2009

Does Morningstar Shine in the Universe of Mutual Funds? A Study on Morningstar Mutual Fund Ratings

Wee Seng NG
Singapore Management University, weeseng.ng.2007@smu.edu.sg

DOES MORNINGSTAR SHINE IN THE UNIVERSE OF MUTUAL FUNDS?
A STUDY ON MORNINGSTAR MUTUAL FUND RATINGS

NG WEE SENG

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN FINANCE

SINGAPORE MANAGEMENT UNIVERSITY

2009

DOES MORNINGSTAR SHINE IN THE UNIVERSE OF MUTUAL FUNDS?
A STUDY ON MORNINGSTAR MUTUAL FUND RATINGS

NG WEE SENG

ABSTRACT

Using data from Morningstar Principia CDs and employing standard methodologies, we examine the extent to which two mutual fund ratings: Morningstar star ratings and Morningstar stewardship grades can predict future fund performance. In particular, we investigate whether the combined predictive power of the two ratings exceeds that of a single rating. We decompose funds into various groups characterized by fund age and fund categories in order to address such issues as whether predictive performance is uniform across characteristic-based groups. Although our analysis shows that none of the ratings alone possesses strong predictive power, there is statistical evidence to support the notion that combined rating is superior to single rating in forecasting future returns. However, the evidence is not overwhelming enough to justify the efficacy of an investment strategy based primarily on Morningstar ratings. Besides studying predictability of ratings, we also construct a logistic regression model to seek potential determinants of the stewardship grades. We find that funds with good stewardship grades are generally those that incur low expenses, possess a large asset base and are managed by experienced fund managers. Finally, we investigate whether the two Morningstar ratings exhibit short-term persistence. Our findings indicate that the degree of persistence (as measured by the percentage of funds that retain their initial rating over a 12-month period) of the stewardship grades is much more pronounced than that of the star ratings.

TABLE OF CONTENTS

Acknowledgment

My thesis supervisor, Professor Jeremy Goh, has demonstrated much tolerance and understanding towards a part-time student like me. We have had discussions on my thesis after working hours on several occasions. The timely submission of this dissertation would not have been possible without his systematic guidance and regular contribution of interesting research ideas.

Professor Aurobindo Ghosh has generously shared with me, as a co-supervisor, the January 2005 – December 2005 Morningstar Principia CDs, which he has acquired through his research grant. I would like to thank Morningstar Corp. for granting me the permission to use their propriety products. I am very grateful to Prof. Ghosh for having permitted me to stay in his office several afternoons to hunt for the data I required despite the inconvenience caused, and for offering valuable comments which have helped enhanced the quality of my thesis.

I am very fortunate to have my parents constantly pushing me to scale new heights and reminding me to take setbacks in my stride. Their love, encouragement and support constitute a formidable driving force behind me. Of course, I will not forget to thank my sister, my brother and many of my great friends from different corners of Singapore for their moral support.

Special thanks are due to my brother, Dr Ng Wee Leng, for his meticulous proofreading of this thesis. Last but not least, I express my sincere gratitude to Miss Shelly Li Hiaw Gee for her efficient administrative support and to the friendly library staff for their spectacular service.

Chapter One

Introduction

Mutual funds are among the fastest growing and most successful financial innovations. In the past decades, the global mutual fund industry has flourished. As of the end of year 2006, there were a total of 61800 mutual funds managing assets worth more than 21 trillions of U.S. dollars. Even against the backdrop of financial turbulence following the emergence of the sub-prime crisis in mid-2007, the number of mutual funds worldwide had jumped to 66300, with the total assets managed increased to 26 trillions of U.S. dollars by the end of 2007 (See Mutual Funds Fact Book by Investment Company Institute (2008)). Apparently, the incessant growth of this booming business is not showing any sign of slowing down. A recent research report published by the Lipper Service[1] reveals that conventional US mutual fund industry attracted a total inflow of U.S. $51 billions in the month of September 2007 alone. Similar magnitudes of growth were also registered in other regional markets.

The explosion of the number of mutual funds traded in the financial markets has left many investors scrambling to seek financial advisory services. Yet there are also a huge number of investors who prefer to take the responsibility of managing personal finance in their own hands. In order to make an informed investment decision, these investors have turned to salient and readily accessible fund information such as publicised fund reports and fund ratings for guidance. Debuted in 1985, the Morningstar star ratings awarded to mutual funds based on historical performance is probably the most popular and influential mutual fund ratings.

[1] Lipper Services Research Series Fund Flow Report 30 September 2007

Indeed, the work by Goetzmann and Peles (1997), Sirri and Tufano (1998), Blake and Morey (2000) and Guercio and Tkac (2008) have all provided empirical evidence to support the assertion that mutual fund ratings have a significant impact on the trading behaviour of mutual fund investors. In particular, Guercio and Tkac (2008) report that when a fund receives the most prestigious five-star rating for the first time, significantly strong abnormal inflow is registered within six months from the initiation date.

Leveraging on the success of the star rating and recognizing the importance of corporate governance to mutual fund investors, Morningstar launched in 2004 the fiduciary grades which was renamed the Morningstar stewardship grade in 2005. In contrast to the seasoned star rating, the young stewardship grading system evaluates funds by considering less tangible factors such as corporate culture of fund sponsors and the extent to which the investment style of fund managers may benefit investors in the long run. A letter grade of A(best) to F(worst) will be assigned to funds under evaluation.

The eruptions of U.S. mutual fund scandals in 2003 had aroused enormous public concern and precipitated the publication of a series of academic studies on the subject of mutual fund governance. The creation of this new fund-rating system can therefore be interpreted as nothing but a natural development in the fund rating business. The primary objective of this new rating is to help investors identify fund managements that take their fiduciary responsibilities seriously.

The notion that fund governance matters for fund performance is supported by the work of many researchers, including Cremers, Driessen, Maenhout and Weinbaum (2005), Chen, Goldstein and Jiang (2006), Ferris and Yan (2007), Khorana, Tufano and Wedge (2007) and Qian (2007). Ferris and Yan (2007) report that funds for which the independent directors receive huge compensation have a higher likelihood of being involved in fund scandals. Chen, Goldstein and Jiang (2006) find that in the absence of other control mechanisms, directors tend to own shares in the funds that offer the highest expected benefits. Cremers, Driessen, Maenhout and Weinbaum (2005) discover a strong positive association between fund performance and directors' stakes in the funds. Qian

(2006) document that investors' ability to withdraw assets as a show of discontent or disapproval of the fund management is an effective fund-monitoring mechanism.

The success of the Morningstar fund ratings has been a puzzle to some researchers since the standard finance paradigm suggests that publicly known information such as fund ratings cannot be exploited profitably. In order to investigate whether the star rating can actually offer some insights into the future performance of funds, many academics embark on a wide range of studies pertaining to Morningstar ratings. Blume (1998) examines Morningstar's method of calculating fund rankings and finds that it suffers from a rounding-error bias. The error is caused by the conversion of the raw score representing a fund's risk-adjusted return to the nearest integer which is then translated into the published star rating. For example, funds that receive scores of 3.2 and 3.4 both receive a 3-star rating while funds scoring 3.5 and 3.9 are awarded a 4-star rating. Sharpe (1998) reports statistical artifacts inherent in the Morningstar's methodology.

Blake and Morey (2000) evaluate the ability of the star rating to forecast future performance. They find that there is little statistical evidence indicating that 5-star funds gain higher future returns than funds with 3-star or 4-star rating, although poor-rated funds continue to suffer dismal performance over a one, three and five post-rating period. Both Khorana and Nelling (1998) and Warshawsky, DiCarlantonio and Mullan (2000) focus on the persistence of the 4-star and 5-star ratings. Their results indicate that the degree of persistence varies across age groups and time horizons. Older funds are likely to show higher persistence. Morey (2002) detects the presence of age bias which is not a consequence of survivorship bias, but rather, of the Morningstar methodology. Age bias results in seasoned funds receiving higher average ratings than younger funds.

Unlike the star rating, the stewardship grade has not captured much attention within the research community. Hitherto, the work by Wellman and Zhou (2007) is probably the only attempt to study the stewardship grade in substantial detail. Despite their effort, many important issues have yet to be explored. For example, while Wellman and Zhou (2007) have obtained statistical evidence to support the contention that the

trading behaviour of mutual fund investors is influenced by changes in stewardship grades, the question of whether funds with good stewardship grades necessarily yield superior ex-post performance has not been properly addressed. This question is an important one as it can help explain the phenomenon of huge abnormal fund flows following fund rating changes. Furthermore, with two ratings now at our disposal, it is natural to examine statistically the degree to which the rating pair can jointly predict future performance. It is also meaningful and instructive to investigate whether joint predictive ability is stronger than when any one of the ratings is considered in isolation. In this dissertation, we aim to explore these issues. Besides predictive ability, we also seek potential determinants of the stewardship grades and investigate whether the stewardship grades exhibit short-term persistence.

The results of our studies should have important economic significance to investors who intend to use the fund ratings in one way or another to construct their investment plan. If it turns out that both persistence and predictive ability of high-rated funds are supported by sound statistical evidence, then investors can consider adopting the strategy of buying and holding a portfolio of funds with top ratings. Otherwise, investors should consider factors beyond fund ratings in making their investment decisions.

This dissertation has contributed to the literature in at least two ways. First, we make the first attempt to compare the stewardship grade with the star rating in several aspects. Since Morningstar claims that the two ratings are independent of each other, our study helps to examine the extent to which the claim is valid. Second, our study of the joint predictive power of the two ratings is also unprecedented. Evidence from various academic works that focus on fund flows pattern in relation to fund ratings have revealed that many mutual fund investors are ardent supporters of the Morningstar star ratings. Our empirical work on joint ratings in this direction can help to shed some light on the question of whether investors will be better off by investing in funds that fare excellently in not just one, but both ratings.

The rest of this dissertation is organized as follows. Chapter 2 presents in detail the methodologies for both the star rating and the stewardship grade. Chapter 3 describes the data and methodology used in this dissertation. Chapter 4 examines the results and provides relevant economic insights. Chapter 5 concludes this dissertation and offers suggestions for further research. Throughout this dissertation, we shall refer to Morningstar fund ratings as both Morningstar star rating and Morningstar stewardship grade.

Chapter Two

Star Ratings and Stewardship Grades

2.1 Methodology For Star Ratings

The Morningstar fund rating has increasingly become one of the most popular and influential mutual funds ratings. It has been widely used by mutual fund distributors such as commercial banks, financial advisory firms and life insurance companies as an advertising tool. The first Morningstar ratings were available to the public via subscription to its publications *Morningstar Mutual Funds*. Morningstar evaluates almost all existing funds which are at least three years old. Funds are first classified into categories. Prior to the revision of the Morningstar methodology in 2002, there were four broad categories for mutual funds: Domestic Stock, International Stock, Municipal Bond and Taxable Bond. After the revision, Morningstar increased the number of categories to sixty four[2], grouping each fund into a more narrowly-defined peer group and thus facilitating a more effective comparison of funds.

A fund's star rating is derived from its historical performance, taking into account both its risk and return within its own category. To determine a fund's star rating, Morningstar first calculates the Morningstar risk-adjusted return, $MRAR_T$ (T = 3, 5 or 10) based on the fund's past 3, 5 and 10-year returns, where applicable.

We now elaborate on how the formula for $MRAR_T$ is derived. First, based on data on the fund's monthly returns $\{r_t : t = 1, 2, \ldots, 12T\}$ over the past T years, the fund's cumulative return R_T, given by

[2] See Appendix A

$$R_T = 100\left[\left(\prod_{t=1}^{12T}(1+r_t)\right)-1\right] \tag{2.1}$$

is computed. In addition, the corresponding T-year cumulative risk-free return, R_f, based on monthly T-bill rates $\{ r_{ft} : t = 1, 2, \ldots, 12T\}$ over the same period is determined. The load-adjusted return, LAR, of the fund is then computed according to the formula

$$LAR = R_T(1-\frac{L}{100}) - R_f \tag{2.2}$$

where L is the front-end load. For example, if the front end load is 5%, $LAR = 0.95R_T - R_f$. Denoting the average LAR of all funds in the same category by AvgLAR, the Morningstar load-adjusted return rate (MLARate) of the fund is defined by

$$MLARRate = \frac{LAR}{\max(AvgLAR, R_f)}. \tag{2.3}$$

Next, the Morningstar risk MRisk of the fund, given by the following formula, is calculated.

$$MRisk = \frac{1}{12T}\sum_{t=1}^{12T} -\min(r_t - r_{ft}, 0). \tag{2.4}$$

The Morningstar risk rate MRiskRate is then determined according to the formula

$$MRiskRate = \frac{MRisk}{AvgMRisk} \tag{2.5}$$

where AvgMRisk is the average of MRisk of all funds within the same category. Finally, the Morningstar risk-adjusted return, $MRAR_T$ of the fund is given by

$$MRAR_T = MLARRate - MRiskRate \tag{2.6}$$

Funds within each investment category are then ranked based on their $MRAR_T$. A star rating SR_T, for the fund's T-year past performance, is then determined according to the following rules:

(i) Funds in the top ten percent receive 5 stars.

(ii) Funds in next 22.5% receive 4 stars.

(iii) Funds in the middle 35% receive 3-stars.

(iv) Funds in the next 22.5% receive 2 stars.

(v) Funds in bottom 10% receive 1 star.

Finally, the overall star rating, MS is the greatest integer less than or equal to the MSScore given by the function

$$\text{MSScore} = \begin{cases} SR_3 & \text{if fund has 3 - 5 years of returns} \\ 0.6\,SR_5 + 0.4\,SR_3 & \text{if fund has 5 - 10 years of returns} \\ 0.5\,SR_{10} + 0.3\,SR_5 + 0.2\,SR_3 & \text{if fund has} > 10 \text{ years of returns} \end{cases} \quad (2.7)$$

provided MSScore does not end in 0.5, in which case the score will be rounded up to the next integer. As an illustration, a fund that receives 4 stars for each of its 3-year and 5-year assessments and 5 star for its 10-year assessment will get MSScore = 4.5 and hence an overall rating of 5-star.

2.2 Methodology For Stewardship Grades

Launched by Morningstar in 2004, the fiduciary grade was engineered to serve as a barometer for the standard of corporate governance of mutual funds. It was renamed the Morningstar stewardship grade in 2005. The first set of stewardship grades was released in August 2004. Going beyond the usual risk and returns assessment, the stewardship grade attempts to capture some of the intangibles such as the degree in which potential conflicts of interests of the fund management company might affect shareholders and the investment culture of portfolio managers.

The Morningstar stewardship grade is calculated based on five Stewardship components:

(i) Board Quality (BQ),

(ii) Corporate Culture (CC),

(iii) Fees (Fees),

(iv) Manager Incentives (MI)

(v) Regulatory Issues (RI).

To derive the final stewardship grade, Morningstar computes a score for each of these criteria. Prior to 2007, each criterion carried a maximum score of 2 points. For Regulatory Issues, the lowest possible score was -2. For each of the other four criteria, the minimum score was 0. Based on these scores, a qualitative grade published in Morningstar Fund Reports would be assigned according to Table 1 below:

Table 1 Qualitative Grade For Stewardship components (Prior to July 2007)

Score	Qualitative Grade
2.0	Excellent
1.5	Good
1	Fair
0.5	Poor
<= 0	Very Poor

The sum of the scores assigned to these five criteria was used to determine the overall stewardship grade as outlined in Table 2.

Table 2 Qualitative Grade For Overall Stewardship Grade

Score	Stewardship Score
9 – 10	A
7 - 8.5	B
5 - 6.5	C
3 - 4.5	D
<= 2.5	F

Since July 2007, a few changes had been made. First, the maximum score for Corporate Culture was increased to 4 in 2007, reflecting Morningstar's belief that corporate culture is, among the five criteria, the most effective means of measuring fund governance. Second, the range of points to be assigned to Regulatory Issues was changed to -2 – 0. Finally, each component score was translated into qualitative terms:

Table 3 Qualitative Grade For Stewardship components (In and after July 2007)

Score	Letter Grade
Full credit	Excellent
¾ credit	Good
½ credit	Fair
¼ credit	Poor
No credit	Very Poor

We now briefly outline the important features of the five Stewardship components for the sake of subsequent discussion. What we present here are methodologies used prior to the revamp mentioned above. The reason for our choice is that we use the 2005 Morningstar data in this dissertation. More details on the changes in methodologies which were effective on and after July 2007 can be obtained from the Morningstar Fact sheets, Morningstar (2007) available on the official website of Morningstar.

Board Quality

The following four factors for board quality, each worth up to 0.5 point, will be examined:

(i) Does the board act consistently to protect the interests of shareholders.

Examples of positive action taken by the board include dismissing or replacing underperforming fund managers and disapproving attempt by fund management to merge poor-performing funds with more successful funds.

Ding and Wermers (2005) document evidence supporting the hypothesis that the replacement of fund managers is beneficial to shareholders. It is found that on the

average, fund managers who are replaced by board directors underperform their peers, and that incoming managers outperform those replaced by one percentage point per year.

(ii) Do independent directors have significant investments in the funds?

The highest score of 0.5 point can be earned if at least 75% of a board's directors invest in the funds they oversee with an amount exceeding his/her aggregate annual compensation for serving on a board.

Chen, Goldstein and Jiang (2006) and Cremers, Driessen, Maenhout and Weinbaum (2005) have independently examined director ownership in funds. The former group of authors report that the optimal contracting hypothesis holds: in the absence of other control mechanisms, directors tend to own shares in the funds that offer high expected benefits. Indeed, directors are found to hold more shares in actively managed funds such as small-cap equity funds than in, for example, bond funds. The latter find that there exists a strong positive association between fund performance and directors' stakes in the funds.

(iii) Is a board overseeing too many funds to the extent that its ability to protect shareholders' interest will be compromised?

Ferris and Yan (2007) show that directors who oversee many funds have a higher chance of being implicated in a fund scandal. Their results support Morningstar's view that board's effectiveness would be adversely affected by "over-burdened" directors.

(iv) Does the fund meet the requirement of the Security and Exchange Commission (SEC) that at least 75% of the board's directors are independent?

Morningstar does not consider current and former employees as well as family members of both fund company and fund services providers to be independent. The belief that board independence is positive for fund performance is affirmed by Khorana (1996) who demonstrates that the degree of independence of a fund's board has a positive association

with the quality of fund governance. It is found that underperforming managers are more likely to be replaced when the board has a higher proportion of independent directors.

Corporate Culture

For this component, Morningstar considers a wide spectrum of factors.

(i) Has the fund management company launched "trendy" funds just to chalk up assets, regardless of whether the timing to launch such funds is appropriate? For example, many funds that were narrowly-focused on technology stocks were launched during the time when the technology sector was at the verge of collapse. Indeed, the bubble burst in 2001, causing many investors to suffer heavy losses.

(ii) Has the fund management company closed funds at an appropriate asset size or has it allowed the size to reach an unacceptable level. This question is crucial because it is difficult to manage a fund with a huge asset base effectively and profitably. Fund managers might be forced, due to liquidity and other consideration, to take large positions in stocks which might not offer the best potential returns. Chen, Hong, Huang and Kubik (2004) find strong evidence that fund size erodes fund performance. Furthermore, the adverse effect that fund size has on fund returns is most pronounced for funds that invest in illiquid assets such as stocks having small capitalization, thus suggesting that liquidity concerns could in part explain this effect. As part of fund managers' compensation is derived from management fees which are in turn a fraction of the asset size of the fund, unscrupulous fund managers might want to continue growing the size of their fund's fund asset base to reap higher monetary gains.

(iii) Does the fund implement measures such as high back-end loads to discourage frequent redemption of funds?

(iv) Does the firm communicate effectively with shareholders? For example, the management is expected to produce comprehensive publications such as updated

fund fact sheets and portfolio managers' reports for all shareholders on a regular basis.

(v) Has the firm used soft dollars which are payments made to the fund service providers? Soft dollars are incorporated into brokerage fees which will neither be reported nor included in the calculation of fund's expense ratio. Control on the use of soft dollars benefits shareholders. Funds paying high soft dollar commissions will be penalised.

Fees

Mutual fund investors pay various levels of fees. Fees related to distribution and redemption, commonly known as front-end loads, or sales charges, are paid at the time of transaction. Back-end loads or redemption fees are paid when investors sell the fund. Management fees are paid on a regular (usually annual) basis via direct deduction from the funds' assets. Funds can come in various share classes[3]. Although all classes hold the same securities and are managed by the same portfolio manager, they have different fees structure. In addition, as mentioned in the preceding paragraph, fund sponsors make soft dollar arrangement in which fund managers pay higher brokerage commissions to research and brokerage services, incurring another layer of fees for shareholders. Morningstar evaluates funds based on two aspects of the fees structure, each worth 1 point.

(i) One aspect is Fees Comparison. A fund receives 0.5 points if its expense ratio is lower than the median expense ratio of all funds within the same category group and having the same share class[3]. An additional 0.5 points is awarded to funds having an expense ratio within the lowest 25th percentile.

(ii) The other aspect is Fees Trends. A fund receives 1 point if its expense ratio decreases as its assets grows, or if there is evidence indicating that the fund managers will lower the expense ratio when the fund size increases. Funds that

[3] See Appendix B

charge additional fees such as performance fees, typically a fraction of excess returns over a certain benchmark, will be viewed less favorably by Morningstar and hence tend to receive lower score.

Manager Incentives

Two aspects, each worth 1 point, will be evaluated:

(i) The first aspect is Fund Ownership. Does a portfolio manager invest a significant amount of money in the fund he oversees?

Managers with at least U.S. $1 million or at least one third of their liquid net worth, whichever is lower, invested in their funds will be given 1 point. For investment of U.S. $500,000 – U.S. $1 million, managers receive 0.5 points. In cases where the fund size is small, fund managers can invest in other funds of the same firm to earn partial credit. As of 2005, fund managers of US mutual funds are required to disclose the amount (in U.S. dollars) of their wealth invested in the fund they manage, in the following seven ranges:

(i) 0,
(ii) 1-10,000,
(iii) 10,001-50,000,
(iv) 50,001-100,000,
(v) 100,001-500,000,
(vi) 500,000-1,000,000
(vii) above 1,000,000.

The above disclosure requirement is one of the series of new regulations enacted by the Security Exchange and Commisions in 2004 in response to fund scandals discovered then.

Fund ownership, according to a recent work by Khorana, Servaes and Wedge (2007), is positively correlated to the risk-adjusted returns of funds, with fund performance improving by as much as three basis points for each basis point increase in managerial ownership. The results of their work support the notion

that managerial ownership gives managers more incentives to generate higher returns for fund's shareholders, and is an important determinant of fund performance. Hence, disclosure on the level of managerial ownership offers investors valuable information to shareholders.

(ii) Does the compensation scheme reward portfolio manager based on long-term performance or short-term asset appreciation? Funds with incentives geared towards short-term growth will be viewed less favourably, and hence given lower score. Conversely, funds whose managers are compensated based on long-term fund performance instead of asset growth will generally receive higher rating.

Morningstar instructs fund companies to complete a survey which details the compensation structure of their fund managers as well the level of their investment in the funds they manage. Morningstar believes that fund managers' incentives have a strong influence on the quality of management. A fund manager whose compensation is tied to short-term out-performance of its benchmark (e.g. performance fees) will have a tendency to take excessive risk, as documented in Brown, Harlow and Starks (1996**)**

Regulatory Issues

At the point of assessment, Morningstar examines regulatory issues at the fund management level over its past three years of history. Funds found with severe breaches of certain regulations might get the lowest score of -2. Funds free from regulatory violations or potential fund indictments receive the highest score of 2. Funds found to have breached certain rules but have remedial actions in place will get a score between -2 and 2, depending on their level of commitment to reform.

2.3 The Effects and Defects of Morningstar Ratings

It is almost indisputable that the Morningstar star rating is popular and influential among mutual fund investors. Evidence that supports this claim can be found not only in high-profile business press such as the Wall-street Journal and Business Week, but also in numerous academic journals. Many recent scholarly works are devoted to an examination of the degree to which the star rating can affect investor's decision. Guercio and Tkac (2008) launch a very detailed study on how fund ratings affect fund flows. They find that the initiation of a 5-star rating of funds results in a spectacular 53% abnormal cash inflow. Furthermore, significant abnormal flow in the right direction (positive for upgrades and negative for downgrades) is detected for various rating changes.

In the same vein, Wellman and Zhou (2007) make an unprecedented effort to examine, among other issues, the influence Morningstar stewardship grade has on fund flows. Using the first release of the stewardship grades dated August 24 2004, they examine flows patterns following a stewardship grade upgrade or downgrade of funds. The results seem to indicate that investors trade funds in response to these events, selling funds with poor grades and buying those with good grades.

Blake and Morey (2000) examine the effectiveness of the star rating as a predictor of future fund performance. Their results show that while poor-rated funds generally show weaker future performance than do funds with better ratings, there is only weak statistical evidence indicating that good-rated funds have superior ex-post performance.

Morey (2005) studies the potential effect that the initiation of a five-star rating has on the future fund performance and fund characteristics such as expense ratio and turnover ratio. Using standard performance metrics to measure out-of-sample performance, Morey (2005) reports that the performance of winning funds falls dramatically over a three-year post-rating period and that the risk level of the funds rises significantly, though expense ratio and turn over ratio do not change notably. This result, which is consistent with those reported in Blake and Morey (2000), can be seen as partly

supporting the tournament hypothesis that fund managers are competing with one another to emerge winners under the star rating system.

The popularity of the star rating has prompted some researchers to explore potential shortcomings underlying the star rating methodology. Morey (2002) discovers that the rounding method by which the overall fund star rating is derived from the three-year, five-year and ten-year ratings creates an asymmetry in ranking. We recall that the overall star rating of a fund with more than ten years of history is derived from a weighted sum of a fund's three-year, five-year and ten-year ratings. Morey (2000) reports that for seasoned funds (funds with ten or more years in age), a downgrade by one to two points in their three-year Morningstar rating does not make any difference to the overall rating in certain cases. For example, if a fund receives 4-star and 5-star for its five-year and ten-year ratings respectively, then the overall rating is 4-star regardless of whether the three-year rating is 1, 2 or 3; similarly, the overall rating is unaffected by change of rating from 4-star to 5-star or vice versa. We illustrate this point with the aid of additional examples in Table 4. This suggests that the recent performance of a fund does not have much influence on the overall rating, resulting in seasoned funds with good long-term historical performance to continue enjoying good ratings despite their poor performance over a shorter pre-evaluation period.

Table 4 Illustration of Effect of 3-year rating Change On Overall Rating of Seasoned Funds

Ten-year rating	Five-year rating	Three-year rating	Overall star rating
5	4	1	4
5	4	2	4
5	4	3	4
5	4	4	5
5	4	5	5
5	5	1	4
5	5	2	4
5	5	3	5
5	5	3	5
5	5	5	5

We also find that the distribution of overall star ratings of seasoned funds has a bias towards higher ratings. Taking all 125 possible triples (SR_3, SR_5, SR_{10}) representing a fund's three-year, five-year and 10-year ratings, we calculate the fund's overall rating and display the frequency distribution in Table 5 below. Assuming it is equally likely for a fund to receive any of the 125 triples, the Morningstar's method of calculating the overall rating results in a small likelihood of funds getting either the worst rating or the best rating.

Table 5 Distribution of Possible Overall Star Ratings of Seasoned Funds

Overall Star Rating	Frequency
1	1
2	18
3	46
4	44
5	16

Blume (1998), Morey (2002) and Adkisson, Fraser and Don (2005) also demonstrate the presence of age bias the Morningstar star rating. Their studies unanimously reveal that young (three to five years) funds have significantly higher chance of getting the top 5-star rating than do seasoned funds. In particular, Blume (1998) finds that seasoned funds are also less likely than younger funds to get a one-star rating.

Using a sample of 1589 funds[4] taken from the Morningstar July 2005 Principia CD, we examine the above phenomenon. We find that 35% of young funds receive 5-star ratings as compared to only 11.7% of seasoned funds. In addition, only 2.63% of seasoned funds are awarded the worst 1-star rating. The corresponding figure for young fund is 18.8%.

Being relatively young as compared to the star rating, the Morningstar stewardship grade has not been a popular subject of research. Wellman and Zhou (2007) is probably the first group of researchers to launch an academic study on this subject. Using the first release of the stewardship grades on August 24 2004, they find that funds

[4] A full description of the data set we use in this dissertation will be furnished in Chapter 3.

that score high on stewardship grade outperform their peers with poor grades by 19 to 23 basis points per month over the post-observation period January 2001 to July 2004. They also find that among the five stewardship components, Fees and Board Quality exhibit the most significant explanatory power for ex-post returns, thus demonstrating the positive relation between corporate governance and fund performance.

Chapter Three

Data and Methodology

3.1 Data Description

We obtain twelve monthly mutual fund data from the January 2005 - December 2005 Morningstar Principia CDs. The data contain the star ratings (comprising 3-year, 5-year, 10-year and overall ratings), the grades for all five stewardship components, the overall stewardship grade as well as the following fund characteristics: Morningstar-style best-fit Alpha (BFAlpha), Morningstar-style best-fit beta (BFbeta), Morningstar-style best-fit R-squared (BFRsq), 2004 annual returns (AR04), annual expense ratio (ER), 12-month prior load-adjusted returns (LAR12m), 3-year prior load-adjusted returns (LAR3yr), average manager tenure (MT), asset size (asset), 1-month return after taxes and sales (RTS1m), 1-year return after taxes and sales (RTS1yr), monthly Sharpe ratio (SR), 3-year Standard deviation (SD), total number of holdings (TH) and turn-over ratio (TR). For estimating out-of-sample performance of our sample funds, we also extract data from the Centre for Research in Security Prices (CRSP) mutual fund database.

We include only mutual funds that receive both star rating and stewardship grade. Hence, funds that do not have at least three years of age will not be considered. In order to assemble a manageable data set and yet maintain a balanced approach to data analysis, we select funds that fall under three broad asset-based categories: Domestic Stock, International Stock and Bond (both Municipal Bond and Taxable Bond). For funds classified as Domestic Stock funds, we select those under the subcategories of Large Blend, Mid-cap Blend and Small-cap blend. For International Stock funds, we discard specialty funds which narrowly focus on sectors such as real estates and commodities.

We retain all bond funds. This gives rise to a portfolio comprising 1300 to 1500 funds for each monthly sample.

For reporting the distribution of fund ratings and basic summary statistics of fund characteristics, we use the June 2005 sample. For studies on persistence of fund ratings, we employ a sample of funds that are continuously graded from December 2004 to November 2005. To examine predictive abilities, we use all the twelve monthly samples. Each month represents an evaluation period, or evaluation month, a term we shall adopt henceforth. We perform, for each fund and in each evaluation month, a time-series regression over a post-evaluation period of predetermined duration using returns information such as market monthly risk-free returns and fund's monthly returns taken from the CRSP database.

In order to perform regressions and other statistical analyses, we convert all categorical variables to numeric variables. For star ratings (3-year, 5-year, 10-year and overall) , the conversion formula is straightforward: 1-star = 1, 2-star = 2, 3-star = 3, 4-star = 4 and 5-star = 5. For stewardship grades, we assign scores on the same scale: Grade F = 1, Grade D = 2 , Grade C = 3, Grade B = 4 and Grade A = 5. The five stewardship components are also quantified as follows: Excellent = 2 , Good = 1.5 , Fair = 1 , Poor = 0.5 and Very Poor = 0 or less.

Based on the above quantitative scores for the star ratings, we compute an estimate of the star rating raw score in accordance with Morningstar's methodology:

Raw score for star rating (SR) =

$$\begin{cases} SR_3 & \text{if fund has } 3-5 \text{ years of returns} \\ 0.6\,SR_5 + 0.4\,SR_3 & \text{if fund has } 5-10 \text{ years of returns} \\ 0.5\,SR_{10} + 0.3\,SR_5 + 0.2\,SR_3 & \text{if fund has } > 10 \text{ years of returns} \end{cases} \quad (3.1)$$

where SR_t is the t-year Morningstar rating.

In like manner, we use the quantitative scores of the five stewardship components to calculate the raw score for stewardship grade by taking half the sum of the five scores. The multiplication factor of half places the raw scores of both ratings on a common scale of 0 – 5.

The use of raw scores rather than the overall ratings should improve the accuracy of our statistical tests. The reason is that under the Morningstar methodology, funds having different raw scores can have a common rating. For example, a raw score of 3.2 and 3.3 for star rating will both lead to an overall rating of 3-star. The same applies to stewardship grades. Raw scores thus contain more information than the actual ratings. Nonetheless, we will use both the actual ratings and the raw cores in our analyses, as it is the actual star rating that investors would be observing. Raw scores are only available to investors who have paid access to specific Morningstar products. Furthermore, the use of both actual ratings and raw scores provides a basis for comparison.

3.2 Methodology

In this dissertation, we examine a wide range of issues. To get an idea of how funds perform under the Morningstar rating systems, we report frequency distributions of both star ratings and stewardship grades. In order to find out if the distribution of ratings varies with fund age, we divide funds into three fund age groups: three-year, five-year and ten-year. Three-year funds are funds that have at least three years but not more than 5 years of age. Five-year funds are funds with age between five and ten years. Ten-year funds have at least ten years of history. We also divide funds into three categories: Domestic Stock, International Stock and Bond.

Morningstar states in the Morningstar Fact Sheet (2007) that "The methodology for the stewardship grade for funds is completely different from the Morningstar Rating for funds (the "star rating") and the stewardship grade has no impact on a fund's star rating." We verify the claim of independence via a standard chi-square contingency table

analysis and t-test. For this purpose, we group sample funds into a 2-by-2 contingency table based on their star rating and stewardship grade and compute the chi-square statistics for test of independence. We also perform correlation analyses. We report both Pearson product-moment correlation coefficient and Spearman-rho rank correlation coefficients based on both original ratings and the raw scores.

Khorana and Nelling (1998) employ a probit regression model to explore influential determinants of the Morningstar star ratings. They find that expense ratio, asset size and alpha are some of the fund characteristics that possess strong explanatory power for the overall star rating. We extend their study to the stewardship grades, using an ordered logit regression approach. The use of logit regression model is appropriate since the dependent variable, namely the Morningstar rating, is a categorical variable. In addition, the dependent variable has a natural ordering of 1 (Grade-F) to 5 (Grade-A). Ordered logit model takes the form

$$\text{logit}\sum_{j=1}^{k} p_j = \log\left(\frac{\sum_{j=1}^{k} p_j}{1 - \sum_{j=1}^{k} p_j p}\right) = \alpha_k + \sum_{i=1}^{I} \beta_i X_i \qquad (3.2)$$

for k = 1, 2, 3 and 4, where p_j is the odds that a fund has a j-rating and the vector $\mathbf{X} = (X_1, X_2, \ldots, X_I)$ is the set of regressors. The highest rating group (k = 5) is used as a reference group as it provides a ceiling for all other groups. A significantly positive β_i is interpreted as an indication that an increment in the independent variable X_i increases the likelihood of a fund falling in a higher stewardship-rating group. Conversely, a significantly negative β_i is interpreted as an evidence that a higher value in X_i increases the chance that the dependent falls within a lower stewardship grade category.

We also perform ordinary least-squared regression using the stewardship raw scores as dependent variable, treating the raw scores as continuous variables. The independent variables are the list of fund-specific variables taken from the Morningstar Principia CDs. The list of variables with their abbreviations is given in Appendix C.

In order to construct a parsimonious regression model with a reasonable goodness-of-fit, we go through a rigorous variable selection process. First, we run a series of twelve cross-sectional regressions based on the twelve monthly samples, using all the available fund characteristics as regressors. From the results that emerge, we then examine the statistics and p-values for each regression, removing insignificant factors which have p-value exceeding 10% in 6 or more out of the 12 regressions. We proceed to examine the correlation matrix for the variables from which we remove highly correlated variables. This helps mitigate the adverse effect that can possibly be caused by the problem of multicollinearity. Next, we regress the dependent variable on the remaining variables to further identify those factors that have statistically significant regression coefficients. Finally, we use the Akaike's Information Criterion to obtain the final model. We report results based on this model.

For our studies on the predictive ability of the ratings, we employ a standard methodology in which the in-sample ratings of funds are compared with their out-of-sample performance as measured by some commonly-used performance metrics over a post-evaluation period. For this purpose, we need to first construct a measure for the out-of-sample performance of the sample funds. For robustness, we use two commonly used (see for example, Blake and Morey (2000)) performance metrics: Sharpe ratio and Carhart's four-factor alpha. We perform the analysis for twelve evaluation samples. The Sharpe ratio of fund i is defined by

$$\text{Sharpe Rato}_i = \frac{\overline{R_i - R_f}}{\sigma_i} \qquad (3.3)$$

where $\overline{R_i - R_f}$ is the mean return of fund i in excess of 30-day Treasury-bill rate and σ_i the standard deviation of the excess returns of fund i, over the post-evaluation period. We compute a twelve-month series of Sharpe ratios for each of the monthly samples over the evaluation period: December 2004 to November 2005.

The four factor alpha is the estimated intercept of the following regression model:

$$R_{it} - R_{ft} = \alpha + \beta_{i1} RMRF_t + \beta_{i2} SMB_t + \beta_{i2} HML_t + \beta_{i4} UMD_t + \varepsilon_{it} \quad (3.4)$$

which is an extension of the celebrated Fama and French (1993) three-factor model. In this model, $RMRF_t$ is the value of the market return in excess of monthly T-Bill rate; SMB_t (small minus big factor) is the difference in returns across small and big portfolios; HML_t (high minus low factor) is the difference in returns between high and low book-to-market equity portfolios; UMD_t (monthly momentum factor) is the difference in average returns on two high ex-ante return portfolios and two low ex-ante return portfolios. The SMB factor which is designed to capture the size effect is based on a portfolio comprising a long position in a portfolio of small-cap stocks financed by a short position in a portfolio of large-cap stocks. The HML factor which is meant to capture the book-to-market factor is calculated by building a portfolio that takes a long position in a portfolio of high book-to-market (value) stocks and a short position in a portfolio of low book-to-market (growth) stocks. The UMD factor, described in Jegadeesh and Titman (1993), is a momentum factor estimated from a portfolio long in high-momentum stocks and short in low-momentum stocks.

For each monthly sample, we merge Morningstar data with those taken from the CRSP database. We remove funds that do not have a complete 30-month post-evaluation data. Nonetheless, we keep track of number and ratings of the funds that are dropped from the sample. About ten percent of the funds in each sample are disqualified and the distribution of ratings is quite uniform among the discarded funds. We therefore do not adopt the usual approach of assuming returns of disappearing funds to be equal to that of a portfolio of surviving funds in the same fund category, as adopted by Blake and Morey (2000). For each eligible fund, we run a time series regression based on model (3), over a 30-month post-evaluation period.

For robustness, we perform the regression on twelve consecutive monthly data over the period December 2004 – November 2005. We thus generate twelve sets of four-factor alphas, each set representing the ex-post performance of funds in a monthly sample. Using the twelve sets of estimates, we can compute a simple estimate of the predictive strength of the Morningstar ratings. Specifically, we construct a table reporting the number of times (out of twelve sample) in which $\mu_k > \mu_l$ occurs, where μ_j denotes the mean/median out-of-sample performance measures of funds in j-star group. The larger this number, the higher the predictive power.

We can also adopt standard tests in the literature to examine the strength of the relationship between ex ante fund rating and ex post fund performance. One type of statistical test for this purpose is to classify funds as winners or losers in the ranking period and repeat the classification in the evaluation period, count the number of winner-winners, winner-losers, loser-winners and loser-losers and conduct a chi-square test of independence.

Another type of test typically involves sorting funds into performance groups (for example, deciles ranked by average performance measure) based on prior performance and computing the within-group average performance over a subsequent evaluation period. The Spearman-rho rank correlation between the in-sample ranking and out-of-sample ranking is a measure of the predictive power of in-sample ratings. We caution that traditional parametric tests such as Student-*t* test as parametric tests typically require the differences to be random samples from a normal distribution. The validity of such tests would be called into question if normality assumption fails to hold. We shall perform normality test on the performance estimates to justify our use of parametric-free tests.

A third test, employed by Blake and Morey (2000), is to perform the following cross-sectional dummy variable regression of out-of-sample performance measure on a set of dummy variables representing the in-sample rating.

$$S_i = \beta_0 + \sum_{j=1}^{4} \beta_j D_{ij} + \varepsilon_i \qquad (3.5)$$

If stewardship grades are used as the in-sample ratings, we define four binary dummy variables D_1, D_2, D_3 and D_4 each representing one level of rating. Specifically, the dummy variable D_j indicates whether the fund has a j-star rating, where j = 1, 2, 3 or 4. The 5-star rating is used as the reference group as it forms the upper bound for all ratings. Under the hypothesis that ratings possess strong predictive ability, we would expect all betas in the regression model to be negative, since funds with ratings 1 to 4 are expected to underperform five-star funds. Furthermore, strong predictive ability will translate into an increasingly negative regression coefficients : $\beta_1 < \beta_2 < \beta_3 < \beta_4 < 0$.

We consider several ways of forming in-sample performance. First, we divide funds into five star-rating groups. The resulting analysis generates results on the predictive ability of the star rating only. Similarly, we divide funds into five groups based on their stewardship grades. This allows us to measure the predictive power of stewardship grades.

Since we have two rating systems, we can also divide funds into groups based on the rating pair. For this purpose, we use both original ratings and raw scores to split the funds. Using the former, we classify funds as 'good', 'fair' or 'poor' based on the following criteria:

Category	Criterion
Good	star rating = 4 or 5 and stewardship grade = ‘A’ or ‘ B’
Poor	Star rating = 1 or 2 and stewardship grade = ‘1’ or ‘2’
Fair	Otherwise

Based on the latter, we sort funds by the sum of their ratings raw scores and organize the funds into deciles. In addition, we also partition funds by fund age and fund type to examine the robustness of our results across various age groups and asset classes respectively.

Finally, to examine the short-term persistence of Morningstar ratings, we assemble a twelve-month series of star ratings and stewardship grades. This is done by

selecting funds that are continuously graded by both star rating and stewardship grade over the twelve-month period December 2004 – November 2005. The final sample contains 1107 funds, comprising 404 bond funds, 361 domestic stock funds and 342 international stock funds.

We first compute the percentage of funds that are able to retain the same rating throughout the twelve-month period as an estimate of rating persistence. We examine the results for the entire sample as well as for sub-samples partitioned by fund age and fund type. Next, we construct and compute two measures of rating volatility. One measure is based on the actual rating. The other is derived from the raw scores of ratings. These measures indicate the degree to which fund ratings fluctuate over the observation period.

When the actual ratings are used, we define volatility of a fund's rating to be the following quantity

$$\sigma_{1i}^{rating} = \frac{1}{12}\sum_{i=1}^{11}\left|\text{grade}_{i+1} - \text{grade}_{i}\right|. \qquad (3.6)$$

When raw scores are used , the corresponding volatility measure is given by

$$\sigma_{2i}^{rating} = \sqrt{\tfrac{1}{12}\sum_{i=1}^{12}(\text{rawscore}_{i} - \overline{\text{rawscore}})^{2}} \qquad (3.7)$$

where grade_i = the rating of the fund in month i, where month 1 is December 2004 and month 12 is November 2005.

Finally, we report a 5-by-5 contingency table where each cell (i, j) displays the frequency of funds that have a rating of i in the first observation month and a rating of j in the last observation month. The diagonal entries in the table show the percentage of funds that retain the same fund rating after 11 months. Off diagonal terms give an indication of the extent to which the sample funds experience rating upgrade or downgrade over the same period.

Chapter Four

Empirical Results

4.1 Introduction

This chapter presents the statistical results of our empirical studies. In Section 4.1, we report the frequency distribution of star ratings and stewardship grades for the sample funds. This is followed by a display of summary statistics of fund characteristics in Section 4.2. We present results of correlation analysis of the two Morningstar ratings in Section 4.3. The output of the logistic regressions for finding determinants of stewardship grade are contained in Section 4.4. We analyse the out of-sample performance measures in Sections 4.5 and examine the results of Spearman-rho correlation test for predictive ability in Section 4.6. Section 4.7 is devoted to the dummy variable regression of out-of-sample performance measures on fund ratings. A discussion of the results on the test of ratings persistence in Section 4.8 concludes the chapter.

4.2 Frequency Distribution of Morningstar Ratings

First, we present in Table 1 the frequency distributions of the various star ratings and stewardship grades of the sample funds. We select only funds that receive both star rating and stewardship grade. Panel A and Panel B present the frequency distributions of star ratings and stewardship grades respectively for the June 2005 sample. We find that a majority of funds in the sample receive the middle star-rating of 3-star (35.9%) and the second best stewardship grade of B (41.7%). This observation is consistent across all fund types. A comparison between Panel A and Panel B also shows that it is more likely for funds to receive the highest star rating than to get the best stewardship grade. To

further substantiate this claim, we compile separately the percentage of funds that are awarded the top and worst star ratings and stewardship grades for the twelve monthly Morningstar data, over the period December 2004 – November 2005. The percentage of funds with 5-star rating exceeds that with A stewardship grade by about 7% for all twelve monthly data, as displayed in Panel E. In addition, we also find evidence from Panel E that the percentage of funds receiving top rating does not fluctuate very much over a short one-year period.

We extend the above analysis by partitioning funds by their age (three-year, five-year or ten-year). The results, given in Panels C and D, indicate that age does matter when it comes to fund ratings. Young (three-year) funds exhibit the highest percentage of receiving the worst 1-star rating (18.8%) and the best star rating (35.4%). In contrast, only respectively 4.9% and 2.6% of five-year and ten-year funds receive the worst rating. The percentage of ten-year and five-year funds in the best rating category are also far below that of three-year funds. This observation is nothing but a manifestation of the age bias in the star rating as posited by Blume (1998) and many others. However, the same conclusion cannot be drawn for the case of stewardship grades. We find no young funds receiving the worst stewardship grade in the June 2005 sample. In fact, this is true for all twelve monthly samples. We conduct, but do not report, results of chi-square tests of association between fund ratings and fund age. The result is highly significant (p-value < 0.0001) for both ratings, suggesting a significant association between age and fund ratings. In Section 4.5, we shall further explore the issue of potential age bias in stewardship grades.

As articulated in chapter two, the stewardship grade is a weighted sum of five scores given to five stewardship components: board quality, corporate culture, fees, manager incentive and regulatory issues. We take a microscopic view of the funds' governance by compiling the distribution of each of these component scores. We report the results in Panel F for not just the entire sample, but also for the three categories of funds, namely Domestic Stock, International Stock and Bond.

TABLE 1

Frequency of Funds Receiving Morningstar Ratings and Stewardship Grades

Panel A. Frequency Distributions of Morningstar Ratings For June 2005 Sample

This panel reports the percentage of funds that receive the various Morningstar star ratings (1-star (Worst) to 5-star(Best)) awarded on the month of June 2005, as reported in the Morningstar July 2005 Principia CD. The sample contains 1589 funds which comprise 494 domestic stock funds, 409 international stock funds and 686 bond funds. We include only funds that receive both Morningstar rating and Stewardship Grade.

	% Frequency			
Star Rating	All funds (N=1589)	Domestic Stock (N = 494)	International Stock (N = 409)	Bond (N = 686)
1-star	3.52	3.04	4.16	3.50
2-star	20.33	28.14	15.40	17.64
3-star	35.93	37.25	37.41	34.11
4-star	28.70	22.87	30.81	31.63
5-star	11.52	8.70	12.22	13.12

Panel B. Frequency Distributions of Stewardship Grades For June 2005 Sample

This panel reports the percentage of funds that receive the various Stewardship Grades (F (Worst) to A(Best)) awarded on the month of June 2005, as reported in the Morningstar July 2005 Principia CD. The sample contains 1589 funds which comprise 494 domestic stock funds, 409 international stock funds and 686 bond funds. We include only funds that receive both Morningstar rating and Stewardship Grade.

	% Frequency			
Stewardship Grade	All funds (N=1589)	Domestic Stock (N = 494)	International Stock (N = 409)	Bond (N = 686)
F	5.92	3.85	0.24	10.79
D	11.64	15.99	8.81	10.20
C	36.44	36.44	35.45	37.03
B	41.66	38.06	49.88	39.36
A	4.34	5.66	5.62	2.62

Panel C. Frequency Distributions of Morningstar Ratings, Partitioned by Age Groups, for June 2005 Sample

This panel reports the percentage of funds in each of the three age groups: 3-year, 5-year and 10-year, that receive the various Morningstar ratings awarded on the month of June 2005, as reported in the Morningstar July 2005 Principia CD. 3-year funds are funds that have at least three year but not more than 5 years of history. 5-year funds are those with at least 5 years but less than 10 years of history. 10-year funds have at least 10 years of history. We include only funds that receive both Morningstar rating and Stewardship Grade.

	% Frequency			
Star Rating	All funds (N =1589)	3-year (N = 48)	5-year (N = 284)	10-year (N =1257)
1-star	3.52	18.75	4.93	2.63
2-star	20.33	25.00	20.07	20.21
3-star	35.93	14.58	36.27	36.67
4-star	28.70	6.25	32.04	28.80
5-star	11.52	35.42	6.69	11.69

Panel D. Frequency Distributions of Stewardship Grades, Partitioned by Age Groups for June 2005 Sample

This panel reports the percentage of funds in each of the three age groups: 3-year, 5-year and 10-year, that receive the various Stewardship Grades awarded on the month of June 2005, as reported in the Morningstar July 2005 Principia CD. 3-year funds are funds that have at least three year but not more than 5 years of history. 5-year funds are those with at least 5 years but less than 10 years of history. 10-year funds have at least years of history. We include only funds that receive both Morningstar rating and Stewardship Grade.

	% Frequency			
Stewardship Grade	All funds (N =1589)	3-year (N = 48)	5-year (N = 284)	10-year (N =1257)
F	5.92	0	4.23	6.52
D	11.64	18.75	12.68	11.14
C	36.44	14.58	29.93	38.74
B	41.66	60.42	44.01	40.41
A	4.34	6.25	9.15	3.18

Panel E. Percentage of Funds Receiving Best and Worst Morningstar Ratings Over 12 Consecutive Months (December 2004 to November 2005).

This panel reports the percentage of funds that receive the best and worst Morningstar star ratings (1-star (Worst) to 5-star(Best)) and stewardship grades (F (Worst) to A(Best)) awarded over 12 consecutive months December 2004 to November 2005, as reported in the Morningstar Principia January 2005 to December 2005 CDs. We include only funds that receive both Morningstar rating and Stewardship Grade.

	% Frequency			
	Star Rating		Stewardship Grade	
Month	5-star	1-star	Grade A	Grade F
Dec 04	4.40	6.49	13.20	5.24
Jan 05	4.44	6.10	12.41	5.34
Feb 05	4.38	6.41	12.07	5.87
Mar 05	4.29	6.57	11.44	4.88
Apr 05	4.27	6.54	11.08	4.47
May 05	4.26	6.26	11.75	3.74
Jun 05	4.34	5.92	11.52	3.52
Jul 05	4.29	5.16	11.44	3.48
Aug 05	4.43	4.85	11.89	4.00
Sep 05	4.58	2.44	12.80	4.23
Oct 05	4.57	1.34	12.79	6.18
Nov 05	4.56	1.32	12.55	5.77

Panel F. Frequency Distributions of Raw Scores Assigned to the Five Stewardship Components for June 2005 Sample

This panel reports the percentage of funds that receive the various scores given to the five Stewardship assessment components: Board Quality (BQ), Corporate Culture (CC), Fees (Fees), Manager Incentives (MI) and Regulatory Issues (RI), on the month of June 2005, as reported in the Morningstar July 2005 Principia CD. With the exception of Regulatory Issues whose minimum attainable score is -2, the worst possible score a fund can receive in each of the other components is 0. A letter grade is assigned to the overall Stewardship Grade as follows: A: 9 - 10 points, B: 7 - 8.5 points, C: 5 - 6.5 points, D: 3 - 4.5 points and F: 2.5 points or less. Funds' scores for each of the five components are reported in qualitative terms: Excellent = 2 points, Good = 1.5 points, Fair = 1 point, Poor = 0.5 point and Very Poor = 0 points and below.

% Frequency

	All funds (N=1589)					Domestic Stocks (N =494)					International Stocks (N =409)					Bond (N = 686)				
Grade	BQ	CC	Fees	MI	RI	BQ	CC	Fees	MI	RI	BQ	CC	Fees	MI	RI	BQ	CC	Fees	MI	RI
Very Poor	0.13	2.58	10.89	16.93	6.23	0.20	4.05	12.96	21.05	4.05	0.24	0.24	9.05	15.89	0.24	0.00	2.92	10.50	14.58	11.37
Poor	5.22	11.58	6.86	24.73	19.19	4.86	7.29	11.34	24.90	21.86	2.44	3.91	3.91	15.16	19.07	7.14	19.24	5.39	30.32	17.35
Fair	17.56	35.56	22.97	42.42	8.56	24.29	37.45	22.87	33.81	8.30	9.78	38.14	23.96	58.19	10.02	17.35	32.65	22.45	39.21	7.87
Good	65.14	34.05	18.75	9.19	4.22	61.74	30.97	22.47	8.70	3.24	58.68	41.81	16.14	3.91	6.36	71.43	31.63	17.64	12.68	3.64
Excellent	11.96	16.24	40.53	6.73	61.80	8.91	20.24	30.36	11.54	62.55	28.85	15.89	46.94	6.85	64.30	4.08	13.56	44.02	3.21	59.77

We make a few important observations based on the output in Panel F. First, a high percentage (60%) of funds get the best qualitative score of 'Excellent' for Regulatory Issue. This observation is consistent across the three category groups. Next, funds in this sample fare poorly in Manager Incentives, with 30 – 50% of the funds graded 'Poor' or 'Very Poor', suggesting that low level of fund ownership of portfolio managers or adoption of compensation schemes that hurt the interests of shareholders might still be prevalent in the industry. We also find that a majority of funds, regardless of their fund type, attain 'Good' or 'Excellent' for Fees, indicating that fund managements are conscientious in keeping fund expenses low. Finally, we observe that funds in the International Stock category surpass their peers in other categories in Board Quality: about 29% of International Stock funds get the best rating whereas only 9% and 4% of domestic stock funds and bond funds have made this achievement. In fact, 87% of the International Stock funds are awarded the best and second best grade. This figure far exceeds that for each of the other two groups. We examine (but do not report results in Table 1) all twelve monthly samples and find that this observation is consistent across all 12 samples. Although we do not have an explanation for this observation, we reckon that US investors should diversify by investing not just in funds that focus on domestic stocks, but also in those that manage assets globally for additional benefits that non-domestic funds might be able to offer.

Panel G displays two-way contingency tables for the frequency distributions of the pair of ratings: star rating and stewardship grade, for Domestic stock funds in the June 2005 sample. Results for International Stock funds and Bond funds are contained in Panels H and I respectively. Funds that do well in one rating but poorly in another (for example, funds with (5-star, Grade F) rating pair) or receive the best grade in both ratings (that is, the (5-star, Grade A) rating pair) are very scarce. Across the three fund types, less than 1% of sample funds fall within each of these categories. About 1% to 3% of funds are awarded the highest grades in both ratings, with funds in the International Stock funds category registering the highest percentage. We also perform (but do not report in Table 1) chi-square test for dependence between the two ratings based on all twelve samples. The test statistics is highly significant in all twelve samples and for all fund

types. These results suggest that we have to interpret Morningstar's claim on the independence of the two ratings with caution.

Panel G. Two-way Contingency Table of Rating Pairs for June 2005 Sample

This panel displays the two-way contingency table for the pair of ratings (MR, SG) where MR is the Morningstar star rating (1-star (Worst) to 5-star(Best)) and SG the corresponding stewardship grade (F (Worst) to A (Best)), that funds received at the end of June 2005, as reported in the Morningstar July 2005 Principia Discs. The sample contains 1589 funds which comprise 494 domestic stock funds, 409 international stock funds and 686 bond funds. We include only funds that receive both Morningstar rating and Stewardship Grade.

		Stewardship Grade					
		F	D	C	B	A	Total
Star Rating	1-Star	8 0.5	15 0.94	21 1.32	12 0.76	0 0	56 3.52
	2-Star	28 1.76	64 4.03	155 9.75	69 4.34	7 0.44	323 20.33
	3-Star	35 2.2	67 4.22	245 15.42	213 13.4	11 0.69	571 35.93
	4-Star	17 1.07	29 1.83	124 7.8	261 16.43	25 1.57	456 28.7
	5-Star	6 0.38	10 0.63	34 2.14	107 6.73	26 1.64	183 11.52
	Total	94 5.92	185 11.64	579 36.44	662 41.66	69 4.34	1589 100

Panel H. Two-way Contingency Table of Rating Pairs for Domestic Stock Funds In June 2005 Sample

This panel displays the two-way contingency table for the pair of ratings (MR, SG) where MR is the Morningstar star rating (1-star (Worst) to 5-star (Best)) and SG the corresponding stewardship grade (F (Worst) to A (Best)) that funds in the "Domestic Stock" category received at the end of June 2005, as reported in the Morningstar July 2005 Principia Discs. We include only funds that receive both Morningstar rating and Stewardship Grade. The sample contains 494 funds.

		Stewardship Grade					
		F	D	C	B	A	Total
Star Rating	1-Star	2 0.4	9 1.82	3 0.61	1 0.2	0 0	15 3.04
	2-Star	6 1.21	29 5.87	75 15.18	25 5.06	4 0.81	139 28.14
	3-Star	7 1.42	28 5.67	75 15.18	70 14.17	4 0.81	184 37.25
	4-Star	3 0.61	11 2.23	20 4.05	66 13.36	13 2.63	113 22.87
	5-Star	1 0.2	2 0.4	7 1.42	26 5.26	7 1.42	43 8.7
	Total	19 3.85	79 15.99	180 36.44	188 38.06	28 5.67	494 100

Panel I. Two-way Contingency Table of Rating Pairs for International Stock Funds In June 2005 Sample

This panel displays the two-way contingency table for the pair of ratings (MR, SG) where MR is the Morningstar star rating (1-star (Worst) to 5-star (Best)) and SG the corresponding Stewardship grade (F (Worst) to A (Best)) that funds in the "International Stock" category received at the end of June 2005, as reported in the Morningstar July 2005 Principia Discs. We include only funds that receive both Morningstar rating and Stewardship Grade. The sample contains 404 funds.

		Stewardship Grade					
		F	D	C	B	A	Total
Star Rating	1-Star	1	5	7	4	0	17
		0.24	1.22	1.71	0.98	0	4.16
	2-Star	0	8	32	22	1	63
		0	1.96	7.82	5.38	0.24	15.40
	3-Star	0	12	63	76	2	153
		0	2.93	15.4	18.58	0.49	37.41
	4-Star	0	8	34	77	7	126
		0	1.96	8.31	18.83	1.71	30.81
	5-Star	0	3	9	25	13	50
		0	0.73	2.2	6.11	3.18	12.22
	Total	1	36	145	204	23	409
		0.24	8.8	35.45	49.88	5.62	100.00

4.3 Descriptive Statistics of Fund Characteristics

Table 2 displays the descriptive statistics of several fund-specific characteristics which include fund size, expense ratio and total returns, for the June 2005 sample. Panel A contains the statistics for the entire sample. Panels B and C display the results for funds the various star rating and stewardship grade groups respectively. We perform Kruskal-Wallis test to gauge the significance of equality of medians for each variable across various rating groups. Of particular interest are expense ratio and fund size. The statistical tests reject the hypotheses of equal medians at the 1% significance level. Both the mean and median expense ratio decreases monotonically with increasing star rating. This is consistent with observation found in various studies on the relation between fund expenses and fund performance. This relation, however, does not hold strongly for stewardship grades.

Fund size, as measured by the natural logarithm of the net asset size of the fund, exhibits a positive relation with star ratings: it increases monotonically with improving star rating. The same relation partially holds between fund size and stewardship grade: Grade-A funds have larger asset base than funds in other rating groups. Grade-F funds are on average smaller than grade-C and grade-D funds. The only case when the relation fails to hold is when we compare grade-C funds with grade-D funds. The differences in median fund size across rating groups are significant at the 1% significance level. We also apply two-sample t-tests to all twelve samples to see if the difference in means of expense ratio and fund size of funds in the best (five-star or Grade-A) and worst (one-star or Grade-F) rating groups are significant. We find that the differences in means are significant at the 1% level for all 48 cases (12 months, two fund characteristics and two ratings).

To further investigate the relation between expense ratio and fund ratings, we compute the Spearman-rho rank correlation coefficients between expense ratio and each of the two ratings for all the twelve monthly data. We do this in two ways. First, we compute the average expense ratio of funds within each rating group. We then rank the five groups by these averages. The Spearman-rho rank correlation coefficient measures the degree to which the two rankings differ. Second, we directly compute the Spearman-rho rank coefficient coefficients between the rating raw scores and the expense ratios for all the funds. As the results based on both methods are similar, we only report the former in Panel D.

For the case of star rating, the results show that for all twelve samples, star ratings are perfectly and negatively correlated with expense ratio at the 1% level of significance. However, no such relationship is found between expense ratio and stewardship grade even though one of the stewardship component is Fees which in turn includes funds' expense ratio as one of the factors of consideration. We shall investigate this issue further in Section 4.5. We also find that there is a perfectly positive correlation between star rating and mean fund size as measured by the logarithm of the total net asset of the funds. A similar relationship between stewardship grade and fund size also emerges.

TABLE 2
Summary Statistics of Fund-specific Variables

Panel A. Summary Statistics of Fund-specific Variables for June 2005 Sample

This panel reports the descriptive statistics, including mean, median, range and skewness, of selected fund variables. The sample contains 1589 funds which comprise 494 domestic stock funds, 409 international stock funds and 686 bond funds. We include only funds that received both the Morningstar rating and Stewardship grades.

	Mean	Median	Standard Deviation	Kurtosis	Skewness	Range	Minimum	Maximum
AR04	10.99	10.31	7.43	0.94	0.76	55.13	-11.56	43.57
BFAlpha	0.26	-0.23	3.55	12.45	2.51	39.28	-9.57	29.71
BFBeta	0.91	0.94	0.42	277.80	-13.23	10.23	-7.97	2.26
BFRsq	90.15	94.00	11.65	9.93	-2.77	88.00	12.00	100.00
BQ	1.42	1.50	0.35	1.12	-0.77	2.00	0.00	2.00
CC	1.25	1.50	0.49	-0.34	-0.26	2.00	0.00	2.00
ER	1.23	0.02	1.20	0.60	-0.46	0.13	3.13	0.00
Fees	1.36	1.50	0.67	-0.65	-0.71	2.00	0.00	2.00
LAR12m	13.12	11.08	10.55	3.21	1.33	85.66	-15.09	70.57
LAR3yr	12.24	12.08	7.77	2.81	1.09	54.92	0.09	55.01
MI	0.82	1.00	0.54	-0.25	0.26	2.00	0.00	2.00
MT	6.49	5.30	4.88	1.30	1.10	28.80	0.10	28.90
MR	3.24	3.00	1.02	-0.59	-0.04	4.00	1.00	5.00
log(asset)	4.89	5.38	2.75	0.31	-0.78	13.59	-2.30	11.28
RI	1.48	2.00	0.71	-0.96	-0.85	2.00	0.00	2.00
RTS1m	0.30	-0.09	2.09	-0.79	0.00	11.10	-4.71	6.39
RTS1yr	8.90	7.39	6.94	3.08	1.32	55.68	-9.81	45.87
RTS3yr	10.29	9.76	6.92	3.46	1.25	48.73	-0.33	48.40
SR	1.06	0.98	0.52	0.53	0.53	3.80	-0.79	3.01
SD3	10.02	11.43	5.62	1.28	0.58	47.19	0.31	47.50
SG	3.27	3.00	0.93	0.14	-0.67	4.00	1.00	5.00
TNH	507.82	174.50	1847.90	203.23	12.74	36030.00	1.00	36031.00
TR07	82.46	45.00	110.98	13.19	3.30	823.00	1.00	824.00

Panel B. Summary Statistics of Fund Variables by Star Ratings for June 2005 Sample

This panel reports the median sample values of selected fund characteristics across different categories of star ratings. The p-value for the Kruskal-Wallis tests of equality of medians are given in parentheses ( ). The sample contains 1589 funds which comprise 494 domestic stock funds, 409 international stock funds and 686 bond funds. We include only funds that received both the Morningstar rating and Stewardship Grade. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

	1-star (N = 56)		2-star (N = 323)		3-star (N = 571)		4-star (N = 456)		5-star (N = 183)	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
AR0407***(<0.0001)	9.25	4.11	9.11	8.43	11.16	10.62	11.75	10.71	12.46	11.76
BFAlpha07**(0.024)	-1.82	-2.42	-1.46	-1.60	-0.09	-0.57	1.10	0.31	2.79	1.32
BFBeta07 (0.2209)	0.82	0.94	0.87	0.94	0.91	0.94	0.93	0.94	0.99	0.95
BFRsq07***(0.0004)	86.45	96.00	90.52	94.00	91.33	95.00	89.87	94.00	88.10	90.00
BQ07***(<0.0001)	1.34	1.50	1.30	1.50	1.39	1.50	1.48	1.50	1.57	1.50
CC07***(<0.0001)	1.00	1.00	1.06	1.00	1.19	1.00	1.39	1.50	1.48	1.50
ER07***(<0.0001)	1.66	1.64	1.48	1.52	1.29	1.25	1.04	0.99	0.95	0.88
Fees07***(<0.0001)	0.84	0.50	1.18	1.00	1.30	1.50	1.56	2.00	1.50	1.50
LAR12m07***(<0.0001)	11.18	3.77	10.98	9.85	13.26	11.56	13.74	11.16	15.57	13.25
LAR3yr07***(<0.0001)	10.01	6.41	9.93	9.99	12.33	12.07	12.95	13.02	14.99	14.97
MI07**(0.0117)	0.67	0.50	0.76	1.00	0.84	1.00	0.85	1.00	0.86	1.00
MT07***(<0.0001)	5.20	4.50	4.91	3.60	6.46	4.70	7.36	6.90	7.65	6.60
MR07***(<0.0001)	1.00	1.00	2.00	2.00	3.00	3.00	4.00	4.00	5.00	5.00
Logasset***(<0.0001)	3.48	3.64	4.08	4.56	4.53	5.18	5.64	6.10	6.00	6.36
RI07***(<0.0001)	1.13	1.00	1.26	1.50	1.42	2.00	1.65	2.00	1.74	2.00
RTS1m07***(<0.0001)	-0.61	-1.17	-0.05	-0.38	0.29	0.08	0.50	0.15	0.77	0.02
RTS1yr07***(<0.0001)	7.45	3.05	7.23	6.29	8.97	7.65	9.48	7.65	10.69	9.27
RTS3yr07***(<0.0001)	8.90	5.84	8.37	8.27	10.31	9.72	10.76	10.73	12.78	12.73
SR07*** (<0.0001)	0.93	0.80	0.82	0.75	1.06	0.98	1.17	1.09	1.24	1.17
SD307 (0.4521)	10.62	12.43	10.25	12.32	10.32	11.78	9.53	10.02	9.72	8.87
SG07***(<0.0001)	2.66	3.00	2.89	3.00	3.17	3.00	3.54	4.00	3.75	4.00
TNH07***(<0.0001)	208.98	137.50	332.01	142.00	548.89	167.00	538.82	221.00	702.30	266.00
TR07***(<0.0001)	109.18	63.50	332.01	142.00	84.87	52.00	67.52	39.00	70.30	34.00

Panel C. Summary Statistics of Fund Variables by Stewardship Grades for June 2005 Sample

This panel reports the median sample value of several fund characteristics across different categories of Stewardship Grades. The p-value for the Kruskal-Wallis tests of equality of medians for selected variables (highlighted) are given in parentheses. The sample contains 1589 funds which comprise 494 domestic stock funds, 409 international stock funds and 686 bond funds. We include only funds that received both the Morningstar rating and Stewardship Grade(the corresponding frequency distributions of which are displayed in Panel B). The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

.	Grade F (N=94)		Grade D (N = 185)		Grade C (N = 579)		Grade B (N = 662)		Grade A (N = 69)	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
AR0407***(<0.0001)	6.68	4.36	9.95	8.37	10.57	10.18	11.94	11.16	14.24	13.76
BFAlpha07**(0.024)	-0.25	-0.79	-1.20	-0.92	-0.51	-0.84	1.13	0.12	2.03	1.17
BFBeta07***(<0.0001)	0.62	0.92	0.97	0.96	0.91	0.94	0.94	0.95	0.87	0.95
BFRsq07***(<0.0001)	86.93	93.00	93.26	96.00	90.72	94.00	89.84	93.00	84.43	89.00
BQ07***(<0.0001)	1.33	1.50	1.07	1.00	1.31	1.50	1.58	1.50	1.86	2.00
CC07***(<0.0001)	0.31	0.50	0.81	1.00	1.01	1.00	1.64	1.50	1.93	2.00
ER07***(<0.0001)	1.32	1.30	1.49	1.53	1.39	1.36	1.04	1.03	1.05	1.00
Fees07***(<0.0001)	0.56	0.50	0.46	0.50	1.32	1.50	1.70	2.00	1.79	2.00
LAR12m07***(<0.0001)	5.70	4.22	11.75	9.61	12.35	10.42	14.89	13.28	16.52	18.26
LAR3yr07***(<0.0001)	7.20	4.97	10.68	10.21	11.27	11.69	13.89	13.36	15.67	16.65
MI07***(<0.0001)	0.37	0.50	0.72	0.50	0.80	1.00	0.84	1.00	1.71	2.00
MT07***(<0.0001)	5.21	3.60	4.70	3.50	6.14	4.10	7.37	6.60	7.85	7.20
MR07***(<0.0001)	2.84	3.00	2.76	3.00	2.99	3.00	3.58	4.00	4.01	4.00
Logasset***(<0.0001)	3.67	4.07	4.24	4.59	4.03	4.75	5.92	6.16	6.36	6.75
RI07***(<0.0001)	0.02	0.00	0.99	0.50	1.25	1.00	1.98	2.00	1.99	2.00
RTS1m07***(<0.0001)	-0.91	-0.71	-0.09	-0.28	0.16	-0.19	0.63	0.38	1.09	1.10
RTS1yr07***(<0.0001)	4.17	3.06	7.89	6.27	8.42	6.92	10.01	8.78	11.02	12.19
RTS3yr07***(<0.0001)	6.60	4.97	9.04	8.43	9.26	9.24	11.54	10.88	13.28	14.30
SR07*** (<0.0001)	0.96	0.91	0.89	0.81	1.02	0.94	1.14	1.03	1.22	1.17
SD307***(<0.0001)	6.73	4.79	9.99	11.89	9.74	10.73	10.64	12.27	10.99	12.21
SG07***(<0.0001)	1.00	1.00	2.00	2.00	3.00	3.00	4.00	4.00	5.00	5.00
TNH07***(<0.0001)	304.96	246.50	246.72	115.00	482.33	151.00	669.32	288.00	143.57	69.00
TR07***(<0.0001)	92.10	44.50	99.59	60.00	85.58	52.00	77.41	42.00	45.74	29.00

Panel D. Spearman-rho Rank Correlation between Expense Ratio/Fund Size and Star Rating/Stewardship Grade for Twelve Monthly Samples: December 2004 to November 2005
This panel reports the Spearman-rho rank correlation coefficient between fund rating and median expense ratio/log of total net asset of funds for 12 consecutive monthly samples from December 2004 to November 2005, as reported in the Morningstar January 2005 to December 2005 Principia CDs. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively

	Spearman-rho Rank Correlation			
Month	Expense Ratio		Fund Size	
	Star Rating	Stewardship Grade	Star Rating	Stewardship Grade
Dec 04	-1.00***	-0.6	1.00***	0.9**
Jan 05	-1.00***	-0.6	1.00***	0.9**
Feb 05	-1.00***	-0.6	1.00***	0.9**
Mar 05	-1.00***	-0.6	1.00***	0.9**
Apr 05	-1.00***	-0.6	1.00***	0.9**
May 05	-1.00***	-0.6	1.00***	0.9**
Jun 05	-1.00***	-0.6	1.00***	0.9**
Jul 05	-1.00***	-0.7	1.00***	0.9**
Aug 05	-1.00***	-0.6	1.00***	0.9**
Sep 05	-1.00***	-0.7	1.00***	0.9**
Oct 05	-1.00***	-0.7	1.00***	0.8
Nov 05	-1.00***	-0.7	1.00***	0.9**

4.4 Correlation Between Star Ratings and Stewardship Grades

We now turn our focus on the correlation between the two Morningstar ratings: star rating and stewardship grade. For this purpose, we compute for each of the twelve monthly samples the mean stewardship grade of funds within each star rating group and rank the funds by these averages. The Spearman rank correlation coefficient and Pearson linear product-moment correlation coefficient between the two rankings are then computed. Table 3 Panel A reports the results.

Based on all twelve monthly data, we find that the two ratings are perfectly and positively correlated when Spearman rank correlation coefficient is used. The results based on Pearson product-moment linear correlation also indicate strong correlation of above 0.97 in all twelve months. When we repeat the correlation analysis by computing the correlation coefficients using the rating raw scores, we obtain the same conclusion.

TABLE 3
Correlation Analysis of Star Ratings and Stewardship Grades

Panel A. Correlation Between Star Ratings and Stewardship Grades for Twelve Monthly Samples December 2004 to November 2005

This panel reports the Spearman-rho Rank and Pearson linear product-moment correlation coefficient between star ratings and stewardship grades of funds for 12 consecutive monthly sample from December 2004 to November 2005, as reported in the Morningstar January 2005 to December 2005 Principia CDs. Funds' ranking by star rating is compared with the ranking induced by the mean stewardship grade within each star rating group. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

Month	Spearman-rho Rank Correlation Coefficient	Pearson Product -moment Correlation Coefficient
Dec 04	1.000*** (<0.0001)	0.99309*** (0.0007)
Jan 05	1.000*** (<0.0001)	0.99051*** (0.004)
Feb 05	1.000*** (<0.0001)	0.99868*** (<0.0001)
Mar 05	1.000*** (<0.0001)	0.99252*** (0.0008)
Apr 05	1.000*** (<0.0001)	0.99081*** (0.0011)
May 05	1.000*** (<0.0001)	0.99984*** (<0.0001)
Jun 05	1.000*** (<0.0001)	0.99598*** (0.0003)
Jul 05	1.000*** (<0.0001)	0.99119*** (0.001)
Aug 05	1.000*** (<0.0001)	0.97671*** (0l0043)
Sep 05	1.000*** (<0.0001)	0.97124*** (0.0058)
Oct 05	1.000*** (<0.0001)	0.98558*** (0.0021)
Nov 05	1.000*** (<0.0001)	0.98587*** (0.002)

To investigate whether there is a significant difference between the raw scores of the two Morningstar ratings for individual funds, we perform paired-sample t-test and Wilcoxon signed-rank test on the June 2005 data, with funds partitioned by age and fund type. The results displayed in Panel B indicate that the differences in means are highly significant for the entire sample. However, when we consider age-based and category-based groups, the differences in raw scores are only found to be significant in seasoned funds (10-year) and Bond funds.

Panel B. Paired-Sample t-test and Wilcoxon Signed-rank Test for Differences in Mean and Median Star Rating Raw Scores and Stewardship Grade Raw Scores

This panel reports the results of paired-sample t-test and Wilcoxon signed-rank test on the difference between star rating raw scores and stewardship grade raw scores of the June 2005 sample funds, as reported in the Morningstar July 2005 Principia CD. Raw scores for star ratings are computed based on funds' risk-adjusted returns over three periods: trailing 3, 5, and 10-years, where applicable. Ratings are recalculated each month. A raw score is assigned for each of these periods and the overall score is a weighted sum of these scores. Funds are then ranked by the overall raw scores and their final star rating assigned as follows: 5-star: top 10%, 4star: next 22.5%, 3-star: middle 35%, 2-star: next 22.5% and 1-star: bottom 10%.. Raw score for stewardship is the sum of the point awarded for each Stewardship components. The final stewardship grade is assigned as follows: A: 9 - 10 points, B: 7 - 8.5 points, C: 5 - 6.5 points, D: 3 - 4.5 points and F: 2.5 points or less. The sample contains 1589 funds which comprise 494 domestic stock funds, 409 international stock funds and 686 bond funds. We include only funds that received both the Morningstar star rating and stewardship grade. The numbers in parentheses ( ) are the p-values. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

Paired-sample t-test and Wilcoxon signed-rank Test for Mean and Median of
(Star Rating Raw Score - Stewardship Grade Raw Score)

	t-test	Wilcoxon Signed-Rank Test
All funds	2.1679**	37539**
	(0.03)	(0.0377)
3-year	-1.2444	-99.00
	(0.2195)	(0.3400)
5-year	-0.91154	-1694.5
	(0.3628)	(0.2218)
10-year	3.449***	40875.5***
	(0.0006)	(0.0012)
Domestic Stock	-1.76124*	-8389.5***
	(0.07)	(0.00015)
International Stock	-0.19535	907
	(0.8452)	(0.704)
Bond	4.701582***	25197.5***
	(<0.0001)	(<0.0001)

4.5 Determinants of Stewardship Grades

Khorana and Nelling (1998) use a probit regression model to determine influential determinants of the star ratings. They show that fund size, expense ratio, manager tenure and turn over ratio possess significant explanatory power for the star ratings. In the same vein, we perform both an ordinary least-squared regression as well as an ordered logit regression to seek determinants of the stewardship grade.

Panel A of Table 4 reports the correlation matrix of a set of Morningstar variables of funds taken from the June 2005 sample. We do not observe any highly correlated pair of variables except for the obvious ones such as stewardship grades with any of the five stewardship components.

Panel B displays the results of regressing stewardship grade raw scores (treated as a continuous variable) on short-listed variables that emerge from a rigorous model selection process that has been carried out. Except for Morningstar-style 12-month load-adjusted returns (LAR12m) and Morningstar-style best-fit R-squared (BFRsq), all other variables exhibit highly significant explanatory power for the dependent variable in all the twelve monthly regressions.

A few interesting conclusions can be made. First, expense ratio is a highly significant variable in this regression model. The negativity of its beta indicates that funds with high expense ratio are more likely to get poor stewardship grades than those with low expenses. This is not surprising, considering that fund expenses is one of the major assessment components in the Stewardship grading system. Next, we find that fund size has a significantly positive beta. This shows that the larger funds are, the more likely it is for the funds to receive good stewardship grades. Similarly, we find evidence based on the significantly positive beta for manager tenure that funds managed by more experienced portfolio managers tend to have better stewardship grades. Finally, the positive relation between stewardship grade and the star rating, controlling for other determinants, is evidenced by the highly significant beta for star rating.

Panel C contains the results of the second regression model in which both BFRsq and LAR12m are removed; the model also controls for fund type and fund age by including the following dummy variable: D_{age3}, D_{age5}, D_{bond} and $D_{domestic}$. D_{age3} and D_{age5} take value one when the fund has 3-5 and 5-10 years of history respectively, and zero otherwise. Similarly, D_{bond} and $D_{domestic}$ take value one when the fund belongs to Bond and Domestic Stock category respectively, and zero otherwise. The 10-year fund group serves as the reference for age. Similarly, the International Stock fund group is the reference for fund type.

The results are mostly consistent with those obtained in the preceding analysis. However, Morningstar-style one year return after tax and sales (RTS1yr) appears highly significant in all regressions once we control for fund age and fund type. An examination on the coefficients of the dummy variables reveals two pieces of information. First, we do not find any perceivable relation between stewardship and age in the presence of other fund characteristics. Neither D_{age3} nor D_{age5} turns out to be significant even at the 10% level, thus providing evidence that fund age does not have significant explanatory power when we control for other influential fund characteristics. Second, the coefficients of both fund-type dummy variables are highly significant and negative, indicating that the type of assets held has an influence on stewardship score. The negative sign also indicates that funds under the International Stock category fare better in stewardship grade than their peers in Bonds and Domestic Stock groups. Furthermore, domestic stocks outperform bonds in stewardship grades when all other variables are held constant.

The use of ordinary least-squared regression with the stewardship grade raw score as dependent variable has an obvious shortcoming as the stewardship grade raw score is not exactly a continuous variable (since it is a weighted sum of 5 component scores, each being a polychotomous variable, which means the raw score is at best another multinomial variable with a large number of levels). We therefore repeat the analysis by performing ordered logit regression of the actual stewardship grades on those fund-characteristics that have been found significant. The results reported in Panel D, however, do not show much deviation from those recorded in Panel C.

TABLE 4
Regression of Stewardship Grade on Potential Determinants

Panel A. Pair-wise Pearson Correlation Coefficient Matrix for Selected Fund Characteristics

This panel reports the linear correlation matrix of a set of fund characteristics that are possible determinants of the stewardship grade. The sample contains 1589 funds which comprise 494 domestic stock funds, 409 international stock funds and 686 bond funds. We include only funds that receive both star rating and stewardship grade. Appendix C contains an explanation of the abbreviated terms used in the table below.

Variables	BQ	CC	Fees	MI	MR	RI	SG	BFAlpha	BFBeta	BFRsq	ER	LAR12m	MT	logasset	RTS1m	RTS1yr	SD3
BQ	1.000	0.444	0.348	0.074	0.232	0.145	0.475	0.223	0.060	-0.044	-0.049	0.201	0.105	0.108	0.143	0.193	0.115
CC	0.444	1.000	0.438	0.158	0.302	0.638	0.839	0.209	0.112	0.011	-0.327	0.234	0.161	0.330	0.251	0.233	0.173
Fees	0.348	0.438	1.000	-0.064	0.240	0.260	0.608	0.200	0.028	-0.025	-0.344	0.064	0.145	0.248	0.070	0.060	-0.032
MI	0.074	0.158	-0.064	1.000	0.071	0.022	0.297	0.066	0.004	-0.102	0.185	0.062	0.117	0.005	0.012	0.052	0.008
MR	0.232	0.302	0.240	0.071	1.000	0.250	0.327	0.376	0.085	-0.031	-0.330	0.122	0.187	0.267	0.142	0.146	-0.052
RI	0.145	0.638	0.260	0.022	0.250	1.000	0.724	0.150	0.061	-0.056	-0.207	0.122	0.214	0.321	0.124	0.127	0.116
SG	0.475	0.839	0.608	0.297	0.327	0.724	1.000	0.234	0.076	-0.064	-0.251	0.210	0.186	0.303	0.203	0.205	0.140
BFAlpha	0.223	0.209	0.200	0.066	0.376	0.150	0.234	1.000	-0.123	-0.399	-0.093	0.298	0.044	0.134	0.061	0.299	0.162
BFBeta	0.060	0.112	0.028	0.004	0.085	0.061	0.076	-0.123	1.000	0.382	0.066	0.194	-0.035	-0.007	0.115	0.198	0.310
BFRsq	-0.044	0.011	-0.025	-0.102	-0.031	-0.056	-0.064	-0.399	0.382	1.000	-0.074	0.021	-0.022	-0.006	0.057	0.027	0.069
ER	-0.049	-0.327	-0.344	0.185	-0.330	-0.207	-0.251	-0.093	0.066	-0.074	1.000	0.244	-0.106	-0.415	-0.006	0.240	0.293
LAR12m	0.201	0.234	0.064	0.062	0.122	0.122	0.210	0.298	0.194	0.021	0.244	1.000	-0.112	0.001	0.694	0.996	0.688
MT	0.105	0.161	0.145	0.117	0.187	0.214	0.186	0.044	-0.035	-0.022	-0.106	-0.112	1.000	0.165	-0.105	-0.098	-0.103
logasset	0.108	0.330	0.248	0.005	0.267	0.321	0.303	0.134	-0.007	-0.006	-0.415	0.001	0.165	1.000	-0.056	0.000	-0.013
RTS1m	0.143	0.251	0.070	0.012	0.142	0.124	0.203	0.061	0.115	0.057	-0.006	0.694	-0.105	-0.056	1.000	0.692	0.539
RTS1yr	0.193	0.233	0.060	0.052	0.146	0.127	0.205	0.299	0.198	0.027	0.240	0.996	-0.098	0.000	0.692	1.000	0.676
SD3	0.115	0.173	-0.032	0.008	-0.052	0.116	0.140	0.162	0.310	0.069	0.293	0.688	-0.103	-0.013	0.539	0.676	1.000

Panel B. Regression of Stewardship Grade on Fund characteristics

This panel reports the regression estimates and summary statistics of the regressions of fund stewardship grade on the following fund characteristics: BERsq, ER, LAR12m, LAR3yr, MT, MR, logaset, RTS1m, RTS1yr. We perform a series of cross-sectional regressions using twelve consecutive monthly data over the period December 2004 to November 2005, using the model

$$S_i = \alpha + \sum_{j=1}^{J} \beta_j F_{ji} + \varepsilon_i$$

where $F_{1i}, F_{2i}, ..., F_{Ji}$ are the J characteristics of fund i.. The characteristics include best fit R-squared (BFRsq), expense ratio (ER), 1-yr load-adjusted return (LAR12m), 3-yr Load-adjusted return (LAR3yr), average manager tenure (MT), Morningstar star rating raw score (MR), log of fund size (logasset), 1-yr return after tax and sales (RTS1yr) ad 3-yr return after tax and sale (RTS3-yr). The model does not control for fund age and find type. Each monthly sample contains more than 1400 funds from the following three categories: (i) Domestic Stock (ii) International Stock and (iii) Bonds. We then select the most appropriate model based on the significance of the regression coefficients, the size of the regression R^2 as well as the Akiade's Information Criterion (AIC). The results below show only the results based on the final model. The numbers within < > are the t-statistics and those within ( ) are the p-values. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively

Month	Estimates of Regression Coefficients, t-statistics and p-values									
	Intercept	BFRsq	ER	LAR12m	LAR3yr	MT	MR	logasset	RTS1m	RTS1yr
Dec04	0.9516***	-0.0004	-0.0903***	0.0300***	0.0164***	0.0094***	0.0354***	0.0357***	0.0105*	-0.0501***
	<8.903>	< -.3738>	< -4.6195>	<3.0459>	<5.7471>	<4.7446>	<3.358>	<6.7877>	<1.699>	< -3.2644>
	(0.0000)	(0.70861)	(0.0000)	(0.00238)	(0.0000)	(0.0000)	(0.00081)	(0.0000)	(0.08961)	(0.00113)
Jan 05	1.0413***	-0.0009	-0.1019***	0.0228**	0.0196***	0.0125***	0.0274***	0.0285***	0.0149*	-0.0460***
	<9.6233>	< -.9241>	< -5.6356>	<2.2128>	<6.6905>	<6.3921>	<2.629>	<5.4347>	<1.9078>	< -2.9>
	(0.0000)	(0.35563)	(0.0000)	(0.02711)	(0.0000)	(0.0000)	(0.00868)	(0.0000)	(0.05667)	(0.0038)
Feb 05	1.0199***	-0.0009	-0.0682***	0.0322***	0.0168***	0.0139***	0.0313***	0.0293***	0.0167***	-0.0489***
	<9.7235>	< -.988>	< -3.5062>	<3.2742>	<6.5003>	<7.3335>	<2.9298>	<5.7175>	<2.6435>	< -3.217>
	(0.0000)	(0.32337)	(0.00047)	(0.00109)	(0.0000)	(0.0000)	(0.00346)	(0.0000)	(0.00831)	(0.00133)
Mar 05	0.9585***	-0.0009	-0.0806***	0.0205**	0.0151***	0.0126***	0.0464***	0.0282***	-0.0056	-0.0294**
	<9.5934>	< -1.0455>	< -4.1179>	<2.1531>	<4.5614>	<6.7298>	<4.3238>	<5.4135>	< -1.0556>	< -2.0269>
	(0.0000)	(0.29602)	(0.00004)	(0.03151)	(0.00001)	(0.0000)	(0.00002)	(0.0000)	(0.29136)	(0.04289)
Apr 05	0.9208***	-0.0007	-0.0747***	0.0253***	0.0172***	0.0137***	0.0499***	0.0287***	0.0100*	-0.0412***
	<9.3949>	< -.7547>	< -4.2541>	<2.6468>	<5.192>	<7.5238>	<4.812>	<5.8576>	<1.8888>	< -2.8378>
	(0.0000)	(0.45057)	(0.00002)	(0.00823)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.05915)	(0.00462)

May 05	1.0129***	-0.0013	-0.0807***	0.0241**	0.0130***	0.0141***	0.0476***	0.0264***	0.0144**	-0.0386***
	<10.5577>	< -1.4563>	< -4.4978>	<2.4671>	<5.1643>	<7.9037>	<4.8778>	<5.5139>	<2.0715>	< -2.6115>
	(0.0000)	(0.14557)	(0.00001)	(0.01375)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.03852)	(0.00912)
Jun 05	1.1402***	-0.0023***	-0.1012***	0.0231**	0.0069***	0.0144***	0.0502***	0.0224***	0.0065**	-0.0327**
	<13.5097>	< -3.1219>	< -5.9399>	<2.4675>	<2.9446>	<8.4927>	<5.5326>	<5.358>	<1.1391>	< -2.3717>
	(0.0000)	(0.00184)	(0.0000)	(0.01373)	(0.00329)	(0.0000)	(0.0000)	(0.0000)	(0.25487)	(0.01785)
Jul 05	1.1060***	-0.0010	-0.0915***	0.0021	0.0144***	0.0110***	0.0242***	0.0231***	0.0180***	-0.0064
	<13.3645>	< -1.3713>	< -5.3617>	<.2206>	<5.4079>	<6.7443>	<2.6276>	<5.5387>	<3.1011>	< -.4697>
	(0.0000)	(0.17053)	(0.0000)	(0.82545)	(0.0000)	(0.0000)	(0.0087)	(0.0000)	(0.00197)	(0.63862)
Aug 05	1.2263***	-0.0017**	-0.1209***	-0.0031	0.0110***	0.0110***	0.0342***	0.0168***	0.0138**	-0.0003
	<14.8066>	< -2.3532>	< -7.0842>	< -.3214>	<5.4416>	<6.7661>	<4.011>	<4.1444>	<2.3026>	< -.0217>
	(0.0000)	(0.01875)	(0.0000)	(0.74797)	(0.0000)	(0.0000)	(0.00006)	(0.00004)	(0.02145)	(0.98265)
Sep 05	1.1621***	-0.0007	-0.1074***	0.0051	0.0168***	0.0123***	0.0185**	0.0194***	0.0282***	-0.0151
	<13.9249>	< -.9792>	< -6.3666>	<.6072>	<8.1812>	<7.768>	<2.2042>	<4.8441>	<4.747>	< -1.1993>
	(0.0000)	(0.32763)	(0.0000)	(0.54384)	(0.0000)	(0.0000)	(0.02767)	(0.0000)	(0.0000)	(0.23059)
Oct 05	1.2991***	-0.0024***	-0.1263***	0.0085	0.0121***	0.0098***	0.0410***	0.0166***	0.0191***	-0.0214*
	<16.3092>	< -3.3515>	< -7.9687>	<1.016>	<6.888>	<6.533>	<5.31>	<4.2855>	<3.7225>	< -1.7088>
	(0.0000)	(0.00083)	(0.0000)	(0.30983)	(0.0000)	(0.0000)	(0.0000)	(0.00002)	(0.00021)	(0.08771)
Nov 05	1.3095***	-0.0023***	-0.1220***	0.0206***	0.0105***	0.0102***	0.0325***	0.0198***	0.0314***	-0.0400***
	<16.9145>	< -3.3358>	< -7.8261>	<3.226>	<7.4771>	<6.8782>	<4.3521>	<5.2054>	<6.2979>	< -4.0479>
	(0.0000)	(0.00087)	(0.0000)	(0.00128)	(0.0000)	(0.0000)	(0.00001)	(0.0000)	(0.0000)	(0.00005)

Month	1	2	3	4	5	6	7	8	9	10	11	12
R^2	0.2956	0.2863	0.3168	0.3009	0.3060	0.2957	0.2867	0.2682	0.2515	0.2728	0.2814	0.2825
F Stats	50.1279	48.3903	57.7468	54.2218	55.8301	53.4268	53.0126	48.9478	46.1243	52.7666	53.0524	54.2558

Month 1 – Dec 04, Month 2 – Jan 05, …, Month 12 – Nov 05

Panel C. Regression of Stewardship Grade on Fund characteristics

This panel reports the regression estimates and summary statistics of the regressions of fund stewardship grade on the following fund characteristics: ER, LAR3yr, MT, MR, logaset, RTS1m, RTS1yr. We perform a series of cross-sectional regressions using twelve consecutive monthly data over the period December 2004 to November 2005, using the model

$$S_i = \alpha + \sum_{j=1}^{J} \beta_j F_{ji} + \sum_{k=1}^{K} \gamma_k D_{ki} + \varepsilon_i$$

where $F_{1i}, F_{2i}, ..., F_{Ji}$ are the J characteristics of fund i. and $D_{1i}, D_{2i}, ..., D_{Ki}$ are dummy variables associated with fund age and fund type. Dummy variables representing age are Dage3 and D_{age5} which take value 1 when the fund has 3-5 and 5-10 years of history respectively; B_{ond} and $D_{omestic}$ are fund-type dummy variables taking value 1 if the fund belong to bond and domestic stock category respectively. The 10-year fund group serves as the reference group for age. Similarly, the International Stock funds group is the reference for fund type. Characteristics include expense ratio (ER), 3-yr Load-adjusted return (LAR3yr), average manager tenure (MT), Morningstar rating (MR), log of fund size (log asset), 1-yr return after tax and sales (RTS1yr) ad 3-yr return after tax and sale (RTS3-yr). Each monthly sample contains more than 1400 funds from the following three categories: (i) Domestic Stock (ii) International Stock and (iii) Bonds. We then select the most appropriate model based on the significance of the regression coefficients as well as on the Akiade's Information Criterion. The results below show only the results based on the final model. The numbers within < > are the t–statistics and those in ( ) are the p-values. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

Month	Estimates of Regression Coefficients, t-statistics and p-values											
	Intercept	ER01	LAR3yr	MT	MR	logasset	RTS1m	RTS1yr	D_{age3}	D_{age5}	D_{bond}	$D_{domestic\ stock}$
Dec04	1.1948***	-0.1107***	0.0248***	0.0136***	0.0408***	0.0242***	0.0431***	-0.0278***	0.0329	0.0414*	-0.26477***	-0.08906***
	<19.571>	< -5.973>	<8.341>	<7.074>	<4.106>	<6.058>	<5.69>	< -5.962>	<.585>	<1.76>	< -7.892>	< -3.345>
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.5586)	(0.0787)	(0.0000)	(0.0008)
Jan 05	1.0939***	-0.1309***	0.0216***	0.0146***	0.0399***	0.0211***	0.0225***	-0.0216***	0.0578	0.0567**	-0.15126***	-0.10685***
	<19.238>	< -7.155>	<7.478>	<7.712>	<4.085>	<5.217>	<2.937>	< -4.677>	<1.006>	<2.374>	< -5.057>	< -3.991>
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0034)	(0.0000)	(0.3146)	(0.0177)	(0.0000)	(0.0001)
Feb 05	1.1502***	-0.0919***	0.0224***	0.0153***	0.0311***	0.0270***	0.0416***	-0.0162***	0.0717	0.0738***	-0.23968***	-0.10333***
	<20.445>	< -5.019>	<8.783>	<8.456>	<3.149>	<6.939>	<6.552>	< -4.808>	<1.368>	<3.223>	< -8.208>	< -3.979>
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0017)	(0.0000)	(0.0000)	(0.0000)	(0.1715)	(0.0013)	(0.0000)	(0.0001)
Mar 05	1.0909***	-0.0899***	0.0248***	0.0145***	0.0430***	0.0264***	0.0295***	-0.0168***	0.0614	0.0741***	-0.24045***	-0.08149***
	<18.487>	< -4.87>	<7.546>	<8.174>	<4.258>	<6.699>	<4.419>	< -4.351>	<1.232>	<3.222>	< -7.635>	< -3.095>
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.2181)	(0.0013)	(0.0000)	(0.002)
Apr 05	1.0748***	-0.1072***	0.0227***	0.0148***	0.0497***	0.0261***	0.0279***	-0.0174***	0.0559	0.0479**	-0.20291***	-0.18335***
	<19.048>	< -5.924>	<7.19>	<8.451>	<5.107>	<6.697>	<4.46>	< -4.782>	<1.111>	<2.101>	< -7.868>	< -6.163>
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.267)	(0.0358)	(0.0000)	(0.0000)

May 05	1.0632***	-0.1122***	0.0163***	0.0144***	0.0458***	0.0257***	0.0261***	-0.0122***	0.2305***	0.0515*	-0.15701***	-0.13765***
	<19.268>	< -6.205>	<6.579>	<8.416>	<4.852>	<6.617>	<3.813>	< -3.878>	<3.552>	<1.731>	< -6.334>	< -5.212>
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0004)	(0.0836)	(0.0000)	(0.0000)
7Jun 05	1.0230***	-0.1041***	0.0118***	0.0133***	0.0464***	0.0272***	0.0244***	-0.0081***	0.1113**	0.0344	-0.09131***	-0.15571***
	<19.054>	< -6.044>	<5.313>	<8.054>	<5.196>	<7.44>	<3.872>	< -2.715>	<2.379>	<1.557>	< -3.216>	< -6.325>
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0067)	(0.0175)	(0.1196)	(0.0013)	(0.0000)
Jul 05	1.1045***	-0.1035***	0.0158***	0.0108***	0.0326***	0.0244***	0.0269***	-0.0101***	0.0256	0.0408*	-0.12936***	-0.04429***
	<20.976>	< -6.099>	<6.542>	<6.806>	<3.722>	<6.742>	<4.495>	< -3.48>	<.544>	<1.921>	< -4.488>	< -1.86>
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0002)	(0.0000)	(0.0000)	(0.0005)	(0.5867)	(0.0549)	(0.0000)	(0.063)
Aug 05	1.1204***	-0.1316***	0.0104***	0.0103***	0.0432***	0.0203***	0.0107***	-0.0073***	0.0848*	0.0401*	-0.09208***	-0.07553***
	<21.379>	< -7.902>	<5.064>	<6.429>	<5.227>	<5.668>	<1.999>	< -2.447>	<1.849>	<1.901>	< -2.95>	< -3.105>
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0458)	(0.0145)	(0.0647)	(0.0575)	(0.0032)	(0.0019)
Sep 05	1.2179***	-0.1120***	0.0161***	0.0117***	0.0239***	0.0228***	0.0382***	-0.0131***	0.0489	0.0520***	-0.14979***	-0.10399***
	<24.592>	< -6.979>	<8.182>	<7.713>	<3.>	<6.569>	<6.594>	< -4.957>	<1.178>	<2.583>	< -5.162>	< -4.666>
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0027)	(0.0000)	(0.0000)	(0.0000)	(0.2391)	(0.0099)	(0.0000)	(0.0000)
Oct 05	1.1497***	-0.1265***	0.0109***	0.0089***	0.0409***	0.0205***	0.0261***	-0.0118***	0.0647	0.0352*	-0.07505***	-0.10752***
	<24.631>	< -8.052>	<5.729>	<6.019>	<5.424>	<5.847>	<4.698>	< -4.405>	<1.576>	<1.784>	< -2.761>	< -4.863>
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.1153)	(0.0746)	(0.0058)	(0.0000)
Nov 05	1.0837***	-0.1261***	0.0117***	0.0096***	0.0324***	0.0221***	0.0286***	-0.0111***	0.1047**	0.0369*	-0.01036	-0.04535**
	<22.475>	< -8.152>	<6.268>	<6.511>	<4.226>	<6.298>	<5.465>	< -4.189>	<2.432>	<1.859>	< -.331>	< -1.965>
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0151)	(0.0632)	(0.741)	(0.0495)

Month	1	2	3	4	5	6	7	8	9	10	11	12
R^2	0.2956	0.2863	0.3168	0.3009	0.3060	0.2957	0.2867	0.2682	0.2515	0.2728	0.2814	0.2825
F Stats	50.1279	48.3903	57.7468	54.2218	55.8301	53.4268	53.0126	48.9478	46.1243	52.7666	53.0524	54.2558

Month 1 – Dec 04, Month 2 – Jan 05, …, Month 12 – Nov 05

Panel D. Ordered Logit Regression of Stewardship Grade on Fund Characteristics

This panel reports the regression estimates and summary statistics of the regressions of fund stewardship grade on the following fund characteristics: ER, LAR3yr, MT, MR, logaset, RTS1m, RTS1yr. We perform a series of 12 cross-sectional ordered logit regressions using twelve consecutive monthly data over the period December 2004 to November 2005. The response variable is the stewardship grade; the regressors include expense ratio (ER), 3-yr Load-adjusted return (LAR3yr), average manager tenure (MT), Morningstar rating (MR), log of fund size (log asset), 1-yr return after tax and sales (RTS1yr) ad 3-yr return after tax and sale (RTS3-yr) and dummy variables associated with fund Age and fund type. Dummy variables representing age are Dage3 and D_{age5} which respectively take value 1 when the fund has 3-5 and 5-10 years of history; B_{ond} and $D_{omestic}$ are fund-type dummy variables taking value 1 if the fund belong to bond and domestic stock category respectively. The dummy variable for 10-year fund serves as the reference for age. Similarly, the dummy for international stock funds is the reference. The numbers within < > are the Wald Chi-square Statistics and those in ( ) are the p-values. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

Month	Estimates of Regression Coefficients, t-statistics and p-values										
	ER	LAR3yr	MT	MR	logasset	RTS1m	RTS1yr	D_{age3}	D_{age5}	D_{bond}	$D_{domestic\ stoc}$
Dec04	-0.7353***	0.1695***	0.0638***	0.2386***	0.1517***	0.3382***	-0.2236***	0.4803	0.4325***	-1.9509***	-0.4608***
	<39.5488>	<75.5497>	<27.2282>	<14.5088>	<36.8001>	<49.747>	<55.7233>	<1.7481>	<8.4768>	<79.9503>	<7.4546>
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(.186)	(.004)	(0.000)	(.006)
Jan 05	-0.8664***	0.1435***	0.0662***	0.2028***	0.1300***	0.2080***	-0.1789***	0.6658*	0.4947***	-0.9832***	-0.5395***
	<56.373>	<59.17>	<30.7132>	<11.0997>	<27.1135>	<18.9881>	<37.2763>	<3.2432>	<10.9027>	<27.2515>	<10.2796>
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(.072)	(.001)	(0.000)	(.001)
Feb 05	-0.6350***	0.1575***	0.0692***	0.1253***	0.1771***	0.3428***	-0.1349***	0.5733*	0.6462***	-1.7069***	-0.5636***
	<29.2533>	<83.3654>	<34.9761>	<3.9609>	<50.721>	<69.6792>	<36.9723>	<2.8131>	<18.9288>	<77.5018>	<11.2207>
	(0.000)	(0.000)	(0.000)	(.047)	(0.000)	(0.000)	(0.000)	(.093)	(0.000)	(0.000)	(.001)
Mar 05	-0.6451***	0.1691***	0.0620***	0.2201***	0.1681***	0.2342***	-0.1267***	0.3101	0.5613***	-1.7208***	-0.4039**
	<29.8688>	<61.0991>	<29.182>	<11.6757>	<44.8115>	<30.5401>	<26.4359>	<.9473>	<14.2647>	<69.2095>	<5.6614>
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(.33)	(0.000)	(0.000)	(.017)
Apr 05	-0.7046***	0.1498***	0.0625***	0.3059***	0.1625***	0.2386***	-0.1337***	0.2804	0.3287**	-1.3938***	-1.2448***
	<37.059>	<52.1552>	<30.6081>	<24.1917>	<42.9529>	<35.4451>	<32.4117>	<.7576>	<5.1262>	<67.7184>	<41.6958>
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(.384)	(.024)	(0.000)	(0.000)
May 05	-0.7346***	0.1055***	0.0590***	0.3028***	0.1555***	0.1951***	-0.0918***	1.4563***	0.3904**	-1.0189***	-0.8351***
	<40.397>	<41.814>	<28.6031>	<25.3126>	<40.143>	<20.3801>	<20.6595>	<11.7081>	<4.1353>	<40.1843>	<24.0435>
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(.001)	(.042)	(0.000)	(0.000)

Jun 05	-0.7136***	0.0704***	0.0567***	0.3423***	0.1625***	0.1585***	-0.0545***	0.7526**	0.2932***	-0.6590***	-0.8975***
	<40.0525>	<22.8621>	<26.8219>	<34.4053>	<46.94>	<15.0664>	<7.8607>	<6.0198>	<4.1507>	<12.4685>	<30.413>
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(.005)	(.014)	(.042)	(0.000)	(0.000)
Jul 05	-0.6846***	0.0973***	0.0566***	0.2832***	0.1847***	0.2232***	-0.0713***	0.3855	0.6199***	-0.8755***	-0.2616
	<34.9912>	<34.2562>	<26.9075>	<22.6952>	<56.2407>	<30.0346>	<12.8039>	<1.4519>	<17.9041>	<19.3333>	<2.5714>
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(.228)	(0.000)	(0.000)	(.109)
Aug 05	-0.8111***	0.0628***	0.0482***	0.3659***	0.1603***	0.1066***	-0.0576***	0.6227**	0.6209***	-0.5565***	-0.5140***
	<51.8146>	<20.5301>	<20.0322>	<43.3387>	<45.015>	<8.6513>	<8.1868>	<4.0042>	<18.8088>	<6.9778>	<9.7685>
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(.003)	(.004)	(.045)	(0.000)	(.008)	(.002)
Sep 05	-0.7161***	0.1017***	0.0661***	0.2281***	0.1754***	0.2777***	-0.0897***	0.3208	0.7335***	-0.9925***	-0.8049***
	<39.7594>	<52.3418>	<37.193>	<16.8025>	<51.6533>	<46.3394>	<22.9329>	<1.2005>	<25.9786>	<23.1865>	<25.7735>
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(.273)	(0.000)	(0.000)	(0.000)
Oct 05	-0.8452***	0.0611***	0.0527***	0.3521***	0.1608***	0.1817***	-0.0763***	0.5023*	0.6248***	-0.4730**	-0.8427***
	<54.2481>	<19.683>	<23.9782>	<41.3046>	<40.6541>	<20.4495>	<15.299>	<2.8392>	<18.8963>	<5.8191>	<27.2929>
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(.092)	(0.000)	(.016)	(0.000)
Nov 05	-0.8172***	0.0734***	0.0548***	0.2754***	0.1738***	0.2336***	-0.0775***	0.5427*	0.5871***	0.0740	-0.3271**
	<52.9436>	<29.0922>	<26.0799>	<24.9853>	<47.5965>	<38.1932>	<16.4137>	<3.0672>	<16.6617>	<.109>	<3.9018>
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(.08)	(0.000)	(.741)	(.048)

In Section 4.2, we report evidence of an association between fund age and stewardship grade. We now proceed to investigate further the issue of age bias in stewardship grade by examining the statistical significance of pair-wise differences in the mean stewardship grades across the three age groups: three-year, five-year and ten-year. We report in Panel E the results based on the overall stewardship grades. We also repeat the analysis using stewardship raw scores. All the results point to the same conclusion: there is strong evidence supporting the hypothesis that both three-year and five-year funds outperform the more seasoned ten-year funds although there is no statistical evidence to indicate that three-year and five-year funds perform differently under stewardship assessment.

We extend our analysis via a dummy variable regression of stewardship grade on fund age. Seasoned (ten-year) fund group is used as the reference group. The regression coefficient corresponding to the dummy variable for three-year and five-year group represent the rating difference between the respective age-group and the reference group. A significantly positive regression coefficient indicates that funds in the younger group earn higher average stewardship grade than those in the reference group. The results shown in Panel F again reveals that younger funds generally receive better stewardship grade than seasoned funds as evidenced by the observation that the regression coefficient for both three-year and five-year dummy variables are positive and significant at the 1% level in almost all twelve monthly regressions.

Panel E. Differences in Mean Stewardship Grades Across Age Groups

This panel reports, for twelve consecutive monthly data over the period: December 2004 to November 2005, the mean stewardship grades of funds in the three age groups: 3-year, 5-year and 10-year and results of t-tests on the pair-wise differences in mean stewardship grades. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

	Average Stewardship Rating			t-test of differences in Stewardship grade		
Month	10-year	5-year	3-year	Between 3-yr & 5-yr	Between 3-yr and 10-yr	Between 5-yr and 10-yr
1	3.6889	3.3712	3.1968	2.47**	4.20***	2.61***
2	3.7045	3.4000	3.1968	2.34**	4.28***	3.04***
3	3.5625	3.4104	3.1945	1.09	2.85**	3.26***
4	3.4681	3.4453	3.1980	0.16	2.06**	3.88***
5	3.4681	3.3808	3.1907	0.62	2.07**	2.87***
6	3.8333	3.4904	3.2105	2.28**	4.70***	3.50***
7	3.5417	3.4120	3.2269	0.94	2.45**	2.95***
8	3.5000	3.5618	3.2348	-0.43	1.92**	6.00***
9	3.5745	3.5619	3.4212	0.09	2.42**	6.06***
10	3.5283	3.5681	3.2775	-0.30	2.01**	5.61***
11	3.4727	3.4914	3.2942	-0.14	1.42*	3.93***
12	3.4167	3.4800	3.2946	-0.42	0.85	3.73***

Panel F. Ordered Logit Regression of Stewardship Grade on Age Dummy Variables

This panel reports the regression estimates and summary statistics of the regressions of fund stewardship grade on two dummy variables representing fund age: D_{age3} and D_{age5}, where Dage3 and D_{age5} which respectively take value 1 when the fund has 3-5 and 5-10 years of history. Funds with age 10 years form the reference group. We perform a series of 12 cross-sectional ordered logit regressions using twelve consecutive monthly data over the period December 2004 to November 2005. The numbers within ( ) are the p-values. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

Month	D_{age3}	D_{age5}
1	1.0791***	0.3653***
	(0.0003)	(0.0042)
2	1.1439***	0.4388***
	(0.0002)	(0.0006)
3	0.8398***	0.4642***
	(0.003)	(0.0003)
4	0.5698***	0.5388***
	(0.0423)	(<0.0001)
5	0.5708**	0.4008***
	(0.0417)	(0.0012)
6	1.2643***	0.5913***
	(0.002)	(0.0002)
7	0.7406***	0.3956***
	(0.0086)	(0.0013)
8	0.6468**	0.6528***
	(0.0269)	(<0.0001)
9	0.8273***	0.6578***
	(0.0039)	(<0.0001)
10	0.6213**	0.6611***
	(0.0198)	(<0.0001)
11	0.4905*	0.4988***
	(0.0584)	(<0.0001)
12	0.3031	0.4655***
	(0.2694)	(<0.0001)

To determine which of the five stewardship components contribute most to the age bias, we run separately five ordered logistic regressions on the age dummy variables, each having one stewardship component as the dependent variable. The results, displayed in Panel G, indicate that regulatory issue and corporate culture are the main contributing factors for the age bias of stewardship grades. In particular, younger funds tend to do better in these two components as reflected in the significantly positive regression coefficients. This conclusion is also strongly supported by the results of t-tests on differences in means as reported in Panel H.

Consolidating the above findings, we believe our conclusion that younger funds fare better than seasoned funds in stewardship grades can be partly accounted for by the manner in which Morningstar evaluate funds for the individual stewardship component. For Regulatory Issue, which is one of the two components that we have found to be a crucial determining factor for the overall stewardship grade, we recall that funds are given penalty points ranging from -2 to 0. The final score depends on the degree to which the funds are involved in such regulatory issues as fund indictments and scandals over the past three-year history of the funds as well as the level of commitment the fund companies show in trying to salvage the situation. We conjecture that older funds could have a higher chance of being involved in regulatory woes due to certain level of managerial entrenchment, for seasoned funds are typically well-performing flagship funds for the fund family to which it belongs. Fund managers who fear being replaced for their failure to at least maintain a good track record of fund performance have the tendency to engage in activities that might lie beyond the regulatory framework. Similarly, our finding that younger funds outperform older ones under Corporate Culture could be ascribed to possible differences in investment culture adopted by managers of old and young funds. Older funds tend to have larger asset base as these funds have a longer time horizon to build their assets. When fund size grows too large and fund managers do not put in a concerted effort to curb the size, a lower score for Corporate Culture will be given.

Panel G. Ordered Logit Regressions of Stewardship Grade Components on Age Dummy Variables

This panel reports the regression estimates and summary statistics of the regressions of each of the five stewardship grade components: Board Quality (BQ), Corporate Culture (CC), Fees (Fees), Manager Incentives (MI) and Regulatory Issue (RI) on two dummy variables representing fund age: D_{age3} and D_{age5}, where D_{age3} and D_{age5} respectively takes value 1 when the fund has 3-5 and 5-10 years of history. Funds with age 10 years form the reference group. We perform a series of 12 cross-sectional ordered logit regressions using twelve consecutive monthly data over the period December 2004 to November 2005. The numbers within ( ) are the p-values. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

Month	BQ		CC		Fees		MI		RI	
	D_{age3}	D_{age5}	D_{age3}	D_{age5}	D_{age3}	D_{age5}	D_{age3}	D_{age5}	D_{age3}	D_{age5}
1	0.3744	-0.0605	1.2795***	0.3682***	-0.1178	-0.2374**	0.1379	0.2065*	1.3449***	0.7922***
	(0.2266)	(0.6592)	(<0.0001)	(0.0031)	(0.6687)	(0.0545)	(0.6173)	(0.0966)	(0.0009)	(<0.0001)
2	0.3907	0.0906	1.3428***	0.4309***	-0.1163	-0.2380**	0.1124	0.2882**	1.5055***	0.9014***
	(0.2124)	(0.5136)	(<0.0001)	(0.0006)	(0.6757)	(0.055)	(0.6868)	(0.0212)	(0.0005)	(<0.0001)
3	0.1364	0.0491	1.1325***	0.4438***	-0.3772	-0.2767**	0.1166	0.2222*	1.6132***	1.0352***
	(0.6475)	(0.7193)	(<0.0001)	(0.0003)	(0.1532)	(0.0235)	(0.6623)	(0.0708)	(0.0002)	(<0.0001)
4	0.1355	0.0693	0.8155***	0.5594***	-0.4377	-0.2221**	0.2494	0.3002**	1.0838***	1.1072***
	(0.6532)	(0.6078)	(0.0027)	(<0.0001)	(0.1002)	(0.066)	(0.3588)	(0.0142)	(0.0025)	(<0.0001)
5	0.1200	-0.0996	0.7786***	0.4754***	-0.5089*	-0.3140***	0.3774	0.3628***	1.0903***	0.9076***
	(0.6894)	(0.4517)	(0.0041)	(0.0001)	(0.0556)	(0.0085)	(0.1652)	(0.0028)	(0.0023)	(<0.0001)
6	1.0135**	0.2764	1.1166***	0.6101***	0.5273	-0.1735	0.8064**	0.5668***	1.1065***	1.0456***
	(0.0147)	(0.1062)	(0.0032)	(0.0001)	(0.1784)	(0.2533)	(0.0333)	(0.0003)	(0.0312)	(<0.0001)
7	0.0406	-0.1846	0.7058***	0.5339***	-0.3621	-0.2949**	0.6106**	0.4351***	1.0180***	0.8193***
	(0.8932)	(0.1669)	(0.0086)	(<0.0001)	(0.1694)	(0.0127)	(0.024)	(0.0003)	(0.0044)	(<0.0001)
8	0.0685	-0.1885	0.5205**	0.4237***	-0.5796**	-0.2092*	0.7482***	0.5374***	0.8572***	1.0251***
	(0.826)	(0.1529)	(0.0647)	(0.0004)	(0.0345)	(0.0778)	(0.0081)	(<0.0001)	(0.0154)	(<0.0001)
9	0.0806	-0.2025	0.8728***	0.4235***	-0.2185	-0.2991***	0.9410***	0.5846***	0.9483***	1.0367***
	(0.791)	(0.118)	(0.0015)	(0.0003)	(0.4138)	(0.0096)	(0.0006)	(<0.0001)	(0.0069)	(<0.0001)
10	0.1629	-0.2016	0.8844***	0.4433***	-0.4726*	-0.2202**	0.6957***	0.4331***	1.1505***	1.1073***
	(0.5699)	(0.1158)	(0.0007)	(0.0002)	(0.0593)	(0.0558)	(0.0072)	(0.0002)	(0.0008)	(<0.0001)
11	0.0355	-0.4291***	0.8002***	0.3230***	-0.3903	-0.3467***	0.4824**	0.4754***	0.9314***	0.8189***
	(0.8967)	(0.0002)	(0.0016)	(0.0032)	(0.1121)	(0.0012)	(0.0582)	(<0.0001)	(0.0035)	(<0.0001)
12	-0.0069	-0.4631	0.7962	0.2926***	-0.5122**	-0.2762***	1.0825***	0.4303***	0.7479**	0.8497***
	(0.9811)	(0.0001)	(0.0033)	(0.0072)	(0.0503)	(0.0097)	(0.0001)	(0.0001)	(0.0209)	(<0.0001)

Panel H Differences in Mean Scores of Stewardship Components: Corporate Culture and Regulatory Issues Across Age Groups

This panel reports, for twelve consecutive monthly data over the period December 2004 to November 2005, the mean scores of two stewardship components: Corporate Culture and Regulatory Issue, of funds in the three age groups: 3-year, 5-year and 10-year and results of t-tests on the differences in mean stewardship grades. The numbers within ( ) are the p-values. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

	Corporate Culture					Regulatory Issue				
	Average Score in Corporate Culture			t-test of differences in means		Average Score in Culture			t-test of differences in means	
Month	10-year	5-year	3-year	Between 3-year and 10-year	Between 5-year and 10-year	10-year	5-year	3-year	Between 3-year and 10-year	Between 5-year and 10-year
1	1.5889	1.3125	1.2404	6.58***	2.02**	1.7667	1.6780	1.4181	4.11***	6.02***
2	1.6023	1.3231	1.2373	6.98**	2.31**	1.7955	1.6981	1.4209	4.60***	6.47***
3	1.5521	1.3231	1.2373	5.81***	2.58**	1.8125	1.7295	1.4199	5.20***	7.56***
4	1.4574	1.3431	1.2386	4.41***	3.53***	1.7234	1.7336	1.3952	3.96***	8.41***
5	1.4468	1.3238	1.2179	4.33***	3.07***	1.7234	1.6815	1.3923	3.99***	6.90***
6	1.5417	1.3758	1.2251	4.25***	3.43***	1.7708	1.7516	1.4189	3.62***	7.15***
7	1.4271	1.3363	1.2224	4.18***	3.30***	1.7292	1.6831	1.4256	3.73***	6.20***
8	1.4205	1.3622	1.2621	2.99***	3.19***	1.7159	1.7739	1.4910	2.68**	8.50***
9	1.5000	1.3629	1.2604	4.10***	3.35***	1.7340	1.7759	1.4908	3.06***	8.81***
10	1.5000	1.3621	1.2545	4.71***	3.56***	1.7642	1.7940	1.5004	3.69***	9.78***
11	1.4727	1.3290	1.2579	4.10***	2.59***	1.7273	1.7339	1.5168	2.86***	7.58***
12	1.4688	1.3200	1.2550	3.17***	2.37**	1.6875	1.7443	1.5134	1.99**	7.90***

4.6 Descriptive Statistics of Out-of-sample Performance Measures

We display and analyse in Table 5 the estimates of two out-of-sample performance: Carhart's four-factor alpha and Sharpe ratio of each fund for each of the twelve monthly samples of mutual funds taken from the Morningstar Principia CDs.

We first perform Shapiro and Wilk's Normality Test on these estimates. Panel A only reports the results for the November 2005 sample, although we conduct the tests for all twelve samples. We reject at the 1% significance level the hypothesis of normality for both performance estimates for all twelve samples. This provides strong evidence that the distribution of the ex-post performance of mutual funds is not normal.

A closer examination of the distribution of the performance estimates reveals that the four-factor alphas has a significantly right-skewed distribution while the distribution of the Share ratio is somewhat symmetrical, as indicated by the estimates of skewness. We therefore remark that in any subsequent tests that we shall conduct using these performance measures, the median value is probably a more appropriate choice than the mean.

We also perform Student-t test to determine whether the mean of each of the performance estimates differs significantly from zero. Results from Panel A show that the mean values of both performance measures are significantly positive, providing an evidence that the sample funds deliver, on average, significantly positive excess returns over a 30-month post-evaluation period

One of our primary reasons for estimating the out-of-sample performance of the sample funds is to examine whether fund ratings have any power to forecast future performance. With this in mind, we first decompose the sample into groups based on the funds' ratings (star rating or stewardship grade) and compare the average out-of-sample performance measures of funds within each rating group. If ratings do possess predictive

ability, then we would expect the average out-of-sample performance to increase monotonically with fund ratings.

We record the statistics for the November 2005 sample in Panels B and C. Although the median is a more appropriate statistic to use than the mean in view of our earlier observation that the performance measures are positively skewed, we report results based on both mean and median. When funds are grouped according to their star ratings, as reported in Panel B, we find that the median within-group Sharpe ratio increases monotonically from 1-star group to 5-star group. The corresponding results for four-factor alpha also indicates that funds in the higher rating groups (4-star and 5-star) outperform funds in the lower rating group, though there is no evidence that 2-star funds fare better than 1-star funds on the average. However, we have to note that the latter could be partly due to the relatively small class size of 1-star funds (67) as compared to the much larger class size (279) of 2-star funds. In addition, we obtain (but do not report) the results when funds are further portioned by their age. We find that the predictive power improves drastically in the three-year and ten-year groups, with the effect manifested almost monotonically across ratings; both mean and median out-of-sample measure in a rating group exceeds the corresponding figures for funds in a lower rating group. However, no improvement has been observed in the five-year group.

Panel C contains the results for funds partitioned by their stewardship grades. The predictive ability of stewardship grades appears to be relatively weaker than that of the star ratings. For example, we find little evidence that Grade-A funds outperform Grade-B funds, and conclusions drawn with regard to funds in C, D and F categories are mixed. When we consider the Sharpe ratio, the expected trend of higher ex-post performance for higher stewardship grade is observed across the grades. However, the results based on the four-factor alpha do not appear to support this observation.

TABLE 5
Summary Statistics of Out-of-sample Performance Measures

For each fund in each of the 12 evaluation months: December 2004 to November 2005, we estimate two out-of-sample performance measures: four-factor alpha and Sharpe ratio. The four-factor alpha of Carhart (1997) is estimated via the following time-series linear regression model:

$$R_{it} - R_{ft} = \alpha + \beta_{i1} RMRF_t + \beta_{i2} SMB_t + \beta_{i3} HML_t + \beta_{i4} UMD + \varepsilon_{it}$$

for t = 1, 2, ..., 30, where t = 0 is the evaluation month.

$R_{it} - R_{ft}$ is the monthly return of fund i in excess of monthly T-Bill rate; $RMRF_t$ is the value of the market return in excess of monthly T-Bill rate; SMB_t (small minus big factor) is the difference in returns across small and big portfolios; HML_t (high minus low factor) is the difference in returns between high and low book-to-market equity portfolios; UMD_t (monthly momentum factor) is the difference in average returns on two high ex-ante return portfolio and two low ex-ante return portfolios. The second performance measure, the Sharpe ratio, of a fund i is defined by

$$\text{Sharpe Ratio}_i = \frac{\overline{R_i - R_f}}{\sigma_i}$$

where $\overline{R_i - R_f}$ is the mean return of fund i in excess of 30-day Treasury-bill rate and σ_i the standard deviation of the excess returns of fund i , over the out-of-sample 30-month post-evaluation period, using data obtained form the Centre for Research in Security Prices (CRSP) database.

Panel A. Results of Statistical Tests of Normality and of Zero Mean on the Cross-sectional Series of Performance Metrics for November 2005 Sample

This panel report results of statistical tests of normality on the two out-of-sample performance metrics: four-factor alpha and Sharpe ratio, of funds calculated over a 24-month period December 2005 to November 2007, for the November 2005 sample. The Shapiro-Wilk test statistic W lies between 0 and 1. Normality is rejected when W is sufficiently close to 1. The numbers in parenthesis ( ) are the p-values. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

Shapiro and Wilk's Normality Test and Test of Zero Mean

Four-factor Alpha			Sharpe Ratio		
Shapiro-Wilk Test H_o: Underlying distribution is Normal	t-test of zero mean	Skewness	Shapiro-Wilk Test H_o: Underlying distribution is Normal	t-test of zero mean	Skewness
0.872965*** (<0.0001)	11.83212*** (<0.0001)	1.60777211	0.988961*** (<0.0001)	24.93664*** (< 0.0001)	-0.2520578

Panel B. Comparison of Out-of-sample Performance Measures Across Various Star-rating Groups for November 2005 Sample

This panel reports the within-group averages of two out-of-sample performance metrics: four-factor alpha and Sharpe ratio, of funds calculated over a 24-month period December 2005 through November 2007 for the November 2005 sample. The groups are formed by partitioning the funds by their star ratings. We perform Kruskal-Wallis tests of equality of median. The test statistics and p-value are indicated below the table. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

	Four-factor Alpha			Sharpe Ratio		
Star-rating Group	N	Mean	Median	N	Mean	Median
1-star	67	0.272277	0.049703	67	0.152771	0.095263
2-star	279	0.080088	-0.00526	279	0.139996	0.138414
3-star	470	0.166234	0.008209	470	0.175315	0.162361
4-star	428	0.178813	0.073447	428	0.185355	0.167683
5-star	185	0.184377	0.119062	185	0.187492	0.195631

Kruskal-Wallis Test of equality of means	
Four-factor Alpha	Sharpe ratio
Chi-Square Statistic	Chi-Square Statistic
46.0393***	20.5301***
(<0.0001)	(<0.0001)

Panel C. Comparison of Out-of-sample Performance Measures Across the Five Stewardship-grade Groups for November 2005 Sample

This panel reports the within-group averages of the two out-of-sample performance metrics: four-factor alpha and Sharpe ratio of funds calculated over a 24-month period December 2005 to November 2007 for the November 2005 sample. The groups are formed by partitioning the funds by their stewardship grades. We perform Kruskal-Wallis tests of equality of median. The test statistic and p-value are indicated below the table. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

	Four-factor Alpha			Sharpe Ratio		
Stewardship Group	N	Mean	Median	N	Mean	Median
F	16	0.031169	0.028614	16	0.061002	0.06291
D	154	-0.00359	-0.04444	154	0.06845	0.091512
C	519	0.12005	-0.00339	519	0.159815	0.133969
B	664	0.228748	0.091028	664	0.202608	0.20582
A	76	0.200234	0.033499	76	0.219964	0.188409

Kruskal-Wallis Test of equality of means	
Four-factor Alpha	Sharpe ratio
Chi-Square Statistic	Chi-Square Statistic
54.5342***	70.9408***
(0.0013)	(0.0004)

In order to investigate whether there is any discernible difference in future performance across different age groups and fund categories, we compile the average out-of-sample performance of funds in each of the three age groups and three fund categories. The results displayed in Panel D show that middle-age (five-year) funds outperform funds in other age groups. Kruskall Walli's test shows that the differences in medians across the three groups are significant at the 1% level. International Stock funds emerge the winner among the three categories based on the figures reported in Panel E. Differences in group medians are also found to be significant at the 1% level.

Panel D. Comparison of Out-of-Sample Performance Measures Across Various Age Groups for June 2005 Sample

This panel reports the within-group averages of the two out-of-sample performance metrics: four-factor alpha and Sharpe ratio of funds calculated over the period December 2005 to November 2007 for the November 2005 sample. The groups are formed by partitioning the funds by their fund age. We perform Kruskal-Wallis tests of equality of median. The test statistic and p-value are indicated. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

	Four-factor Alpha			Sharpe Ratio		
Age Group	N	Mean	Median	N	Mean	Median
3-year	41	0.104198	0.053173	41	0.132667	0.151605
5-year	257	0.308343	0.159591	257	0.264438	0.24819
10-year	1131	0.12895	0.021549	1131	0.152352	0.146235

Kruskal-Wallis Test of equality of means	
Four-factor Alpha	Sharpe ratio
Chi-Square Statistic	Chi-Square Statistic
43.4401 ***	47.8280***
(<0.0001)	(<0.0001)

Panel E. Comparison of Out-of-sample Performance Measures Across Various Fund-type Categories

This panel reports the within-group averages of the two out-of-sample performance metrics: four-factor alpha and Sharpe ratio of funds calculated over the period December 2005 to November 2007 for the November 2005 sample. The groups are formed by partitioning the funds by their respective fund types: Bonds, Domestic Stock and International Stock. We perform Kruskal-Wallis tests of equality of median. The test statistic and p-value are indicated. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

	Four-factor Alpha			Sharpe Ratio		
Type	N	Mean	Median	N	Mean	Median
Bond	581	-0.03278	-0.0516	581	-0.00300	0.01763
Domestic Stock	434	-0.08386	-0.07351	434	0.169272	0.172768
International Stock	414	0.68792	0.646718	414	0.420265	0.428467

Kruskal-Wallis Test of equality of means	
Four-factor Alpha	Sharpe ratio
Chi-Square Statistic	Chi-Square Statistic
798.7697 ***	928.7181***
(<0.0001)	(<0.0001)

We proceed to examine the joint ability of the two Morningstar ratings in forecasting future returns. There are a few ways in which funds can be divided into rating-based performance groups using the two ratings. One approach would be to simply group the funds into a 25 subgroups based on rating pair (SR, SG) of the funds, where SR is the star rating and SG the stewardship grade and compute the average (mean or median) out-of-sample performance measures of funds for each group. We display the results in Panel F. For subsequent discussion, we shall refer to group (i, j) the group in which funds receive ratings i and j for star rating and stewardship grade respectively. We comment that grouping funds by their rating pairs has a serious limitation in achieving our objective of making a comparison: we do not have a proper way of ordering the groups. For example, there is difficulty with comparing group (4, 5) with group (5, 4). Nonetheless, we can draw some conclusions for cases that are less ambiguous. We observe that funds in groups with good rating pairs (for example, those with both $i \geq 4$ and $j \geq 4$) fare better, in terms of the average Sharpe ratio, than all their peers in groups with poor ratings (for example, those with $i \leq 2$ and $j \leq 2$). The results based on four-factor alpha turn out to be less assuring.

Panel F. Comparison of Out-of-sample Performance Measures Across Twenty Five (Star rating, Stewardship grade) Groups for November 2005 Sample

This panel reports the within-group averages of two out-of-sample performance metrics: four-factor alpha and Sharpe ratio of funds calculated over the period December 2005 to November 2007 for the November 2005 sample. The groups are formed by partitioning the funds by their rating pairs : (star rating , stewardship grades).

		Four-factor Alpha			Sharpe Ratio		
Star-rating	Stewardship Grade	N	Mean	Median	N	Mean	Median
1-star	F	0	NA	NA	0	NA	NA
1-star	D	21	0.157791	0.029225	21	0.108832	0.246889
1-star	C	20	-0.00414	-0.10474	20	0.026146	-0.01164
1-star	B	26	0.577374	0.721149	26	0.285663	0.458391
1-star	A	0	NA	NA	0	NA	NA
2-star	F	5	0.011244	0.024734	5	0.062064	0.07502
2-star	D	46	-0.07643	-0.08545	46	0.068603	0.051444
2-star	C	155	0.106503	0.010794	155	0.155402	0.211368
2-star	B	65	0.145719	0.007981	65	0.166288	0.177621
2-star	A	8	-0.02191	-0.01691	8	0.087112	0.133897
3-star	F	6	0.042587	0.044135	6	0.0892	0.06291
3-star	D	55	-0.07258	-0.07074	55	0.050499	0.106593
3-star	C	189	0.088376	-0.06746	189	0.137034	0.098969
3-star	B	205	0.30827	0.143076	205	0.244739	0.229823
3-star	A	15	0.131193	0.050783	15	0.200961	0.156575
4-star	F	4	0.03953	0.042578	4	0.025682	0.028008
4-star	D	24	0.129088	0.088647	24	0.098764	0.117114
4-star	C	107	0.191316	0.055066	107	0.214195	0.199866
4-star	B	262	0.179305	0.087285	262	0.182363	0.173076
4-star	A	31	0.187974	-0.01367	31	0.198741	0.125379
5-star	F	1	0.02885	0.02885	1	0.027786	0.027786
5-star	D	8	0.067993	-0.00886	8	-0.00595	0.008044
5-star	C	48	0.1814	0.099626	48	0.198239	0.139841
5-star	B	106	0.162563	0.134679	106	0.173068	0.187342
5-star	A	22	0.345365	0.228397	22	0.311138	0.249797

The average out-of-sample performance (be it four-factor alpha or Sharpe ratio) in the best rating group (5,5) is superior to all other groups except group (1, 4) (1-star and grade B). In fact, the latter exhibits the best ex-post performance. This applies to both performance measures. One possible explanation for this anomaly is that managers of funds in this group changed their investing strategy, taking higher risk to improve the subsequent short-term fund returns in order to gain a higher star rating. This is a possible manifestation of the so-called tournament phenomenon in the mutual fund literature: fund managers, in their attempt to compete for good fund ratings, tend to make drastic changes to their investment style to improve fund performance.

We remark that the preceding analysis based on rating-pair groups is problematic in that the size of rating-pair group is not uniform. Top and bottom rating-pair groups generally have much smaller group size than the other groups. In order to have a meaningful comparison of out-of-sample performance based on the two ratings, we devise a method of grouping the funds into three ordered groups 'Good', 'Fair' or 'Poor'. A fund is in the 'Good' category if both its star rating and stewardship grade are either 4 or 5. A fund is classified as 'Poor' if both its star rating and stewardship grade are 1 or 2. Funds with other ratings combinations are labeled as 'Fair'. It turns out that the effect of rating-based performance on future performance is pronounced. Panel G shows that the average out-of-sample performance, be it the four-factor alpha or the Sharpe ratio, exhibits the expected monotone increasing trend from 'Poor' to 'Good'. Results of Krusal-Wallis tests provide evidence that the differences in medians across the three groups are significant at the 1% level, while the outcome of two-sample t-test also indicates that the differences in means between any two groups are significant at the 1% significant level.

Another way to group funds based on their rating pair is to sort the funds by the sum of the raw scores for their two ratings. By this ordering, we are assuming it is not the individual rating, but the combined score for both ratings, that determines the performance of a fund. We organise funds into deciles and report the results in Panel H.

Panel G. Comparison of Out-of-sample Performance Measures Across Three Performance Groups Defined by Fund's (Star-rating, Stewardship-grade) Pairs

This panel reports the within-group averages of two out-of-sample performance metrics: four-factor alpha and Sharpe ratio of funds calculated over the period December 2005 to November 2007 for the November 2005 sample. The three performance groups formed by partitioning the funds by their rating pairs: (star rating, stewardship grades) are defined as follows: A fund is in the 'Good' category if both SR and SG are 4 or 5, in the 'Poor' category if both SR and SG are 2 or 1, and in the 'Fair' category otherwise. We perform Kruskal-Wallis tests of equality of median. The test statistic and p-value are indicated. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

	Four-factor Alpha			Sharpe Ratio		
Performance Group	N	Mean	Median	N	Mean	Median
Poor	128	0.118723	-0.00801	128	0.112028	0.061301
Fair	880	0.155145	0.017924	880	0.173001	0.162838
Good	421	0.184406	0.090605	421	0.187958	0.183214

Kruskal-Wallis Test of equality of means	
Four-factor Alpha	Sharpe ratio
Chi-Square Statistic	Chi-Square Statistic
13.5412 ***	13.9031***
(0.0011)	(0. 0.001)

Panel H. Comparison of Out-of-sample Performance Measures Across Deciles Formed by Ranking Funds Based on Sum of Raw Scores

This panel reports the within-decile averages of two out-of-sample performance metrics: four-factor alpha and Sharpe ratio of funds calculated over the period December 2005 to November 2007 for the November 2005 sample. Funds are ranked by the sum of the raw scores of star rating and stewardship grade and organized into deciles. We perform Kruskal-Wallis tests of equality of median. The test statistic and p-value are indicated. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

	Four-factor Alpha			Sharpe Ratio		
Decile	N	Mean	Median	N	Mean	Median
(Worst)1	143	0.06734	-0.04007	143	0.08954	0.07271
2	143	0.106366	-0.01281	143	0.14660	0.16188
3	143	0.096343	-0.01532	143	0.14836	0.15248
4	143	0.110674	-0.02014	143	0.14952	0.13904
5	143	0.236849	0.073535	143	0.16554	0.13143
6	143	0.234225	0.002373	143	0.20627	0.20928
7	143	0.131252	0.05678	143	0.19142	0.17669
8	143	0.210226	0.114394	143	0.19477	0.20536
9	143	0.2196	0.104262	143	0.21327	0.17345
(Best)10	142	0.19238	0.157738	142	0.21445	0.21476

Kruskal-Wallis Test of equality of means	
Chi-Square Statistic	Chi-Square Statistic
43.4401 ***	36.8411***
(<0.0001)	(<0.0001)

For robustness, we repeat the above analysis on 12 consecutive monthly data over the period December 2004 – November 2005. We document the statistics based on star ratings in Panel I. In order to make sense out of these voluminous statistics, we construct a table reporting the number of times (out of 12) $\mu_k > \mu_l$ occurs, where μ_j denotes the mean/median out-of-sample performance measures of funds in j-star group. The results are given in Panel J.

In general, top-rated funds outperform lower-rated funds, with 5-star funds enjoying higher ex-post returns than 3-star finds in all twelve evaluation months. In at least 10 out of 12 months, 4-star funds are also found to gain higher average returns than 3-star funds. In addition, 5-star funds also appear to do better than 4-star funds in most of the cases. On the flip side, the predictive ability of lower-rated funds is relatively weaker than that of high-rated ones. When we consider the four-factor alpha, only in 3 out of 12 cases do we get $\mu_2 > \mu_1$. The corresponding number for Sharpe ratio is 7. To determine

whether the differences in out-of-sample performance of 4-star funds and 5-star funds are significant, we perform a series of twelve two-sample t-tests on the difference in mean performance between the two groups. We find that in the case of four-factor alpha, only 5 out 12 of the test-statistics are significant at the 10% level. When Sharpe ratio is used instead, the difference in mean out-of-sample performance is found significant in 10 out of 12 samples.

We report the corresponding results for stewardship grades in Panels K and L. Consistent with our earlier report, stewardship grades are at best on par with the star rating in forecasting future fund returns. If we use the four-factor alpha as performance metric, then Grade A funds do not seem to fare better than Grade B funds, although Grade B funds outperform lower grade funds in at least 10 out of 12 sample. On the positive side, lower ratings exhibit strong predictive power, with $\mu_3 > \mu_2 > \mu_1$ in at least 8 out of 12 samples.

Tuning to joint predictive ability of the two Morningstar ratings, we divide funds into three categories: 'Good', 'Fair' and 'Poor', according to each fund's (SR SG) rating pair, based on the following.

Fund i is classified as

$$\begin{cases} \text{good} & \text{if } SR_i \geq 4 \text{ and } SG_i \geq 4 \\ \text{poor} & \text{if } SR_i \leq 3 \text{ and } SG_i \leq 3 \text{ and } SR_i + SG_i < 6 \\ \text{fair} & \text{otherwise} \end{cases}$$

where SR_i and SG_i are the star rating and stewardship grade of fund i, with stewardship grades quantified by: A = 5, B = 4, C = 3, D = 2 and F = 1. The choice of the above criteria is based on the assumption that investors regard funds that fail to attain more than half the maximum total score of 10 (5-star + Grade A) as poor performers and funds that attain at least the second best rating as good performers. We denote by μ_k the mean/median of out of-sample performance measure in group k where 'Good' = group 3, 'Fair' = group 2 and 'Poor' = group 1.

Panel I. Comparison of Out-of-sample Performance Measures Across Various Star-rating Groups for Twelve Consecutive Monthly Samples December 2004 to November 2005

This panel reports the within-group averages of two out-of-sample performance metrics: four-factor alpha and Sharpe ratio, of funds calculated over a 30-month post-evaluation period for twelve evaluation months December 2004 to November 2005. For each sample, groups are formed by partitioning the sample funds by star ratings.

		Summary statistics (Mean and Median)									
		Four-factor Alpha					Sharpe Ratio				
Month	Statistics	1-star	2-star	3-star	4-star	5-star	1-star	2-star	3-star	4-star	5-star
Dec04	N	31	133	299	232	103	50	212	460	364	184
	Mean	0.0492	0.1119	0.0903	0.1606	0.2496	0.1037	0.0954	0.1674	0.2001	0.2015
	Median	0.0584	0.0555	0.0136	0.1076	0.2526	0.1846	0.2020	0.2214	0.2607	0.2523
Jan 05	N	39	141	312	236	100	49	221	468	372	171
	Mean	0.1204	0.1419	0.1062	0.2245	0.2006	0.1213	0.0883	0.1462	0.1746	0.1844
	Median	-0.2101	0.0579	-0.0253	0.1500	0.2969	0.1533	0.1752	0.1780	0.2060	0.2261
Feb 05	N	44	141	330	236	97	59	227	492	375	170
	Mean	0.1246	0.0607	0.0863	0.1486	0.1885	0.0935	0.0650	0.1327	0.1493	0.1670
	Median	0.0247	0.0021	-0.0545	0.0909	0.2489	0.1374	0.1477	0.1669	0.1806	0.2329
Mar 05	N	32	150	327	237	98	46	255	488	384	166
	Mean	0.2168	0.0841	0.1144	0.2094	0.2734	0.2007	0.1296	0.2119	0.2320	0.2628
	Median	0.0801	-0.0206	-0.0410	0.0924	0.3289	0.1757	0.1817	0.2351	0.2565	0.3250
Apr 05	N	33	156	334	242	86	45	249	504	381	161
	Mean	0.1819	0.0628	0.1766	0.1710	0.3005	0.2635	0.1771	0.2594	0.2764	0.2741
	Median	-0.1663	-0.0296	0.0209	0.0461	0.3695	0.2565	0.2630	0.3112	0.3306	0.3703
May 05	N	28	178	325	244	96	38	267	483	396	174
	Mean	0.1646	0.1073	0.1927	0.1856	0.2983	0.1412	0.1241	0.1758	0.1831	0.2066
	Median	-0.1173	0.0175	0.0702	0.0716	0.3588	0.1077	0.1550	0.1868	0.1867	0.2214
Jun 05	N	29	178	325	245	94	40	267	493	421	172
	Mean	0.4075	0.0681	0.1815	0.1874	0.2498	0.1611	0.0962	0.1386	0.1544	0.1560
	Median	0.1610	-0.0288	0.0652	0.0869	0.2506	0.1365	0.1078	0.1565	0.1507	0.1735
Jul 05	N	30	185	321	249	89	38	265	501	430	171
	Mean	0.4151	0.0804	0.1349	0.1982	0.2716	0.0724	0.0500	0.0631	0.0987	0.0952
	Median	-0.0008	-0.0412	-0.0116	0.0688	0.2944	0.0361	0.0480	0.0538	0.1060	0.1216

Aug 05	N	39	176	324	267	93	47	260	503	445	185
	Mean	0.6209	0.0195	0.1286	0.1998	0.2110	0.1370	0.0024	0.0361	0.0658	0.0412
	Median	0.4156	-0.0658	0.0172	0.1120	0.2482	0.1350	0.0039	0.0330	0.0844	0.0908
Sep 05	N	40	174	325	268	112	54	258	512	443	200
	Mean	0.2811	0.0189	0.1037	0.1444	0.1555	0.0367	0.0159	0.0379	0.0644	0.0536
	Median	0.0416	-0.0645	0.0226	0.0783	0.1775	0.0577	0.0135	0.0406	0.0662	0.0656
Oct 05	N	48	179	308	279	117	68	252	465	423	184
	Mean	0.3054	-0.0098	0.1089	0.1697	0.2078	0.0992	0.0736	0.1081	0.1327	0.1357
	Median	0.0916	-0.0246	0.0351	0.0659	0.3040	0.0966	0.0819	0.1124	0.1302	0.1497
Nov 05	N	49	197	315	271	114	67	276	463	424	183
	Mean	0.3525	0.0079	0.1042	0.1531	0.2020	0.1030	0.0715	0.0989	0.1201	0.1215
	Median	0.1641	-0.0047	0.0421	0.0768	0.2623	0.0822	0.0788	0.0949	0.1136	0.1392

Panel J. Summary of Ability of Star Ratings to Forecast Future Fund Performance

This panel reports the number of times (out of 12) the mean or median out of-sample performance measure in one rating group exceeds that in a lower rating group for several cases of comparison. We denote by μ_k the mean/median of out of-sample performance measures in rating group k.

Out of-sample Performance	Average	# time (out of 12) $\mu_5 > \mu_4$	# time (out of 12) $\mu_4 > \mu_3$	# time (out of 12) $\mu_5 > \mu_3$	# time (out of 12) $\mu_3 > \mu_2$	# time (out of 12) $\mu_2 > \mu_1$
Four-factor alpha	mean	11	10	12	10	2
	median	12	12	12	8	3
Sharpe ratio	mean	8	12	12	12	0
	median	10	10	12	12	7

Panel K. Comparison of Out-of-sample Performance Measures Across Various Stewardship-Grade Groups for Twelve Consecutive Monthly Sample December 2004 to November 2005

This panel reports the within-group averages of two out-of-sample performance metrics: four-factor alpha and Sharpe ratio, of funds calculated over a 30-month post-evaluation period for twelve evaluation months December 2004 to November 2005. For each sample, groups are formed by partitioning the funds by their stewardship grades.

		Summary statistics (Mean and Median)									
		Four-factor Alpha					Sharpe Ratio				
Month		F	D	C	B	A	F	D	C	B	A
Dec04	N	24	90	283	354	47	68	138	430	572	62
	mean	-0.1352	0.1076	0.0582	0.2061	0.2237	0.0095	0.1462	0.1667	0.1797	0.2746
	median	-0.1945	0.0669	-0.0379	0.1586	0.2249	-0.0002	0.2338	0.2194	0.2379	0.3265
Jan 05	N	26	89	292	373	48	68	141	437	573	62
	mean	-0.1478	0.1568	0.0833	0.2293	0.2272	-0.0194	0.1144	0.1435	0.1701	0.2478
	median	-0.2077	0.0140	-0.0674	0.1980	0.0973	-0.0382	0.1807	0.1774	0.2054	0.2515
Feb 05	N	26	96	296	382	48	74	148	443	595	63
	Mean	-0.1404	0.0866	0.0432	0.1812	0.1909	-0.0399	0.1038	0.1164	0.1545	0.2225
	Median	-0.1504	-0.0052	-0.0776	0.1512	0.0867	-0.0850	0.1752	0.1511	0.1899	0.2320
Mar 05	N	26	78	299	393	48	74	133	462	607	63
	Mean	-0.1085	0.0821	0.0928	0.2327	0.2206	0.0405	0.1564	0.1948	0.2412	0.2891
	Median	-0.1237	-0.0229	-0.0625	0.2062	0.0682	-0.0169	0.2231	0.2130	0.2771	0.3154
Apr 05	N	26	88	303	387	47	74	143	461	599	63
	Mean	-0.0836	0.0451	0.1024	0.2539	0.2321	0.0490	0.1917	0.2412	0.2870	0.3490
	Median	-0.1296	-0.0425	-0.0424	0.2298	0.0946	-0.0460	0.2905	0.2821	0.3576	0.3705
May 05	N	27	88	316	391	49	77	144	465	609	63
	Mean	-0.0623	0.0787	0.1305	0.2617	0.2339	-0.0065	0.1179	0.1533	0.2109	0.2489
	Median	-0.0977	-0.0189	0.0335	0.1959	0.0937	-0.0300	0.1589	0.1585	0.2094	0.2350
Jun 05	N	27	88	311	396	49	74	147	472	635	65
	Mean	-0.0530	0.0652	0.1332	0.2404	0.2321	-0.0316	0.0839	0.1167	0.1791	0.2083
	Median	-0.0622	-0.0371	0.0569	0.1699	0.0699	-0.0488	0.1285	0.1179	0.1861	0.1852
Jul 05	N	12	66	333	414	49	63	125	502	650	65
	Mean	-0.1155	0.0565	0.1144	0.2313	0.1628	-0.0460	0.0203	0.0541	0.1107	0.1164
	Median	-0.0737	-0.0365	-0.0068	0.1386	0.0188	-0.0290	0.0433	0.0499	0.1020	0.1151

Aug 05	N	12	72	338	425	52	64	130	508	669	69
	Mean	-0.1132	0.0665	0.1162	0.2181	0.1328	-0.0936	-0.0078	0.0216	0.0783	0.0845
	Median	-0.0776	0.0045	0.0172	0.1279	-0.0040	-0.0645	0.0277	0.0334	0.0841	0.0781
Sep 05	N	8	76	348	434	53	31	163	527	673	73
	Mean	0.0308	0.0068	0.0676	0.1712	0.1094	-0.0252	-0.0212	0.0273	0.0736	0.0688
	Median	0.0344	-0.0559	0.0211	0.0951	0.0243	-0.0057	-0.0231	0.0295	0.0745	0.0449
Oct 05	N	8	79	340	450	54	16	150	492	659	75
	Mean	0.0225	0.0199	0.0605	0.2024	0.0878	0.0911	0.0301	0.0950	0.1437	0.1231
	Median	0.0251	-0.0443	0.0296	0.1536	-0.0335	0.1074	0.0524	0.0965	0.1474	0.0987
Nov 05	N	4	83	356	447	56	16	154	506	661	76
	Mean	0.0246	0.0275	0.0702	0.1856	0.1041	0.0240	0.0202	0.0931	0.1308	0.1127
	Median	0.0248	-0.0163	0.0496	0.1520	-0.0069	0.0263	0.0236	0.0913	0.1266	0.0932

Panel L. Summary of Ability of Stewardship Grades to Forecast Future Fund Performance.

This panel reports the number of times (out of 12) the mean and median out of-sample performance measure in one stewardship grade category exceeds that in a lower grade category for several cases of comparison. We denote by μ_k the mean/median of out of-sample performance measures in rating group k.

Out of-sample Performance	Average	# time (out of 12) $\mu_5 > \mu_4$	# time (out of 12) $\mu_4 > \mu_3$	# time (out of 12) $\mu_5 > \mu_3$	# time (out of 12) $\mu_3 > \mu_2$	# time (out of 12) $\mu_2 > \mu_1$
Four-factor alpha	mean	2	12	12	9	10
	median	1	12	9	8	9
Sharpe ratio	mean	9	12	12	12	10
	median	7	12	12	5	9

We report the summary statistics in Panel M and the table indicating the strength of predictive ability in Panel N. We find evidence to support the hypothesis that the predictive ability of the combined ratings is superior to that of a single rating. With only one exception, namely mean four-factor alpha between 'Poor' and 'Fair' funds, the event $\mu_3 > \mu_2 > \mu_1$ occurs in 11 out of 12 samples.

We also perform Kruskall Wallis tests for equality of means across the three groups. The null hypothesis is rejected at the 1% significance level, indicating strong evidence that the differences in performance across the three performance groups are significant. However, when we carry out sample t-tests for the difference in means of any pair of performance groups, we are unable to reject the hypothesis of equality of means for 'Good' and 'Fair' groups, indicating that it is impossible to distinguish between the best and second best funds. On the positive side, results are significant for the other combinations ('Good' versus 'Bad') and ('Fair' versus 'Bad'). We remark that there is a small percentage of funds having SR = 1 and SG = 5 or SR = 5 and SG = 1. Treating these funds as outliers, we remove them from the sample and repeat the above analysis. We find that the conclusions remain in tact.

In summary, our statistical analyses have provided evidence that each of the Morningstar ratings possesses some ability, albeit a weak one, to forecast funds' future returns. The stewardship grade is at best on par with the star rating in terms of predictive ability. Evidence based on our studies also show that low-ratings on the whole indicate poor future performance and funds with good star ratings tend to deliver better ex-post returns than funds with poor ratings. In addition, we find evidence to support the contention that the two Morningstar ratings, when jointly used, exhibit superior predictive performance than each individual rating. On the flip slide, we are unable to conclude that the best rated funds necessarily outperform funds with second best rating when funds are ranked according to predetermined criteria governed by the fund's joint ratings.

Panel M. Comparison of Out-of-sample Performance Measures Across Three Performance Groups Defined by Fund's (Star-rating, Stewardship-grade) Pairs For 12 consecutive Evaluation Months

This panel reports the within-group averages of the two out-of-sample performance metrics: four-factor alpha and Sharpe ratio of funds calculated over a 30-month post-evaluation period for the 12 evaluation months Dec 2004 to Nov 2005. The three performance groups formed by partitioning sample funds by their rating pairs: (star rating , stewardship grades) are defined as follows: A fund is in the 'Good' category if both SR and SG are 4 or 5, in the 'Poor' category if both SR and SG are 2 or 1, and in the 'Fair' category otherwise. We perform Kruskal-Wallis tests of equality of median. Hypothesis of equal medians is rejected at 1% significance for all 12 samples.

		Summary statistics (Mean and Median)					
		Four-factor alpha			Sharpe Ratio		
Month		Poor	Fair	Good	Poor	Fair	Good
Dec04	N	189	666	415	149	754	367
	mean	0.0506	0.0314	0.1274	0.0650	0.1721	0.1986
	median	-0.0487	-0.0411	0.0313	0.1322	0.2251	0.2664
Jan 05	N	189	689	403	148	778	355
	mean	0.0417	0.0764	0.1333	0.0387	0.1578	0.1742
	median	-0.0644	-0.0337	0.0127	0.0720	0.1963	0.2090
Feb 05	N	185	733	405	165	801	357
	mean	0.0024	0.0423	0.1229	0.0348	0.1319	0.1641
	median	-0.0655	-0.0556	0.0330	0.0234	0.1739	0.2149
Mar 05	N	205	728	406	161	812	366
	mean	0.0118	0.0787	0.1661	0.1067	0.2076	0.2530
	median	-0.0685	-0.0375	0.0442	0.0911	0.2382	0.2881
Apr 05	N	214	725	401	159	820	361
	mean	0.0002	0.0867	0.1432	0.1373	0.2555	0.2902
	median	-0.0825	-0.0456	0.0187	0.1890	0.3112	0.3639
May 05	N	223	715	420	154	826	378
	mean	0.0103	0.1064	0.1367	0.0613	0.1754	0.2051
	median	-0.0662	-0.0143	0.0471	0.0764	0.1886	0.1942
Jun 05	N	228	723	442	153	837	403
	mean	-0.0280	0.0940	0.1268	0.0445	0.1379	0.1740
	median	-0.0863	-0.0298	0.0225	0.0282	0.1618	0.1649
Jul 05	N	218	730	457	134	861	410
	mean	-0.0005	0.1375	0.1713	0.0077	0.0683	0.1134
	median	-0.0395	-0.0028	0.0802	-0.0009	0.0634	0.1190
Aug 05	N	227	737	476	136	873	431
	mean	-0.0251	0.1189	0.1356	-0.0035	0.0348	0.0747
	median	-0.0675	0.0008	0.0712	-0.0210	0.0425	0.0969
Sep 05	N	245	739	483	129	910	428
	mean	-0.0315	0.1081	0.1120	-0.0086	0.0380	0.0730
	median	-0.0718	0.0183	0.0631	-0.0220	0.0368	0.0766
Oct 05	N	241	662	489	126	836	430
	mean	-0.0191	0.1365	0.1537	0.0662	0.1069	0.1371
	median	-0.0195	0.0251	0.1210	0.0470	0.1130	0.1425
Nov 05	N	255	676	482	128	864	421
	mean	0.0035	0.1335	0.1477	0.0595	0.0993	0.1239
	median	-0.0008	0.0463	0.1251	0.0184	0.0972	0.1230

Panel N. Summary of Joint Ability of Star Ratings and Stewardship Grades To Forecast Ex-post Fund Performance

This panel reports the number of times (out of 12) the mean or median out of-sample performance measure in one rating-pair group exceeds that in an inferior rating-pair group for several comparison cases. We denote by μ_k the mean/median of out of-sample performance measure in group k where "Good" = group 3, "Fair" = group 2 and "Poor" = group 1.

Out of-sample performance	Average	# time (out of 12) $\mu_3 > \mu_2$	# time (out of 12) $\mu_2 > \mu_1$
Four-factor alpha	By mean	12	11
	By median	12	12
Sharpe Ratio	By mean	12	12
	By median	12	12

Remarks

"Outliers" , defined to be funds that receive top grade in one rating and the worst in the other (e.g. 5 star and Grade F), are very rare. We repeat the above analysis with these funds removed. The results were hardly affected.

Panel P. Comparison of Out-of-sample Performance Measures Within Deciles Ranked by Sum of Raw Scores of Star Ratings and Stewardship Grades for 12 Consecutive Evaluation Months

This panel reports the decile averages of the two out-of-sample performance metrics: four-factor alpha and Sharpe ratio of funds calculated over a 30-month post-evaluation period for the 12 evaluation months December 2004 to November 2005. Deciles are formed by ranking funds by the sum of the raw scores of star rating and stewardship grade.

		Four-factor Alpha									
Month	Decile	1	2	3	4	5	6	7	8	9	10
1	N	127	127	127	127	127	127	127	127	127	127
	mean	-0.0009	0.0249	-0.0511	0.0552	0.0727	0.1236	0.0404	0.0877	0.1500	0.1542
	median	-0.0487	-0.0541	-0.0744	-0.0420	-0.0243	0.0050	-0.0499	0.0523	0.0367	0.0392
2	N	128	128	128	128	128	128	128	128	128	128
	mean	0.0251	0.0047	-0.0192	-0.0075	0.2358	0.1150	0.1184	0.1201	0.1216	0.1790
	median	-0.0661	-0.0710	-0.0678	-0.0857	0.0102	-0.0197	-0.0157	0.0303	0.0082	0.0549
3	N	132	132	132	132	132	132	132	132	132	132
	mean	-0.0143	-0.0177	-0.0496	-0.0301	0.1927	0.1089	0.0683	0.0903	0.1186	0.1478
	median	-0.0964	-0.0653	-0.0665	-0.0792	0.0242	-0.0086	-0.0583	0.0256	0.0171	0.0687
4	N	133	133	133	133	133	133	133	133	133	133
	mean	0.0200	0.0014	-0.0209	0.0272	0.1764	0.1521	0.0942	0.1232	0.1587	0.2104
	median	-0.0793	-0.0489	-0.0666	-0.0419	-0.0222	-0.0026	-0.0433	-0.0047	0.0567	0.0886
5	N	134	134	134	134	134	134	134	134	134	134
	mean	0.0061	-0.0014	-0.0326	0.0375	0.2346	0.1267	0.1116	0.0741	0.1309	0.2103
	median	-0.1048	-0.0629	-0.0831	-0.0579	-0.0125	-0.0097	-0.0230	-0.0246	0.0343	0.0630
6	N	135	135	135	135	135	135	135	135	135	135
	mean	-0.1097	0.0696	0.0597	-0.0423	0.1760	0.1670	0.2057	0.0926	0.1203	0.2497
	median	-0.1189	-0.0532	-0.0319	-0.0551	0.0017	0.0114	0.0293	-0.0241	0.0667	0.1457
7	N	139	139	139	139	139	139	139	139	139	139
	mean	0.0351	-0.0602	0.0067	0.0437	0.2090	0.1459	0.0810	0.0799	0.0819	0.2166
	median	-0.0732	-0.1056	-0.0623	-0.0251	0.0008	0.0009	-0.0310	-0.0157	0.0256	0.1423
8	N	140	140	140	140	140	140	140	140	140	140
	mean	0.0772	-0.0385	0.0565	0.1062	0.2473	0.1389	0.1471	0.1455	0.1605	0.2297
	median	-0.0088	-0.0614	-0.0415	-0.0073	0.0783	0.0032	0.0207	0.0696	0.1228	0.2037

9	N	144	144	144	144	144	144	144	144	144	144
	mean	0.0240	-0.0221	0.1228	0.0825	0.1056	0.1373	0.1555	0.1205	0.1826	0.1088
	median	-0.0371	-0.0765	-0.0014	-0.0073	-0.0319	0.0061	0.0712	0.0897	0.1194	0.0457
10	N	146	146	146	146	146	146	146	146	146	146
	mean	-0.0228	-0.0006	0.0622	0.1096	0.0725	0.1787	0.0600	0.1308	0.1724	0.1025
	median	-0.0600	-0.0504	-0.0043	0.0248	-0.0297	0.0028	0.0619	0.1157	0.1206	0.0689
11	N	139	139	139	139	139	139	139	139	139	139
	mean	0.0195	0.0334	0.0751	0.0508	0.1403	0.1879	0.1404	0.1713	0.1718	0.1655
	median	-0.0304	-0.0042	0.0122	-0.0379	0.0872	0.0399	0.0861	0.1434	0.1171	0.2098
12	N	141	141	141	141	141	141	141	141	141	141
	mean	0.0471	0.0733	0.0305	0.0837	0.1120	0.1269	0.1610	0.1725	0.1802	0.1669
	median	-0.0006	-0.0007	-0.0020	-0.0008	0.0715	0.0421	0.1261	0.1461	0.1449	0.1948

4.7 Spearman-Rho Rank Correlation Test For Predictive Ability

To conduct a further test for joint predictive ability of rating, we adopt the approach of Blake and Morey (2000). The main idea is to examine the Spearman-rho rank correlations between the star ratings and the ranking induced by the out-of-sample performance measures. More precisely, we organise funds into various fund rating-based groups, either by the star ratings, the stewardship grades or the respective rating raw scores. We compute the average out-of-performance measures of funds within each group and organise the funds into deciles. We can then compare the decile rankings determined by the average out-of-sample performance measures with the rating-based rankings. We compute both mean and median out-of-sample performance measures. The results are presented in Table 6.

Panels A reports the results for the star ratings based on the November 2005 sample when funds are ranked by the star rating. For the entire sample, correlation is significant when the Sharpe ratio is used to measure out-of-sample performance. No evidence of significant correlation is found in the case of four-factor alpha. However, significantly high correlation emerge within the '3-year' group and the Bond category. Corresponding results for the Stewardship grades, contained in Panel B, show that for the case of Sharpe ratio, correlation is significantly positive on the whole, although for groups formed by fund age or fund category, only International Stock funds exhibit significant correlation. For four factor alphas, no significant correlation is found for the entire sample, but in the International Stock category, significantly perfect positive correlation is observed. Summing up, the conclusion that individual rating possesses rather limited predictive power based on correlation analysis is in line with that collated in the preceding sections.

TABLE 6
Spearman-rho Rank Correlation Between In-sample Fund Ratings and Out-of-sample Performance Metrics

We organise sample funds into rating-induced performance groups. For each of the methods of partitioning, we compute the average out-of-performance measures. We then rank the groups based on theses averages and evaluate the Spearman-rho rank coefficient, given by the formula

$$1-\frac{6\sum_{i=1}^{N}(R_i^{\text{in-sample}}-R_i^{\text{out-of-sample}})^2}{N^2(N-1)}$$

where $R_i^{\text{in-sample}}$ and $R_i^{\text{out-of-sample}}$ are respectively the in-sample rank of the i^{th} group based on the ratings and out-of-sample ranking based on the group averages of performance measures; N is the number of groups being ranked.

Panel A. Rank Correlation Using Star-Rating As In-sample Performance Measures for November 2005 Sample

This panel reports the Spearman-rho rank correlation coefficients between ranking by funds' star ratings received in November 2005 and ranking based on subsequent 24-month post-evaluation performance measures: four-factor alpha and Sharpe ratio, for the full sample as well as for sub-samples organised by fund age and fund type. The numbers within parentheses ( ) are the p-values for the test of significance of correlation. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

	Four-factor Alpha		Sharpe Ratio	
Star Rating	Mean	Median	Mean	Median
All	0.0 (1.00)	0.7 (0.1881)	0.9** (0.0374)	1.00*** (<0.0001)
3-year	0.9** (0.0374)	0.9** (0.0374)	0.700 (0.18881)	0.6 (0.2848)
5-year	-0.1 (0.8729)	0.0 (1.00)	-0.1 (0.8729)	-0.1 (0.8729)
10-year	0.6 (0.2848)	0.9** (0.0374	0.700 (0.18881)	0.9** (0.0374)
Bond	0.9** (0.0374	0.9** (0.0374	0.9** (0.0374	0.700 (0.18881)
Domestic Stock	0.1 (0.8729)	0.1 (0.8729)	-0.3 (0.6238)	-0.3 (0.6238)
International Stock	-0.1 (0.8729)	-0.1 (0.8729)	-0.5 (0.391)	-0.2 (0.7471)

Panel B. Spearman-rho Rank Correlation Using Stewardship Grades As In-sample Performance Measures for November 2005 Sample

This panel reports the Spearman-rho rank correlation coefficients between ranking by stewardship grades of funds in an evaluation month and ranking based on a subsequent post-evaluation performance measures: Four-factor alpha and Sharpe ratio, for the full sample as well as for sub samples partitioned by fund age and fund type. The numbers within parentheses are the p-values for the test of significance of correlation. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

	Four-factor Alpha		Sharpe Ratio	
Stewardship Grade	Mean	Median	Mean	Median
All	0.8 (0.1041)	0.6 (0.2848)	1.00*** (<0.0001)	0.9** (0.0374)
3-year	0.8 (0.1041)	0.8 (0.2)	0.4 (0.6)	0.4 (0.6)
5-year	-0.2 (0.8)	0.8 (0.2)	-0.2 (0.8)	-0.2 (0.8)
10-year	0.6 (0.2848)	0.2 (0.7471)	0.8 (0.1041)	0.8 (0.1041)
Bond	-0.10000 (0.8729)	-0.10000 (0.8729)	0.4 (0.5046)	0.7 (0.1881)
Stock	0.6 (0.4)	-0.4 (0.6)	-0.8 (0.2)	-0.4 (0.600)
International Stock	1.00*** (<0.0001)	1.00*** (<0.0001)	0.8 (0.2)	1.00*** (<0.0001)

To examine joint ability of the two ratings in predicting future performance, we consider groups formed by the sum of rating raw scores. The results in Panel C present evidence that for the November 2005 sample, there is a highly significant and positive correlation between joint rating as measured by the sum of raw scores and post-evaluation returns using either four-factor alpha or Sharpe ratio as performance measure. We also find that Bond funds and seasoned (ten-year) funds appear to possess the strongest predictive power among their peers.

Panel C. Spearman-rho Rank Correlation Using Sum of Rating Raw Scores As In-sample Performance Measures for November 2005 Sample

This panel reports the Spearman-rho rank correlation coefficients between ranking by the sum of the raw scores for ratings received in November 2005 and ranking based on a subsequent 30-month post-evaluation performance measures: Four-factor alpha and Sharpe ratio, for the full sample as well as for sub-samples partitioned by fund age and fund type. Funds are sorted by the sum and raw scores and divided into ten deciles. The decile average of out-of-sample performance measures are computed and new ranking based on these averages determined. The numbers within parentheses are the p-values for the test of significance of correlation. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

	Four-factor Alpha		Sharpe Ratio	
Decile	Mean	Median	Mean	Median
All	0.64848**	0.90303***	0.96364***	0.76970***
	(0.0425)	(0.0003)	(<.0001)	(0.0092)
3-year	0.58333*	0.53333	0.40000	0.40000
	(0.0992)	(0.1392)	(0.2861)	(0.2861)
5-year	0.05455	0.35758	0.05455	0.06667
	(0.8810)	(0.3104)	(0.8810)	(0.8548)
10-year	0.68485**	0.76970***	0.97576***	0.80606***
	(0.0289)	(0.0092)	(<.0001)	(0.0049)
Bond	0.91515***	0.84242***	0.86667***	0.74545**
	(0.0002)	(0.0022)	(0.0012)	(0.0133)
Domestic Stock	0.66061**	0.62424*	0.01818	-0.05455
	(0.0376)	(0.0537)	(0.9602)	(0.8810)
International Stock	-0.26061	-0.34545	0.23636	0.36970
	(0.4671)	(0.3282)	(0.5109)	(0.2931)
Bottom 5 Deciles	0.90**	0.60	0.4	0.4
	(0.0374)	(0.2848)	(0.6)	(0.6)
Top 5 Deciles	-0.30	0.9000**	-0.2	-0.2
	(0.6238)	(0.0374)	(0.8)	(0.8)

For robustness, we repeat the analysis for all twelve monthly samples: December 2004 to November 2005. Out-of-sample performance measures are estimated over a 30-month post-evaluation period. We report the results in Panel D. Parallel to what we have previously gathered, we detect highly significant Spearman coefficients in all twelve samples, regardless of which out-of-sample performance measure is being used, thus offering yet another justification for the conjecture that the joint predictive ability of the two Morningstar ratings exceeds that of an individual rating.

Panel D. Spearman-rho Rank Correlation Using Sum of Rating Raw Scores as In-sample Performance Measures for 12 Consecutive Evaluation Months

This panel reports the Spearman-rho rank correlation coefficient between ranking by sum of raw scores for star rating and stewardship grade in an evaluation month and ranking by a subsequent 30-month post-evaluation period performance metrics: Four-factor alpha and Sharpe ratio for full sample. The numbers in the parenthesis are the p-values. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

	Four-factor Alpha		Sharpe Ratio	
Month	Mean	Median	Mean	Median
1	0.8667***	0.7697***	0.7091**	0.8909***
	(0.00117)	(0.00922)	(0.02167)	(0.00054)
2	0.7091**	0.7939***	0.8909***	0.7939***
	(0.02167)	(0.0061)	(0.00054)	(0.0061)
3	0.6727**	0.8424***	0.8182***	0.7697***
	(0.03304)	(0.00222)	(0.00381)	(0.00922)
4	0.7939***	0.8909***	0.7939***	0.7697***
	(0.0061)	(0.00054)	(0.0061)	(0.00922)
5	0.7212**	0.8788***	0.6970**	0.6364**
	(0.01857)	(0.00081)	(0.0251)	(0.04791)
6	0.7576**	0.8909***	0.6727**	0.5758*
	(0.01114)	(0.00054)	(0.03304)	(0.08155)
7	0.7576**	0.8545***	0.8182***	0.7091**
	(0.01114)	(0.00164)	(0.00381)	(0.02167)
8	0.7697***	0.8909***	0.9515***	0.9394***
	(0.00922)	(0.00054)	(0.00002)	(0.00005)
9	0.6485**	0.8667***	0.8303***	0.9273***
	(0.04254)	(0.00117)	(0.00294)	(0.00011)
10	0.6485**	0.9030***	0.8303***	0.8909***
	(0.04254)	(0.00034)	(0.00294)	(0.00054)
11	0.8303***	0.8788***	0.8667***	0.7818***
	(0.00294)	(0.00081)	(0.00117)	(0.00755)
12	0.9273***	0.8667***	0.9394***	0.8909***
	(0.00011)	(0.00117)	(0.00005)	(0.00054)

4.8 Regression Analysis of Out-of-sample Performance Measures

As a final test of the predictive ability of fund ratings, we regress out of-sample performance on the fund ratings. First, we perform separate analysis on each of the following combinations of in-sample rating and out-of-sample performance:

(i) regression of four-factor alpha on star ratings,

(ii) regression of Sharpe ratio on star ratings,

(iii) regression of four-factor alpha on stewardship grades,

(iv) regression of Sharpe Ratio on stewardship grades.

For each of the above case, we perform a series of twelve cross-sectional regressions, each based on a monthly data from the period December 2004 to November 2005. The results are contained in Panels A to D of Table 7.

To better interpret the massive statistics generated, we construct and display in Panel E a table reporting the number of times (out of 12) $\beta_k > \beta_l$, where β_j denotes the regression coefficient of the dummy variable D_j which takes the value 1 if the fund has a j-rating and 0 otherwise. The higher the number of occurrences of these inequalities, the stronger the evidence supporting the hypothesis that fund ratings possess predictive ability. The results are presented in Panel E.

A few important inferences can be made from these statistics. First, we find the intercept term, which corresponds to the reference group (5-star or Grade A) significantly positive in all 48 cases (12 months and four combinations). This indicates that the best-rating group outperforms all other groups. Second, we see that 4-star funds do not differ significantly from 5-star funds as the coefficient β_4 is significant and negative in at most 3 out of the 12 regressions for each of the four combinations we analyse.

TABLE 7
Dummy Variable Regression To Examine Predictive Ability of Morningstar Ratings and Stewardship Grades

We perform the following dummy variable cross-sectional regression

$$S_i = \beta_0 + \sum_{j=1}^{4} \beta_j D_{ij} + \varepsilon_i$$

where S_i is the out-of-sample performance measure of funds measured over a 30-month post-evaluation period, for twelve consecutive evaluation months December 2004 to November 2005. The dummy variable D_{ij} takes the value 1 when a fund i has stewardship grade j (j is the numeric grade of stewardship where Grade A = 5, B = 4, C = 3, D = 2, F =1) and 0 otherwise. In-sample funds, comprising domestic stock, international stock and bond funds, are taken form the January 2005 to December 2005 Principia CDs. Out-of-sample data are obtained from the Centre for Research in Security Prices (CRSP) mutual fund database

Panel A Dummy Variable Regression Of Four-Factor Alpha On Star Ratings

This panel displays the results of cross-sectional dummy variable regression of out-of-sample four-factor alpha on dummy variables representing the star ratings (Grade 4 to 1) of funds. The numbers within < > are the t–statistics and those in ( ) are the p-values. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively

Dummy Variable Regression Of Four-factor Alpha on Star Rating

Month	Intercept	β_1 (1-star)	β_2 (2-star)	β_3 (3-star)	β_4 (4-star)	R^2	F-Stat
1	0.1363***	-0.1766***	-0.0919***	-0.1036***	-0.0378	0.0211	6.8098
	<6.1436>	< -3.6792>	< -3.0309>	< -3.9459>	< -1.3886>		
	(0.0000)	(0.0002)	(0.0025)	(0.0001)	(0.1652)		
2	0.1212***	-0.0858	-0.0507	-0.0707**	0.0202	0.0119	3.8560
	<4.2818>	< -1.4302>	< -1.3463>	< -2.1397>	<.5919>		
	(0.0000)	(0.1529)	(0.1784)	(0.0326)	(0.554)		
3	0.1149***	-0.0875*	-0.1003***	-0.0713**	-0.0205	0.0111	3.6999
	<4.5345>	< -1.7525>	< -2.9941>	< -2.4257>	< -.6729>		

	(0.0000)	(0.0799)	(0.0028)	(0.0154)	(0.5011)		
4	0.1796***	-0.1139*	-0.1448***	-0.1115***	-0.0436	0.0170	5.7528
	<6.2534>	< -1.8477>	< -3.9244>	< -3.3536>	< -1.2687>		
	(0.0000)	(0.0649)	(0.0001)	(0.0008)	(0.2048)		
5	0.1657***	-0.1205*	-0.1479***	-0.0722**	-0.0605*	0.0122	4.1343
	<5.5115>	< -1.873>	< -3.8348>	< -2.0917>	< -1.6866>		
	(0.0000)	(0.0613)	(0.0001)	(0.0367)	(0.0919)		
6	0.1715***	-0.1436**	-0.1252***	-0.0647**	-0.0682**	0.0101	3.4366
	<6.1098>	< -2.1655>	< -3.4713>	< -1.9754>	< -2.0242>		
	(0.0000)	(0.0305)	(0.0005)	(0.0484)	(0.0431)		
7	0.1182***	0.0878	-0.1027***	-0.0319	-0.0176	0.0115	4.0537
	<4.2367>	<1.3681>	< -2.8722>	< -.9852>	< -.5303>		
	(0.0000)	(0.1715)	(0.0041)	(0.3247)	(0.596)		
8	0.1602***	0.1194	-0.1043**	-0.0377	-0.0106	0.0094	3.3060
	<4.6868>	<1.4896>	< -2.3788>	< -.9526>	< -.2633>		
	(0.0000)	(0.1366)	(0.0175)	(0.341)	(0.7924)		
9	0.0913***	0.3774***	-0.0955**	0.0109	0.0374	0.0312	11.5663
	<2.7277>	<5.0741>	< -2.1812>	<.2777>	<.9387>		
	(0.0065)	(0.0000)	(0.0293)	(0.7813)	(0.3481)		
10	0.0928***	0.0753	-0.0708**	-0.0047	0.0154	0.0079	2.8959
	<3.4739>	<1.3004>	< -1.9914>	< -.1486>	<.478>		
	(0.0005)	(0.1937)	(0.0466)	(0.8819)	(0.6328)		
11	0.1582***	0.0623	-0.1510***	-0.0425	-0.0135	0.0209	7.3938
	<5.5623>	<1.1375>	< -4.0366>	< -1.2652>	< -.396>		
	(0.0000)	(0.2555)	(0.0001)	(0.206)	(0.6922)		
12	0.1513***	0.0970*	-0.1276***	-0.0428	-0.0070	0.0219	7.8744
	<5.6629>	<1.8786>	< -3.703>	< -1.3548>	< -.2185>		
	(0.0000)	(0.0605)	(0.0002)	(0.1757)	(0.8271)		

Panel B Dummy Variable Regression of Sharpe Ratio on Star Ratings

This panel displays the results of cross-sectional dummy variable regression of out-of-sample four-factor alpha and Sharpe ratio alpha on dummy variables representing the star ratings (Grade 4 to 1) of funds. The numbers within < > are the t–statistics and those in ( ) are the p-values. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively

Dummy Variable Regression of Four-factor Alpha on Star Rating

Month	Intercept	β_1 (1-star)	β_2 (2-star)	β_3 (3-star)	β_4 (4-star)	R^2	F-Stat
1	0.2015***	-0.0978**	-0.1062***	-0.0341	-0.0014	0.020061	6.474215
	<10.0758>	< -2.2606>	< -3.884>	< -1.4426>	< -.0589>		
	(0.0000)	(0.024)	(0.0001)	(0.1494)	(0.953)		
2	0.1844***	-0.0631	-0.0961***	-0.0382	-0.0098	0.014041	4.543011
	<8.971>	< -1.4495>	< -3.5108>	< -1.5888>	< -.3963>		
	(0.0000)	(0.1474)	(0.0005)	(0.1124)	(0.692)		
3	0.1670***	-0.0736**	-0.1020***	-0.0344	-0.0178	0.018483	6.204726
	<9.1378>	< -2.0435>	< -4.2196>	< -1.6213>	< -.8057>		
	(0.0000)	(0.0412)	(0.0000)	(0.1052)	(0.4205)		
4	0.2628***	-0.0621	-0.1332***	-0.0508	-0.0307	0.028727	9.863706
	<14.0298>	< -1.5443>	< -5.5342>	< -2.3446>	< -1.3717>		
	(0.0000)	(0.1227)	(0.0000)	(0.0192)	(0.1704)		
5	0.2741***	-0.0106	-0.0970***	-0.0146	0.0024	0.01508	5.11011
	<11.9355>	< -.2155>	< -3.2915>	< -.5545>	<.0859>		
	(0.0000)	(0.8294)	(0.001)	(0.5793)	(0.9316)		
6	0.2066***	-0.0655	-0.0825***	-0.0309	-0.0236	0.01084	3.706846
	<11.0006>	< -1.4757>	< -3.4183>	< -1.4082>	< -1.0459>		
	(0.0000)	(0.1403)	(0.0006)	(0.1593)	(0.2958)		
7	0.1560***	0.0051	-0.0598**	-0.0174	-0.0016	0.007704	2.694054
	<8.2903>	<.1176>	< -2.4772>	< -.7949>	< -.0718>		
	(0.0000)	(0.9064)	(0.0134)	(0.4268)	(0.9427)		
8	0.0952***	-0.0228	-0.0452***	-0.0321**	0.0034	0.013456	4.7739
	<7.3589>	< -.7529>	< -2.724>	< -2.1424>	<.2255>		
	(0.0000)	(0.4517)	(0.0065)	(0.0323)	(0.8216)		
9	0.0412***	0.0957***	-0.0388**	-0.0051	0.0246	0.020847	7.638205

	<2.9551>	<3.089>	< -2.126>	< -.3156>	<1.4834>		
	(0.0032)	(0.002)	(0.0337)	(0.7523)	(0.1382)		
10	0.0536***	-0.0169	-0.0377**	-0.0156	0.0108	0.011911	4.405814
	<4.8616>	< -.7081>	< -2.5689>	< -1.2035>	<.8123>		
	(0.0000)	(0.479)	(0.0103)	(0.229)	(0.4168)		
11	0.1357***	-0.0365	-0.0621***	-0.0276**	-0.0030	0.019213	6.792732
	<11.6902>	< -1.6326>	< -4.0684>	< -2.0092>	< -.2178>		
	(0.0000)	(0.1028)	(0.0000)	(0.0447)	(0.8276)		
12	0.1215***	-0.0185	-0.0500***	-0.0226*	-0.0014	0.013767	4.913468
	<10.6412>	< -.8407>	< -3.397>	< -1.6756>	< -.1009>		
	(0.0000)	(0.4007)	(0.0007)	(0.094)	(0.9197)		

Panel C Dummy Variable Regression of Four-Factor Alpha on Stewardship Grades

This panel displays the results of dummy variable cross-sectional regression of out-of-sample four-factor alpha on dummy variables representing the stewardship grade (Grade B to F) of funds. The numbers within < > are the t –statistics and those in ( ) are the p-values. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively.

Dummy Variable Regression of Four-factor Alpha on Star Rating

Month	Intercept	β_1 (1-star)	β_2 (2-star)	β_3 (3-star)	β_4 (4-star)	R^2	F-Stat
1	0.1711***	-0.2948***	-0.1190*	-0.1525***	-0.0558	0.0471	15.6315
	<4.5386>	< -5.6547>	< -2.6205>	< -3.7802>	< -1.4064>		
	(0.0000)	(0.0000)	(0.0089)	(0.0002)	(0.1598)		
2	0.1666***	-0.2904***	-0.0934*	-0.1281***	-0.0179	0.0376	12.4535
	<3.592>	< -4.529>	< -1.679>	< -2.5852>	< -.3676>		
	(0.0003)	(0.0000)	(0.0934)	(0.0098)	(0.7132)		
3	0.1441***	-0.2681***	-0.1157**	-0.1386***	-0.0185	0.0480	16.5991
	<3.5293>	< -4.8257>	< -2.3744>	< -3.1765>	< -.4308>		
	(0.0004)	(0.0000)	(0.0177)	(0.0015)	(0.6667)		
4	0.1734***	-0.2717***	-0.1448***	-0.1287***	-0.0103	0.0415	14.4556
	<3.768>	< -4.339>	< -2.5912>	< -2.6222>	< -.214>		
	(0.0002)	(0.0000)	(0.0097)	(0.0088)	(0.8306)		
5	0.1813***	-0.2890***	-0.1867***	-0.1382***	-0.0180	0.0455	15.9169
	<3.837>	< -4.4968>	< -3.2929>	< -2.7435>	< -.3625>		
	(0.0001)	(0.0000)	(0.001)	(0.0062)	(0.717)		
6	0.1761***	-0.2683***	-0.1598***	-0.1191**	-0.0071	0.0426	15.0483
	<3.8386>	< -4.3369>	< -2.9049>	< -2.4373>	< -.1466>		
	(0.0001)	(0.0000)	(0.0037)	(0.0149)	(0.8835)		
7	0.1675***	-0.2751***	-0.1728***	-0.1210**	-0.0203	0.0402	14.5167
	<3.7464>	< -4.4895>	< -3.2189>	< -2.5374>	< -.432>		
	(0.0002)	(0.0000)	(0.0013)	(0.0113)	(0.6658)		
8	0.1380**	-0.1901**	-0.1043	-0.0554	0.0577	0.0254	9.1224
	<2.5081>	< -2.4241>	< -1.5382>	< -.947>	<1.001>		
	(0.0123)	(0.0155)	(0.1242)	(0.3438)	(0.317)		

9	0.1105**	-0.2010**	-0.0942	-0.0489	0.0559	0.0226	8.2836
	<2.0065>	< -2.5315>	< -1.3832>	< -.8336>	<.9659>		
	(0.045)	(0.0115)	(0.1668)	(0.4047)	(0.3342)		
10	0.0884**	-0.1281	-0.1033*	-0.0389	0.0564	0.0246	9.2310
	<2.0163>	< -1.5961>	< -1.9586>	< -.8313>	<1.2223>		
	(0.044)	(0.1107)	(0.0504)	(0.4059)	(0.2218)		
11	0.0842*	-0.0321	-0.0806	-0.0203	0.1007**	0.0308	11.0200
	<1.8992>	< -.304>	< -1.4838>	< -.4268>	<2.1519>		
	(0.0577)	(0.7612)	(0.1381)	(0.6696)	(0.0316)		
12	0.0943**	-0.0228	-0.0797	-0.0194	0.0779*	0.0244	8.80651
	<2.2773>	< -.2294>	< -1.5744>	< -.4358>	<1.7821>		
	(0.0229)	(0.8186)	(0.1156)	(0.6631)	(0.0749)		

Panel D Dummy Variable Regression of Sharpe Ratio on Stewardship Grades

This panel displays the results of cross-sectional dummy variable regression of out-of-sample Sharpe Ratio on dummy variables representing the stewardship grade (Grade B to F) of funds. The numbers within < > are the t-statistics and those in ( ) are the p-values. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively

	Dummy Variable Regression Of Sharpe Ratio On Stewardship Grade						
Month	Intercept	β_1 (Grade F)	β_2 (Grade D)	β_3 (Grade C)	β_4 (Grade B)	R^2	F-Stat
1	0.2746***	-0.2651***	-0.1284***	-0.1079***	-0.0949***	0.026907	8.744605
	<7.9972>	< -5.584>	< -3.1068>	< -2.9371>	< -2.6253>		
	(0.0000)	(0.0000)	(0.0019)	(0.0034)	(0.0088)		
2	0.2478***	-0.2672***	-0.1335***	-0.1043***	-0.0777**	0.031778	10.46991
	<7.326>	< -5.7134>	< -3.2876>	< -2.8857>	< -2.1823>		
	(0.0000)	(0.0000)	(0.001)	(0.004)	(0.0293)		
3	0.2225***	-0.2623***	-0.1187***	-0.1061***	-0.0679**	0.042113	14.48641
	<7.4993>	< -6.4993>	< -3.3514>	< -3.3458>	< -2.1779>		
	(0.0000)	(0.0000)	(0.0008)	(0.0008)	(0.0296)		
4	0.2891***	-0.2486***	-0.1328***	-0.0943***	-0.0480	0.044936	15.6911
	<9.5913>	< -6.0614>	< -3.6281>	< -2.9354>	< -1.5146>		
	(0.0000)	(0.0000)	(0.0003)	(0.0034)	(0.1301)		
5	0.3490***	-0.3000***	-0.1573***	-0.1079***	-0.0620	0.043026	15.00561
	<9.646>	< -6.0941>	< -3.6221>	< -2.796>	< -1.6292>		
	(0.0000)	(0.0000)	(0.0003)	(0.0052)	(0.1035)		
6	0.2489***	-0.2554***	-0.1309***	-0.0956***	-0.0380	0.051548	18.38383
	<8.1408>	< -6.1962>	< -3.5719>	< -2.9339>	< -1.1825>		
	(0.0000)	(0.0000)	(0.0004)	(0.0034)	(0.2372)		
7	0.2083***	-0.2399***	-0.1244***	-0.0916***	-0.0292	0.048906	17.84291
	<6.9501>	< -5.8407>	< -3.4562>	< -2.866>	< -.9282>		
	(0.0000)	(0.0000)	(0.0006)	(0.0042)	(0.3535)		
8	0.1164***	-0.1624***	-0.0960***	-0.0623**	-0.0057	0.060432	22.51173
	<5.6817>	< -5.5633>	< -3.8028>	< -2.8603>	< -.2632>		
	(0.0000)	(0.0000)	(0.0001)	(0.0043)	(0.7924)		
9	0.0845***	-0.1781***	-0.0923***	-0.0629***	-0.0062	0.051446	19.45727
	<3.7583>	< -5.4962>	< -3.3195>	< -2.6256>	< -.264>		

	(0.0002)	(0.0000)	(0.0009)	(0.0087)	(0.7918)		
10	0.0688***	-0.0940***	-0.0900***	-0.0415**	0.0048	0.045198	17.30173
	<3.8366>	< -2.8625>	< -4.172>	< -2.1687>	<.2559>		
	(0.0001)	(0.0043)	(0.0000)	(0.0303)	(0.798)		
11	0.1231***	-0.0321	-0.0931***	-0.0281	0.0205	0.052047	19.038
	<6.8903>	< -.7522>	< -4.2514>	< -1.4645>	<1.0876>		
	(0.0000)	(0.4521)	(0.0000)	(0.1433)	(0.277)		
12	0.1127***	-0.0887**	-0.0925***	-0.0197	0.0181	0.050591	18.75689
	<6.4837>	< -2.1282>	< -4.3544>	< -1.0553>	<.9845>		
	(0.0000)	(0.0335)	(0.0000)	(0.2915)	(0.3251)		

Panel E. **Statistics Summarising Predictive Ability of Fund Ratings Based on Regression Coefficient of Dummy Variable Regression of Out-of-sample Performance on Fund Ratings**

This panel displays the number of statistically significant and negative coefficients $\beta_1, \beta_2, \beta_3$ and β_4 as well as the number of significant occurrences (out of 12) of $\beta_j < 0$, $\beta_j > \beta_i$ for certain $j > i \geq 1$

Out-of-sample Performance	**In-sample Rating**	No. of times β_1 <0 and is significant	No. of times β_2 <0 and is significant	No. of times β_3 <0 and is significant	No. of times β_4 <0 and is significant	No. of times $\beta_1 < \beta_2 < \beta_3$ <0 and is significant
Four-factor alpha	Stewardship Grade	9	8	7	0	5
Four-factor alpha	Star Rating	5	11	6	2	0
Sharpe Ratio	Stewardship Grade	11	12	10	3	10
Sharpe Ratio	Star Rating	2	12	4	0	0

The results for lower-rating coefficients, β_1 and β_2, turn out to be more affirmative, with β_2 significantly negative in all 12 regressions for two of the cases and in 8 and 11 regressions for the other cases (iii) and (i) respectively. Finally, we find predictive ability more pronounced for stewardship grade. This conclusion applies to both out-of-sample measures.

From the regression of out-of-sample performance on dummy variables representing the three performance groups, we again find strong evidence supporting the conclusion drawn so far that joint forecasting power of the two ratings is significantly strong. When the out-of-sample performance used is the Sharpe ratio, the inequality $0 > \beta_2 > \beta_1$ occur with high statistical significance for all 12 samples. For the case when the performance is measured by the four-factor alpha, we have $\beta_1 < 0$ significantly for all 12 samples and $\beta_2 < 0$ for 5 out of 12 samples. The constant term which captures the performance of the reference (best performing) group is significantly positive, regardless of the performance measure used. These results are robust to fund age and fund type. We repeat (but do not display the results of) the above analysis by controlling for fund type and fund age. The conclusions remain the same.

Panel F Dummy Variable Regression of Four-Factor Alpha on Performance Levels Determined by Rating Pairs of Funds

This panel displays the results of a cross-sectional dummy variable regression of out-of-sample performance measures (Four-factor alpha/Sharpe Ratio) on dummy variables representing the performance groups 'Good', 'Fair' or 'Poor' where, a fund is in the 'Good' category if both SR and SG are 4 or 5, in the 'Poor' category if both SR and SG are 2 or 1 and in the 'Fair' category otherwise. The regression model is

$$S_i = \beta_0 + \beta_1 D_{i1} + \beta_2 D_{i2} + \varepsilon_i$$

where D_1 takes the value 1 when the fund is under 'Poor' category and $D_2 = 1$ when the fund is classified as 'Fair'. The 'Good' category is the reference group. The numbers within < > are the t–statistics and those in ( ) are the p-values. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively

Dummy Variable Regression of Out-of-sample Performance Measures on Performance Groups Determined Jointly by Star Ratings & Stewardship Grades

	Four-factor Alpha				Sharpe Ratio			
Month	Intercept	β_1	β_2	R^2	Intercept	β_1	β_2	R^2
1	0.1274***	-0.0769***	-0.0960***	0.020556	0.2054***	-0.0803***	-0.0500***	0.010887
	<8.6303>	< -2.9118>	< -5.1023>		<15.3607>	< -3.3617>	< -2.9373>	
	(0.00000)	(0.00366)	(0.00000)		(0.00000)	(0.00080)	(0.00337)	
2	0.1333***	-0.0916***	-0.0570**	0.007491	0.1832***	-0.0994***	-0.0370**	0.013677
	<7.2219>	< -2.8047>	< -2.4511>		<13.6873>	< -4.1951>	< -2.1956>	
	(0.00000)	(0.00511)	(0.01438)		(0.00000)	(0.00003)	(0.02830)	
3	0.1229***	-0.1205***	-0.0806***	0.016793	0.1689***	-0.0900***	-0.0503***	0.015561
	<7.5161>	< -4.1256>	< -3.954>		<14.2475>	< -4.2516>	< -3.4027>	
	(0.00000)	(0.00004)	(0.00008)		(0.00000)	(0.00002)	(0.00069)	
4	0.1661***	-0.1543***	-0.0874***	0.019736	0.2552***	-0.1136***	-0.0549***	0.023163
	<9.0656>	< -4.8787>	< -3.8227>		<21.2632>	< -5.4842>	< -3.667>	
	(0.00000)	(0.00000)	(0.00014)		(0.00000)	(0.00000)	(0.00026)	
5	0.1432***	-0.1430***	-0.0565**	0.014585	0.2863***	-0.0943***	-0.0377**	0.010849
	<7.5312>	< -4.4369>	< -2.3833>		<19.6523>	< -3.8168>	< -2.0787>	
	(0.00000)	(0.00001)	(0.01730)		(0.00000)	(0.00014)	(0.03783)	
6	0.1367***	-0.12648***	-0.0303	0.012752	0.2024***	-0.0883***	-0.0327**	0.013549
	<7.581>	< -4.1287>	< -1.3329>		<16.7786>	< -4.3114>	< -2.1483>	
	(0.00000)	(0.00004)	(0.18279)		(0.00000)	(0.00002)	(0.03187)	

7	0.1268***	-0.1548***	-0.0328	0.019911	0.1696***	-0.0909***	-0.0320**	0.01459
	<7.3214>	< -5.2154>	< -1.492>		<14.504>	< -4.5358>	< -2.1564>	
	(0.00000)	(0.00000)	(0.13593)		(0.00000)	(0.00001)	(0.03122)	
8	0.1713***	-0.1719***	-0.0338	0.016012	0.1087***	-0.0807***	-0.0394***	0.025219
	<8.226>	< -4.6888>	< -1.2744>		<13.8228>	< -5.8373>	< -3.9286>	
	(0.00000)	(0.00000)	(0.20274)		(0.00000)	(0.00000)	(0.00009)	
9	0.1356***	-0.1607***	-0.0167***	0.014376	0.0689***	-0.0750***	-0.0272**	0.016449
	<6.4474>	< -4.3413>	< -.6197>		<7.9108>	< -4.8932>	< -2.4353>	
	(0.00000)	(0.00002)	(0.53553)		(0.00000)	(0.00000)	(0.01500)	
10	0.1120***	-0.1435***	-0.0039	0.019377	0.0659***	-0.0754***	-0.0183**	0.026246
	<6.5617>	< -4.8776>	< -.1774>		<9.3716>	< -6.22>	< -2.022>	
	(0.00000)	(0.00000)	(0.85924)		(0.00000)	(0.00000)	(0.04336)	
11	0.1537***	-0.1728***	-0.0172	0.025467	0.1370***	-0.0808***	-0.0219**	0.030325
	<8.8366>	< -5.7074>	< -.75>		<19.3604>	< -6.5664>	< -2.3507>	
	(0.00000)	(0.00000)	(0.45338)		(0.00000)	(0.00000)	(0.01888)	
12	0.1477***	-0.1443***	-0.0142	0.02084	0.1229***	-0.0674***	-0.0160*	0.022765
	<8.9727>	< -5.1543>	< -.6588>		<17.5542>	< -5.6615>	< -1.7482>	
	(0.00000)	(0.00000)	(0.51011)		(0.00000)	(0.00000)	(0.08065)	

Panel G Statistics Summarising Predictive Ability of Fund Ratings Based on Regression Coefficients of Dummy Variable Regression of Out-of-sample Performance on Fund Ratings

This panel displays the number of statistically significant and negative coefficients β_1 and β_2 as well as the number of significant occurrences of $0 > \beta_2 > \beta_1$, out of 12 regressions.

Out-of-sample Performance	No. of times $\beta_1 < 0$ and is significant	No. of times $\beta_2 < 0$ and is significant	No. of times $\beta_1 < \beta_2$ and is significant
Four-factor alpha	12	5	4
Sharpe Ratio	12	12	12

We proceed to run separately the regression of the out-of-sample performance measures on each of the following: the raw scores of stewardship grade, the raw scores of ratings and the product of the two raw scores. We believe results based on the raw scores are more informative than those based on the overall rating as the raw scores contain more information about the performance of funds. However, we remark that unlike the actual ratings which are more salient and easily accessible to the public, the raw scores are only available to mutual fund investors who subscribe to products such as the Morningstar Principia CD Roms. Hence, we still see the need to base our empirical work on both the raw scores and the actual ratings.

The regression coefficients in the regression on the raw scores of either rating are found to be highly significant for all twelve samples. However, when we consider the model which includes both rating raw scores and their interaction term, significance of the interaction term disappears, as displayed in Panel H. This indicates that each rating and hence the combined rating as measured by the product of raw scores, exhibits some degree of predictive power, but the combined rating does not significantly enhance predictive power. This result, though in contrast to our earlier findings based on other tests, is not surprising because the two independent variables representing the two raw scores are highly correlated. Multicollinearity makes it difficult to separate the effects of the two correlated variables and thus discounts the significance of the interaction term.

As the stewardship grade is a function of five component scores, it would be instructive to find out which of the five components possess the highest explanatory power for future performance. For this purpose, we run for all twelve monthly samples, cross-sectional regressions of out-of-sample performance measures on the five stewardship component scores. We report the results in Panels J and K. We find that Board Quality is the single most influential component for ex-post fund performance. It is surprising that Fees does not turn out to have a significant impact on future performance. One possible explanation is that Morningstar assigns scores in increment of 0.5. Funds are given scores ranging from 0 to 2.0 according to the percentile within which their

expense ratio falls. The categorical nature of the Stewardship component might attenuate its correlation with future fund returns.

Panel H. Regression of Out-of-sample Four-factor Alpha on Raw Rating Scores and Interaction Term

This panel displays the results of cross-sectional regression of out-of-sample Four-factor alphas and Sharpe ratio on raw scores of stewardship grade and star ratings, including the interaction term. The numbers within < > are the t–statistics and those in ( ) are the p-values. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively

Month	Intercept	SG	MR	SG*MR
1	-0.2847***	0.2180***	0.0469	-0.0195
	<-2.7258>	<2.6096>	<1.422>	<-.7865>
	(0.0065)	(0.0092)	(0.1553)	(0.4318)
2	-0.4421***	0.3779***	0.0943**	-0.0609**
	<-3.4885>	<3.7242>	<2.3518>	<-2.008>
	(0.0005)	(0.0002)	(0.0188)	(0.0448)
3	-0.3937***	0.3132***	0.0700**	-0.0407
	<-3.6307>	<3.6066>	<2.0148>	<-1.5551>
	(0.0003)	(0.0003)	(0.0441)	(0.1202)
4	-0.3325***	0.2687***	0.0609	-0.0273
	<-2.6189>	<2.6828>	<1.4902>	<-.8958>
	(0.0089)	(0.0074)	(0.1364)	(0.3705)
5	-0.3840***	0.3270***	0.0661	-0.0378
	<-2.9147>	<3.1401>	<1.567>	<-1.1995>
	(0.0036)	(0.0017)	(0.1173)	(0.2306)
6	-0.3864***	0.3607***	0.0739*	-0.0466
	<-2.8309>	<3.3597>	<1.7048>	<-1.4447>
	(0.0047)	(0.0008)	(0.0885)	(0.1488)
7	-0.3262***	0.2916***	0.0449	-0.0262
	<-2.5847>	<2.9405>	<1.1245>	<-.8835>
	(0.0098)	(0.0033)	(0.261)	(0.3771)
8	-0.2130	0.2257*	0.0195	-0.0049
	<-1.2472>	<1.7193>	<0.3679>	<-.1267>
	(0.2125)	(0.0858)	(0.713)	(0.8992)
9	-0.3749**	0.3637***	0.0640	-0.0483
	<-2.1804>	<2.7658>	<1.2015>	<-1.2399>
	(0.0294)	(0.0058)	(0.2297)	(0.2152)
10	-0.4271***	0.3555***	0.0900**	-0.0570*
	<-3.0765>	<3.3281>	<2.1251>	<-1.827>
	(0.0021)	(0.0009)	(0.0337)	(0.0679)
11	-0.3366**	0.2871**	0.0776*	-0.0415
	<-2.3237>	<2.581>	<1.6924>	<-1.238>
	(0.0203)	(0.01)	(0.0908)	(0.2159)
12	-0.2937**	0.2552	0.0753*	-0.0400
	<-2.166>	<2.454>	<1.7506>	<-1.2721>
	(0.0305)	(0.0142)	(0.0802)	(0.2036)

Panel I. Regression of Out-of-sample Sharpe Ratio on Raw Rating Scores and Interaction Term

This panel displays the results of cross-sectional regression of out-of-sample Sharpe Ratio and Sharpe ratio alpha on raw scores of stewardship grade and star ratings, including the interaction term. The numbers within < > are the t–statistics and those in ( ) are the p-values. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively

Month	Intercept	SG	MR	SG*MR
1	0.0029	0.0574	0.0212	0.0045
	<0.031>	<0.753>	<0.706>	<0.201>
	(0.9756)	(0.4514)	(0.4802)	(0.841)
2	-0.0516	0.0839	0.0254	0.0020
	<-.558>	<1.131>	<0.866>	<0.092>
	(0.5772)	(0.2581)	(0.3865)	(0.9267)
3	-0.0258	0.0538	0.0086	0.0132
	<-.326>	<0.851>	<0.339>	<0.691>
	(0.7441)	(0.395)	(0.7343)	(0.4899)
4	0.0811	0.0252	-0.0045	0.0251
	<0.976>	<0.384>	<-.167>	<1.261>
	(0.3292)	(0.7008)	(0.8672)	(0.2076)
5	0.0870	0.0671	-0.0027	0.0198
	<0.86>	<0.84>	<-.082>	<0.821>
	(0.3898)	(0.4012)	(0.9345)	(0.412)
6	-0.0069	0.0825	-0.0031	0.0179
	<-.073>	<1.121>	<-.104>	<0.808>
	(0.9414)	(0.2624)	(0.9168)	(0.4191)
7	0.0077	0.0631	-0.0148	0.0222
	<0.091>	<0.947>	<-.553>	<1.113>
	(0.9279)	(0.344)	(0.5805)	(0.266)
8	-0.0087	0.0247	-0.0244	0.0298
	<-.138>	<0.508>	<-1.25>	<2.07>
	(0.8899)	(0.6115)	(0.2113)	(0.0386)
9	-0.1279*	0.1115**	0.0000	0.0052
	<-1.824>	<2.079>	<0.001>	<0.329>
	(0.0684)	(0.0378)	(0.9989)	(0.7425)
10	-0.1117**	0.0838*	0.0074	0.0046
	<-1.977>	<1.927>	<0.427>	<0.359>
	(0.0482)	(0.0542)	(0.6691)	(0.7195)
11	-0.0659	0.0978**	0.0137	0.0005
	<-1.136>	<2.195>	<0.746>	<0.04>
	(0.2562)	(0.0283)	(0.4556)	(0.9684)
12	-0.1026*	0.1230***	0.0239	-0.0084
	<-1.809>	<2.829>	<1.33>	<-.636>
	(0.0706)	(0.0047)	(0.1837)	(0.525)

Panel J. Regression of Out-of-sample Four-factor Alpha on Five Stewardship Components

This panel displays the results of cross-sectional regression of out-of-sample four-factor alpha on the scores of the five stewardship components: Board Quality (BQ), Corporate Culture (CC), Fees (Fees), Manager Incentives (MI) and Regulatory Issue (RI). The numbers within < > are the t–statistics and those in ( ) are the p-values. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively

Month	Intercept	BQ	CC	Fees	MI	RI
1	-0.29802***	0.19799***	0.08334***	-0.03648***	0.039224**	-0.00684
	<-7.9415>	<6.9708>	<3.2109>	<-2.6348>	<2.4998>	<-.4259>
	(<0.0001)	(<0.0001)	(0.0014)	(0.0085)	(0.0126)	(0.6702)
2	-0.33825***	0.257496***	0.055068*	-0.05395***	0.044873**	0.016589
	<-7.4617>	<7.5136>	<1.7265>	<-3.1742>	<2.3594>	<.8409>
	(<0.0001)	(<0.0001)	(0.0845)	(0.0015)	(0.0185)	(0.4006)
3	-0.34533***	0.229691***	0.03493	-0.02549*	0.030796*	0.02811
	<-8.6381>	<7.6086>	<1.2544>	<-1.7016>	<1.8464>	<1.6202>
	(<0.0001)	(<0.0001)	(0.2099)	(0.0891)	(0.0651)	(0.1054)
4	-0.33838***	0.261772***	0.035843	-0.04306**	0.032797*	0.030503
	<-7.5461>	<7.9816>	<1.162>	<-2.5652>	<1.7443>	<1.6059>
	(<0.0001)	(<0.0001)	(0.2454)	(0.0104)	(0.0813)	(0.1085)
5	-0.3873***	0.295674***	0.02226	-0.04385**	0.033402*	0.039772**
	<-8.5151>	<8.8886>	<.7068>	<-2.5656>	<1.7621>	<2.0572>
	(<0.0001)	(<0.0001)	(0.4798)	(0.0104)	(0.0783)	(0.0399)
6	-0.35755***	0.283363***	0.036531	-0.04108**	0.030007	0.025756
	<-8.1852>	<8.9139>	<1.1891>	<-2.4903>	<1.6419>	<1.3792>
	(<0.0001)	(<0.0001)	(0.2346)	(0.0129)	(0.1008)	(0.1681)
7	-0.37981***	0.268269***	0.033147	-0.02694*	0.027006	0.034285*
	<-8.5044>	<8.7858>	<1.1025>	<-1.6613>	<1.5229>	<1.8828>
	(<0.0001)	(<0.0001)	(0.2704)	(0.0969)	(0.128)	(0.0599)
8	-0.34013***	0.293721***	-0.05322	-0.00297	0.001041	0.077154***
	<-6.0894>	<7.8232>	<-1.4204>	<-.1444>	<.0472>	<3.2101>
	(<0.0001)	(<0.0001)	(0.1557)	(0.8852)	(0.9624)	(0.0014)
9	-0.36356***	0.301403***	-0.04642	-0.0195	-0.0021	0.07949***
	<-6.3321>	<7.8013>	<-1.2136>	<-.9317>	<-.094>	<3.2338>
	(<0.0001)	(<0.0001)	(0.2251)	(0.3517)	(0.9251)	(0.0012)
10	-0.30231***	0.252863***	-0.02142	-0.00872	-0.00711	0.048272**
	<-6.4376>	<8.1026>	<-.6962>	<-.5232>	<-.3886>	<2.287>
	(<0.0001)	(<0.0001)	(0.4864)	(0.6009)	(0.6977)	(0.0223)
11	-0.25131***	0.239379***	-0.04727	0.014127	-0.0268	0.057508**
	<-4.9534>	<7.4747>	<-1.455>	<.7816>	<-1.3959>	<2.5312>
	(<0.0001)	(<0.0001)	(0.1459)	(0.4346)	(0.163)	(0.0115)
12	-0.21201***	0.231083***	-0.06999**	0.025456	-0.03723**	0.055566***
	<-4.5056>	<7.6432>	<-2.3092>	<1.5206>	<-2.0946>	<2.6475>
	(<0.0001)	(<0.0001)	(0.0211)	(0.1286)	(0.0364)	(0.0082)

Panel K. Regression of Out-of-sample Sharpe Ratio on Five Stewardship Components

This panel displays the results of cross-sectional regression of out-of-sample Sharpe ratio on the scores of the five stewardship components: Board Quality (BQ), Corporate Culture (CC), Fees (Fees), Manager Incentives (MI) and Regulatory Issue (RI). The numbers within < > are the t–statistics and those in ( ) are the p-values. The symbols * , ** and *** indicate statistical significance at the 10, 5 and 1 percentage levels respectively

Month	Intercept	BQ	CC	Fees	MI	RI
1	-0.1082***	0.2073***	-0.0091	-0.0350***	0.0031	0.0223
	<-3.159>	<7.9941>	<-.3834>	<-2.7669>	<.2162>	<1.5203>
	(0.0016)	(<0.0001)	(0.7015)	(0.0057)	(0.8289)	(0.1287)
2	-0.1535***	0.2217***	-0.0040	-0.0464***	0.0062	0.0312***
	<-4.6583>	<8.897>	<-.1715>	<-3.7518>	<.4495>	<2.1759>
	(<0.0001)	(<0.0001)	(0.8638)	(0.0002)	(0.6532)	(0.0297)
3	-0.1519***	0.1885***	-0.0014	-0.0337***	0.0030	0.0364***
	<-5.248>	<8.6242>	<-.071>	<-3.1069>	<.251>	<2.8999>
	(<0.0001)	(<0.0001)	(0.9434)	(0.0019)	(0.8019)	(0.0038)
4	-0.0941***	0.2027***	-0.0129	-0.0332***	0.0163	0.0405***
	<-3.2206>	<9.4895>	<-.6424>	<-3.0343>	<1.3272>	<3.2719>
	(0.0013)	(<0.0001)	(0.5207)	(0.0025)	(0.1847)	(0.0011)
5	-0.1014***	0.2383***	-0.0172	-0.0436***	0.0255**	0.0489***
	<-2.9149>	<9.3689>	<-.7147>	<-3.335>	<1.7561>	<3.3049>
	(0.0036)	(<0.0001)	(0.4749)	(0.0009)	(0.0793)	(0.001)
6	-0.1518***	0.2079***	0.0074	-0.0330***	0.0187	0.0309**
	<-5.2248>	<9.8323>	<.3642>	<-3.0026>	<1.5386>	<2.4838>
	(<0.0001)	(<0.0001)	(0.7158)	(0.0027)	(0.1241)	(0.0131)
7	-0.1899***	0.1999***	0.0159	-0.0233**	0.0147	0.0271**
	<-6.3522>	<9.7791>	<.7918>	<-2.1502>	<1.2378>	<2.2234>
	(<0.0001)	(<0.0001)	(0.4286)	(0.0317)	(0.216)	(0.0263)
8	-0.1723***	0.1332***	0.0066	0.0053	0.0182**	0.0171**
	<-8.3602>	<9.6126>	<.4779>	<.6979>	<2.2374>	<1.9296>
	(<0.0001)	(<0.0001)	(0.6328)	(0.4854)	(0.0254)	(0.0539)
9	-0.2190***	0.1457***	0.0102	-0.0048	0.0185**	0.0202**
	<-9.4034>	<9.2961>	<.6591>	<-.5703>	<2.0431>	<2.0268>
	(<0.0001)	(<0.0001)	(0.5099)	(0.5686)	(0.0412)	(0.0429)
10	-0.1575***	0.1148***	0.0214*	0.0049	0.0091	-0.0020
	<-8.2805>	<9.0777>	<1.7205>	<.7198>	<1.2225>	<-.2321>
	(<0.0001)	(<0.0001)	(0.0855)	(0.4718)	(0.2217)	(0.8165)
11	-0.1004***	0.1271***	0.0006	0.0096	0.0098	0.0053
	<-4.9786>	<9.9918>	<.0482>	<1.3392>	<1.2908>	<.5877>
	(<0.0001)	(<0.0001)	(0.9616)	(0.1807)	(0.197)	(0.5568)
12	-0.1044***	0.1238***	-0.0136	0.0162**	0.0086	0.0116
	<-5.3298>	<9.8329>	<-1.0766>	<2.3207>	<1.1597>	<1.3325>
	(<0.0001)	(<0.0001)	(0.2819)	(0.0204)	(0.2464)	(0.1829)

4.9 Short-term Persistence of Morningstar Ratings

In this section, we examine the short-term persistence of the Morningstar ratings based on our twelve-month data. Our first measure of rating persistence is the percentage of funds that are continuously awarded the same rating over the 12 month period December 2004 – November 2005. We report in Panel A and Panel B of Table 8 the corresponding statistics for star ratings and stewardship grades respectively.

TABLE 8
Persistence of Morningstar Ratings and Stewardship Grades

Panel A. Percentage of Funds that Are Continuously Awarded the Same Star Ratings from December 2004 - November 2005.

This panel displays the frequency and percentage frequency of funds that receive the same star rating continuously throughout the 12-month period: December 2004 to November 2005. The sample contains 1107 funds which comprise 361 domestic stock funds, 342 international stock funds and 404 bond funds. We include only funds that received both the Morningstar rating and stewardship grade (the corresponding frequency distributions of which are displayed in Panel B) and which are continuously recorded in the January 2005 to December 2005 Morningstar principia disc series. Numbers in parentheses ( ) are percentages.

	Frequency and Percentage Frequency					
All Funds	1-Star N=49 12 (24.49)	2-Star N=176 78 (44.32)	3-Star N=397 154 (38.79)	4-Star N=329 113 (34.35)	5-Star N=156 68 (43.59)	Total N=1107 425 (38.39)
Bond	1-Star N=10 3 (30.0)	2-Star N=62 26 (41.93)	3-Star N=147 64 (43.53)	4-Star N=125 49 (39.2)	5-Star N=60 29 (48.3)	Total N=404 171 (42.33)
Domestic Stock	1-Star N=31 6 (19.35)	2-Star N=69 30 (43.47)	3-Star N=135 45 (33.33	4-Star N=83 33 (39.76)	5-Star N=43 34 (79.07)	Total N=361 138 (38.22)
International Stock	1-Star N=8 3 (37.5)	2-Star N=45 22 (48.88)	3-Star N=115 45 (39.13	4-Star N=121 31 (25.62)	5-Star N=53 15 (28.30	Total N=342 116 (38.39)

Panel B. Percentage of Funds That Are Continuously Awarded the Same Stewardship Grade from December 2004 - November 2005.

This panel displays the frequency and percentage frequency of funds that receive the same stewardship grade continuously throughout the 12-month period: December 2004 to November 2005. The sample contains 1107 funds which comprise 361 domestic stock funds, 342 international stock funds and 404 bond funds. We include only funds that received both the Morningstar rating and stewardship grade and which are continuously recorded in the January 2005 to December 2005 Morningstar principia disc series. Numbers in parenthesis ( ) are percentages.

Frequency and Percentage Frequency						
All Funds	Grade F N=49	Grade D N=176	Grade C N=397	Grade B N=329	Grade A N=156	Total N=1107
	10	116	334	321	151	932
	(20.51)	(65.99)	(89.78)	(97.55)	(96.61)	(84.19)
Bond	Grade F N=35	Grade D N=52	Grade C N=130	Grade B N=173	Grade A N=14	Total N=404
	8	31	117	165	14	171
	(22.9)	(59.61)	(90.00)	(95.37)	(100.00)	(42.33)
Domestic Stock	Grade F N=4	Grade D N=45	Grade C N=145	Grade B N=141	Grade A N=26	Total N=361
	0	42	123	141	25	331
	(0.00)	(93.33)	(84.83)	(100.00)	(96.15)	(91.69)
International Stock	Grade F N=0	Grade D N=50	Grade C N=97	Grade B N=176	Grade A N=19	Total N=342
	0	24	94	172	18	308
	(37.5)	(48.00)	(96.91)	(97.73)	(94.73)	(90.05)

We find that for the star rating, less than 50% of the sample funds in any rating group are able to keep their ratings consistently over a 12-month period. In contrast, approximately 96% of funds with top two stewardship grades continuously receive the same grade for twelve months and about 90% of funds with Grade C do not experience any stewardship grade changes. Across fund categories, domestic stock funds have the highest percentage (79%) of 5-star funds keeping their top star rating, reflecting the less volatile nature of bonds. There is no discernible difference in persistence of stewardship grades across categories.

A second way to measure the 12-month volatility of fund ratings is to compute the sample standard deviation of the raw scores associated with the rating, as given by (3.7)

in Chapter Three. When the actual ratings are used, we compute the mean absolute difference between consecutive-month ratings given by (3.6) in Chapter Three. The mean absolute difference as compared to the usual standard deviation can more accurately capture changes in ratings. As an illustration, if the evolution of the ratings of fund A over three consecutive months is 3, 4, 3 while that of fund B is 3, 3, 4, both funds record the same standard deviation for rating change but the mean absolute difference of consecutive rating change for fund A is 2/3 as compared to 1/3 for B, thus correctly reflecting the higher fluctuation of rating changes experienced by fund A.

Panel C contains the statistics for both methods of measuring volatility. We observe that both the mean and median volatility of star ratings are higher than those for stewardship grades. Results of a Wilcoxon signed-rank test indicate that the differences in volatility are indeed statistically significant.

Finally, we display in each of Panels D and E a contingency table that summaries the relationship between the Morningstar ratings of funds as of December 2004 and the corresponding ratings as of November 2005. The diagonal terms dominate both tables, suggesting a high degree of persistence. The most drastic rating changes from the best to the worst or vice versa does not occur. When comparing the persistence of the two ratings, we examine the percentage of funds that experience rating changes of more than two grades. It is obvious from the contingency tables that the stewardship grades show greater degree of persistence. While no funds have stewardship upgrade or downgrade by more than two grades, four funds have star ratings changed by three to four stars. The number of funds having their ratings altered by two stars is 44, which is eight more than the corresponding figure for stewardship grade. Furthermore, winners in stewardship grade are more likely to remain winners than those in star ratings. While only 54% of five-star funds retain top rating, 97% of Grade-A funds do not suffer any rating downgrade. Results of chi-square tests for both Morningstar ratings reject the hypothesis that the two ratings are independent and thus provide another statistical evidence for persistence of ratings.

Panel C. Volatility of Star Ratings and Stewardship Grades over 12-month Period December 04 to November 05

This panel displays the descriptive statistics of the volatility measures of both the star ratings and the stewardship grades of a sample of the 1107 funds which comprise 361 domestic stock funds, 342 international stock funds and 404 bond funds. We include only funds that received both ratings continuously over the 12-month period December 2004 to November 2005, as recorded in January 2005 to December 2005 Morningstar Principia CDs. We use both the final grades (star rating: 1-star to 5-star and stewardship grade: A to F) as well as the estimated raw scores for each rating to calculate the volatility. When actual grades are used, we define volatility of a fund to be the quantity $\frac{1}{12}\sum_{i=1}^{11}\left|grade_{i+1} - grade_i\right|$; when raw scores are used, the volatility is given by $\sqrt{\frac{1}{12}\sum_{i=1}^{12}(rawscore_i - \overline{rawscore})^2}$ where $grade_i$ = grade of the fund in month i (for example, grade i = 3 if the fund receives a 3-star rating or a C grade. We perform paired-sample t-test and Wilcoxon signed rank test to test equality of mean and median volatility of both rating methods and report the test statistics and p-values in brackets [ ] and ( ) respectively.

Volatility of Ratings Using Raw Scores (N-1107)

Rating	Summary Statistics				
	mean	Median	Min	Max	Standard deviation
Star rating raw score (0 – 5)	0.18695	0.17321	0	0.92421	0.12592
Stewardship Grade raw score (0 – 10)	0.06329	0.00	0.00	0.45146	0.0099456
Difference	0.12366*** [24.36371] (<0.0001)	0.17321*** [173773.5] (<0.0001)	-	-	-

Volatility of Ratings Using Actual Rating (N=1107)

Rating	Summary Statistics				
	mean	Median	Min	Max	Standard deviation
Star rating (1– 5)	0.13353	0.09091	0	0.72727	0.14459
Stewardship Grade (1 – 5) A=5, B=4, .. , F=1	0.011333	0	0	0.18182	0.03126
Difference	0.122198*** [27.38453] (<0.0001)	0.09091*** [115165.5] (<0.0001)			

Panel D. Contingency Tables of December 2004 Star Ratings and November Star Ratings.
This panel displays the contingency table that summarises the star ratings of funds in December 2004 and November 2005. The sample contains 1107 funds which comprise 361 domestic stock funds, 342 international stock funds and 404 bond funds. We include only funds that received both ratings continuously over the 12-month period December 2004 to November 2005, as recorded in January 2005 to December 2005 Morningstar Principia CDs. Numbers in parenthesis ( ) are percentages.

Frequency Percent		Star Rating as of December 2004					
		1-Star	2-Star	3-Star	4-Star	5-Star	Total
Star Rating as of November 2005	1-star	26 (2.35)	21 (1.90)	1 (0.09)	1 (0.09)	0 0	49 (4.43)
	2-star	20 (1.81)	125 (11.29)	26 (2.35)	4 (0.36)	1 (0.09)	176 (15.90)
	3-star	10 (0.90)	70 (6.32)	240 (21.68)	67 (6.05)	10 (0.90)	397 (35.86)
	4-star	2 (0.18)	11 (0.99)	102 (9.21)	186 (16.80)	18 (2.53)	329 (29.72)
	5star	0 (0.00)	0 (0.000)	8 (0.72)	64 (5.78)	84 (7.59)	156 (14.09)
	Total	58 (5.24)	227 (20.57)	377 (34.06)	322 (29.09)	123 (11.11)	1107 100

Chi-square Statistics 1196.8470 ($p < 0.001$)

Panel E. Contingency Tables of December 2004 Stewardship Grades and November Star Ratings.
This panel displays the contingency table that summarises the stewardship grades of funds in December 2004 and November 2005. The sample contains 1107 funds, which comprise 361 domestic stock funds, 342 international stock funds and 404 bond funds. We include only funds that received both ratings continuously over the 12-month period December 2004 to November 2005, as recorded in January 2005 to December 2005 Morningstar Principia CDs. Numbers in parenthesis ( ) are percentages.

Frequency Percent		Stewardship Grades of December 2004					
		1-Star	2-Star	3-Star	4-Star	5-Star	Total
Stewardship Grade as of November 2005	1-star	8 (0.72)	31 (2.80)	0 (0.000	0 (0.00)	0 (0.00)	39 (3.52)
	2-star	0 (0.00)	97 (8.76)	46 (4.16)	4 (0.36)	0 (0.00)	147 (13.28)
	3-star	0 (0.00)	3 (0.27)	335 (30.260	34 (3.07)	0 (0.00)	372 (33.60)
	4-star	0 (0.00)	0 (0.00)	2 (0.18)	478 (43.18)	10 (0.90)	490 (4.26)
	5star	0 (0.00)	0 (0.00)	0 (0.00)	2 (0.18)	57 (5.15)	59 (5.33)
	Total	8 (0.72)	131 (11.83)	38 (34.60)	518 (46.79)	67 (6.05)	1108 (100)

Chi-square Statistics = 2699.718 ($p < 0.000$)

Chapter Five

Conclusions And Further Research

5.1 Conclusions

Using both ordinary least-squared regressions and multinomial ordered logit regressions, we seek influential determinants of the Morningstar stewardship grades. Parallel to results reported by Khorana and Nelling (1998) in their study on determinants of the Morningstar star ratings, we report that funds that receive good stewardship grades are generally those managed by experienced managers, incur low expenses and possess a large asset base.

In examining the predictive power of the Morningstar ratings, we find that neither the star rating nor the stewardship grade possesses a strong ability to forecast future fund performance. Although funds with poor ratings generally continue to exhibit poor future performance and underperform their peers with better ratings, we do not find strong enough evidence to support the contention that funds with the best rating (5-star in star rating or A in stewardship grade) will outperform funds with second best rating (4-star in star rating or B in stewardship grade). Our results on the star ratings are by and large in line with those reported by Blake and Morey (2000). However, when we consider joint ratings as measured either by the sum of the raw scores of the two ratings or by some ranking criteria based on the pair of ratings, we find that predictive performance improves dramatically. This seems to suggest that some degree of synergy is created when the two ratings are simultaneously examined. However, the lack of statistical significance in the difference in mean future returns of funds in the top and second best groups undermines the efficacy of the investment strategy of buying only funds with best ratings. Any conclusive results about predictive power can at best be drawn from a

comparison between the top (for example, top two ratings or top two deciles) and the bottom (for example, lowest rating or worst two deciles) groups.

The hypothesis that there exists short-term persistence in mutual fund ratings has been strongly supported by our statistical results. We observe that both star ratings and stewardship grades exhibit strong short-term persistence, with an overwhelmingly high percentage of funds continuously receiving the same rating over a twelve-month time horizon. We also find that the degree of persistence is much more pronounced in stewardship grades than in star ratings. Furthermore, the volatility in stewardship grades is much lower than that in the star ratings. For mutual fund investors who seek funds with good corporate governance but have a short investment horizon, our results suggest that they should consider selling funds with poor stewardship grades as the likelihood of a poor rated fund "turning around" is small.

In summary, our results support the view that a blend of quantitative and qualitative information of mutual funds contained in the two Morningstar fund ratings can jointly provide reasonably good insights into the future performance of funds. However, neither of the two ratings is capable of forecasting future returns with a high accuracy. In addition, our results, as well as those reported by many authors, indicate that funds with poor ratings are likely to continue to perform badly in the future, thus implying that investors holding poor-rated funds should consider selling them. As for good-rated funds, we do not have sufficiently strong statistical evidence to justify the investment strategy of buying and holding only the best rated funds. Specifically, our analysis suggests that there is no significant difference in the future performance of funds in the top two rating groups. The prospect of a portfolio of best-rated funds delivering superior returns to investors is thus questionable.

5.2 Suggestions For Further Research

The current work can be expanded in several dimensions. To begin with, we can obtain results that are robust over different periods of assessment by using a Morningstar data set that spans over a longer period of time. Since the stewardship grades were first released in 2004, we can assemble a thirty-six month (August 2004 – August 2007) in-sample data on Morningstar ratings. We can then compare in-sample ratings with out-of-sample performance on time intervals of various durations. For instance, if we consider intervals of one-year, we will have 25 overlapping periods of 24 months, with 12-month in-sample ratings and 12-month out-of-sample performance, the first being August 2004 to July 2005 and the last August 2006 to July 2007. We can assess predictive power of ratings by using standard methodology. Similarly, a longer period of assessment will allow us to conduct a more robust analysis of persistence and other characteristics of ratings. Furthermore, we can perform a comprehensive examination of the effect that the initiation of top and bottom stewardship grades might have on future performance. We can also examine how important fund characteristics such as fund expenses, portfolio turnover and fund size change with ratings downgrades or upgrades.

Next, we can embark on a study of the relation between mutual fund governance (as measured by the stewardship grades) and the corporate governance of the firms in the fund's portfolio holdings. Such a study can be designed to test the hypothesis that fund managements that value corporate governance are more likely to focus on, and hence include in their portfolio, shares of companies with good corporate governance. In addition, a cross examination of stewardship grades and corporate governance of fund holdings can shed some light on how stewardship grades can predict future performance, since a positive relation between the quality of corporate governance and fund performance has long been affirmed by many researchers.

Finally, we can investigate whether geography makes any difference in the corporate governance of mutual funds as measured by the fund stewardship grades. To widen the scope of this study, we can interpret 'geography' in the two different ways: the

location of the fund management and the location of the fund assets. The results of such an academic exercise will not only complement existing works on the relation between geography and mutual fund performance, but also provide mutual fund investors a criterion, which is based on geographical considerations, for selecting mutual funds.

Bibliography

Adkisson, J. A., and Fraser, Don R., 2003. Reading the stars: age bias in Morningstar ratings. Financial Analysts Journal 59, 5, 24 – 27.

Blake, C., and Morey, M., 2000. Morningstar ratings and mutual fund performance. Journal of Financial and Quantitative Analysis 35, 451-483.

Blume, Marshall, 1998. An anatomy of Morningstar ratings. Financial Analysts Journal 54, 1-9.

Bollen, Nicholas P. B., and Busse Jeffrey A., 2001. On the timing ability of mutual fund managers. Journal of Finance 56, 3, 1075-1094.

Brown K. C., Harlow W. V., and Starks L. T., 1996. Of tournaments and temptations: an analysis of managerial incentives in the mutual fund industry. Journal of Finance 51, 1, 85-110.

Carhart, M., 1997. On persistence in mutual fund performance. Journal of Finance 51, 1, 57-82.

Chen, Q., Goldstein, I., and Jiang, W., 2008. Directors' ownership in the U.S. mutual fund industry. Journal of Finance 63, 6, 2629 - 2702.

Chen J., Hong H., Huang M., and Kubik J. D., 2004. Does fund size erode mutual fund performance? Role of liquidity and organization. The American Economic Review 94, 5, 1276 – 1302.

Cremers M., Driessen, J., Maenhout, P., and Weinbaum D., 2005. Does skin in the game matter? Director incentives and governance in the mutual fund industry. Working Paper. Yale University.

Daniel, K., Grinbatt M., Titman, S. and Wermers, R., 1997. Measuring mutual fund performance with characteristic-based benchmarks. Journal of Finance 52, 3, 1035-1058.

Ding, B., and Wermers, R., 2005. Mutual fund performance and governance structure: the role of portfolio manager and boards of directors. Working Paper. University of Maryland.

Fama, E., and French, K. R., 1993. The common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 1, 3-56.

Ferris S. P. and Yan. X., 2007. Do independent directors and chairmen matter? The role of boards of directors in mutual fund governance. Journal of Corporate Finance 13, 392-420.

Gottezmann W. N. and Peles N., 1997. Cognitive dissonance and mutual fund investors. Journal of Financial Research, 2, 145-158.

Guercio, D., Dann. L., and Partch, M. Y., 2003. Governance and boards of directors in close end investment companies. Journal of Financial Economics 69, 111-152

Guercio D., and Tkac P. A., 2008. Star power: The effect of Morningstar ratings on mutual fund flows. Journal of Financial and Quantitative Analysis 43, 3, 907 - 936.

Investment Company Institute, 2008. Mutual Fund Fact Book. Washington DC.

Jegadeesh N., and Titman S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 48, 1, 65-91.

Khorana A., 1996. Top management turnover. An empirical investigation of mutual fund managers. Journal of Financial Economics 40, 403-427.

Khorana, A, and Nelling E., 1998. The determinants and predictive ability of mutual fund ratings. The Journal of Investing 7, 61-66

Khorana, A., Servaes, H., and Wedge, L., 2007. Portfolio manager ownership and fund performance. Journal of Financial Economics 85, 179 – 204.

Khorana, A., Tufano P., and Wedge L., 2007. Board structure, merger and shareholder wealth.: a study of mutual fund industry. Journal of Financial Economics 85, 571-598.

Morey R. M., 2005. The kiss of Death: A 5-star Morningstar fund rating? The Journal of Investment Management 3, 2, 41-52.

Morey, R. M., 2002. Mutual fund age and Morningstar ratings. Financial Analysts Journal 58, 2, 56 – 63.

Morningstar 2007. Fact sheet: The Morningstar stewardship grade for funds.

Qian M., 2007. Whom Can You Trust? A study of mutual fund governance. Working Paper. National University of Singapore.

Sharpe, William., 1998. Morningstar's risk-adjusted ratings. Financial Analysts Journal 54, 21-32.

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

Tufano, P., and Sevick, M., 1997. Board structure and fee-setting in the U.S. mutual fund industry. Journal of Financial Economics 46, 321-255.

Warshawsky M., DiCarlantonio M., and Mullan L., 2000. The persistence of Morningstar ratings. Journal of Financial Planning 13, 9, 110-128.

Wellman, J., and Zhou, J., 2007. Corporate governance and mutual fund performance: a first look at the Morningstar stewardship grades. Working Paper. Binghamton University.

Appendix A Morningstar Categories

The 64 comparison groups into which funds are classified by Morningstar in determining the star rating of funds.

Large Growth
Mid-cap Growth
Small Growth
Large Blend
Mid-cap Blend
Small Blend
Large Value
Mid-cap Value
Small Value
Specialty Natural Resources
Specialty Technology
Specialty Utilities
Specialty Health
Specialty Financial
Specialty Real Estate
Specialty Communications
Specialty Precious Metals
Bear-Market
Conservative Allocation
Moderate Allocation
Convertibles

World Stock
Europe Stock
Diversified Pacific/Asia
Pacific/Asia ex-Japan Stk
Japan Stock
Diversified Emerging Mkts
Latin America Stock
World Allocation
Foreign Large Value
Foreign Large Blend
Foreign Large Growth
Foreign Small/Mid Value
Foreign Small/Mid Growth
Long Government
Intermediate Government
Short Government
Long-term Bond
Short-term Bond
Ultrashort Bond
Bank Loan
High Yield Bond
Multisector Bond
World Bond

Emerging Markets Bond
Stable Value
Municipal Bond
High-Yield Muni
Muni National Long
Muni National Interm
Muni National Short
Muni Single State Long
Muni Single State Int/Sh
Muni New York Long
Muni New York Int/Sh
Muni California Long
Muni California Int/Sh
Muni Florida
Muni Pennsylvania
Muni Massachusetts
Muni New Jersey
Muni Ohio
Muni Minnesota
Muni Single State Short

Appendix B: Mutual Fund Share Classes

Many mutual funds are made available in several share classes, typically Class A, Class B, and Class C. The differences among these share classes typically lie in the fee structure. Such a multi-class structure offers investors an opportunity to select the share class that best fits their investment goals.

Here is a brief description of Class A, B and C. Class A shares are ideal for investors who are considering holding large number of units over a long period of time. Investors of Class A shares are expected to pay a sales charge or a front-end load, typically 200 to 500 basis points, which is deducted upfront. For example, a 3 percentage sales charge reduces a $10000 investment to $97000.

Class B shares appeal to investors who wish to invest a smaller amount of money over the long term. Unlike Class A chares, there is no up-front load, which means all the money invested will be converted into fund shares or units. However, investing in Class B shares incurs a back-end load, also known as a contingent deferred sales charge (CDSC) or simply, redemption fees, payable upon selling the fund units. Back-end load usually decreases over time, although this varies from fund to fund and no charges apply if the fund is held till the end of the time period.

As for Class C shares, there is normally no front-end load. CDSC is typically higher for Class C than for Class B shares. This charge is reduced to zero if investors hold the shares beyond the CDSC period. For Class C shares, this period is typically 12 months. The catch, however, is that fund expenses such as management fees, is higher for Class C shares than for the other two share classes. As such, Class C shares are suitable for investors who intend to invest and hold the fund for a shorter period of time, say less than 5 years.

Appendix C List of Abbreviations

Variable	Definition
AR04	Annual return 2004
BFAlpha	Best-fit alpha
BFBeta	Best-fit beta
BFRsq	Best-fit R-squared
BQ*	Board quality
CC*	Corporate culture
ER	Expense ratio
Fees*	Fees
LAR12m	12-month Load-adjusted returns
LAR3yr	3-year Load-adjusted returns
MI*	Manager incentives
MT	Manager tenure (Average)
MR*	Morningstar star rating
Log asset	Log of total net asset
RI*	Regulatory issue
RTS1m	1-mth returns after tax and sales
RTS1yr	1-year Returns after tax and sales
RTS3yr	3-year returns after tax and sales
SR	Sharpe ratio
SD3	Standard deviation of fund's return
SG*	Stewardship grades
TNH	Total number of holdings
TR	Turnover ratio

*These are categorical variables that have been assigned the appropriate numeric values. For example, 'Excellent' = 2.0 and 'Fair' = 1.0 for the five Stewardship components, while 'A' = 5 and 'C' = 3 for overall Stewardship grade.