly, will underperform in the future. This momentum effect suggests that there is some persistence in stock price trends that can impact future returns (Carhart, 1997).

$$r_{i,t} = r_{f,t} + \beta_{1,i}(r_{m,t} - r_{f,t}) + \beta_{2,i}SMB_t + \beta_{3,i}HML_t + \beta_{4,i}WML_t + \alpha_{i,t} \tag{5.4}$$

WML: This is the momentum factor. It is measured by subtracting the equal weighted average of the lowest performing firms from the equal weighed average of the highest performing firms, lagged one month.

Figure 5.1: Factor returns



5.2 Fixed Effects Panel Data Regressions

A panel data regression, is a statistical analysis method used to analyze data collected over time from multiple individuals, entities, or subjects. It combines aspects of both cross-sectional and time series data, allowing for the examination of both individual and time-related effects (Stock and Watson, 2018). Previous research examining the predictive power of Morningstar ratings has commonly used cross-sectional dummy variable regressions to estimate the model (Blake and Morey, 2000)(Morey and Gottesman, 2006). However, cross-sectional data only captures variables at a specific moment in time. In a panel data regression, observations are made on the same entities (such as individuals, firms, or countries) over multiple time periods. This type of data structure provides more information and allows for the analysis of both within-entity and across-entity variations, thereby offering several advantages over cross-sectional regression. In the context of our analysis, the model accounts for the fact that individuals have the opportunity to invest in different mutual funds at various time periods.

We can perform our panel data regression by using either a fixed effects (FE) or random effects (RE) model. Fixed effects models account for individual-specific factors that are constant over time, while random effects models assume that these factors are uncorrelated with the independent variables and estimate the average effects across individuals. Choosing between these models typically relies on intuition, as well as statistical tests such as the Hausman test. (Stock and Watson, 2018)

It is reasonable to assume that there are time-invariant fixed effects in the error-term that are correlated with the rating. This can for example be investment strategy, fees or fund manager. By using fixed effects, we can control for these types of time-invariant factors that may be present within each fund. This intuition is supported by the results of the Hausman test, as shown in section A2. Specifically, the null hypothesis of the Hausman test is that the random effects assumption is consistent and efficient, indicating zero correlation between unique errors and explanatory variables. If the test fails to reject the null, RE models should be used; if it rejects the null, FE models should be used (Stock and Watson, 2018). The results of the Hausman test indicate that the null hypothesis can be rejected for all of our models. Therefore, FE models are preferred over RE models in all cases.

Both the analyses used in our study involve the use of panel data regression models that incorporate dummy variables for each Morningstar rating. The aim of this approach is to investigate whether there exists statistically significant differences in performance levels across various Morningstar rating categories. We will use two approaches, explained in the following sub-sections.

When trying to establish a causal relationship between a specific fund characteristic and mutual performance, one should consider including control variables in the analysis. In previous research on mutual funds, control variables have been commonly used to account for the influence of other factors that may impact the relationship between the variables of interest. Additionally, control variables can help mitigate omitted variable bias by capturing relevant factors that may impact the relationship being studied. This reduces the risk of obtaining biased and inaccurate results due to the omission of important variables. Including relevant factors such as fund characteristics, investment strategy, expense ratios, fund size and turnover ratios are considered factors that could influence the performance of mutual funds. By including these as control variables, you can better evaluate whether the variable of interest has a significant impact on fund performance. (Stock and Watson, 2018)

Despite these advantages, in our analysis, we have chosen not to include control variables. The purpose of our analysis is not to establish a causal relationship between Morningstar ratings and future performance, but rather to examine potential differences in future performance based on the construction of the ratings themselves. We are specifically interested in evaluating whether there are variations in risk-adjusted returns among different Morningstar ratings, independent of other factors that may influence performance. Our focus is on the impact of the rating system itself, assuming that investors choose a 5-star rated fund based on its rating, regardless of other factors. By excluding control variables, we can better isolate the specific effects of Morningstar ratings and assess their relative impact on risk-adjusted returns. This approach allows us to focus on the fundamental question of whether the rating system leads to significantly different performance outcomes across the different Morningstar ratings.

5.2.1 Regression on Monthly Performance with Lagged Dummies

The first analytical approach involves the use of monthly alphas as the dependent variable, and dummy variables for each Morningstar rating. The aim of this approach is to investigate whether there exists statistically significant differences in performance levels across the different ratings. In order to provide a more nuanced assessment of the findings, this analysis is further extended by introducing lagged dummy variables for one and two months. This expanded approach aims to assess whether the higher-rated funds consistently exhibit superior performance in the short term compared to their lower-rated counterparts.

The fixed effects model is estimated as shown in the following equation:

$$\alpha_{i,t} = \beta_1 D1_{i,t-l} + \beta_2 D2_{i,t-l} + \beta_3 D3_{i,t-l} + \beta_4 D4_{i,t-l} + u_i + \epsilon_i \tag{5.5}$$

We do the regression with different lags for the dummies, where $l = 0, 1, 2$ is the monthly lag period length. u_i is the time-invariant omitted variable and ϵ_i, t is the classic error term in the model.

The selection of lag periods is based on the rationale that, given the usage of monthly alphas, extended lags may not yield informative outcomes. For instance, employing a 12-month lag would solely consider the alpha attained in a year's time by a 3-star rated fund, disregarding the multitude of factors that may have influenced the alpha, including the presence of anomalous, favorable or unfavorable months for the fund in question. Instead, the lags selected intend to exhibit the fund's performance in the immediate months following, given the fund's Morningstar rating at time $t - l$. While longer lags could be utilized, this thesis has opted for an alternative approach in the subsequent segment of the analysis, which will be elaborated upon in the following sub-section.

The 5-star rated funds serve as the reference group in the regression. On the other hand, the coefficients $beta_4$ through $beta_1$ will represent the difference between each of the rest of the ranking groups and the reference group.The reason for using the 5-star rated funds as the reference group is that since this is the highest ranked group, it is appropriate to compare with the other groups. If the 5-star rated funds outperform the other groups, and if the star ratings accurately forecast performance, we should observe increasingly negative and significant coefficients when moving from $beta_4$ through $beta_1$.

5.2.2 Regression on Performance over Different Time Horizons

Blake and Morey (2000) performed a cross-sectional regression on a sample of funds to determine the Morningstar ratings' ability to forecast future performance over two distinct time periods. In contrast to their approach, our analysis introduces a panel data regression and employs an adjusted version of the performance metrics used in our previous analysis to ensure consistency. Specifically, we utilize alphas obtained through the CAPM, FF 3-factor model, and Carhart 4-factor model, just like in the first analysis. However, the present analysis differs from the previous, in its method of alpha calculation, employing short-term (6 months), medium (12 months), and long-term (36 months) time horizons. This is done to evaluate the Morningstar rating system's ability to predict performance over different investment periods.

In contrast to our first analysis, which only examined performance over discrete periods, this analysis aims to replicate more realistic investment horizons, as mutual funds are generally considered a long-term investment option, rather than a short-term investment tool for investors seeking quick liquidity. To examine the relationship between Morningstar ratings and mutual fund performance, we regress fund performance over different time periods against the Morningstar rating assigned at the beginning of the relevant period, using panel data regression techniques. Thus, we consider different time periods for purchase and account for the varying performance and rating that a fund may achieve over time. In contrast to Blake and Morey (2000) and Morey and Gottesman (2006), which limited its analysis to only the two time periods considered, our approach evaluates the impact of Morningstar ratings on mutual fund performance over an array of different time periods. As such, we are able to evaluate the consistency of the predictive ability (if any) of the Morningstar rating system across various time horizons, thereby enhancing the rigor of our findings.

$$\alpha_{i,t}^{n} = \beta_1 D1_{i,t} + \beta_2 D2_{i,t} + \beta_3 D3_{i,t} + \beta_4 D4_{i,t} + \epsilon_i \tag{5.6}$$

In this model, $\alpha_{i,t}^{n}$ is fund i's n-month alpha at time t. We estimate this for each of the factor models (CAPM, Fama-French 3-factor and Carhart 4-factor) and for 6, 12 and 36 month periods. $D1$, $D2$, $D3$ and $D4$ are the dummy variables that represent the Morningstar rating of the fund. For example $D4 = 1$ indicates a 4-star rating.

5.2.3 Time Periods

The analyses are initially conducted on the full sample period, and then split into two time periods. This is done to strengthen the robustness of the analysis, and to check whether there is consistency in the results across two different time periods. The first period is from January 2011 to December 2016, and the second from January 2017 to December 2022.

5.2.4 Investment Category

We also run the second analysis on an additional dataset, specifically a different Morningstar category. The Morningstar category is “U.S. Small-Cap Growth”, which differs from the original sample as a more volatile and uncertain investment style. The reason for running the regression on a different investment style is to check whether there is consistency across different Morningstar categories.

5.2.5 Missing Observations

Our panel data is an unbalanced data set. This means that there are missing observations for some points in time. The choice of how to handle missing values depends on the specific characteristics of the data set, the reason why they are missing, and the assumptions of the analysis. There are three main reasons for missing values in our data set.

Firstly, some funds may have missing returns prior to a certain date because they did not exist before that date. Secondly, there might be missing Morningstar ratings for certain funds before a specific date. This is because Morningstar requires at least a three-year return history for a fund to be rated. In both cases, we remove the missing values for the following reasons:

1. Investors cannot invest in a mutual fund that has not yet been established.
2. Since our analysis focuses on differences between Morningstar ratings, funds that have not been rated do not contribute to the purpose of our analysis.

Lastly, there are missing observations of mutual fund returns after a certain date. This is mainly due to liquidation of the fund, or because the fund is merged with another fund from the same company. In the context of our analysis, we assume that missing observations indicate fund termination. In the United States, mutual funds are generally required to provide notice to shareholders at least 30 days before liquidation, although this is not always the case. This allows investors to make informed decisions regarding their investments and plan accordingly. Based on this assumption, we consider that an investor can select a mutual fund with return observations for the next two months. If a mutual fund has only one observation remaining at a given point in time, it is excluded from the analysis. In the panel data regression for investment horizon analysis, we only include funds with at least two monthly observations of returns for a specific month. For example, even if the investment horizon is one year, the representative alpha reflects the alpha achieved in the next two months of available data. This assumption represents an individual investing in a particular fund that is subsequently terminated. After this, the investor is free to invest in any other available fund, at this point in time.

The assumption that missing values in the data set are solely due to fund termination and not mergers with other funds of the same company may not be entirely accurate. In some cases, mutual funds may merge with another fund for reasons other than underperformance, such as optimizing product offerings or achieving cost efficiencies. Therefore, it is reasonable to consider an alternative approach to handling missing values.

One alternative is to investigate each individual fund to determine the specific reason for the missing values. This can involve conducting online research or reaching out to the fund directly for information. However, this approach can be time-consuming and may not be feasible due to the size of the data set and time constraints. Considering the limitations and practical constraints, we consider it reasonable to opt for the first choice of assuming missing values are primarily due to fund termination. While this assumption may not capture all cases of fund mergers accurately, it provides a practical and efficient way to handle missing data in the analysis.

6 Analysis

In this section, we will describe and discuss the empirical analysis and the results from the methods described in the previous section. First, we will present the results from the panel data regression that uses monthly performance. After that, we will present the results from the panel data regression with performance measured over different time horizons. Both analyses are based on the alphas obtained through the different models in section 5.1. We also conduct analyses for two different time periods, and an additional Morningstar category, to assess the robustness of the results.

6.1 Monthly Performance with Lags

In this subsection, we present the outcomes of the random effects panel data regression conducted on the monthly performance data. The regression analysis was performed using lagged dummy variables, as explained in section 5.2.1, with lags ranging from zero to two months. The regression models investigate whether there are performance differences across the Morningstar ratings, in the month of the rating, and the two months following.

6.1.1 Results for the Full period: 01/2011 - 12/2022

Table 6.1 presents the regression results covering the entire period from 2011 to 2022. When examining the coefficients without any lags, we find significant results at the 1% level for all coefficients, except for 4-star ratings in the CAPM model. This indicates that there are performance differences based on Morningstar ratings, and Morningstar appears to accurately assess the funds performance when comparing them with alternative performance metrics. However, when introducing lags of one and two months, we see notable differences. With a one-month lag, 4 out of the 12 coefficients show negative and significant results at the 1% level. With a two-month lag, 6 out of the 12 coefficients exhibit significant results at the 1% level. Focusing on the 4-star rated funds, we observe a mix of positive and negative coefficients, although none reach significance at the 1% level. Similar patterns are observed for the 3-star rated funds, with one coefficient showing negative and significant results at the 1% level for the two-month lag in the Carhart model. These findings suggest that 5-star funds do not consistently outperform 4- and

3-star rated funds within the two months following the rating period.

For the 1- and 2-star rated funds, we find a higher number of significant results. All β_1 coefficients are negative and significant, indicating that 5-star rated funds consistently outperform 1-star rated funds within the same category. Similar trends can be observed, to some extent, for the 2-star rated funds, with all coefficients being negative and 3 out of 6 coefficients reaching significance at the 1% level.

The same analysis is split into the two different time periods and the results can be found in appendix A3. The results show some of the same patterns, but with weaker results.

6.1.2 Main results

Morningstar's fund rating system is based on utility theory. Their ranking of funds focuses on three key factors: strong historical returns, low volatility as a measure of risk (with a greater emphasis on downward volatility). As the returns are calculated after management fees are accounted for, high management fees are also penalized. When running the regressions without incorporating monthly lags, the results indicate persistent performance differences across the Morningstar ratings, as measured by the CAPM, Fama-French 3-factor, and Carhart 4-factor model. Each rating appears to outperform the others in terms of risk-adjusted returns, with the findings robust across different time periods. There is consistency in the Morningstar ratings, and the alternative performance metrics used in the analysis. These outcomes can be explained by Morningstar's practice of dividing funds into peer groups based on their market of operation and investment style plays. By grouping funds with similar market conditions and investment strategies, the effects of market capitalization (CAPM) and investment style factors, such as investing in large/small stocks or growth/value funds (Fama-French), are indirectly accounted for. Consequently, the observed performance differences can be attributed to other factors beyond market conditions and investment styles.

Table 6.1: Predictive Regression for the Full Sample Period: 201101-202212

Table 6.1 presents the fixed effects panel data regression for the full sample period from January 2011 to December 2022. The coefficients are reported in percentages, and represents the difference in monthly alpha compared to 5-star funds. I.e. 0.04 means 0.04%. Standard errors for each estimate are reported in parenthesis.

	No lag			Lag = 1 month			Lag = 2 months		
	CAPM	FF	Carhart	CAPM	FF	Carhart	CAPM	FF	Carhart
β_4 (4-star)	-0.057^{*}	-0.074^{***}	-0.086^{***}	0.050	0.014	-0.002	-0.029	-0.053^{*}	-0.059^{**}
	(0.032)	(0.027)	(0.027)	(0.032)	(0.027)	(0.027)	(0.032)	(0.027)	(0.027)
β_3 (3-star)	-0.155^{***}	-0.164^{***}	-0.172^{***}	0.010	-0.016	-0.035	-0.035	-0.059^{**}	-0.069^{***}
	(0.031)	(0.026)	(0.026)	(0.031)	(0.027)	(0.026)	(0.031)	(0.026)	(0.026)
β_2 (2-star)	-0.309^{***}	-0.313^{***}	-0.317^{***}	-0.001	-0.055^{*}	-0.073^{***}	-0.062^{*}	-0.104^{***}	-0.110^{***}
	(0.033)	(0.028)	(0.027)	(0.034)	(0.029)	(0.028)	(0.034)	(0.028)	(0.027)
β_1 (1-star)	-0.541^{***}	-0.517^{***}	-0.503^{***}	-0.119^{***}	-0.147^{***}	-0.161^{***}	-0.197^{***}	-0.230^{***}	-0.231^{***}
	(0.043)	(0.036)	(0.035)	(0.044)	(0.037)	(0.036)	(0.044)	(0.037)	(0.035)
Observations	27,987	27,987	27,987	27,711	27,711	27,711	27,437	27,437	27,437
R^2	0.009	0.012	0.012	0.001	0.001	0.001	0.001	0.002	0.002

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

6.2 Monthly Performance over Different Time Horizons

In this section, we present the outcomes of the fixed effects panel data regression conducted on the different performance metrics across different investment horizons of 6-, 12- and 36 months. The regression model aids to investigate whether there are performance differences across the Morningstar ratings, for different investment horizons. The regression model takes into account both the length of the investment and the timing of the investment. All results are presented as the difference in monthly alpha, achieved in the investment horizon. To increase the robustness of the findings, the analysis is also divided into time periods, to assess potential variations in the results.

6.2.1 Full period: 01/2011 - 12/2022

Table 6.2 presents the regression results for the entire period, covering 2011 to 2022. The results, presented as monthly values and percentages, indicate that no significant coefficients are observed for any of the performance metrics at the 1% level over a 6-month investment period. Notably, there are variations in the sign of the coefficients. Specifically, all coefficients for the CAPM model are negative, suggesting that 5-star funds outperform other funds when adjusting for market risk. Conversely, all coefficients for the Carhart model are positive, indicating underperformance of 5-star rated funds relative to their peers. However, these results are not statistically significant, except for the 3-star rated funds which exhibit a negative coefficient of -0.033% at the 5% level. This implies that 5-star funds generally fail to outperform funds in other rating groups over a 6-month investment period, with varying outcomes depending on the performance metric used.

Extending the investment horizon to one year yields slightly different results. Here, we observe 4 out of 12 significant coefficients at the 1% level and 1 at the 5% level. All three coefficients for the 1-star rated funds are negative and significant at the 1% level, indicating consistent outperformance of 5-star funds over 1-year periods across all three performance metrics. A similar pattern is observed for the 2-star rated funds, where all coefficients are negative, with one significant at the 1% level and one at the 5% level. However, the results for 3- and 4-star rated funds are less conclusive. Although most coefficients are negative, one coefficient for the 4-star rated funds in the Carhart model is positive. As none of these coefficients are statistically significant, it suggests that

5-star funds also fail to consistently outperform 4- and 3-star rated funds over a one year investment horizon.

When extending the investment horizon further to three years, we find similar patterns in the results as for the one year horizon. 5-star funds appear to outperform both 1- and 2-star funds across all three performance metrics over a three-year period, with stronger evidence indicated by more negative coefficients for 1- and 2-star funds, all of which are significant at the 1% level. Notably, 1-star rated funds consistently exhibit the most negative coefficients, indicating that low Morningstar ratings exhibit poor future performance. For 3- and 4-star rated funds, the results are still inconclusive, although one significant coefficient is observed. All coefficients are negative for these funds, and for the 4-star rated funds, the coefficient for Jensen's alpha is negative and significant at the 5% level, indicating that 5-star funds outperform 4-star rated funds when accounting for market factors. On the other hand, no result reaches significance at the 5% level for the 3-star funds, suggesting that 5-star funds fail to outperform 3-star rated funds when considering all three performance metrics.

Table 6.2: Predictive Regression for the Full Sample Period 201101-202212

Table 6.2 presents the fixed effects panel data regression for the full sample period from January 2011 to December 2022. The coefficients are reported in percentages, and represents the difference in the monthly average alpha compared to 5-star funds. I.e. 0.04 means 0.04%. Standard errors for each estimate are reported in parenthesis.

	6 months			12 months			3 years		
	CAPM	FF	Carhart	CAPM	FF	Carhart	CAPM	FF	Carhart
β_4 (4-star)	−0.020	−0.0003	0.010	−0.020	−0.010	0.002	−0.023**	−0.017*	−0.015*
	(0.014)	(0.012)	(0.011)	(0.014)	(0.011)	(0.010)	(0.010)	(0.009)	(0.009)
β_3 (3-star)	−0.019	0.0001	0.012	−0.025*	−0.023*	−0.015	−0.017*	−0.015*	−0.009
	(0.015)	(0.012)	(0.012)	(0.015)	(0.012)	(0.012)	(0.010)	(0.009)	(0.009)
β_2 (2-star)	−0.033**	−0.0004	0.009	−0.086***	−0.029**	−0.024*	−0.119***	−0.115***	−0.114***
	(0.016)	(0.013)	(0.013)	(0.016)	(0.013)	(0.013)	(0.015)	(0.012)	(0.012)
β_1 (1-star)	−0.035	0.004	0.005	−0.165***	−0.115***	−0.108***	−0.194***	−0.192***	−0.191***
	(0.022)	(0.018)	(0.017)	(0.021)	(0.018)	(0.017)	(0.019)	(0.018)	(0.016)
Observations	26,650	26,650	26,650	25,136	25,136	25,136	19,618	19,618	19,618
R^2	0.001	0.004	0.004	0.006	0.006	0.007	0.037	0.050	0.051

Note: *p<0.1; **p<0.05; ***p<0.01

6.2.2 First Period: 01/2011 - 12/2016

Table 6.3 presents the regression results for the first period analyzed, spanning from 2011 to 2017. As mentioned in the data section of the paper, this period was characterized by relatively stable growth in the S&P 500 and global markets, despite occasional negative events like the European debt crisis.

The results from the first period exhibit similar patterns to the overall period but with stronger outcomes. We observe a notable increase in significant results compared to the whole period analysis. Across all investment periods, all coefficients are negative.

For the 6-month investment period, we now observe 6 out of 12 significant coefficients, a substantial improvement compared to none in the whole period. In the CAPM model, all coefficients for different ratings are significant at the 1% level. The coefficients for 4-, 3-, and 2-star rated funds are similar, while the 1-star fund stands out with significantly larger coefficients. This indicates that, measured by Jensen's alpha, 5-star funds outperform their peers, with 1-star rated funds performing the worst. Similar patterns are observed for the 3-factor model and the Carhart model, although with fewer or no significant results except for the 1-star funds. 4-, 3-, and 2-star funds exhibit negative coefficients significant at the 5% level in the 3-factor model, while their coefficients are negative but insignificant in the Carhart model. All coefficients for 1-star funds are negative and significant. This suggests that 5-star funds outperformed 1-star rated funds across the different metrics for a 6-month investment horizon. Although 5-star funds outperformed the other rating groups in the CAPM model, they did not significantly outperform them when accounting for additional risk factors.

Consistent with the whole period analysis, expanding the investment horizon to one and three years yields more significant results. For the one-year investment horizon, we observe negative and significant coefficients for 8 out of 12 coefficients at the 1% level. These results also exhibit a decreasing trend in size. This suggests that 5-star rated funds outperformed other ratings with a one-year investment horizon in both the CAPM and 3-factor model. However, in the Carhart model, 5-star rated funds failed to outperform 4- and 3-star rated funds. Among the rating groups, 1-star funds performed the worst, showing an even larger difference compared to the other groups than what was observed for the 6-month horizon. Additionally, 2-star funds also underperformed compared to 3- and 4-star rated funds.

For the 3-year investment horizon, although all coefficients for 2- and 1-star rated funds remain negative and significant, the difference from the reference group is smaller. This indicates that while 5-star rated funds still outperformed lower rated funds, the margin of outperformance decreased. In the CAPM and 3-factor model, 5-star funds also outperformed 4-star rated funds, but there was no evidence of outperformance in the Carhart model. Regarding 3-star rated funds, there is only weak or no evidence of outperformance across the three different performance metrics.

Table 6.3: Predictive Regression for the First Period 201101-201612

Table 6.3 presents the fixed effects panel data regression for the first period from January 2011 to December 2016. The coefficients are reported in percentages, and represents the difference in the monthly average alpha compared to 5-star funds. I.e. 0.04 means 0.04%. Standard errors for each estimate are reported in parenthesis.

		6 months			12 months			3 years	
	CAPM	FF	Carhart	CAPM	FF	Carhart	CAPM	FF	Carhart
β_4 (4-star)	-0.065^{***}	-0.037^{**}	-0.019	-0.068^{***}	-0.038^{**}	-0.026	-0.035^{**}	-0.033^{**}	-0.027^{*}
	(0.017)	(0.016)	(0.016)	(0.017)	(0.016)	(0.016)	(0.015)	(0.015)	(0.015)
β_3 (3-star)	-0.067^{***}	-0.037^{**}	-0.018	-0.061^{***}	-0.034^{**}	-0.021	-0.027^{*}	-0.029^{*}	-0.021
	(0.018)	(0.017)	(0.017)	(0.018)	(0.016)	(0.017)	(0.015)	(0.015)	(0.015)
β_2 (2-star)	-0.072^{***}	-0.043^{**}	-0.025	-0.136^{***}	-0.122^{***}	-0.122^{***}	-0.117^{***}	-0.115^{***}	-0.115^{***}
	(0.020)	(0.019)	(0.019)	(0.020)	(0.019)	(0.019)	(0.018)	(0.018)	(0.018)
β_1 (1-star)	-0.163^{***}	-0.102^{***}	-0.080^{***}	-0.324^{***}	-0.278^{***}	-0.278^{***}	-0.198^{***}	-0.200^{***}	-0.205^{***}
	(0.028)	(0.027)	(0.026)	(0.027)	(0.027)	(0.026)	(0.022)	(0.021)	(0.022)
Observations	14,499	14,499	14,499	12,985	12,985	12,985	7,467	7,467	7,467
R^2	0.004	0.003	0.003	0.023	0.019	0.017	0.058	0.056	0.062

Note: *p<0.1; **p<0.05; ***p<0.01

6.2.3 Second Period: 01/2017 - 12/2022

Table 6.4 shows the regression results for the second period analyzed, from 2017 to 2022. The second period, spanning from 2017 to 2022, was marked by increased volatility. The COVID-19 pandemic, the Russia-Ukraine war, and fears of recession have had their impact in the global markets, as well as the American market. With that being said, the U.S. market still had substantial growth in this period.

The findings from the second period have similarities to the previous periods, but with weaker evidence supporting the outperformance of the reference group. We observe a notable decrease in the number of significant results compared to the analysis of the entire period. Additionally, a few positive coefficients indicate underperformance from the 5-star rated funds. Considering the 6-month investment horizon, most coefficients are positive, except for the 3-star funds in the CAPM and 3-factor model, and the 2-star rated funds in the CAPM model. The coefficients for the 4-star rated funds and the remaining coefficients for the 3- and 2-star funds display a positive sign, contrary to expectations. However, these coefficients are insignificant, suggesting that the performance of 5-star funds did not significantly differ from the 4-, 3-, and 2-star rated groups within a 6-month investment horizon. In contrast, all β_1 coefficients are positive and significant at the 1% level. This indicates that 1-star rated funds outperformed not only 5-star rated funds but also the other rating groups within this investment horizon.

For the 1-year and 3-year investment horizons, the results indicate that 5-star funds also failed to outperform funds with 3- and 4-star ratings. Notably, some coefficients exhibit opposite signs to what one would expect. In the 3-factor and Carhart models for both horizons, the β_4 coefficients are positive, indicating that the reference group underperformed compared to the 4-star rated funds, although these results are insignificant. Only once, in the Carhart model for the 3-year investment horizon, the β_3 coefficient is negative and significant at the 1% level.

However, there is stronger evidence supporting the outperformance of 5-star funds compared to lower-rated funds within these investment horizons. Five out of six β_1 coefficients are negative and significant at the 1- or 5% level. Similarly, three out of six β_2 coefficients are negative and significant at the 1% level, and three at the 5% level.

Table 6.4: Predictive Regression for the Second Period 201701-202212

Table 6.4 presents the fixed effects panel data regression for the second period from January 2017 to December 2022. The coefficients are reported in percentages, and represents the difference in the monthly average alpha compared to 5-star funds. I.e. 0.04 means 0.04%. Standard errors for each estimate are reported in parenthesis.

	6 months			12 months			3 years		
	CAPM	FF	Carhart	CAPM	FF	Carhart	CAPM	FF	Carhart
β_4 (4-star)	0.004	0.015	0.020	-0.045^{*}	0.010	0.010	-0.050^{**}	0.018	0.023
	(0.027)	(0.018)	(0.018)	(0.027)	(0.018)	(0.018)	(0.024)	(0.016)	(0.016)
β_3 (3-star)	−0.002	−0.005	0.001	−0.039	-0.034^{*}	−0.010	−0.014	−0.021	0.012
	(0.027)	(0.019)	(0.019)	(0.026)	(0.018)	(0.018)	(0.024)	(0.016)	(0.016)
β_2 (2-star)	−0.002	0.017	0.015	-0.120^{***}	-0.042^{**}	-0.050^{**}	-0.072^{***}	-0.056^{***}	-0.063^{***}
	(0.030)	(0.022)	(0.021)	(0.029)	(0.021)	(0.021)	(0.024)	(0.020)	(0.020)
β_1 (1-star)	0.122^{***}	0.098^{***}	0.073^{***}	-1.230^{***}	-0.059^{**}	-0.079^{***}	-0.069^{**}	-0.109^{***}	-0.111^{***}
	(0.038)	(0.028)	(0.027)	(0.036)	(0.027)	(0.027)	(0.032)	(0.025)	(0.025)
Observations	11,206	11,206	11,206	10,100	10,100	10,100	5,895	5,895	5,895
R^2	0.002	0.002	0.001	0.003	0.001	0.002	0.010	0.016	0.018

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

6.2.4 U.S. Small-Cap Growth - Full Period Regression

Table 6.5 shows the regression results for the whole period, from 2011 to 2022, for the alternative mutual fund group investing in small-cap growth funds in the US. The results are presented as monthly values and percentages, where 0.4 represents 0.4%. This approach aims to assess the consistency of results across different investment styles, thereby enhancing the robustness of the findings. While some patterns remain consistent with the original sample, there are notable differences.

Firstly, longer investment periods exhibit stronger and more significant results. The 6-month investment horizon reveals minimal significant findings, except for the β_1 coefficients at the 1% level. The positive β_3 coefficients, although insignificant, suggest underperformance from the 5-star group compared to the 3-star group. In contrast, the 1-year and 3-year horizons yield more significant results.

Secondly, 5-star rated funds appear to outperform lower-rated funds (1 and 2 stars). All coefficients for β_1 and β_2 are negative and significant at the 1% level. These results are even stronger than those observed in the original sample. Specifically, the coefficients indicate around -0.5% (-6% annualized) for one-star funds and around -0.25% (3% annualized) for two-star funds. These figures suggest that, on an annual basis, 5-star funds have achieved approximately 6% and 3% higher risk-adjusted returns than the lower-rated funds, respectively.

Thirdly, there is limited statistical evidence supporting the outperformance of 5-star rated funds over the second-best group. Only one of the β_4 coefficients is negative and significant for the 3-year investment horizon in the 3-factor model. Similar results are observed for the 3-star funds in the one-year investment horizon, with no significant coefficients. However, all three β_3 coefficients are negative and significant for the 3-year horizon. This implies that, for this particular investment style and horizon, 5-star rated funds managed to outperform the middle group.

A clear difference from the original sample is the difference in the standard errors. All standard errors for the different groups are higher for the small-cap growth funds than for the large-cap funds. This indicates a higher uncertainty in performance for each individual rating group. Due to the higher volatility in the small cap growth market, the returns of small cap growth mutual funds can exhibit larger fluctuations, resulting in higher standard errors. This means that the performance estimates for these funds may be less precise or more uncertain compared to funds in the less volatile large-cap blend market.

Table 6.5: U.S. Small Cap Growth - Full Period

Table 6.5 presents the fixed effects panel data regression for the full sample period from January 2011 to December 2022. The coefficients are reported in percentages, and represents the difference in the monthly average alpha compared to 5-star funds. I.e. 0.04 means 0.04%. Standard errors for each estimate are reported in parenthesis.

	6 months			12 months			3 years		
	CAPM	FF	Carhart	CAPM	FF	Carhart	CAPM	FF	Carhart
β_4 (4-star)	−0.019	−0.023	−0.023	−0.020	−0.023	0.004	−0.027	−0.036*	−0.003
	(0.027)	(0.023)	(0.023)	(0.025)	(0.022)	(0.021)	(0.021)	(0.020)	(0.016)
β_3 (3-star)	0.011	0.021	0.002	−0.031	−0.037*	−0.031	−0.074***	−0.069***	−0.051***
	(0.027)	(0.024)	(0.024)	(0.025)	(0.022)	(0.022)	(0.019)	(0.017)	(0.017)
β_2 (2-star)	−0.044	−0.023	−0.026	−0.264***	−0.207***	−0.211***	−0.280***	−0.244***	−0.239***
	(0.030)	(0.026)	(0.026)	(0.028)	(0.025)	(0.025)	(0.022)	(0.019)	(0.019)
β_1 (1-star)	−0.086**	−0.090**	−0.088**	−0.358***	−0.329***	−0.326***	−0.570***	−0.492***	−0.481***
	(0.043)	(0.040)	(0.040)	(0.041)	(0.037)	(0.037)	(0.032)	(0.029)	(0.028)
Observations	11,206	11,206	11,206	10,100	10,100	10,100	5,895	5,895	5,895
R^2	0.002	0.002	0.001	0.003	0.001	0.002	0.010	0.016	0.018

Note: *p<0.1; **p<0.05; ***p<0.01

6.2.5 Main Results

The fixed effects panel data regression conducted on the different performance metrics across different investment horizons can be summarized in a few main findings. Firstly, low Morningstar ratings generally indicate poor future performance. This finding holds consistently across different time splits and investment styles. The coefficients associated with low ratings are predominantly negative and statistically significant, with the B1 coefficient being particularly noteworthy. For investment periods of 6, 12 and 36 months, low rated funds generally underperformed the higher rated funds. A notable observation from Table 6.6 is that a significant percentage of 1-star (75%) and 2-star (60%) funds end up being merged or liquidated. This finding suggests that when rating funds, Morningstar has the ability to identify funds that are more likely to face termination. An alternative explanation could be that the low rating itself leads to the liquidation of the fund. This is because poor ratings are typically associated with larger outflows, as investors tend to withdraw their investments from funds that receive unfavorable ratings due to concerns about underperformance or perceived higher risk. The inclusion of these funds in the analysis may explain some of the observed results, as their poor performance and subsequent termination impact the overall performance of their respective rating categories.

Secondly, the analysis reveals inconclusive or insignificant differences in performance between high-rated funds and medium-rated funds. Specifically, 5-star rated funds fail to consistently outperform the 4- and 3-star rated funds. The findings exhibit inconsistency across different time periods, investment styles and performance metrics. For example, for the period before 2017, the coefficients for β_4 and β_3 were all negative, but no coefficient reached significance at the 1% level. For the period after 2017, three of the coefficients were positive but also insignificant.

In the alternative data set, however, 5-star rated funds consistently outperformed 3-star rated funds across all three performance metrics for the 3-year investment horizon. This observation may be attributed to the higher volatility of the small-cap growth market, resulting in larger standard errors and potentially amplifying the significance of the results. Since the 3-star rated funds constitute the largest group, variations in performance could be attributed to certain funds underperforming.

Considering the varying and inconsistent results, we can conclude that there is a lack of substantial evidence indicating that 5-star rated funds consistently differentiate themselves from the next highest (4-star) and medium (3-star) rated funds.

These two findings are consistent with previous research conducted by Morey (2000). Morey‘s study also indicated that low rated funds tend to exhibit poor future performance. Additionally, the study found little statistical evidence that high-rated funds outperform the next to highest and lowest rated funds. However, our study does provide stronger evidence of underperformance of the low rated funds.

Although 5-star rated funds do not consistently outperform 4- and 3-star rated funds, table 6.6 show they do have the highest likelihood of maintaining a high rating among the five groups. If an investor chooses a fund solely based on Morningstar ratings, opting for a 5-star rated fund does offer a better chance of the fund surviving or maintaining a higher rating. However, it is important to note that the probability it remains at the 5-star rating is relatively low. Only 13% of the 5-star funds managed to maintain the 5-star rating throughout the entire period, and 34% maintained a rating of 4 or above. Therefore, it is clear that a high rating does not guarantee success and should not be the sole factor in making investment decisions. Nevertheless, Morningstar ratings can serve as a useful tool for identifying funds with potential, indicating that further investigation is warranted.

Table 6.6: Change in Morningstar Rating

	Rating End of Period					Merged/
Initial Rating	5-star	4-star	3-star	2-star	1-star	Liquidated
5-star	13.04%	21.74%	17.39%	21.74%	0.00%	26.09%
4-star	2.86%	20.00%	22.86%	7.14%	4.29%	42.86%
3-star	1.85%	16.67%	19.44%	16.67%	7.41%	37.96%
2-star	0.00%	7.84%	17.65%	11.76%	7.84%	54.90%
1-star	10.00%	10.00%	0.00%	0.00%	5.00%	75.00%

6.3 Limitations

The findings of this study have to be seen in light of some limitations.

Firstly, the length of the investment horizons used in the second regression model is adequately short considering that it is often recommended to hold such investments for 5 years or more. Due to the length of our sample periods it became a trade-off between the number of observations and the length of investment horizons. For the purpose of this study, we considered it sufficient to look at periods up to 3 years in order to evaluate the fund's performance against its peers. While three years may not fully align with the typical long-term investment horizon, it is considered a reasonable period to assess performance. For instance, it allows for a fair evaluation of the fund's performance in comparison to others, and it also provides an opportunity to consider rebalancing strategies if the fund has underperformed its peers during this timeframe.

Secondly, if the main assumption of the fixed effects model is violated, it might impact the validity of our results. Having that said, the Hausman tests (appendix A2) indicate that this is not the case, and our paper therefore assumes that the assumption of the model holds.

Thirdly, the funds‘ exposure to different risk factors in the CAPM, 3-factor, and Carhart models, measured by beta, is calculated as a single value for the entire period. This approach is common in research, but it may not necessarily provide a completely accurate representation. Over the years, the fund's exposure to these factors can vary, and relying solely on average factors may not yield the most precise results when measuring performance. However, accurately estimating the beta for each specific time period within the sample is often impractical and time-consuming.

Finally, our analysis assumes that missing values are primarily due to fund termination rather than mergers with other funds. However, this assumption may not accurately capture all cases of fund mergers, which can occur for reasons other than underperformance. An alternative approach would involve investigating each individual fund to determine the specific reason for missing values, but this may be time-consuming and impractical given the dataset size. Considering these limitations and practical constraints, the decision was made to assume missing values are primarily due to fund termination, providing a practical and efficient way to handle the missing data in the analysis.

7 Conclusion

This thesis addresses the Morningstar ratings' ability to predict future performance of mutual funds. The purpose of this analysis is not to establish a causal relationship between Morningstar ratings and future performance, but rather to examine potential differences in future performance based on the construction of the ratings themselves. Specifically, the study evaluates whether there are variations in future risk-adjusted returns among different Morningstar ratings, independent of other factors that may influence performance. Our focus is on the predictability of the rating system itself, assuming that investors choose a 5-star rated fund based on its rating, regardless of other factors.

We conducted our analysis using a dataset of 272 mutual funds within the Morningstar category "U.S. Large-Cap blend" over a 12-year period from January 2011 to December 2022. To ensure consistency with the Morningstar rating methodology, we compared the future performance of mutual funds only within the same Morningstar category. Additionally, we divided the analysis into two time periods and examined an alternative Morningstar category to assess potential differences across different time periods and investment styles. For our analysis, we employed fixed effects panel data regressions, utilizing various measures of alphas. Specifically, we regressed different performance measures against the funds' ratings, represented by dummy variables. The performance measures were estimated using three-factor models: CAPM, Fama-French 3-factor, and Carhart 4-factor models. The two analyses differed in how they calculated the alpha. The first analysis was based on the simple monthly alpha, while the second analysis examined cumulative alphas achieved over different investment horizons.

Firstly, we find that low Morningstar ratings generally indicate poor future performance. This holds consistently across different time splits and investment styles. For investment periods of 6, 12 and 36 months, low rated funds generally underperformed the higher rated funds. A large percentage of the lower rated funds also end up merged or liquidated, which can be an inconvenience for investors.

Secondly, we find that 5-star rated funds fail to consistently outperform the 4- and 3-star rated funds. The findings exhibit inconsistency across different time periods, investment styles and performance metrics. The results also find that 5-star rated funds do offer a slightly better chance of the fund surviving or maintaining a higher rating than the other groups, but the probability remains significantly low. It is therefore crucial to recognize that a high rating does not guarantee success, emphasizing the need for additional factors in investment decision-making.

Based on the findings, it can be concluded that the ratings are more effective in identifying potentially poor investments rather than potentially good ones. Although 5-star rated funds have a higher likelihood of survival compared to other groups, they do not consistently outperform the middle-rated funds. Considering that the majority of funds fall within the 5-, 4- and 3-star ratings categories, a 5-star rating only guarantees potential inclusion among the top 67% of funds. Therefore, while the ratings do possess some informational value, they should not be the sole determinant when making investment decisions in mutual funds.

References

Ang, A. (2014). *Asset Management: A Systematic Approach to Factor Investing*. Oxford University Press.

Blake, C. R. and Morey, M. R. (2000). Morningstar ratings and mutual fund performance. *Journal of Financial and Quantitative Analysis*, 35(3):451–483.

Brown, S. J. and Goetzmann, W. N. (1995). Performance persistence. *The Journal of Finance*, 50(2):679–698.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82.

Del Guercio, D. and Tkac, P. A. (2002). The determinants of the flow of funds of managed portfolios: Mutual funds vs. pension funds. *Journal of Financial and Quantitative Analysis*, 37(4):523–557.

Del Guercio, D. and Tkac, P. A. (2008). Star power: The effect of monrningstar ratings on mutual fund flow. *Journal of Financial and Quantitative Analysis*, 43(4):907–936.

Fama, E. F., Fisher, L., Jensen, M. C., and Roll, R. (1969). The adjustment of stock prices to new information. *International Economic Review*, 10(1):1–21.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465.

French, K. (2023). Data Library. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. [Retrieved April 03, 2023].

Grinblatt, M. and Titman, S. (1992). The persistence of mutual fund performance. *The Journal of Finance*, 47(5):1977–1984.

Grossman, S. J. and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, 70(3):393–408.

Investment Company Institute (2022). Characteristics of Mutual Fund Investors. https://www.ici.org/system/files/2022-10/per28-10.pdf. [Online; Retrieved May 19 2023].

Investment Company Institute (2023). What US Households Consider When They Select Mutual Funds. https://www.ici.org/system/files/2023-04/per29-04.pdf. [Online; Retrieved May 19 2023].

Ippolito, R. A. (1992). Consumer reaction to measures of poor quality: Evidence from the mutual fund industry. *The Journal of Law and Economics*, 35(1):45–70.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945-1964. *The Journal of Finance*, 23(2):389–416.

Kirkeby, S. J. (2020). Morningstar ratings and norwegian mutual fund performance: an empirical study of the morningstar rating system as a predictor of performance for mutual funds investing primarily in norwegian equity. Master's thesis.

Morey, M. R. and Gottesman, A. (2006). Morningstar mutual fund ratings redux. *Pace University Finance Research Paper*, (2006/03).

Morningstar (2021). https://www.morningstar.com/content/dam/marketing/shared/research/methodology/771945_Morningstar_Rating_for_Funds_Methodology.pdf. [Online; Retrieved March 30 2023].

Poterba, J. M. and Samwick, A. A. (1997). Household portfolio allocation over the life cycle. Working Paper 6185, National Bureau of Economic Research.

Røed, E. and Høiden, J. (2022). Bør investor velge verdipapirfond basert på morningstar-rangering? en empirisk studie av rangeringens evne til å predikere fondsprestasjon. Master's thesis.

Shiller, R. J. (2003). From efficient markets theory to behavioral finance. *Journal of Economic Perspectives*, 17(1):83–104.

Sirri, E. R. and Tufano, P. (1998). Costly search and mutual fund flows. *The Journal of Finance*, 53(5):1589–1622.

S&P Dow Jones Indices (2023). S&P 500. https://www.spglobal.com/spdji/en/indices/equity/sp-500/#overview. [Retrieved May 31, 2023].

Stock, J. H. and Watson, M. W. (2018). *Introduction to Econometrics*. Pearson Education Limited.

Wharton Research Data Services (2023). WRDS. https://wrds-www.wharton.upenn.edu/. [Online; Retrieved Apr 19 2023].

Appendix

A1 Visualization of the Sample

Figure A1.1: First period

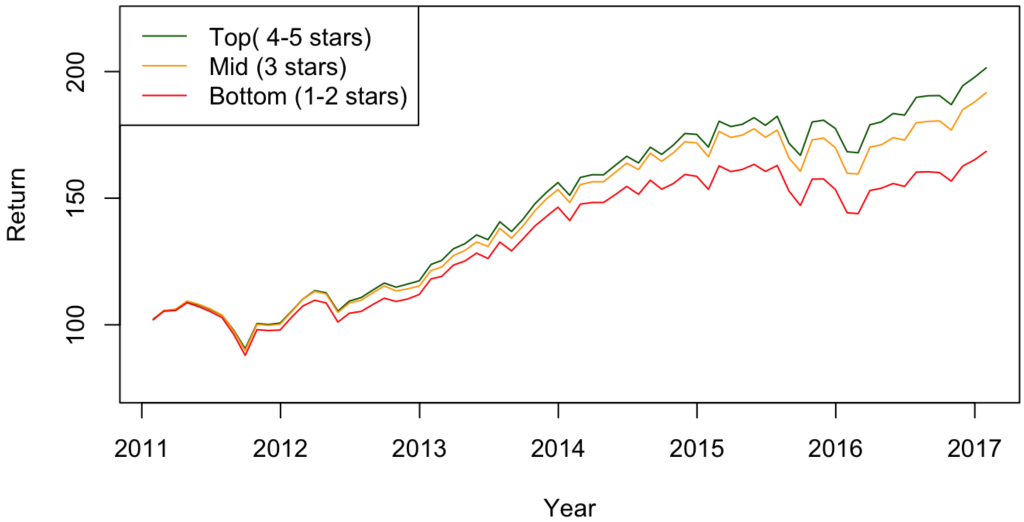
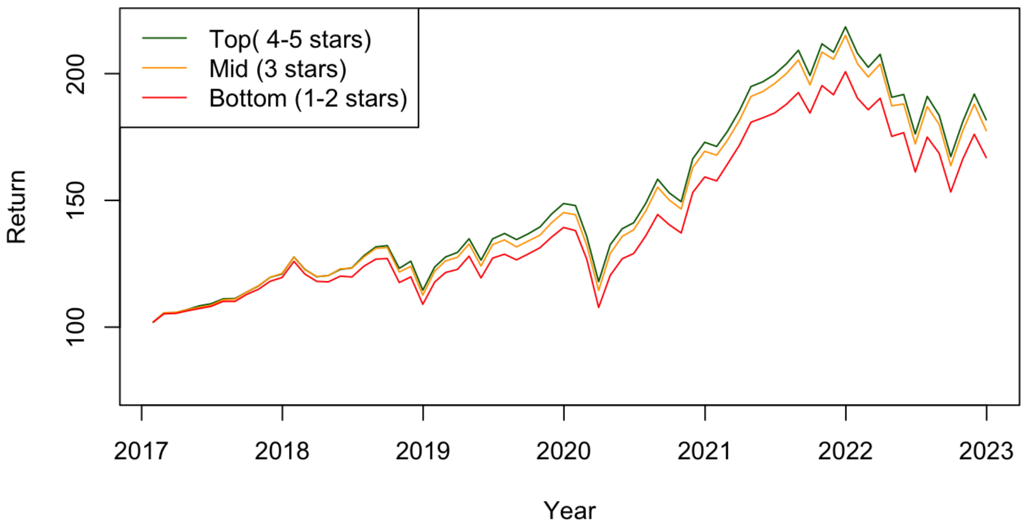


Figure A1.2: Second period



A2 Hausman Test

Table A2.1: CAPM - monthly alpha

data: $\alpha \sim D_4 + D_3 + D_2 + D_1$
chisq = 43.181, df = 4, p-value = $9.49e - 09$
alternative hypothesis: one model is inconsistent

Table A2.2: CAPM - 6 month alpha

data: $\alpha \sim D_4 + D_3 + D_2 + D_1$
chisq = 191.26, df = 4, p-value = $2.2e - 16$
alternative hypothesis: one model is inconsistent

Table A2.3: CAPM - 12 month alpha

data: $\alpha \sim D_4 + D_3 + D_2 + D_1$
chisq = 624.3, df = 4, p-value $< 2.2e - 16$
alternative hypothesis: one model is inconsistent

Table A2.4: CAPM - 36 month alpha

data: $\alpha \sim D_4 + D_3 + D_2 + D_1$
chisq = 2563.7, df = 4, p-value $< 2.2e - 16$
alternative hypothesis: one model is inconsistent

Table A2.5: FF - monthly alpha

data: $\alpha \sim D_4 + D_3 + D_2 + D_1$
chisq = 29.245, df = 4, p-value = $6.97e - 06$
alternative hypothesis: one model is inconsistent

Table A2.6: FF - 6 month alpha

data: $\alpha \sim D_4 + D_3 + D_2 + D_1$
chisq = 443.21, df = 4, p-value $< 2.2e - 16$
alternative hypothesis: one model is inconsistent

Table A2.7: FF - 12 month alpha

data: $\alpha \sim D_4 + D_3 + D_2 + D_1$
chisq $= 613.6$, df $= 4$, p-value $< 2.2e - 16$
alternative hypothesis: one model is inconsistent

Table A2.8: FF - 36 month alpha

data: $\alpha \sim D_4 + D_3 + D_2 + D_1$
chisq $= 3265.8$, df $= 4$, p-value $< 2.2e - 16$
alternative hypothesis: one model is inconsistent

Table A2.9: Carhart - monthly alpha

data: $\alpha \sim D_4 + D_3 + D_2 + D_1$
chisq $= 26.784$, df $= 4$, p-value $= 2.198e - 05$
alternative hypothesis: one model is inconsistent

Table A2.10: Carhart - 6 month alpha

data: $\alpha \sim D_4 + D_3 + D_2 + D_1$
chisq $= 370.94$, df $= 4$, p-value $< 2.2e - 16$
alternative hypothesis: one model is inconsistent

Table A2.11: Carhart - 12 month alpha

data: $\alpha \sim D_4 + D_3 + D_2 + D_1$
chisq $= 628.55$, df $= 4$, p-value $< 2.2e - 16$
alternative hypothesis: one model is inconsistent

Table A2.12: Carhart - 36 month alpha

data: $\alpha \sim D_4 + D_3 + D_2 + D_1$
chisq $= 6378.1$, df $= 4$, p-value $< 2.2e - 16$
alternative hypothesis: one model is inconsistent

A3 Regressions on Monthly Performance with Lags

Table A3.1: Predictive Regression for the First Period 201101-201612

Table A3.1 presents the fixed effects panel data regression for the first period from January 2011 to December 2016. The coefficients are reported in percentages, and represents the difference in monthly alpha compared to 5-star funds. I.e. 0.04 means 0.04%. Standard errors for each estimate are reported in parenthesis.

	No lag			Lag = 1 month			Lag = 2 months		
	CAPM	FF	Carhart	CAPM	FF	Carhart	CAPM	FF	Carhart
β_4 (4-star)	−0.040	−0.069*	−0.085**	0.080**	0.032	0.008	0.001	−0.039	−0.050
	(0.038)	(0.035)	(0.035)	(0.038)	(0.036)	(0.035)	(0.039)	(0.036)	(0.035)
β_3 (3-star)	−0.149***	−0.178***	−0.190***	0.044	−0.017	−0.043	−0.023	−0.073**	−0.085**
	(0.038)	(0.035)	(0.034)	(0.039)	(0.036)	(0.035)	(0.040)	(0.036)	(0.035)
β_2 (2-star)	−0.337***	−0.345***	−0.351***	0.006	−0.060	−0.088**	−0.064	−0.105***	−0.118***
	(0.042)	(0.038)	(0.037)	(0.043)	(0.039)	(0.038)	(0.043)	(0.039)	(0.038)
β_1 (1-star)	−0.536***	−0.527***	−0.514***	−0.036	−0.103**	−0.133***	−0.155***	−0.203***	−0.212***
	(0.056)	(0.051)	(0.050)	(0.057)	(0.052)	(0.050)	(0.058)	(0.052)	(0.051)
Observations	15,836	15,836	15,836	15,561	15,561	115,561	15,288	15,288	15,288
R^2	0.011	0.012	0.012	0.001	0.001	0.001	0.001	0.002	0.002

Note: *p<0.1; **p<0.05; ***p<0.01

Table A3.2: Predictive Regression for the Second Period 201701-202212

Table ?? presents the fixed effects panel data regression for the second period from January 2017 to December 2022. The coefficients are reported in percentages, and represents the difference in monthly alpha compared to 5-star funds. I.e. 0.04 means 0.04%. Standard errors for each estimate are reported in parenthesis.

	No lag			Lag = 1 month			Lag = 2 months		
	CAPM	FF	Carhart	CAPM	FF	Carhart	CAPM	FF	Carhart
β_4 (4-star)	−0.063	−0.075*	−0.083**	0.031	0.004	−0.003	−0.062	−0.067	−0.069
	(0.053)	(0.043)	(0.042)	(0.054)	(0.043)	(0.042)	(0.054)	(0.044)	(0.042)
β_3(3-star)	−0.158***	−0.143***	−0.149***	−0.024	−0.013	−0.025	−0.044	−0.037	−0.044
	(0.049)	(0.039)	(0.038)	(0.050)	(0.040)	(0.039)	(0.050)	(0.040)	(0.039)
β_2 (2-star)	−0.297***	−0.280***	−0.281***	−0.037	−0.060	−0.063	−0.082	−0.099**	−0.093**
	(0.052)	(0.042)	(0.041)	(0.053)	(0.043)	(0.042)	(0.054)	(0.043)	(0.042)
β_1 (1-star)	−0.623***	−0.538***	−0.521***	−0.307***	−0.248***	−0.234***	−0.341***	−0.299***	−0.280***
	(0.064)	(0.051)	(0.050)	(0.065)	(0.052)	(0.051)	(0.066)	(0.053)	(0.052)
Observations	12,151	12,151	12,151	11,959	11,959	11,959	11,769	11,769	11,769
R^2	0.011	0.013	0.013	0.003	0.003	0.003	0.003	0.004	0.003

Note: *p<0.1; **p<0.05; ***p<0.01