The Predictive Performance of Morningstar's Mutual Fund Ratings

Roman Kräussl[1] **Ralph Sandelowsky**

Faculty of Economic Sciences and Business Administration

Vrije Universiteit Amsterdam

First Version: 31 August 2006

This Version: 15 December 2006

Abstract

In this study, the predictive performance of mutual fund ratings given by Morningstar is examined over the course of a 10 year period starting March 1995, by analysing out of sample performance for different out of sample periods up to 10 years, based on Ordinary Least Squares regression analysis. From this analysis it becomes clear that the predictive performance of the different rating systems used by Morningstar do not beat a random walk. Furthermore, research shows that the latest amendment to the rating system, the introduction of 64 categories over four different asset classes, has reduced the predictive performance of the rating system as a whole. Finally, analysis on potential biases and limitations concludes that the comparison of the latest two Morningstar rating systems is not subject to a bias, thereby heavily contradicting results presented by Morningstar itself.

[1] Corresponding author: Roman Kräussl, Free University of Amsterdam and Center for Financial Studies Frankfurt/Main. Address at FU: Department of Finance and Financial Sector Management, FEWEB, de Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, Tel.: +31 20 5986102, Fax: +31 20 5986020, Email: rkraeussl@feweb.vu.nl

1 Introduction

This paper analyses the mutual fund rating system developed by Morningstar. Morningstar assigns up to five stars to a mutual fund to indicate its past performance. Throughout this paper, the predictive performance of the rating system will be analysed and different versions of the rating system will be compared, as the methodology used by Morningstar has changed over the years. The main tool for doing so is an OLS dummy regression that shows whether funds with different star ratings indeed perform differently, and, if so, what the difference in return is. Furthermore, several potential bias introducing factors are analysed.

Morningstar (Kinnel, 2005) has published a study, which concludes that in terms of predictive performance, the rating system introduced by Morningstar in July of 2002 outperforms its predecessor. However, the analysis in that paper is based on June 2002 and June 2003. This introduces a bias into their results as exogenous changes (e.g. in market conditions and number of available mutual funds) could influence the results of their analysis. In order to refrain from making similar mistakes, this paper will compare both rating systems by using samples that are no longer than six month apart. Furthermore, the methodology of Kinnel (2005) is unclear as the paper only contains the results of his study. Blake and Morey (2000) use a very clear methodology. It is for that reason that a similar methodology will be used throughout this paper.

Overall, the paper published by Morningstar (2005) is not very clear on how the research is conducted. Therefore, since previous studies have failed to offer a complete picture of the rating system, (e.g. Blake and Morey (2000) only analyse the U.S. Stock category and Morningstar does not provide their methodology for arriving at the found results) this paper is written in order to offer individuals with interests in mutual fund ratings a complete overview of both the predictive performance of Morningstar's mutual fund rating systems and the change in predictive performance between the last two rating systems.

The importance of Morningstar's star ratings is highlighted by the fact that the bulk of investors have little knowledge about the funds that they are investing in and historical performance is the leading source of information for mutual fund investors (Capon, Fitzsimons and Prince; 1996). This put a great weight on Morningstar's rating system when investors decide upon which fund to invest in. Nevertheless, whether this great weight is justified is under scrutiny in this paper.

Over time, Morningstar had employed two different rating systems based on the same methodology. First, in 1985 Morningstar introduced a rating system based on four broad asset classes. In October of 1996, Morningstar assigned funds into a certain category in order to give a better overview of the mutual fund's area of investment. The system based on four broad asset classes was in effect until July 2002, when these asset classes were replaced by 64 categories as a basis for classifying and analysing mutual funds. The analysis found in this paper consists of two parts. First the predictive performance of the two previous mentioned Morningstar's rating systems is analysed by using out of sample periods of up to ten years. After this analysis, the changes between the two rating systems are further analysed in order to come to a conclusion whether the revision was actually an improvement.

In the dataset obtained from Morningstar it is unclear to which asset class mutual funds were assigned prior to October 1996. In order to be able to use the data for analysis it is assumed that all funds are assigned to a single category from March 1995 until October 1996. This period will be considered as a different rating system.

Results indicate that the rating system in effect up to October 1996 is good at predicting severe underperformance, but fails to discriminate between three, four and five star rated funds. The predictive performance of this system is similar to the predictive performance of the system where the ratings are based on four broad asset classes. However, in terms of predictive performance, the rating system is at best equal to a random walk. Due to changing market demands, mutual funds had to be substitutes for one another. In order to comply with these demands Morningstar revised its rating system in July 2002. Ratings were now based on 64 categories instead of four broad asset classes in order to further indicate the characteristics of individual mutual funds. The results in terms of predictive performance of this last rating system are ghastly. Hardly any categories provide significant regression results. This implies that there is no significant performance difference between one and five star rated funds and undermines the entire usage of the rating system constructed by Morningstar.

In order to compare the two rating systems, the mutual fund out of sample performance of samples three months previous to the rating system change and three months after this event are compared. In order to compare results from samples in the four broad asset classes system (200204 – 200206) with results from samples in the category system (200207-200209), the samples from the four broad asset classes system had to be re-estimated according to the new rating system. Since Morningstar provided category figures since October 1996, this was not a problem. Doing so resulted in ratings based on the four broad asset classes system organised in categories. These new samples were compared to the samples both based on and organised

in categories by comparing the F-stat figures of the overall regressions. From this analysis it can be concluded that the latest revision to the rating system has not improved its predictive abilities. Furthermore, additional analysis shows that there are no biases in the initial comparison.

The results found in this paper are largely in line with those found by Blake and Morey (2000), but contradict the results published by Morningstar (2005). Section 2 contains information on mutual funds, Morningstar and the star rating methodology. Section 3 analyses the previous literature, section 4.1 discusses the dataset and sample construction, 4.2 discusses the methodology used to obtain the results. Section 4.3 contains the analysis on predictive performance for three different rating system methodologies and 4.4 analyses the difference in terms of predictive performance between the last two rating systems. Section 5 concludes the analysis.

2 Morningstar and Mutual Funds

2.1 Mutual Funds

According to Pozen (1998), mutual funds are a type of financial intermediary. They pool investors' assets for collective investment. In other words, investors buy shares of a mutual fund, which in turn invests the money in various types of securities. It is called a *mutual* fund as all of its returns, minus the expenses, are shared by the fund's shareholders. But why should investors invest their well earned money into mutual funds?

Investors have a basic choice, they can invest directly in individual securities, or they can invest indirectly through the use of a financial intermediary. There are several advantages and disadvantages associated to the usage of mutual funds. The SEC (2006) lists professional management, diversification, affordability and liquidity as advantages, while the disadvantages associated to the use of mutual funds are: Costs despite negative returns, lack of control and price uncertainty.

- *Professional Management*. Professional money managers research, select, and monitor the performance of the securities the fund purchases. Due to their increased experience over general investors and the economies of scale obtained through the mutual fund, professional management greatly reduces costs to the investor.
- *Diversification*. Diversification is an investing strategy that can be neatly summed up as "Don't put all your eggs in one basket." Spreading investments across a wide range of companies and industry sectors can help lower risk if a company or sector fails. Some investors find it easier to achieve diversification through ownership of mutual funds rather than through ownership of individual stocks or bonds, especially since the former requires a smaller investment than the latter.
- *Affordability*. Some mutual funds accommodate investors who don't have a lot of money to invest by setting relatively low dollar amounts for initial purchases, subsequent monthly purchases, or both.
- *Liquidity*. Mutual fund investors can readily redeem their shares at the current NAV — plus any fees and charges assessed on redemption — at any time, where the holding of less liquid shares might involve trading against prices of a liquidity provider, or worse, not being able to close a position at all.
- *Costs despite negative returns*. Investors must pay sales charges, annual fees, and other expenses regardless of how the fund performs. And, depending on the timing of

their investment, investors may also have to pay taxes on any capital gains distribution they receive — even if the fund went on to perform poorly after they bought shares.

- *Lack of control*. Investors typically cannot ascertain the exact make-up of a fund's portfolio at any given time, nor can they directly influence which securities the fund manager buys and sells or the timing of those trades.
- *Price Uncertainty*. With an individual stock, one can obtain real-time pricing information with relative ease by checking financial websites or by calling a broker. One can also monitor how a stock's price changes from hour to hour — or even tick by tick. By contrast, with a mutual fund, the price at which you purchase or redeem shares will typically depend on the fund's NAV, which the fund might not calculate until many hours after you've placed your order. In general, mutual funds must calculate their NAV at least once every business day, typically after the major U.S. exchanges close. Some exchange traded funds (ETFs) offer the same characteristics as regular stocks (e.g. tick by tick pricing) nevertheless, the traded volume of ETFs is likely to be lower than that of the stock it is investing in, thereby increasing the possibility of having to buy and sell against the less favourable prices of a liquidity provider.

2.2 Morningstar

Morningstar, Inc. is a leading provider of independent investment research in the United States and in major international markets. They offer an extensive line of Internet, software, and print-based products for individual investors, financial advisors, and institutional clients.

Morningstar is a source for insightful information on stocks, mutual funds, variable annuities, closed-end funds, exchange-traded funds, separate accounts, hedge funds, and 529 college savings plans. With operations in 13 countries, they currently provide data on more than 145,000 investment offerings worldwide.

Morningstar has developed a number of proprietary research and analytical tools that support their fundamental approach to investing. Examples include:

- Morningstar Rating: popularised the concept of risk-adjusted returns among the general investing public;
- Morningstar Style Box: classifies investment offerings based on their underlying size and investment style;

- Morningstar Ownership Zone: graphical tool that plots each stock in a fund's portfolio within the Morningstar Style Box.

In the early 1980s, the mutual fund industry experienced dramatic growth. Individual investors, however, could not readily access comprehensive information about fund performance. Believing that such fundamental information ought to be widely available, Morningstar was established in 1984. One year later, the star rating for mutual funds was introduced. The rating system is subject to continuous improvement, with milestones in 1996 and 2002 where the categories were introduced and the rating system was based on the previously introduced categories.

2.3 Five Star Mutual Fund Rating System

The original Morningstar rating was launched in 1985. It was often used to help investors and advisors choose one or a few funds from a wide array within broadly defined asset classes[2]. However, over time, mutual funds moved from a 'stand alone' investment to being part of a larger portfolio. Due to this development, it was important that funds within a particular rating group be valid substitutes for one another, something the current rating system was unable to do. Therefore, Morningstar changed the methodology in 1996 to assign ratings based on comparisons of all funds within a specific Morningstar category, instead of all funds in a broad asset class. An adjustment to this methodology change was made in 2002, when Morningstar enhanced its star rating with new peer groups and a new measure of risk-adjusted return, in which the ratings were based on the categories to which the funds were assigned. These categories were present since October 1996, but were not used as a basis for the Star rating until July 2002.

Morningstar U.S. places a fund in one of 64 fund categories. These categories are listed in the table 1 on the next page.[3]

---

[2] These asset classes comprised of: U.S. stock funds, international stock funds, taxable bond funds and municipal bond funds.

[3] For more information on the Morningstar categories, see the Appendix.

Table 1: Morningstar Categories

Large Value	Conservative Allocation	Specialty Precious Metals	High Yield Muni
Large Blend	Moderate Allocation	Long Government	Muni Single State Long
Large Growth	Convertibles	Intermediate Government	Muni Single State Interm
Mid-Cap Value	European Stock	Short Government	Muni Single State Short
Mid-Cap Blend	Latin America Stock	Long-Term Bond	Muni California Long
Mid-Cap Growth	Diversified Emerging Mkts.	Intermediate-Term Bond	Muni California Int/Sh
Small Value	Diversified Pacific/Asia	Short-Term Bond	Muni Florida
Small Blend	Pacific/Asia (ex Japan) Stock	Ultrashort Bond	Muni Massachusetts
Small Growth	Japan Stock	Bank Loan	Muni Minnesota
Spec. Communications	Foreign Large Value	High Yield Bond	Muni New Jersey
Specialty Financial	Foreign Large Blend	Multisector Bond	Muni New York Long
Spec. Natural Resources	Foreign Large Growth	World Bond	Muni New York Int/Sh
Specialty Real Estate	Foreign Small/Mid Value	Emerging Markets Bond	Muni Ohio
Specialty Technology	Foreign Small/Mid Growth	Muni National Long	Muni Pennsylvania
Specialty Utilities	World Stock	Muni Natl. Intermediate	Specialty Health
Bear Market	World Allocation	Muni National Short	Stable Value

The table above lists the 64 categories that Morningstar uses to classify mutual funds. Funds placed in the 'Bear Market' category do not receive a rating as their strategies for shorting the market vary widely. Furthermore, not all categories contain funds at all times, as Morningstar has added and changed categories over time.

2.3.1 Morningstar Risk Adjusted Return

Morningstar uses the Morningstar Risk-Adjusted Return (MRAR) to rate funds. In order to obtain this return, one first has to calculate the fund's total return as given below.

$$TR = \left\{ \frac{P_e}{P_b} \prod_{i=1}^{n} \left(1 + \frac{D_i}{P_i} \right) \right\} - 1, \tag{1}$$

where TR is the total return for the month, P_e is the end of the month Net Asset Value (NAV), P_b is the NAV at the beginning of the month, D_i is the per share distribution at time i, P_i is the reinvestment NAV per share at time i and n is the number of distributions during the month. Distributions include dividends, distributed capital gains and return of capital.

Another important aspect of the MRAR is the cumulative value. If there were no loads or redemption fees, the cumulative value of a $1 investment over a period of t months would be:

$$V_u = \prod_{t=1}^{T} \left(1 + TR_t \right), \tag{2}$$

where V_u is the cumulative value, unadjusted for loads and redemption fees and TR_t is the total return for month t.

In the case of loads or redemption fees, this formula changes into the following.

$$V = (1-F)(1-R)V_u - D(1-F)\frac{\min(P_0, P_t)}{P_0}, \quad (3)$$

where V is the cumulative value, adjusted for loads and redemption fees, F is the front load, D is the deferred load, R is the redemption fee, P_0 is the NAV per share at the start of the period and P_t is the NAV per share at the end of the period.

MRAR is defined as follows:

$$MRAR(\gamma) = \left[\frac{1}{T}\sum_{t=1}^{T}(1+r_{Gt})^{-\gamma}\right]^{-\frac{12}{\gamma}} - 1, \quad (4)$$

where r_{Gt} is the geometric excess return in month t expressed as:

$$r_{Gt} = \frac{1+TR_t}{1+R_{bt}} - 1.$$

R_{bt} is the return on a risk-free asset in month t and γ is a parameter that describes the degree of risk aversion. A rating system that would be based only on performance instead of both on performance and risk would rate funds based on their geometric mean return or MRAR(0). A rating system that does account for risk taken by funds requires MRAR(>0). Fund analysts have concluded that for a typical investor, γ=2

Since MRAR is an annualised return, it consists of a return component, MRAR(0) and a risk component MRAR(0)-MRAR(2). Where MRAR(0), the annualised geometric mean of the geometric excess return is:

$$MRAR(0) = \left[\prod_{t=1}^{T}(1+r_{Gt})\right]^{\frac{12}{T}} - 1 \quad (5)$$

This calculation of MRAR assumes no loads and redemption fees. When fees and redemption fees are present, the monthly total returns (TR_t) must be adjusted according to the following formula:

$$ATR_t = a(1+TR_t) - 1, \quad (6)$$

with a being:

$$a = \left(\frac{V}{V_u}\right)^{\frac{1}{T}},$$

where ATR_t is the adjusted total return for month t, a is the adjustment factor, TR_t is given in equation 1, V_u is given in equation 2 and V is given in equation 3. In order to integrate the loads and fees into the MRAR (equation 4), ATR_t should be used instead of TR_t.

2.3.1.1 Weights

Funds are rated for up to three periods, three, five and 10-years. For a fund that does not change categories during the evaluation period, the overall rating is calculated by using the weights in the following table. When a fund does change categories, its historical information is given less weight. This minimises the incentive for fund companies to change a fund's style in order to receive a better rating.

Table 2: Mutual Fund Weights

Fund Age	**Overall Rating**
At least three years, but less than five	100% three-year rating
Between five and 10 years	60% five-year rating 40% three-year rating
At least 10 years	50% 10-year rating 30% five-year rating 20% three-year rating

While this table seems to give the most weight to the 10-year rating, the three-year rating is actually the most important as it is included in all rating periods. When a fund does change categories, the weights in the table above change, based on the similarity between the category the fund belonged to and the category the fund changed to. This is done by first applying the following formulae:

$$D_3 = \frac{\sum_{s=1}^{36} D_s}{36},\ D_5 = \frac{\sum_{s=1}^{60} D_s}{60} \text{ and } D_{10} = \frac{\sum_{s=1}^{120} D_s}{120},$$

where D_n is the average degree of similarity for the n year period and D_s is the degree of similarity between the fund's category in month one and the fund's category in month s, where s=1 is the current month and s=2 is the previous month. Values for D_s can be found in table 3 and table 4. When a category pair is not listed, the similarity is 0.00.

Table 3: Similarity Matrix

Category	1	2	3	4	5	6	7	8	9	10	11
1 Large Value	1.00	0.50	0.00	0.50	0.25	0.00	0.00	0.00	0.00	0.50	0.25
2 Large Blend	0.50	1.00	0.50	0.25	0.50	0.25	0.00	0.00	0.00	0.50	0.25
3 Large Growth	0.00	0.50	1.00	0.00	0.25	0.50	0.00	0.00	0.00	0.50	0.25
4 Mid-Cap Value	0.50	0.25	0.00	1.00	0.50	0.00	0.50	0.25	0.00	0.50	0.25
5 Mid-Cap Blend	0.25	0.50	0.25	0.50	1.00	0.50	0.25	0.50	0.25	0.50	0.25
6 Mid-Cap Growth	0.00	0.25	0.50	0.00	0.50	1.00	0.00	0.25	0.50	0.50	0.25
7 Small Value	0.00	0.00	0.00	0.50	0.25	0.00	1.00	0.50	0.00	0.50	0.25
8 Small Blend	0.00	0.00	0.00	0.25	0.50	0.25	0.50	1.00	0.50	0.50	0.25
9 Small Growth	0.00	0.00	0.00	0.00	0.25	0.50	0.00	0.50	1.00	0.50	0.25
10 World Stock	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	1.00	0.00
11 Mod Allocation	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.00	1.00

The matrix above shows the similarity between certain category pairs. This data is used to calculate the weightings for funds that switched between similar categories.

Table 4: Category Similarity

Category A	Category B	Similarity
Conservative Allocation	World Allocation	0.25
Moderate Allocation	World Allocation	0.25
Conservative Allocation	Multisector Bond	0.25
Moderate Allocation	Conservative Allocation	0.50
Specialty Technology	Specialty Communications	0.25
Foreign Large Value	World Stock	0.50
Foreign Large Blend	World Stock	0.50
Foreign Large Growth	World Stock	0.50
Foreign Small/Mid Value	World Stock	0.50
Foreign Small/Mid Growth	World Stock	0.50
Foreign Large Value	Foreign Large Blend	0.50
Foreign Large Blend	Foreign Large Growth	0.50
Foreign Small/Mid Value	Foreign Small/Mid Growth	0.25
Foreign Small/Mid Value	Foreign Large Value	0.25
Foreign Small/Mid Value	Foreign Large Blend	0.25
Foreign Small/Mid Growth	Foreign Large Blend	0.25
Foreign Small/Mid Growth	Foreign Large Growth	0.25
Long Government	Intermediate Government	0.50
Intermediate Government	Short Government	0.50
Long-Term Bond	Intermediate-Term Bond	0.50
Intermediate-Term Bond	Short-Term Bond	0.50
Short-Term Bond	Ultrashort Bond	0.50
Muni National Long	Muni National Intermediate	0.50
Muni National Intermediate	Muni National Short	0.50
High Yield Muni	Muni National Long	0.50
High Yield Muni	Muni National Intermediate	0.50
High Yield Muni	Muni National Short	0.50
Muni Single State Long	Muni Single State Int/Sh	0.50
Muni New York Long	Muni New York Int/Sh	0.50
Muni California Long	Muni California Int/Sh	0.50

Table 4 shows the similarity between category pairs. It extends table 3 with mutual funds that fall outside the similarity matrix.

When a fund has five years of data available, the three-year and five-year ratings are combined with the following weights.

$$W_5 = \frac{0.60D_5}{0.40D_3 + 0.60D_5} \text{ and } W_3 = \frac{0.40D_3}{0.40D_3 + 0.60D_5},$$

For a fund that has 10 years of available data, its three-year, five-year and 10-year ratings are combined with the following weights:

$$W_{10} = \frac{0.50D_{10}}{0.20D_3 + 0.30D_5 + 0.50D_{10}}, \qquad W_5 = \frac{0.30D_5}{0.20D_3 + 0.30D_5 + 0.50D_{10}} \qquad \text{and}$$

$$W_3 = \frac{0.20D_3}{0.20D_3 + 0.30D_5 + 0.50D_{10}}.$$

The actual star rating is calculated by first placing each fund it the category it belongs to according to the most recent monthly record and then calculating the three-year star rating for all funds that have at least 36 continuous months of data. Funds are ranked based on their MRAR(2), where funds with the highest scores receive the most stars. For those funds where data is available on a five-year and 10-year periods, ratings are assigned for those periods as well and the final rating is a weighted average according to the weights in table 2, but only if the funds remain in the same category over the five- or 10-year period of time. Otherwise the weights are changed according to the formulae described above.

The resulting performance figures are assigned to star groups according to the following table. This distribution is the same for all 64 categories, with the Bear Market category as exception as funds in that category vary widely in their risk factor exposures and receive no rating at all.

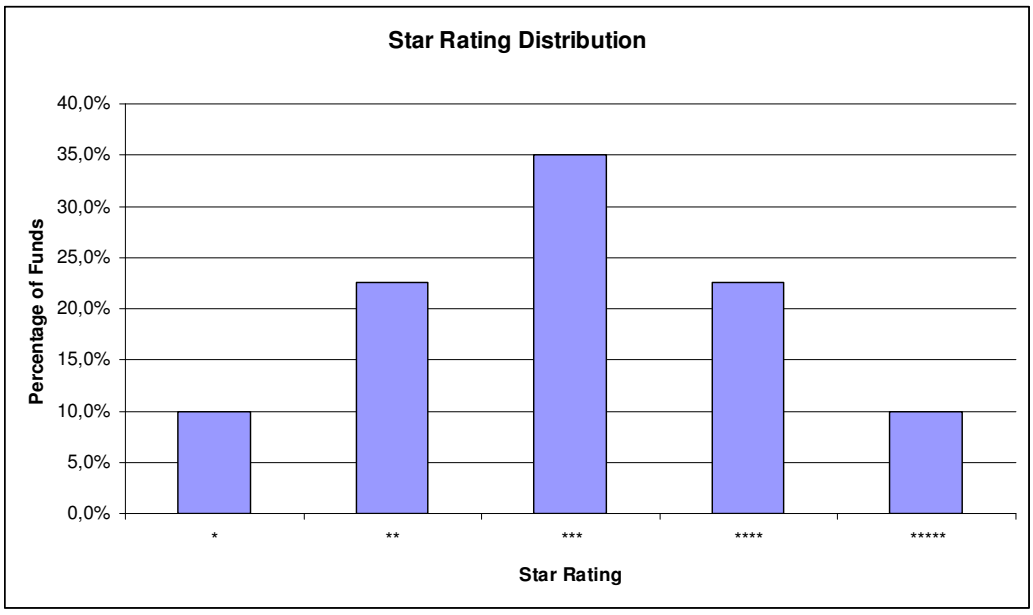

Figure 1: Star Rating Distribution

The figure above shows the distribution of Morningstar's mutual fund rating. From this figure it becomes clear that the three star category is most common. This is an important factor to consider as this distribution implies that is harder to distinguish between the four and the five

star group, than it is to distinguish between the three and the four star group. This is mainly due to the fact that there are simply less fund-months to estimate the performance of the five star group on than there are to estimate the performance of the four star group on.

2.4 Mutual Fund Categories

Many funds, in their prospectus, claim to be seeking 'growth' with some of these funds investing heavily in blue-chip companies while others where mainly investing in small-cap firms. The risk borne by these two funds is obviously different and for that reason, they should not be placed in the same 'growth' category.

In order to eliminate this category allocation problem, Morningstar introduced 64 categories in 1996[4] to help investors to better compare different mutual funds. The creation of the categories is based on five arguments:

- Funds in the same category invest in similar types of securities and consequently, share the same risk factors.
- Funds in the same category are more likely to behave in the same way to one another than to funds in another category.
- The performance of different categories differs substantially over time.
- Categories contain enough funds to form the basis for peer group comparisons.
- The differences between categories are meaningful to investors.

[4] Morningstar's database lists multiple categories since 1996, but these were not used to base the rating system on until July 2002.

3 Previous Studies

3.1 Predictive Performance

Khorana and Nelling (1998) examine the determinants and predictive ability of Morningstar's mutual fund rating system, in order to better understand the extent to which ratings are related to various fund characteristics. Apart from that, they also examine the degree of persistence in fund ratings. Their analysis is based on a dataset obtained from the Morningstar OnDisc CD-ROM as of June 1995. For each fund, the dataset includes the following: The Morningstar rating; the alpha, beta, and R-square values from a market model regression using monthly returns over the period July 1992 through June 1995; expenses; portfolio turnover; front-end load charges; and the tenure of the current fund manager as of June 1995. In all, the dataset provides data on 2871 funds.

Based on descriptive statistics, Khorana and Nelling find that higher-rated funds tend to be larger. Furthermore, they argue that it is likely that highly rated funds performed well in the past and, therefore, attracted more capital, which, in turn resulted in a lower expense ratio, but only when the fund's costs were largely fixed, instead of variable. Moreover, Khorana and Nelling find that funds with higher star ratings exhibit lower portfolio turnover than lower rated funds, and that managers of higher-rated funds tend to serve longer tenures[5]. In order to come to these findings, Khorana and Nelling use a multinomial probit model based on the sample of all funds, and a sample organised by investment objective. Contrary to the analysis performed on all funds, the analysis of funds organised by investment objective does not show longer management tenures for higher-rated funds.

In order to check for persistence in fund ratings, Khorana and Nelling compare the ratings of 848 funds on June 1995 with the ratings of those funds on December 1992. They find that 61% of the four-and five-star funds maintained or improved their rating over the course of the selected period. Furthermore, Khorana and Nelling find that funds with higher ratings tend to have higher risk-adjusted performance, lower systematic risk, a greater degree of diversification, a larger asset base, managers with longer tenures, and lower front-load charges and expense ratios. Furthermore, according to Khorana and Nelling, fund performance is persistent over a short-term horizon.

---

[5] This result seems counterintuitive as successful managers are more likely to receive offers to manage other funds, usually resulting in a better performance structure for the manager.

Blake and Morey (2000) examine the Morningstar rating system as a predictor of mutual fund performance for U.S. domestic equity funds. They compare the future performance of mutual funds against both their Morningstar rating and four alternative predictors: a naïve predictor (in-sample mean monthly returns), the Sharpe ratio and Jensen's single-index and four-index alphas.

In order to cope with the size of their database, Blake and Morey define two samples which they name (1) seasoned funds 1992 – 1997 and (2) complete funds 1993. For the first sample, they select funds classified as domestic equity and that are categorised as aggressive growth, equity-income, growth, growth-income, or small company from the Morningstar On-Disk or Principia programs from 1992 to 1997. They refine this sample by only selecting those funds that have at least 10 years of return data, and were open to new investors at the time the fund was rated by Morningstar.

In order to reduce survivorship bias by only selecting funds that have over 10 years of return data, Blake and Morey select virtually all open aggressive growth, equity-income, growth, growth-income and small company funds that received a Morningstar rating in January 1993 for their second sample.

In order to examine out-of-sample performance of the selected funds, Blake and Morey use two methods: dummy variable regression analysis and the non-parametric Spearman-Rho rank correlation test. The dummy variable regression on the seasoned funds sample shows that the performance of five star funds differs over time and secondly, that the out-of-sample performance of four- and three-star funds does not differ from that of five-star funds. Third, the regression shows that Morningstar ratings are able to predict underperforming funds to a certain extent. The results of the Spearman-Rho correlation tests further prove the point that low scores predict poor performance, while high scores have, at best, only mixed ability to predict future performance. The results are similar for the complete funds 1993 sample, except for the growth and growth-income funds, where there is evidence of ability to predict winning funds.

When considering the different performance metrics, Blake and Morey find that Morningstar‘s rating system is mediocre in terms of predicting future performance. The naïve predictor without adjustment for styles, and the four-index alpha do worse, while the Sharpe ratio does considerably better. Nevertheless, for each predictor, the ability to predict high-performing funds is weak, while the ability to predict low-performing funds is quite high.

Results on the complete funds 1993 sample indicate that the alternative predictors do worse in predicting performance than the Morningstar star method.

It is interesting to note this difference in predictive abilities of the Morningstar system over the alternative predictors when considering the different samples. This is largely due to the fact that the alternative predictors in the complete fund 1993 sample have only three years of return data available, whereas for 545 out of 635 funds, Morningstar uses more data to allocate their stars. When looking at young funds, the difference in predictive abilities compared to the alternative predictors disappears.

Although the results are impressive, this study has several shortcomings. First, Blake and Morey only analyse domestic equity funds, and only those classified as aggressive growth, equity-income, growth, growth-income, or small company. Second, they only analyse seasoned funds. Although the complete funds 1993 sample is constructed to show a more complete picture, Blake and Morey are cautious about its result due to the limited sample size of the complete funds 1993 sample. Third, the seasoned funds 1992 – 1997 sample only covers five years. It would be interesting to see what the predictive performance of Morningstar's star rating system is over a longer time period (i.e. 10 years)

3.2 Investor Behaviour

Sirri and Tufano (1998) find that equity mutual fund investors invest in funds with the highest recent returns, while they fail to disinvest from poor performing funds. Meanwhile, investors are sensitive to fees charged by mutual funds, as funds charging lower fees and funds that have reduced their fees, grow faster. Conversely, funds that receive greater media attention attract greater inflows, resulting in a stronger performance-flow relationship amongst funds that are more active in marketing, thereby charging higher fees. Another interesting finding by Sirri and Tufano is that funds in larger complexes (such as Fidelity, Vanguard and T. Rowe) grow more quickly. This implies that mutual funds receiving a five star rating from Morningstar experience a high inflow of funds, while on disappearance of this five star rating, there is a smaller outflow.

Del Guercio and Tkac (2001) use a sample obtained from Morningstar on mutual funds classified as domestic equity, running from November 1996 to October 1999 containing star ratings for mutual funds present in this domestic equity group. In order to adjust for survivorship bias, they fill in star ratings for funds that disappeared by using the monthly editions of Morningstar's Principia CD-ROM totalling 4,040 fund-months. In order to link the star ratings to performance and order flow, the star ratings were supplemented with data on

returns, total net assets and other characteristics from the 1999 Survivorship Bias Free Mutual Fund Database constructed by the Center for Research in Security Prices (CRSP). From this combined dataset, all fund-months in which a merger took place were removed as these events could distort the flow data. The final dataset consists of 111,715 fund-months from 3,388 funds. From analysis on this database, Del Guercio and Tkac conclude that the average standardised abnormal flow is significantly positive for funds receiving an initial rating of five stars in months one and six after the rating is announced. This means that due to the issuance of the five star rating, the funds attracted abnormal high amounts of money in the first and sixth month after the rating was issued. Furthermore, the same analysis shows that a same effect is found when funds are issued a two star rating, although in this scenario the average standardised abnormal flow is significantly negative: investors disinvest from a fund when its first rating is a two star rating[6].

When looking at the average cumulative standardised abnormal flow (ACSAF), the effect described above is found over more periods; Del Guercio and Tkac find that five star funds have a positive ACSAF for six months after the issuance of the initial five star rating, while two star funds experience a negative ACSAF for three months after the fund receives its two star rating, including the month the rating was issued in. This implies that the effects of an initial rating are measurable up to six months after the rating was issued. When looking at the ACSAF resulting from rating up- and downgrades, the situation is somewhat similar. When funds are upgraded from a one star rating to a two star rating, the ACSAF there is a minor positive effect in months five and six after the rating upgrade. However, an upgrade from two to three stars ensures for a significantly positive ACSAF in the six months after the rating upgrade, with the exception of month two. An upgrade from three tot four stars is similar to an upgrade from one to two stars, only a significantly positive ACSAF in months four five and six. Last but not least, an upgrade from four to five stars ensures a positive ACSAF in all six months after the rating is issued, including the month the rating is issued in.

Rating downgrades result in the following ACSAFs. When a fund is downgraded from five to four stars, there is no significantly negative ACSAF in any of the six months following the rating downgrade. This is different for a rating downgrade from four to three stars; as here there is a significantly negative ACSAF in all six months following the rating downgrade. When a fund is downgraded from three to two stars, there is a significantly negative ACSAF in months three, four five and six after the rating downgrade. A downgrade from two to one

[6] It is remarkable that when a fund receives an initial one star rating, this effect is not present.

star does not result in a significantly negative ACSAF in any of the six months after the rating downgrade.

The above clearly shows that while investors choose to invest in funds receiving a five star rating, they choose not to disinvest from these funds once they receive a rating downgrade. This is in-line with the findings of Sirri and Tufano (1998). However, Del Guercio and Tkac state that Morningstar's master data file contains historical star ratings for each fund that reflects the rating algorithm currently in place, rather than the one in place at that time. If this were true, the dataset used in this paper would not contain any information on fund ratings prior to October 1996, as before that date; the Morningstar Categories were not introduced. Since these categories form the basis for rating the mutual funds in the present rating system, recalculating historical rating based on the current algorithm would not be possible as is no information on the exact category a fund is in, prior to October 1996. This could seriously dilute the results found in their study.

Based on a telephone survey of 3386 mutual fund investors, Capon, Fitzsimons and Prince (1996) find that the bulk of investors have little knowledge about the funds that they are investing in, as 72.3% does not know whether their funds invest in domestic or international investments and, more strikingly, 75.0% of the mutual fund investors did not know whether their fund invested in equity or fixed income. Furthermore, in that same survey, Capon et al. find that published performance rankings are the leading source of information on mutual funds, while historical performance is used as the most important selection criterion. The fact that published performance rankings achieve a score of 4.57 out of 5 further illustrates the importance of a rating system such as Morningstar's to investors. This importance puts a high weight on the validity of Morningstar's ratings, as Morningstar is amongst the leading raters of mutual funds.

All the papers described above highlight the importance of the ratings assigned by Morningstar. A five star rating heavily increases the flow of funds towards a mutual fund, while a rating downgrade does not immediately result in a large outflow of funds. The fact that investors stick to the fund ratings is not at all surprising, as research shows that investors choose their mutual funds based on historical performance, which these ratings are an indication of.

3.3 Survivorship Bias

When analysing a sample mutual funds over longer periods of time, some mutual funds in the sample are sure to cease to exist. As a consequence of this, the results of analysis performed

on the sample are biased towards better performing funds, as funds that cease to exist generally do so due to poor performance or a low total market value (Elton, Gruber and Blake, 1996), based on which managers choose to no longer maintain the fund. In other words, by overlooking the funds that 'die', either by merger or liquidation, the results of analysis performed on a sample of mutual funds are too optimistic. Prior studies show that survivorship bias ranges from 10 to 150 basis points (e.g. Grinblatt and Titman, 1989 and Malkiel, 1995). According to Elton et al. (1996), the three-index model based on the S&P 500, the smallest two deciles of CRSP NYSE stocks and the Leman Brothers Aggregate Bond Index is the appropriate way to measure excess returns and bias.

Morey (2002) investigates the relationship between the age of mutual funds and their Morningstar ratings. Data is obtained from the quarterly Morningstar On-Disk or Principia programs from September 1991 to September 2000. From these disks, Morey selects all funds in the domestic equity category. These funds are then placed in one of three age related categories: young funds (36-59 months of return data), middle-aged funds (60-119 months of return data) and seasoned funds (>119 months of return data).

Using descriptive statistics, Morey finds that the average overall star rating of seasoned funds is almost always higher than that of young funds, and to a lesser extent, that of middle-aged funds[7], thereby concluding that when the age of the fund increases, the average overall star rating increases as well. When looking at the standard deviation, Morey finds that the standard deviation of young funds is higher than that of seasoned funds. Middle-aged funds also have a higher standard deviation than do seasoned funds, but when analysing middle-aged and young funds, there is no clear pattern when focussing on standard deviation. When organising the sample by age, Morey shows that young and middle-aged funds are more likely to receive very high or very low ratings than do seasoned funds.

In order to check whether the results are subject to survivorship bias, Morey uses the time specific three- and five-year ratings to compare funds, instead of the overall rating. When using these ratings, Morey finds that the higher overall star ratings of seasoned funds are not caused by better performance in the three-year time specific ratings. Interesting to note is that when comparing young and middle-aged funds on the three-year time specific rating, the middle-aged funds have a higher rating in 22 out of 30 cases.

---

[7] Perhaps the situation is no different from hedge funds, where, when a fund after a certain period of time still does not perform as required, one might as well liquidate the fund and start a new one.

While survivorship bias does play a role in the rating process, Morey shows that once a fund receives a 10-year time specific rating, its overall rating is less likely to decline, due to the weights that Morningstar uses to calculate the overall rating. While on the other hand, a fund with a 10-year time specific rating will see its overall rating increase more easily.

3.4 Earnings Persistence

Morey and Vinod (2001) examine the estimation risk in the Morningstar mutual fund star rating system. They find that the estimates upon which younger funds ratings are based have significantly higher estimation risk than the estimates upon which the ratings of older funds are based. They use data from the January 2001 Morningstar Principia Data Disk. From this disk, they select all funds that are in Morningstar's International Equity Fund Category and have received an overall star rating (thereby having at least three years of return data) and collect the excess non-load adjusted monthly returns for the 10-year period from January 1991 to December 2000, or the entire history of the fund, if it has less than 10 years of return data. This gives them a sample of 1281 funds, of which 508 are classified as young funds (three to five years of return data), 619 are middle-aged funds (five to 10 years of return data) and 154 are seasoned funds (over 10 years of return data).

The fact that Morningstar uses a discrete interval to measure performance requires Morey and Vinod to use a methodology that does not directly determine the estimation risk in the star ratings themselves, but computes the estimation risk in the estimates that are used by Morningstar to calculate its star ratings. Morey and Vinod subsequently compute a confidence interval on the difference between a fund's load-adjusted return and a fund's risk. Due to the increased amount of available data points for the seasoned funds, the confidence interval of the measure that Morningstar uses to calculate time specific star ratings for seasoned funds is narrower than that of young and middle-aged funds. While this seems impressive at first, it only concerns the time specific Morningstar rating, the overall rating is a weighting of the time specific ratings and the effect of the narrower confidence interval on the overall rating is therefore, much smaller, nevertheless existent. This implies that a young fund that as received a three star rating could actually be a four- or two star fund, while a seasoned fund with a four star rating has a much higher probability of actually being a four star fund, as the confidence interval for the ten year rating is smaller. This makes perfect sense, as, all other things being equal, an average based on three years of data is more volatile than an average based on 10 years of data.

Morey (2003) examines the effect that an initial 5-star Morningstar mutual fund rating has on future fund performance, strategy, risk-taking, expenses and portfolio turnover by using 33

Morningstar mutual fund quarterly data disks from July 1993 till July 2001. From these disks, Morey selected all funds that had inception dates after March 31st, 1990 and had received a 5-star overall Morningstar rating for the first time (e.g. a fund that received an overall 5-star rating on the July 1993 disk would no longer be selected when using the October 1993 data disk). This results in a sample of relatively young funds, as the dataset only spans eight years. In order to make the sample more manageable, Morey selects only those funds that are classified as diversified domestic equity funds, resulting in a sample of 273 funds. For these 273 funds, Morey examines the mean and median performance levels before and after the issuance of the initial 5-star rating. Moreover, he defines three sub samples: (1) All funds, with the exception of index funds in order to only select actively managed portfolios; (2) All actively managed portfolios that do not have multiple share classes; and (3) All actively managed portfolios, that are defined as growth funds, without multiple share classes, resulting in a total of four samples to be tested. Using the following four out-of-sample performance metrics: (1) the Fama-French-Momentum 4-factor alpha; (2) the Elton, Gruber and Blake 4-factor alpha; (3) the Sharpe ratio; and (4) a single-index alpha, for all four samples, Morey shows a steep drop in performance, thereby concluding that a 5-star Morningstar rating does not persist three years out-of-sample.

An explanation for this phenomenon can be found in the fact that managers significantly move towards higher value stocks in attempt to maintain the 5-star rating. Furthermore, Morey's results show that after receiving the initial 5-star rating, funds do a poorer job of loading on momentum stocks. Apart from the previous mentioned factors, a separate test on the risk-taking behaviour of the 5-star rated fund shows that the risk level of the average initial 5-star rated fund increases substantially (i.e. both sigma and beta rise). Finally, there is always the argument of mean reversion: given enough time, most winning funds will revert to the mean in terms of performance.

When taking a good look at the dataset used, it is evident that Morey's larges sample consists of only 273 funds, with even smaller sub samples. Furthermore, by selecting only diversified domestic equity funds, Morey limits the generalisation ability of the results. It would be interesting to perform an identical analysis on a much larger dataset.

4 Empirical Analysis

This section describes the analysis that has been performed on the dataset obtained from Morningstar. The analysis consists of two parts: an analysis on the predictive performance of both Morningstar's rating systems and a comparison of these two rating systems. Section 4.1 describes the dataset and section 4.2 discusses the methodology that was used to analyse the predictive performances (Section 4.3) and to compare the two rating systems (Section 4.4).

4.1 Data

The dataset used for the analysis in this paper was obtained from Morningstar Inc. and consists of data on 25.202 funds, ranging from March 1995 till September 2005. For each fund, it contains data on fund name, fund return, Morningstar overall star rating, Morningstar star rating based on three years of data, inception date, end date, category, equity style box position, and fixed income style box position, on a monthly basis. Moreover, for a selection of funds, return data is available from September 1924 onwards. Furthermore, for each fund, the start date, end date, name, age and education of the fund manager are supplied.

Although Morningstar has changed it rating methodology in July 2002, past ratings are not recalculated as it is the data available to investors at a particular moment in time that is used by those investors to decide upon which fund to invest in. If this takes place before July 2002, the data will be based on the old rating method.

Since the Morningstar Risk Adjusted Return accounts for loads and redemption fees[8], this research should do the same. However, obtaining load information on over 25.000 funds, of which some ceased to exist as long as 10 years ago is an unmanageable quest and would be an excellent suggestion for further research. For the sake of manageability of the already massive database, it is assumed that all funds refrain from charging loads and redemption fees, or rather, that there is no difference in loads and fees charged by funds with different ratings.[9]

In the event that the fund has merged, or was liquidated, it is assumed that investors randomly invest their funds in funds of the same category as the liquidated or merged fund was in, at the

---

[8] See equation 3.

[9] On average, one would expect no difference between a) the loads and redemption fees charged by one star rated and five star rated funds, and b) the number of one and five star rated funds charging loads and redemption fees. If this were to be true, the assumption made should not result in a bias. However, that would be something that further research could point out.

time of liquidation or merger. This means that from the merger or liquidation date onwards, the fund will have the average return of the category it was last listed in. However, since the Morningstar Categories were introduced in October 1996, a fund that ceases to exist between March 1995 and October 1996 will have to be excluded from the sample. In total, the following funds were removed from the database:

- Funds that have a category listing of 'NA' at the time of disappearance from the database[10];
- Funds that ceased to exist prior to the introduction of the Morningstar categories (October 1996)

Apart from funds disappearing from the database, several categories disappeared or changed name. In the event of a name change, the name of the category was changed in the database. When a category ceased to exist, its returns were recreated by using the weighted average of the returns of the categories the original category split into; with the weights being based on the number of funds in the 'new' category. There were only two categories for which this procedure was necessary: Domestic Hybrid and Foreign Stock. The former split into Conservative Allocation and Moderate Allocation, while the latter split into Foreign Small/Mid Value, Foreign Large Blend, Foreign Large Growth, Foreign Small/Mid Growth and Foreign Large Value. Analysis has shown that around 1% of the values are missing (e.g. When performing an analysis on a sample with 240 funds for a one year out of sample window, the regression was based on 2592 fund-months).

There are sixteen samples in this study, running from 199503 (the first four digits indicate the base year, the last two digits indicate the base month) through 200403 and including 200204 - 200209. The annual xxxx03 samples will be used to analyse the predictive performance of the Morningstar rating system, while the 200204 – 200209 samples will be used to compare the rating system used until 200206 with the rating system used from 200207 onwards.

4.2 Methodology

To test for predictive abilities of Morningstar ratings, the equation below was estimated using a one year, three year, five year and 10 year out of sample period.

---

[10] There were 28 funds that had a rating for a period of no more than five months when they ceased to exist. Morningstar assigned these funds to category 36: NA. These funds were removed from the database.

$$R_{ij} = C_0 + \beta_1 * DGR1_{i0} + \beta_2 * DGR2_{i0} + \beta_3 * DGR3_{i0} + \beta_4 * DGR4_{i0} + \varepsilon_{ij}\,, \qquad (7)$$

where: C_0 is the constant (the performance of a five-star fund at time *0*), R_{ij} is a return for fund *i* at time *j*; $DGR4_{i0}$ is a binary dummy variable indicating whether fund *i* is a four-star fund at time *0* ; $DGR3_{i0}$ is a similar dummy variable, only signifying whether fund *i* is a three-star fund at time *0*; the same is true for $DGR2_{i0}$ as it indicates whether fund *i* is a two-star fund at time *0*; $DGR1_{i0}$ is a binary dummy variable as well, that points out whether fund *i* is a one-star fund at time *0* and ε_{ij} is the error term.

In the equation above, the five star fund group is taken as a reference for the other fund groups as the five star fund group forms a ceiling which no fund can surpass. The return of the five star fund group in equation 17 is similar to C_0 (since all the betas are 0), where, in the case of a fund belonging to a different star group, the applicable beta is added to C_0. Intuitively, if the rating system developed by Morningstar were to be completely correct, all betas should be negative, as a four star fund should not outperform a five star fund. Furthermore, assuming that Morningstar's rating system is flawless in it predictions of future performance, $\beta_1<\beta_2<\beta_3<\beta_4$ should hold, as a one star fund should not outperform a two star fund either.

While the methodology described in the paragraph above in itself could yield some interesting results, these results become far more interesting when different out of sample periods are used. In doing so, one is not only able to assess the performance of Morningstar's rating system, but is also capable of seeing shifts in this performance, as the out of sample period increases. Thus, it becomes evident whether the ratings assigned by the rating system, are valid for a certain period. This can be of great value to investors, as when a fund receives a five star rating, that fund is likely to outperform others for the coming period.

4.3 Predictive Performance

In this section, the predictive performance of Morningstar's mutual fund rating system will be tested. Due to its high importance to investors (see e.g. Sirri and Tufano, 1998; Del Guercia and Tkac, 2001; and Capon, Fitzsimons and Prince, 1996) the predictive abilities of a rating system such as Morningstar's should be as high as possible.

4.3.1 Descriptive Statistics

From the dataset, all funds that have a star rating on xxxx03 were selected. For March 1995, this amounts to 2431 funds. These funds were divided into five sub-groups according to the star rating on the sample creation date. This implies the creation of the following groups:

Table 5: Sample Characteristics

199503	**1-star**	**2-star**	**3-star**	**4-star**	**5-star**
N=2431	184	501	864	658	224
100%	7.57%	20.61%	35.54%	27.07%	9.21%
199603	**1-star**	**2-star**	**3-star**	**4-star**	**5-star**
N=3364	293	695	1203	869	304
100%	8.71%	20.66%	35.76%	25.83%	9.04%
199703	**1-star**	**2-star**	**3-star**	**4-star**	**5-star**
N=4755	438	1080	1713	1099	428
100%	9.21%	22.71%	36.03%	23.11%	9.00%
199803	**1-star**	**2-star**	**3-star**	**4-star**	**5-star**
N=6039	534	1435	2145	1386	539
100%	8.84%	23.76%	35.52%	22.95%	8.93%
199903	**1-star**	**2-star**	**3-star**	**4-star**	**5-star**
N=6901	664	1583	2455	1532	667
100%	9.62%	22.94%	35.57%	22.20%	9.67%
200003	**1-star**	**2-star**	**3-star**	**4-star**	**5-star**
N=7991	777	1775	2835	1788	816
100%	9.72%	22.21%	35.48%	22.38%	10.21%
200103	**1-star**	**2-star**	**3-star**	**4-star**	**5-star**
N=8989	829	2022	3152	2098	888
100%	9.22%	22.49%	35.07%	23.34%	9.88%
200203	**1-star**	**2-star**	**3-star**	**4-star**	**5-star**
N=9879	862	2179	3631	2227	980
100%	8.73%	22.06%	36.75%	22.54%	9.92%
200303	**1-star**	**2-star**	**3-star**	**4-star**	**5-star**
N=11673	1110	3005	4148	2475	935
100%	9.51%	25.74%	35.53%	21.20%	8.01%
200403	**1-star**	**2-star**	**3-star**	**4-star**	**5-star**
N=13104	1226	3260	4661	2906	1051
100%	9.36%	24.88%	35.57%	22.18%	8.02%

For each fund in the group, the return and the category were selected on a monthly basis. The table above shows that number of mutual funds has quintupled over the course of a 10 year period. Furthermore, the distribution as shown in figure 1 does not hold for the samples described in table 5, as the tails of the distribution of the samples described in table 5 are not as fat as they should be. An interesting note regarding the distribution found in the table above is that there is no difference between bull and bear markets. One might expect the distribution to be skewed to the right in the case of a bull market and skewed to the left in the case of a bear market, but according to table 5, this is not the case.

Over the course of the sample (199503-200403), Morningstar has changed its rating system. Up to 199609, when Morningstar introduced the Morningstar Categories, all funds were treated as if belonging to a single group[11]. In 199610, the funds were grouped according to the four broad asset classes that formed the basis of the ratings, until this changed to the 64 categories in 200207. The results of the predictive performance analysis on the three rating methods will be discussed in turn.

4.3.2 Single Group

This section will analyse the predictive performance of Morningstar's rating system if it were to consider all mutual funds as a single group. Although Morningstar classified funds by using their four broad asset classes during 1995 and 1996, the lack of data in the database bought from Morningstar requires for an analysis based on a single group. Even though this classification never found its way into practice, it does serve as a proper benchmark to compare results of other rating systems against.

Table 6: Regression Results on Single Group Sample

Sample	C_0	β_1	β_2	β_3	β_4	R^2	F-Stat
One year							
1995	1.5125***[12]	-.1861*	-.0868*	-.0226	.0007	.0006	4.300***
1996	.8329***	-.3118***	-.1138**	.0728	.1416***	.0024	24.483***
Three Year							
1995	1.2543***	-.5948***	-.1513***	.0130	.0432	.0033	73.361***
1996	.8414***	-.6041***	-.1036***	.1008***	.1077***	.0020	59.423***
Five Year							
1995	1.0461***	-.4638***	-.0843**	.0188	-.0151	.0008	30.007***
1996	.7984***	-.4512***	-.1425***	-.0324	-.0041	.0008	37.946***
Ten Year							
1995	.7349***	-.1590***	-.0942***	-.0499*	-.0296	.0001	6.492***

The table above lists the base year of the sample for three different out of sample periods. The 10 year out of sample period only has the 1995 sample, as there is not enough data for such an out of sample period for a sample that starts in 199603. Each combination shows the constant and the four betas estimated by using equation 7. Furthermore, the R^2 and the absolute F-Statistic are stated in order to provide data on the significance of the regression as a whole.

[11] The database obtained from Morningstar does not provide information on which of the four broad asset classes a fund is in prior to October 1996

[12] *** indicates significance at the 1% level, ** indicates significance at the 5% level and * indicates significance at the 10% level.

When looking at the table above, it becomes clear that the rating system employed by Morningstar is excellent at predicting underperformance. All β_1s are significant at the 1% level and have the correct sign. The same can be said for the β_2s; they are all significant (albeit some at the 5% and 10% level), have the correct sign, and are larger than the respective β_1s. This however, is where the praising ends. Of the β_3s, only three have the correct sign, of which only one is significant at the 10% level. This implies that, according to the rating system, there is hardly a difference in performance between three star and five star rated funds. When looking at four star rated funds, the situation becomes even worse. There are three β_4s that have the correct sign, none of which is significant at even the 10% level, furthermore, the β_4s that are significant (at the 1% level), are those with an incorrect sign. This would mean that in two occasions, four star rated funds achieve significantly higher returns than five star rated funds. These results are in line with those found by Blake and Morey (2000)

In addition to the qualities of the rating system as a whole, table 6 shows the performance of the rating system across different out of sample periods. When looking at the last column of table 6, one can see that all regressions are highly significant, but that the out of sample period of three years (and to a lesser extent, the out of sample period of five years) exceeds all others in terms of absolute F-stat values.

When rating mutual funds as a single group, the rating system is perfectly able to predict the future underperformance of one and two star rated funds, while it cannot distinguish between three, four and five star rated funds.

4.3.3 Four Broad Asset Classes

After introducing the 64 categories in 199610, the funds could be traced back to the asset class they belonged to. This changed the situation found in the analysis above. The results of the predictive performance analysis on the four broad asset classes will be discussed per asset class.

U.S. Stock

This section analyses the predictive performance of mutual funds classified as U.S. Stock, or domestic equity. Table 8 shows the regression results of equation 7 on different out of sample periods for mutual funds belonging to the U.S. Stock asset class.

Table 7: Regression Results on U.S. Stock Asset Class

Sample	C_0	β_1	β_2	β_3	β_4	R^2	F-Stat
One Year							
1997	2.3490***	-.6312***	-.3053**	-.2161*	-.1484	.0011	6.036***
1998	.8189***	-.9677***	-.3182	-.2731	-.1771	.0007	5.381***
1999	2.1064***	2.2227***	.5768***	-.7747***	-.5849***	.0237	213.563***
2000	-3.3745***	4.5315***	4.2107***	3.3611***	2.0110***	.0349	384.193***
2001	-.6348***	-.6682***	.0325	.1336	.1766*	.0014	17.892***
2002	1.596***	-.7928***	-.3570***	-.0457	.0658	.0019	28.387***
Three Year							
1997	1.1560***	1.1113***	.8081***	.2637***	.1624	.0027	47.169***
1998	.7007***	.3034***	.1701*	-.1046	-.0824	.0004	8.465***
1999	-.1029	.9437***	.8305***	.3720***	.1601**	.0023	60.014***
2000	-2.4561***	2.5625***	2.1402***	1.7579***	1.0063***	.0132	424.489***
2001	.2169***	-.2196***	-.0079	-.0350	-.0036	.0001	4.436***
2002	.9664***	-.6871***	-.5515***	-.4149***	-.2101***	.0017	75.442***
Five Year[13]							
1997	.8152***	-.1042	-.0570	-.1049	-.0794	.0000	.608
1998	-.0096	.0314	.0094	-.0442	-.0755	.0000	1.066
1999	.0623	.8202***	.6960***	.3004***	.1254**	.0020	88.948***
2000	-.7548***	1.7596***	1.3379***	.9602***	.5283***	.0078	420.076***

Where the previous rating system proved excellent in predicting underperformance, the rating system that diversifies funds amongst four broad asset classes has great difficulties in achieving the same result, as can be seen in the table above. Despite the fact that out of 16 β_1s, 14 are significant at the 1% level, only 6 of these 14 bear the correct sign. This means that in different base years, over several out of sample periods, 50% of all one star rated funds significantly outperform five star rated funds; Whereas five star rated funds significantly outperform one star rated funds in 37.5% of the cases. A result such as this makes the rating system just as accurate as flipping a coin, unless the coin has two heads of course.

When looking at β_2, the situation does not improve. Out of 16 betas, 11 are significant, but only three have the correct sign, resulting in two star rated funds significantly outperforming five star funds in, again, 50% of the time, in different base years, over different out of sample periods. Moreover, the regression results in table 8 show that five star funds significantly outperform two star funds in only 18.75% of the cases.

Out of the 16 β_3s, nine are significant, with only three bearing the correct sign, meaning that five star funds significantly outperform three star funds 18.75% of the time. The opposite

[13] The five year period runs from 1997 through 2000. As the database contains information up to 200509, this makes it impossible to construct five year periods for 2001 and further.

holds for 37.5% of the cases, the results of the remaining 43.75% show no significant difference between three star and five star rated funds.

Of the 16 β_4s, eight are significant, with only two bearing the correct sign. This implies that five star funds outperform four star funds in 12.5% of the occasions, the opposite being true for 37.5% of the cases.

When taking the different out of sample periods into consideration, table 7 illustrates that the results of all one- and three year regressions are significant at the 1% level. For the five year out of sample period, only two out of four regressions are significant, nevertheless at the 1% level. Further distinguishing between the one- and three year out of sample periods, it shows that the absolute F-Statistic of the one year out of sample is higher than its thee year out of sample counterpart, although this difference does not imply much, as they are both significant at the 1% level.

The above clearly shows that for the U.S. Stock category, the rating system employed by Morningstar offers no added value in terms of predicting mutual fund returns, as there is no occasion where the rating system would outperform a random walk. Unlike the system that classifies all mutual funds as a single group, the system using the U.S. Stock classified funds to base a rating upon is even not able to predict underperformance.

International Stock

Since the results of the U.S. Stock category are not impressive, perhaps Morningstar's rating system is better at predicting performance for mutual funds classified as International Stock. Table 8 lists the results of equation 7 for this broad asset class.

The table below shows the estimates of equation 7 for different base years with multiple out of sample periods. A quick glance shows that for the International Stock category, the rating system is perfectly able at estimating betas significantly different form zero for one and two star rated funds. A closer look however, learns that the signs of these coefficients are not always correct.

Table 8: Regression Results on International Stock Asset Class

Sample	C_0	β_1	β_2	β_3	β_4	R^2	F-Stat
One Year							
1997	1.7024***	-4.0776***	-3.0026***	-.9852***	-.3018	.0415	60.882***
1998	.3889	-1.1582***	-1.5913***	-.3974	-.1150	.0066	14.183***
1999	3.7043***	-.5471*	1.0378***	-.5424**	-.9852***	.0124	34.168***
2000	-3.4923***	1.6272***	1.3848***	1.9241***	1.5771***	.0093	31.086***
2001	-1.2119***	1.7216***	.5577***	-.3215**	-.0059	.0101	39.353***
2002	-1.1401***	-.4279*	-.3377**	-.4023***	-.1957	.0005	2.239*
Three Year							
1997	1.5301***	-1.8700***	-.9465***	-.2213	.0370	.0059	25.038***
1998	.6096***	-.4772**	-.3190*	-.2342	.0011	.0005	2.963**
1999	.0612	.8712***	.4884***	-.1241	-.0972	.0025	20.472***
2000	-2.4061***	1.8882***	1.1627***	.9502***	.7006***	.0061	60.269***
2001	.2511***	1.2107***	.4213***	-.1209	-.0316	.0048	55.331***
2002	1.3348***	-.2674**	-.1493*	-.1518*	-.1162	.0001	1.582
Five Year							
1997	.4298***	-.7511***	-.6300***	-.3579***	-.0013	.0015	10.812***
1998	-.2530**	.5294***	.0969	-.1098	.0195	.0006	6.549***
1999	.3590***	1.0423***	.5689***	-.0125	-.0465	.0032	42.674***
2000	-.4942***	1.2015***	.9304***	.5404***	.3957***	.0036	59.496***

For the International Stock category, all 16 β_1s are significant, with 8 of these betas showing the correct sign. This implies that just as with the U.S. Stock category, in only 50% of the cases, five star funds significantly outperform one star rated funds. This situation deteriorates for the β_2s, where from the 15 significant betas; seven have the correct sign, implying that in only 43.75% of the cases, five star funds outperform two star funds. However, out of the nine significant β_3s, six show the correct sign, thereby making sure that in 37.5% of all occasions, five star funds outperform three star funds, whereas in only 18.75% of the cases, three star funds outperform five star funds. Although this is slightly better than the previous results (at least the percentage of five star funds outperforming three star funds is higher than the percentage of three star funds outperforming five star funds), it still does not beat a random walk, as in 43.75% of the cases there is no significant difference between the performance of five- and three star rated funds. Continuing with the β_4s, four out of 16 betas are significant, of which only one has the correct sign. This means that in 6.25% of the occasions, five star funds significantly outperform four star funds, whereas the opposite holds in 18.75% of the cases. This shows that there is hardly a difference between four and five star rated funds in terms of performance, as in 75% of the cases there is no significant difference.

Despite the fact that almost all regressions in all out of sample periods are significant at the 1% level (the exception being the three year 2002 regression), table 8 shows that the rating system best predicts performance for International Stock mutual funds for an out of sample

period of one year. In this period, 12 out of 19 significant betas bear the correct sign. Once again, out of a total of 24 betas, this in only 50%.

The analysis above concludes that the rating system used by Morningstar does, at best, equal the performance of a random walk. However, it not able at outperforming this random walk.

Taxable Bond

This section discusses the results of equation 7 on those mutual finds classified as Taxable Bond. An important difference between this category and the previous two is that the funds in the Taxable Bond category mainly have a fixed income portfolio, whereas the portfolios of funds in the U.S. Stock and International Stock categories are largely made up out of equity.

Table 9: Regression Results on Taxable Bond Asset Class

Sample	C_0	β_1	β_2	β_3	β_4	R^2	F-Stat
One Year							
1997	.8391***	-.2411***	-.1405***	-.1571***	-.1853***	.0016	5.695***
1998	-.3340***	.7821***	.7171***	.7184***	.5524***	.0112	47.194***
1999	.2637***	.2396***	-.2872***	-.2019***	-.0844***	.0186	85.798***
2000	.6863***	-.5006***	-.0696	.1093**	.0745	.0121	61.197***
2001	.4876***	-.8825***	-.4255***	.0884	-.0054	.0093	49.768***
2002	.4586***	.0072	.1710	.3947***	.1182	.0008	4.262***
Three Year							
1997	.4136***	-.1153***	-.0690**	-.0441	-.0318	.0003	3.579***
1998	.0899***	.1758***	.3216***	.3519***	.2597***	.0043	54.002***
1999	.3973***	-.1280***	-.0708**	.0340	.0563*	.0010	12.956***
2000	.4683***	-.0710	.1502***	.1387***	.1095**	.0005	7.857***
2001	.3689***	.2464***	.2393***	.2205***	.1028*	.0005	8.174***
2002	.3824***	.5827***	.3894***	.0941**	-.0367	.0038	63.292***
Five Year							
1997	.2719***	.1088***	.2127***	.2067***	.1445***	.0010	18.293***
1998	.1001***	.3274***	.4605***	.3934***	.2712***	.0024	50.922***
1999	.4299***	.1759***	.0324	.0784**	.0283	.0003	7.427***
2000	.4080***	.1999***	.1762***	.1081***	.0664**	.0005	12.551***

When comparing the table above with the table of the previous two equity categories, it becomes clear when predicting performance of Taxable Bond funds; Morningstar's rating system is able to predict the performance of three and four star rated funds, whereas in the previous two equity categories, the number of significant β_3s and β_4s were limited.

For the β_1s, the Taxable Bond category does not prove to be different from the previous categories. Out of 14 significant betas, only five (31.25%) bear the correct sign, three of which are in the one year out of sample period. The β_2s show the same story, out of 13 significant betas, five (31.25%) show the correct sign. Once again, three of these are in the

one year out of sample period. When looking at the β_3s, the situation worsens a bit. Out of 13 significant betas, only two (12.5%) show the correct sign. Both these betas are in the one year out of sample period. Out of 10 significant β_4s, two (12.5%) bear the correct sign. It comes as no surprise that these two are in the one year out of sample period.

Despite the fact that all regressions are significant at the 1% level, it is clear that the one year out of sample period is the best investment period to use when considering investing based upon Morningstar's rating system. The absolute F-Stat figures as shown in the last column of table 9 are also the highest for the one year out of sample period. Nevertheless, as indicated by the small number of correct betas, the predictive performance for the Taxable Bond category based upon Morningstar's ratings is extremely weak. Where the U.S. Stock and International Stock asset classes were at least able to equal the performance of a random walk, investing based upon Morningstar's ratings in funds classified as Taxable bond will never exceed the returns yielded by a random walk. This is odd, as the number of significant betas is higher for the Taxable Bond category than it is for the equity categories, but the number of betas with the correct sign is lower for the Taxable Bond category, compared to the two equity categories. Perhaps that, due to the characteristics of a bond, the earnings of portfolios consisting of bonds are less volatile, resulting in a more significant beta. This would explain the high number of significant betas. Moreover, in order for funds with a one star rating to raise performance, more risk has to be taken on. This implies adding fixed income products with a lower credit rating (and therefore, a higher return) to the portfolio. When these firms do not default on their loans, a higher return is realised. This would explain the betas with incorrect signs.

Municipal Bond

The last of the four broad asset classes is the Municipal Bond asset classes. The main difference between the Taxable Bond category and the Municipal Bond category is that returns from the Taxable Bond category are subject to capital gains taxes, while municipal bonds have an exemption from this sort of tax. This gives investing in municipalities a tax advantage over investing in corporations. The table below lists the regression results of equation 7 on the Municipal Bond asset class.

Table 10: Regression Results on Municipal Bond Asset Class

Sample	C_0	β_1	β_2	β_3	β_4	R^2	F-Stat
One Year							
1997	.6609***	.0358	.0293	.0181	-.0028	.0002	.619
1998	.4275***	-.0457*	-.0317	-.0189	-.0044	.0004	1.611
1999	-.2918***	-.1149***	-.0591**	-.0181	.0243	.0021	9.988***
2000	.6749***	.2365***	.2280***	.1950*	.1759***	.0025	12.535***
2001	.4770***	-.0511	-.0314	-.0064	.0068	.0003	1.607
2002	.5167***	-.0262	-.0200	.0069	.0076	.0001	.467
Three Year							
1997	.2960***	-.0788***	-.0546***	-.0350**	-.0151	.0005	5.429***
1998	.3131***	-.0213	-.0021	.0126	.0244	.0002	2.282*
1999	.3357***	-.0440**	-.0028	.0067	.0139	.0002	2.838**
2000	.5418***	.0696***	.0823***	.0766***	.0779***	.0003	5.073***
2001	.4693***	.0004	.0017	.0395	.0186	.0001	.901
2002	.3736***	.0179	.0012	.0337	.0047	.0000	.645
Five Year							
1997	.4221***	-.0167	-.0050	.0036	.0086	.0001	.918
1998	.3796***	-.0136	.0025	.0122	.0179	.0001	1.5935
1999	.3963***	-.0311*	-.0018	.0055	.0099	.0001	1.949*
2000	.4413***	.0555***	.0634***	.0569***	.0556***	.0002	4.532***

When looking at table 10, it is immediately clear that hardly any coefficients are significantly different from zero. This would imply that the division of funds classified as Municipal Bond into the five star groups makes no sense, as the performance of all Municipal Bond funds is alike, thereby immediately questioning the existence of the Municipal Bond asset class.

Despite the lack of predictive performance in the previous categories, at least they had a large number of significant betas. For the Municipal Bond category this no longer holds. Out of only eight significant β_1s, five (31.25%) bear the correct sign. For the β_2s, this is even worse, out of five significant betas, there are only two (12.5%) negative ones.

Just as with the previous categories, the predictive performance as a whole of the rating system decreases whenever funds get closer to a five star rating. Out of four significant β_3s, only one (6.25%) shows the correct sign. Where other categories had some correct β_4s, the Municipal Bond category only has three significant β_4s, with zero of them being correct.

When looking at the F-stat values, it becomes clear that the three year out of sample period is the best period to use when investing in Municipal Bond funds based upon Morningstar's rating system. This three year out of sample period has four significant regression results out of a total of six regressions.

The results of tables 7 through 10 are quite disastrous for Morningstar's rating system. However, this is not the first time that the predictive performance of the Morningstar's rating system is analysed. When Blake and Morey (2000) analysed the predictive performance of Morningstar's mutual fund rating system, they only analysed the U.S. Stock[14] category. They came to the conclusion that the rating system was able to predict underperformance, but the system was unable to predict superior performance amongst funds. Nevertheless, the above results, even for the U.S. stock category, suggest otherwise. Morningstar's rating system is neither able to predict superior performance, nor is it able to predict inferior performance. The fact that at best[15] 50% of all β_1s are significant and bear the correct sign illustrates this point.

Regarding the different out of sample periods, the analysis concludes that the one year out of sample period is the period for which the rating is valid. The three and five year out of sample periods produce lower absolute F-Stat values. This implies that, in line with Morey (2003), funds are not able to maintain a high star rating. Furthermore, it is a signal to investors to adjust their investment horizon to this one year period when considering an investment in mutual funds based on Morningstar's rating system.

Morningstar was not unaware of the low predictive performance of its system based on four broad asset classes and introduced an improved rating system in 200207[16]. The predictive performance of this system will be analysed in the next section.

4.3.4 Rating Funds Based on Categories

The analysis on the rating system introduced in 200207 consists of two samples, both with a one year out of sample period. Due to the lack of available data, it is impossible to construct longer out of sample periods. The results of equation 7 on this new rating system are grouped according to the four broad asset classes that contain the newly introduced categories. The results can be found in the sections below. Not all 64 categories are present at all times as Morningstar continuously adds and changes categories. Due to the fact that 200303 only has 48 categories, there is no point in adding the other 18 listed in 200403, as that would remove a comparison of categories over different samples from the analysis.

---

[14] They named this category "domestic equity"
[15] See the β_1s of the International Stock category in table 9
[16] Improved according to Morningstar that is.

U.S. Stock

This section will discuss the results of equation 7 based on the categories that together make up the U.S. Stock asset class. From table 11 on the next page, it becomes clear that overall; the ratings of the new rating system do not have any predictive performance abilities. Exceptions are there for the Large Blend and Large Growth categories, and, to a lesser extent, Small Blend, Small Growth and Specialty Health. The fact that both Small and Large Mid-Cap and Growth categories produce coefficients that are significantly different from zero does not come as a big surprise. As the earnings of Value funds are rather stable[17], there is not a lot of difference between the different funds in terms of performance. For Blend and moreover, Growth funds, this difference in earnings is present. It is due to this difference that a rating system is better at: a) assigning funds over the five star groups and b) predicting the performance of these star groups, largely due to the increased volatility of the Blend and Growth portfolios over the volatility of the Value portfolios. Nevertheless, only the coefficients of the Large Growth and Small Growth funds in the 200403 sample show the correct sign.

Overall, out of 34 β_1s, 11 are significantly different from zero, with five of these (14.71%) bearing the correct sign. This means that in the remaining 85.29%, there is no difference in the performance of one and five star rated funds, or that one star rated funds significantly outperform five star rated funds, over the course of a one year out of sample period. The same can be said for the β_2s, out of 34, only 11 are significantly different from zero, with again, five of these (14.71%) bearing the correct sign. As expected, the number of significant coefficients is lower for the β_3s. Out of 34 betas, six are significantly different from zero, with one of these six (2.9%) bearing the correct sign. This further confirms the conclusion found in the previous section and in Blake and Morey (2000), that Morningstar's rating system is at best able to predict performance for lower rated funds. The β_4s show a situation in which out of the 34 betas, two are significant, with only one (2.9%) bearing the correct sign.

When looking at the absolute F-Stat values, the results confirm that the rating system is able to predict the out of sample performance of mutual funds belonging the Large Blend, Large Growth, Small Blend, Small Growth and Specialty Health categories.

[17] The CAPM β of Value Funds tends to be lower than 1.

Table 11: Regression Results on Categories in U.S. Stock Asset Class

Category	Sample	C_0	β_1	β_2	β_3	B_4	R^2	F-Stat
Large Value	2003	2.7036***	.1575	.1982*	.1942*	.1537	.0004	.900
	2004	.8721***	-.1005	-.1251	-.0879	-.0469	.0002	.599
Large Blend	2003	2.1925***	.4760***	.3972***	.4367***	.4324***	.0016	5.266***
	2004	.6485***	-.2290**	-.1894**	-.1153	-.0753	.0006	2.212*
Large Growth	2003	2.5555***	.5059***	.0926	-.0413	-.0895	.0030	8.452***
	2004	.5293***	-.6232***	-.3787***	-.3355***	-.2429*	.0020	6.544***
Mid-Cap Value	2003	2.9502***	.2212	.3301	.4242*	.1786	.0019	.988
	2004	1.2205***	-.0206	-.1764	-.1962	-.2078	.0006	.409
Mid-Cap Blend	2003	3.2151***	.6507*	.2423	.0725	-.0128	.0035	1.922
	2004	1.1387***	-.4024	-.2687	-.2560	-.1510	.0009	.654
Mid-Cap Growth	2003	3.1396***	.1185	.1644	.1136	-.0676	.0006	1.030
	2004	.7977***	-.2458	-.2073	-.1858	-.0385	.0004	.909
Small Value	2003	3.5874***	.5901	.5197*	.3937	.2319	.0021	1.162
	2004	1.0883**	.0533	.1243	.1116	.1844	.0001	.079
Small Blend	2003	3.3646***	.8237**	.7873***	.5109*	.4741	.0030	2.467**
	2004	1.2058***	-.4264	-.3310	-.0898	-.0521	.0012	1.207
Small Growth	2003	3.8303***	.7381**	.1868	.0875	-.0020	.0018	2.703**
	2004	.9354***	-.9841***	-.5170**	-.3963	-.2173	.0021	3.686***
Specialty Communication	2003	3.6417***	.7242	.5243	-.3037	-.1737	.0093	.780
	2004	.4563	.1521	.1726	-.0437	.5307	.0018	.1883
Specialty Financial	2003	3.1019***	-.1073	.4621	.2410	-.0502	.0033	.852
	2004	.5365*	-.2724	-.2432	-.1471	.3256	.0049	1.392
Specialty Health	2003	2.8706***	1.2183*	.2605	-.5024	.1253	.0189	5.516***
	2004	.2474	-1.1592***	-.6505*	-.1862	-.0230	.0078	3.665***
Specialty Natural Resources	2003	3.4017***	-.9569	-.4894	-.7125	-.6973	.0013	.281
	2004	3.4688***	-.1946	-.5387	-.5947	-.3539	.0012	.250
Specialty Real Estate	2003	2.9872***	.3635*	.4018**	.3023*	.1555	.0049	1.920
	2004	1.6226***	.0104	.0464	-.0596	.0223	.0000	.021
Specialty Technology	2003	3.2377***	.9303	.9005*	.5952	.4237	.0023	1.278
	2004	.2946	-.7671	-.9762**	-.6379	-.4917	.0017	1.412
Specialty Utilities	2003	2.5804***	.1589	-.0463	.0707	-.1435	.0010	.218
	2004	1.6788	.0007	-.1167	-.1112	.2739	.0044	.998
Convertibles	2003	2.4946***	-1.1717***	-.2703	-.3336	-.3413	.0099	1.815
	2004	.2152	-.5181	.0958	.0782	.1167	.0026	.483

International Stock

In this section, the results of equation 7 on the categories within the International Stock asset class will be discussed. Table 12 on the next page shows these results. It immediately shows that some categories do not have values for all betas. This is due to the fact that about 1% of the fund-months are missing. When using equation 7 on the four broad asset classes, this 1% can be ignored as each sample contains enough fund-months. Nevertheless, when basing the regressions on the multiple smaller categories, each containing less fund-months, the missing of a single fund month has a larger impact on the results. The fact that certain coefficients cannot be estimated is a clear example of this.

Despite the missing values, table 12 clearly shows that the only category in the International Stock asset class for which the rating system can produce meaningful results is the Europe Stock category. A quick analysis of the betas in the International Stock category shows that out of 18 β_1s, two are significantly different from zero, with one (5.56%) bearing the correct sign. The results are a little better for the β_2s, out of 18 betas; two are significantly different from zero, both having the correct sign (11.11%). There is only one (5.56%) significant β_3 out of the group of 18, but this beta has the correct sign. The results of equation 7 do not show a significant β_4. When looking at the absolute F-Stat values, the results indicate that Morningstar's rating system is excellent at predicting the returns of mutual funds classified as Europe Stock, almost all coefficients are significantly different from zero and bear the correct sign. Furthermore, the absolute F-Stat values indicate that the system has some predictive performance capabilities for funds classified as World Stock.

Table 12: Regression Results on Categories in International Stock Asset Class

Category	Sample	C_0	β_1	β_2	β_3	β_4	R^2	F-Stat
World Stock	2003	3.499*	.0558	-.1125	-.2133	-.1238	.0007	.632
	2004	1.1147*	-.3093	-.4346**	-.1479	.0286	.0027	2.640**
Diversified Emerging Markets	2003	5.0221*	-.2509	-.3800	-.3851	-.5065	.0012	.525
	2004	2.4458*	-.6502	-.6306	-.5004	-.2070	.0018	.885
Latin America Stock	2003	4.8383*	.6189	.3626	.3006	.2285	.0015	.123
	2004	3.363***	.0608	.1629	.1335	-.1850	.0005	.036
Europe Stock	2003	3.5350*	.4601	.4744	.2386	.4258	.0011	.483
	2004	2.3125*	-1.1096**	-1.0177**	-.9337**	-.5155	.0065	2.411**
Japan Stock	2003	3.2919*	.4908	-.3885	-.4728	NA[18]	.0035	.636
	2004	1.7758	-1.0023	-.8710	-.5266	-.4069	.0020	.266
Pacific/Asia ex-Japan Stock	2003	3.9552*	NA[18]	.6320	.4963	.7924	.0013	.406
	2004	1.5264**	-.4362	-.4331	-.7066	-.5440	.0012	.274
Diversified Pacific/Asia	2003	3.1956*	.1378	-.0097	.3208	.5557	.0037	.4409
	2004	1.1733*	.0431	-.1305	.0900	NA[18]	.0005	.064
Specialty Precious Metals	2003	4.2713*	-.8672	-.2119	-.4545	.6113	.0021	.243
	2004	.3633	.4242	-.0274	.0748	.1378	.0001	.013
International Hybrid	2003	1.9533*	.9425*	.4497	.3919	.2185	.0065	1.158
	2004	1.0842*	-.1077	-.5927	-.1555	-.0577	.0049	.763

* indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicates significance at the 1% level

[18] There is no fund with such a rating in this category

Taxable Bond

This section discusses the result of equation 7 on the categories contained in the Taxable Bond asset class. Table 13 on the next page shows the coefficients for each of the categories. When looking at the β_1s, it becomes clear that out of a total of 22 betas, 10 are significantly different from zero up to a 10% significance level. Out of these 10 betas, 7 (31.81%) bear the correct sign; a heavy improvement over the last two asset classes. For the β_2s, the situation is even better; out of 22 betas, 12 are significantly different from zero, with nine (40.91%) of these showing the correct sign. As expected, the situation deteriorates a bit for the β_3s, where out of 22 betas; eight are significantly different from zero, with six (27.27%) of these bearing the correct sign. Despite the amount of significant betas in the first three groups, the β_4s show two significant betas out of a total of 22, where one (4.55%) of these two has the correct sign.

When looking at the absolute F-Stat values, it becomes clear that out of the Taxable Bond sample, Morningstar's rating system has excellent predictive performance abilities for the Intermediate Government, Intermediate-Term Bond, Ultrashort Bond, High Yield Bond, Multisector Bond and International Bond categories. When considering the individual betas, this list changes to Intermediate Government, Short Government, Long-Term Bond, Intermediate-Term Bond, Multisector Bond and International Bond, as for those categories, almost all betas are significant, and $\beta_1<\beta_2<\beta_3$ holds. This makes Morningstar's rating system an excellent tool for investing in funds classified in one of these categories.

Table 13: Regression Results on Categories in Taxable Bond Asset Class

Category	Sample	C_0	β_1	β_2	β_3	β_4	R^2	F-Stat
Long Government	2003	.4680	-.5561	-.3650	-.2957	-.1011	.0022	.355
	2004	.3215	-.1440	.1247	.0331	.0150	.0009	.087
Intermediate Government	2003	.3139***	-.1582	-.1489*	-.1289	-.0869	.0010	.919
	2004	.2569***	-.2391***	-.1541***	-.1227**	-.0870	.0030	2.987**
Short Government	2003	.2454***	-.1773**	-.1433**	-.1123*	-.0663	.0045	1.760
	2004	.1063**	-.1037	-.1057*	-.0763	-.0544	.0023	1.011
Long-Term Bond	2003	.8125***	.1132	-.2489	-.2896	-.2076	.0051	1.348
	2004	.6325***	-.4103*	-.3933*	-.3751*	-.1605	.0071	1.474
Intermediate-Term Bond	2003	.4646***	.0837	-.0429	-.0578	-.0582	.0009	1.549
	2004	.2997***	-.1551***	-.1591***	-.1232***	-.0943**	.0015	3.459***
Short-Term Bond	2003	.2421***	.0342	-.0545	-.0521	-.0020	.0024	1.521
	2004	.0858**	-.0735	-.0686*	-.0421	-.0088	.0024	1.701
Ultrashort Bond	2003	.1698**	.9352***	.6366***	.0828	-.0375	.2208	61.573***
	2004	.1267***	-.0898**	-.0505	-.0261	-.0026	.0189	3.566***
High-Yield Bond	2003	1.323***	.4286***	.4611***	.4292***	.2978***	.0061	6.095***
	2004	.7524***	.0622	.0437	.0425	.0124	.0003	.3316
Multisector Bond	2003	.6380***	.7274***	.4439**	.4891***	.2149	.0116	5.563***
	2004	.7992***	-.3088*	-.2919**	-.2689**	-.1326	.0055	2.235*
Emerging Markets Bond	2003	1.9612***	-.1929	-.2275	-.0925	-.1229	.0008	.103
	2004	1.4480***	-.3733	-.4257	-.3050	-.3814	.0022	.515
International Bond	2003	.8131***	.0205	.1665	.1014	.0261	.0006	.2303
	2004	.9432***	-.5773**	-.4890***	-.3576**	-.1619	.0077	3.160**

Municipal Bond

This section discusses the results of equation 7 performed on the categories belonging to the Municipal Bond asset class. The estimated coefficients are depicted in table 15 on the next page. The results found in that table indicate that there is no predictive performance whatsoever for any categories in the Municipal Bond asset class, with the exception of the Muni Short and Muni Single State Intermediate categories. This might be due to the effect that the credit ratings of the individual states are the same, and, therefore, the interest paid on the bonds is the same. This would imply similar portfolios for all mutual funds classified in categories belonging to the Municipal Bond asset class. When the returns of the portfolios are similar, it is extremely difficult for the rating system to distinguish between a one star and a five star rated fund.

When looking at the betas, the following situation emerges. Out of 18 β_1s, three (16.67%) are significant, all bearing the correct sign. Out of 18 β_2s, just two (11.11%) are significantly different from zero, both showing the correct sign. There are no significant β_3s and β_4s. The only regression with an absolute F-Stat value significantly different from zero at the 10% level is that of the Muni Short category. Next to the Muni Short category, the Muni Single State Intermediate category has two significant betas as well, with, for both categories $\beta_1<\beta_2$.

When considering tables 11, 12, 13 and 15 as a whole, one of the distinguishing results found is the fact that out of 92 regressions, only 20 are significant up to a 10% level. When considering the different betas, the situation is as shown in the table below

Table 14: Summarised Regression Results of Ratings Based on Categories

Beta	Significant (-)	Significant (+)	Not Significant	% Correct
β_1	16	11	64	17.58%
β_2	18	9	65	19.57%
β_3	8	7	77	8.70%
β_4	2	2	86	2.22%

Table 15 shows the number of significantly positive and significantly negative betas for each rating group. Furthermore, it shows the number of betas that are not significantly different from zero and the percentage of betas with the correct sign. This value is based on the total number of betas, not just the number of significant betas.

Table 15: Regression Results on Categories in Municipal Bond Asset Class

Category	Sample	C_0	β_1	β_2	β_3	β_4	R^2	F-Stat
Muni National Long	2003	.6166***	.0412	-.0925	-.1117	-.0626	.0010	.889
	2004	.2685***	-.1474	-.1011	-.0680	-.0364	.0011	.835
Muni National Intermediate	2003	.5239	-.1060	.5977	-.0831	-.0897	.0011	.612
	2004	.1499**	-.1315	-.0586	-.0646	-.0208	.0013	.078
Muni Single State Ling	2003	.5271***	-.0537	-.0533	-.0426	-.0314	.0001	.141
	2004	.2478***	-.1127	-.0953	-.0503	-.0118	.0012	.988
Muni Single State Intermediate	2003	.5025***	-.1960*	-.1230	-.0795	-.0616	.0011	1.286
	2004	.2007***	-.1860**	-.1527**	-.0917	-.0841	.0021	1.623
Muni California Long	2003	.5231***	.0285	-.0690	-.0367	-.0228	.0003	.093
	2004	.2665**	-.0871	-.0551	-.0049	.0232	.0008	.295
Muni New York Long	2003	.5061***	.0664	-.0569	-.0328	-.0015	.0006	.155
	2004	.2583**	-.1293	-.0668	-.0501	-.0637	.0007	.173
Muni California Intermediate	2003	.5025***	-.1596	-.1117	-.0764	-.0829	.0006	.072
	2004	.1540	-.1052	-.0942	-.0539	-.0666	.0009	.137
Muni Short	2003	.3825***	-.2375**	-.1822*	-.1120	-.0456	.0083	2.328*
	2004	.0722	-.0873	-.0778	-.0612	-.0294	.0024	.650
Muni New York Intermediate	2003	.5494***	-.1266	-.0899	-.1477	-.1056	.0008	.069
	2004	.1838	-.1542	-.1043	-.0639	-.0604	.0014	.226

Notwithstanding the fact that the percentage of betas with a correct sign is appalling, the fact that the number of significant betas with the correct sign (first column) is greater than the number of significant betas with the incorrect sign (second column) can be seen as a positive result for the new rating system at first. However, constructing a similar table for the previous rating system shows the following:

Table 16: Four Broad Asset Classes: 1Y Summarised Regression Results

Beta	Significant (-)	Significant (+)	Not Significant	% Correct
β_1	13	7	4	54.17%
β_2	9	7	8	37.5%
β_3	8	6	10	33.33%
β_4	4	5	15	16.67%

Table 15 and 16 clearly show that, when looking at the same out of sample period, the old rating system surpasses the new rating system in terms of percentage of correct significant betas. When confronted with such a result, it is only natural that the question of rating system superiority arises. It is this question that will be answered in the next section.

4.4 Comparing Rating Systems

Due to criticism (e.g. Blake and Morey, 2000; Khorana and Nelling, 1998), Morningstar adjusted its rating methodology at the start of July 2002. Morningstar recently published a study (Kinnel, 2005) in which they analyse the difference between the two different rating systems. However, the article published by Morningstar is not clear on how the analysis was set up. Furthermore, Kinnel (2005) uses two samples that are 12 months apart to base his conclusion on. During these 12 months, exogenous changes might influence the ratings assigned by Morningstar and, therefore, make the analysis by Kinnel (2005) less accurate. In order to give investors a proper comparison both rating systems, this section will analyse and compare the rating system based on four broad asset classes with the rating system based on 64 categories.

4.4.1 Descriptive Statistics

This study will analyse whether the change of methodology has indeed increased the predictive abilities of the new rating system compared to the old. In order to do so, the predictive performance of the 200204 – 200206 samples is compared with the predictive performance of the 200207 – 200209 samples. Due to the law of averages, it is important to compare samples that are close to each other in time, as otherwise, the increasing number of funds result in a bias towards the younger samples. (i.e. Regressions based on more fund-months are, with all other things being equal, likely to yield results that are more significant

than regressions that are based on fewer fund-months.) The specifics of each sample can be found in the table below.

Table 17: Sample Characteristics

200204					
Total	1-star	2-star	3-star	4-star	5-star
N=9935	888	2220	3565	2270	992
100%	8.94%	22.35%	35.88%	22.85%	9.98%
200205					
Total	1-star	2-star	3-star	4-star	5-star
N=9901	906	2182	3548	2269	996
100%	9.15%	22.04%	35.83%	22.92%	10.06%
200206					
Total	1-star	2-star	3-star	4-star	5-star
N=9901	940	2548	3517	2086	810
100%	9.49%	25.73%	35.52%	21.07%	8.18%
200207					
Total	1-star	2-star	3-star	4-star	5-star
N=9950	933	2573	3548	2112	784
100%	9.38%	25.86%	35.62%	21.23%	7.87%
200208					
Total	1-star	2-star	3-star	4-star	5-star
N=9999	947	2591	3587	2084	790
100%	9.47%	25.91%	35.87%	20.84%	7.90%
200209					
Total	1-star	2-star	3-star	4-star	5-star
N=10064	957	2638	3562	2130	777
100%	9.51%	26.21%	35.39%	21.16%	7.72%

As can be seen in table 17, all samples contain information on a rather similar number of funds. As a comparison, the 200303 sample contains information on 11,673 funds. Furthermore, the ratings are consistently distributed across the samples, with the tails of the distribution being a little more flat than they should be according to figure 1. Just as with the 199503 sample, the return and the category for each fund are selected on a monthly basis and it is assumed that investors randomly invest into funds of the same category once a fund disappears from the database.

4.4.2 Four Broad Asset Classes vs. 64 Categories

Since the analysis on the four broad asset classes contains four regressions per out of sample period, per sample, whereas the analysis on the rating system based on categories contains 48 regressions per out of sample period, per sample, these two cannot be compared like that. In order to be able to check for a shift in predictive performance, the samples prior to 200207 have been grouped using the categories of the new rating system (which were present since October 1996, but not used as a basis for the ratings), while the ratings itself were based on

the output of the old rating system. Since this might result in categories not having all star ratings in a certain sample, the only way to compare the different rating systems is by focussing on the F-statistic. The absolute values of the F-Statistic are of no use here, as there is a tremendous absolute difference between an F-Stat value of 5 and 100, while they are both significant at the 1% level. Therefore, in order to properly compare the rating systems, the probability of the F-Statisctic (pF-stat) will be used. Equation 7 is used to estimate the results.

A summary of the equation results can be found in table 18. Results are obtained by using equation 7 for each category on both a one year and a three year out of sample periods. This implies that all results are based on 96 regressions.

Table 18: Average p(F-stat) Values for Different Samples

Sample	200204	200205	200206	200207	200208	200209
Average P(F-Stat)	0.5558	0.6149	0.6280	0.6522	0.6275	0.6267

The table above shows the average p(F-stat) values for the samples used to compare the two rating systems. This table illustrates that both systems would, on average, not yield results significantly different from zero. Nevertheless, a quick glance shows that the average p(F-stat) values for the regressions based on the old system are a little lower. While table 20 shows the average p-values for the different samples, the effect found is further illustrated in table 21, which lists the differences between the different samples, by subtracting the average value of the regressions based on the new system from the average value of the regressions based on the old system. A negative value implies that the results of the old rating system are more accurate, while a positive value means that the new system provides more accurate results.

Table 19: Analysis of p(F-Stat) Values (Old-New)

New	Old	200204	200205	200206
200207		-0.0964	-0.0717	-0.0709
200208		-0.0373	-0.0126	-0.0118
200209		-0.0242	0.0005	0.0013

When subtracting the average p(F-stat) values of the new rating system from the old, the situation described in table 19 emerges. In this situation, the majority of the values are negative, while positive values are almost equal to zero. This clearly shows that the rating system introduced in 200207 is, in terms of predictive performance, at best equal to its predecessor. In most occasions, the results yielded by the old rating system are more significant than the results produced by the new rating system. However, there can be

numerous reasons for such a result. Several of these possibilities will be discussed in the following sections.

4.4.3 Potential Biases and Limitations

Number of Fund-Months

The argument made by Morey and Vinod (2001) can be applied to the old/new rating system comparison as well. In the event that the old rating system contains fund-months than the new rating system, one might argue that, since the coefficients are close to zero, the results are biased in favour of the old system. Nevertheless, as table 17 shows, the number of fund-months in all samples is about equal.

While the overall fund-months do not suggest a bias, the regressions are based on individual categories and, therefore, fund-months should be analysed based on category and year. Table 20 shows this analysis.

Table 20: Number of Funds per Category, per Sample

Category	200204	200205	200206	200207	200208	200209	Hi-Lo	%
Large Value	690	689	607	615	625	628	83	13.5
Large Blend	915	916	921	917	924	932	17	1.9
Large Growth	679	675	736	743	747	754	79	11.7
Mid-Cap Value	204	205	184	188	186	188	21	11.4
Mid-Cap Blend	150	150	160	163	157	158	13	8.7
Mid-Cap Growth	438	446	471	471	471	471	33	7.5
Small Value	187	188	201	203	206	207	20	10.7
Small Blend	187	186	179	182	182	184	8	4.5
Small Growth	376	369	390	399	401	402	33	8.9
Specialty Communication	21	22	20	20	20	24	4	20.0
Specialty Financial	64	64	71	72	72	72	8	12.5
Specialty Health	57	57	61	64	68	69	12	*21.1*
Specialty Natural Resources	64	64	64	65	65	65	1	1.6
Specialty Real Estate	114	115	113	113	113	118	5	4.4
Specialty Technology	104	106	117	120	123	131	27	*26.0*
Specialty Utilities	81	81	81	82	83	83	2	2.5
Convertibles	56	60	60	61	61	61	5	8.9
Domestic Hybrid	664	664	663	669	677	674	14	2.1
World Stock	232	234	236	237	238	239	7	3.0
Diversified Emerging Markets	146	147	145	141	138	137	10	7.3
Latin America Stock	31	30	26	25	25	25	6	*24.0*
Europe Stock	126	128	120	122	122	123	8	6.7
Japan Stock	38	38	38	39	39	39	1	2.6
Pacific/Asia ex-Japan Stock	72	72	76	77	75	75	4	5.6
Diversified Pacific/Asia	44	44	43	44	41	41	3	7.3
Specialty Precious Metals	29	30	29	29	29	29	1	3.4
Foreign Stock	615	609	618	623	624	627	18	3.0
International Hybrid	51	51	50	50	49	49	2	4.1

Long Government	58	58	57	57	54	59	5	9.3
Intermediate Government	266	267	273	275	280	279	14	5.3
Short Government	116	114	109	111	112	113	7	6.4
Long-Term Bond	91	88	79	75	77	78	16	*21.3*
Intermediate-Term Bond	493	494	496	500	507	513	20	4.1
Short-Term Bond	199	199	185	184	185	187	15	8.2
Ultrashort Bond	42	44	46	49	51	57	15	*35.7*
High Yield Bond	289	284	284	286	281	283	8	2.8
Multisector Bond	151	152	150	152	154	154	4	2.7
Emerging Markets Bond	37	37	34	34	37	37	3	8.8
International Bond	118	112	114	116	116	116	6	5.4
Muni National Long	294	297	303	301	303	303	9	3.1
Muni National Intermediate	138	137	134	134	137	138	4	3.0
Muni National Short	594	585	564	545	545	544	49	9.0
Muni Single State Intermediate	279	258	258	263	264	261	21	8.1
Muni California Long	107	107	107	108	108	107	1	0.9
Muni New York Long	86	86	87	89	89	89	3	3.4
Muni California Intermediate	28	28	27	27	28	29	2	7.4
Muni Short	92	93	93	89	90	89	4	4.5
Muni New York Intermediate	22	22	21	21	22	22	1	4.7

Table 20 shows the number of funds in each category for the months used to compare both systems. The Hi-Lo column shows the difference between the highest number of funds in a category and the lowest number of funds in that category. The % column shows this number in a percentage of the lowest number of funds in a category. Basing the percentage on the lowest number of funds provides a conservative analysis compared to basing the percentage on the highest number of funds in a category. The top ten percent of funds that have the greatest percentage difference between highest and lowest number of funds in a category are printed in italics.

In order to conclude whether the number of fund-months in a category influences the regression results, the monthly differences in p(F-stat) values are analysed to see how these results influence the averages stated in table 19. This analysis can be found in table 21 on the next page.

Table 21: p(F-Stat) Differences for Categories with High Difference in Number of Funds

Category		**200204-200207**	**200204-200208**	**200204-200209**	**200205-200207**	**200205-200208**	**200205-200209**	**200206-200207**	**200206-200208**	**200206-200209**	**Category Average**	**Period Average**
Specialty Health	1y	0.0405	0.3262	0.6122	0.3206	0.6063	0.8923	0.0192	0.3006	0.5866	0.2452	0.4116
	3y	-0.107	0.0214	0.2292	-0.041	0.0874	0.2952	-0.152	0.0871	0.2949		0.0795
Specialty Technology	1y	0.0526	-0.020	0.4313	0.0587	-0.014	0.4374	0.0811	0.0087	0.4598	0.0571	0.1662
	3y	-0.065	-0.052	-0.057	-0.114	-0.102	-0.107	0.0025	0.0149	0.0098		-0.0522
Latin America Stock	1y	-0.000	0.002	-0.001	-0.003	-0.000	-0.003	0.0005	0.003	0.0002	0.0000	0
	3y	0	0.0003	-0.000	-0.001	-0.000	-0.001	0.0008	0.0011	0.0006		0
Long-Term Bond	1y	0.6022	0.4554	0.7456	0.2928	0.146	0.4362	0.3162	0.1694	0.4596	0.3627	0.4026
	3y	0.0319	0.0799	0.1622	0.0727	0.1207	0.203	0.6852	0.7332	0.8155		0.3227
Ultrashort Bond	1y	-0.626	-0.014	0.0001	-0.616	-0.003	0.0102	-0.136	0.4766	0.4902	-0.0232	-0.464
	3y	0	0	0	0	0	0	0	0	0		0
Sample Average		-0.007	0.0800	0.2123	-0.003	0.0841	0.2164	0.0811	0.1795	0.3117	0.1283	

The table above shows the differences of the p(F-stat) values for both regression on the one year and the three year out of sample period for categories with a high percentage difference in fund-months. The differences are obtained by the subtractions stated in bold. The value in the second to last column is the average of all subtractions in a category, while the values in the bottom row are obtained by taking the averages of all values in a column. Table 21 clearly illustrates that when there is a large difference in number of funds in a category over time, the result of the regression of this category is likely to be more significant for the new rating system than for its predecessor as almost all figures in the bottom row are positive values. Furthermore, this effect has a greater presence in the one year out of sample period than it has in the three year out of sample period, as can be concluded from the last column, where the value for the one year out of sample period is always greater than that of the value for the three year out of sample period. The result does not come as a surprise, as in most cases the new rating system has more funds in a category than its old counterpart. The only exceptions to this are the Latin America Stock and Long-Term Bond categories, and even in the latter category, the results of the new rating system are more significant than the results of the old rating system, as can be concluded from the positive value in the second to last column. This indicates that the number of available fund-months does not lead to a bias in our conclusion as: a) The new rating system outperforms the old rating system when the analysis of the new rating system contains more fund-months and

b) In the event that the old rating system contains more fund-months than the new rating system, the new rating system still outperforms the old rating system in terms of predictive performance. This means that the number of funds in a category does not lead to a bias favouring the old rating system.

The analysis on the number of funds proves that, although there is a small bias due to the difference in number of funds in a category over time, this bias is certainly not in favour of the old rating system, further strengthening the results stated in table 19.

Performance across Categories

While the section above shows that the number of funds in a sample does not lead to a bias towards the drawn conclusion, there could however, be other factors influencing the results. It could very well be that the two rating systems perform different across categories (i.e. the old system yielding significant results for the Large Value and Large Blend categories, while the new system performs better on the Long Government and Long-Term Bond categories). In order to analyse this potential occurrence, the top 10 best estimated categories of both rating systems will be compared. This comparison can be found in table 23 on the next page. It shows the 10 regressions with the most significant results per sample. Although there are some categories present in all samples (Ultrashort Bond, Foreign Stock and to a lesser extent International Bond, Long Government and Short-Term Bond), a more interesting result is found when the categories are grouped according to the groups of the old rating system. The old rating system uses the following grouping.

Table 22: Colour Coding of Category Grouping in Four Broad Asset Classes

Group.	Asset Class	Colour
1	US Stock	Red
2	International Stock	Green
3	Taxable Bond	Blue
4	Municipal Bond	Pink

When taking this grouping into account, it becomes clear that both the old and the new rating system provide very accurate results for mutual funds in the Taxable Bond group. This however, does not come as a surprise as, the Taxable Bond category scored very well in the previous predictive performance analysis based on categories. However, there is one category in which the new rating system performs better than the old: Multisector Bond (and to a lesser extent, High Yield Bond), but this is more an anomaly than a real difference between the rating systems. Since there are no major differences between the two rating systems in terms of performance across categories, this does not bias the result found in table 19.

Table 23: Best Estimated Categories per Sample

Pos.	200204	200205	200206	200207	200208	200209
1	Domestic Hybrid (3y)	Ultrashort Bond (3y)	Short Government (1y)	Short Government (1y)	Ultrashort Bond (3y)	Ultrashort Bond (1y)
2	Long Government (1y)	International Bond (1y)	Short-Term Bond (1y)	Ultrashort Bond (3y)	Multisector Bond (3y)	Ultrashort Bond (3y)
3	Short-Term Bond (1y)	International Bond (3y)	Ultrashort Bond (3y)	Foreign Stock (3y)	Foreign Stock (3y)	Multisector Bond (3y)
4	Ultrashort Bond (3y)	Long Government (1y)	Short-Term Bond (3y)	Short Government (3y)	Multisector Bond (1y)	Foreign Stock (3y)
5	International Bond (1y)	Domestic Hybrid (3y)	International Bond (1y)	Multisector Bond (3y)	Short Government (3y)	Multisector Bond (1y)
6	International Bond (3y)	Foreign Stock (3y)	Short Government (3y)	Short-Term Bond (1y)	Long Government (3y)	Long Government (3y)
7	Foreign Stock (3y)	Short-Term Bond (1y)	Foreign Stock (3y)	International Bond (1y)	High Yield Bond (3y)	High Yield Bond (1y)
8	Ultrashort Bond (1y)	Short Government (3y)	Intermediate-Term Bond (1y)	International Bond (3y)	High Yield Bond (1y)	High Yield Bond (3y)
9	Intermediate-Term Bond (1y)	Long Government (3y)	Muni Short (3y)	Muni Short (3y)	Ultrashort Bond (1y)	Large Growth (1y)
10	Domestic Hybrid (1y)	Ultrashort Bond (1y)	Intermediate-Term Bond (1y)	Short-Term Bond (3y)	Intermediate-Term Bond (3y)	Short Government (3y)

Performance over Years

Apart from the number of funds and the performance across categories, the results found in table 19 could be influenced by the performance of the different rating systems over the different out of sample periods (i.e. the old rating system outperforms the new rating system at the 1 year out of sample period, while the new rating system outperforms the old at a sample consisting of three years of return data). In order to check for such an effect, both time windows will be compared for both rating systems in order to see whether one rating system outperforms the other on a certain timeframe.

Table 24: Average p(F-stat) Values for Different Out of Sample Estimation Periods

Period	200204	200205	200206	200207	200208	200209	Avg. Old	Avg. New
1 year	0.5455	0.6424	0.6501	0.6902	0.6825	0.6763	0.6127	0.6830
3 years	0.5661	0.5874	0.6059	0.6142	0.5724	0.5771	0.5865	0.5879

Table 24 shows the average p(F-stat) value for regressions based on a one year out of sample period, and those based on a three year out of sample period. With the exception of 200204, the results of the regressions based on three years of data are more significant than the results of the regressions based on one year of data. This was to be expected as regressions based on more fund-months tend to yield more significant results.

What is interesting to note about table 24 is that the old rating system is superior to the new rating system when looking at estimation periods of 1 year. The p(F-stat) values are, on average, 7% lower for the old rating system than that they are for the new rating system. This means that the old rating system is superior in predicting short term performance, while both systems are about equal in predicting longer term performance. Nevertheless, the averages of both rating systems are far from being significant at the 10% level, but if one were to use an alpha of 0.65[19], the average results of the old rating system on the one year out of sample period would be significant, whereas that would not be true for the average results of the new rating system on the same out of sample period.

[19] An alpha of 0.65 is nothing short from absurd, but the example illustrates the point of the old rating system providing better results than the new rating system.

5 Conclusion

The analysis on predictive performance shows that while both the rating system based on categories and the rating system based on four broad asset classes fail to outperform a random walk, this does not hold for one large sample of mutual funds as seen in table 6. This table shows that a rating system using just one category is perfectly able at distinguishing poor performance from superior performance. However, this system cannot properly discern three and four star rated funds from five star rated funds.

The comparison between the two latest Morningstar rating system methodologies concludes that the old Morningstar Mutual Fund rating system is, in terms of predictive performance, superior to the new rating system. Even after analysing potential biases in the analysis, the conclusion holds. This implies that the results found by Morningstar (Kinnel, 2005) are largely incorrect. The only advantage of the new rating system is that it shows in which exact categories it is able to predict performance. It is for those, and only those categories that the new rating system should be used as a source to base the investment decision upon. For all other categories, the Morningstar's rating system does not offer any value in terms of predicting future performance and is degraded to an excellent source of information about a specific fund (fund manger, top five holdings etc.)

The results of these analyses in this paper indicate that Morningstar is an excellent source for obtaining information on mutual funds as they offer very detailed information on a wide array of funds. However, the results of their rating system prove that, once more, past performance does not guarantee future results. Morningstar is very clear on this by saying that their ratings are based on the past. Nevertheless, as previous research has indicated, investors choose to use the ratings as an indicator for future performance. Albeit a very interesting topic, the reasons for investors to use the mutual fund ratings as a guide for selecting the mutual fund(s) to invest in, falls beyond the scope of this research.

While the conclusion itself is straightforward, there could very well be a rational explanation behind the found results. The results found in the predictive performance section do not applaud Morningstar for the predictive performance of their mutual fund ratings. However, this may be due to one crucial assumption. As it was not possible to gain information on loads and redemption fees for over 25.000 funds, it had to be assumed that all funds did not charge any fees. Nevertheless, Morningstar incorporates the possible loads and fees into the rating. This might result in a bias towards overestimating the performance of funds that charge loads

and fees as their out of sample performance has not been lowered due to the fees, while these same fees are used to reduce the performance in the estimation period used by Morningstar to base the rating upon. Therefore, this might be a reason why lower rated funds achieve higher than expected levels of performance. Testing whether this assumption does result in the theoretical bias is a topic for further research.

Apart from the argument stated above, it would be wise to test the latest rating method on multiple out of sample periods. Furthermore, repeating this study with an even larger database, resulting in more funds per category for the latest rating system, might change the results in favour of Morningstar's new rating system.

Regarding the comparison of the rating systems, the ratings given by the old system are based on four categories whereas the new rating system bases the ratings on 48 categories. When looking at 200206, there are 9901 funds listed in the database. Dividing this number by 4, results in about 2475 funds per category on average, whereas in the 200207 sample, 9950 funds have to be divided over 48 categories, resulting in about 207 funds per category, on average.

When comparing analyses based on 207 funds with analyses based on 2475 funds, the latter analyses will produce more significant results compared to the former. So by adding more categories, Morningstar has reduced the power of the results. While this does not mean that all analyses should be based on just one category as shown in table 6, it does raise the question of the use of the different categories. Morningstar states that the differences between categories have to be meaningful, but differentiating between municipal bonds issued by the state of Florida and municipal bonds issued by the state of California could very well be seen as overkill. The appendix shows the complete list of categories used by Morningstar. In order to overcome this situation, further research could test whether the Morningstar Style Box could be used as a basis for assigning ratings, as all funds have a position somewhere in the Style Box.

References

Blake, C. R. and M. R. Morey, 2000, Morningstar Ratings and Mutual Fund Performance, *Journal of Financial and Quantitative Analysis*, 35:3 pp. 451 - 483

Capon, N., G. Fitzsimons, and R. Prince, 1996, An Individual Level Analysis of the Mutual Fund Investment Decision, *Journal of Financial Services Research*, 10, 59-82

Chevalier, J. and G. Ellison, 1999, Are Some Mutual Fund Managers Better Than Others? Cross-Sectional Patterns in Behaviour and Performance, *Journal of Finance*, 54:3, pp. 875 - 899

Del Guercio, D. and P. A. Tkac, 2001, Star Power: The Effect of Morningstar Ratings on Mutual Fund Flows, *Federal Reserve Bank of Atlanta Working Paper* 2001-15

Elton, E. J., M. J. Gruber, and C. R. Blake, 1996, Survivorship Bias and Mutual Fund Performance, *Review of Financial Studies*, 9, 1097-1120

Grinblatt, M., and S. Titman, 1989, Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings, *Journal of Business*, 62, 393-416

Kinnel, R., 2005, Rating the Star Rating, *Morningstar FundInvestor*, December 2005

Khorana, A., and E. Nelling, 1998, The Determinants and Predictive Ability of Mutual Fund Ratings, *Journal of Investing*, 7:3, pp 61 - 66

Malkiel, B., 1995, Returns from Investing in Equity Funds 1971-1991, *Journal of Finance*, 50, 549-572

Morey, M. R., 2002, Mutual Fund Age and Morningstar Ratings, *Financial Analysts Journal* 58:2 pp. 56-63

Morey, M. R, 2003, The Kiss of Death: A 5-Star Morningstar Mutual Fund Rating?, *SSRN Working Paper*, ID: 455240

Morey, M. R. and H. D. Vinod, 2001, Estimation Risk in Mutual Fund Ratings: The Case of Morningstar, *SSRN Working Paper,* ID: 270234

Morningstar, 2003, The Morningstar Rating Methodology, *Morningstar Research Report,* 1 October 2003

Morningstar, 2003, Fact Sheet: The Morningstar Rating for Funds

Morningstar, 2005, The Morningstar Category Classifications, *Morningstar Data Point Explanation,* 31 January 2005

Pozen, R. C., 1998, The Mutual Fund Business, MIT Press, Cambridge, Massachusetts

SEC, 2006, Flyer: Invest Wisely: An Introduction to Mutual Funds

Sirri, E. R., and P. Tufano, 1998, Costly Search and Mutual Fund Flows, *Journal of Finance*, 53, 1589-1622