research dialogue

64

▞ TIAA-CREF **INSTITUTE**

The Morningstar Mutual Fund Star Ratings: What Investors Should Know

In this *Research Dialogue*, Professor Matthew Morey of Pace University presents a detailed analysis of the Morningstar star rating system for mutual funds. Professor Morey reviews how these ratings are constructed and illustrates several important implications of the way the ratings are generated. His discussion highlights aspects of these calculations that may be of great interest to many investors. He also examines an issue of extreme importance for individual investors: Can the Morningstar star ratings help predict future performance of mutual funds? Professor Morey's research illustrates that when it comes to choosing mutual funds (or indeed any investment), there are few easy answers. While it would be nice to have a simple, universal system to identify lucrative investments, there is just no substitute for thorough research, thoughtful consideration, and an individual's own unique evaluation of their personal goals and circumstances. While summary rating systems can provide some limited information to help with decision making, these data are only part of what is required to make an informed choice.

With some 11,000 mutual funds now available, many investors are interested in methods to identify the best funds in which to invest. Indeed, numerous magazines and newspapers, like *Fortune, Kiplingers, Money, USA Today,* and *The Wall Street Journal,* all now provide comprehensive fund ratings/rankings in order to help investors navigate their way through the numerous funds that are advertised each day in leading publications.

Despite the entry of these popular publications into the mutual fund rating business, the most well-known mutual fund rating system is currently provided by Morningstar Inc. Started in the mid-1980s, the company has grown largely as a result of the success of its now famous 5-star rating system. Similar to the ratings given to hotels, movies, or restaurants, Morningstar rates mutual funds on a scale of 1 to 5 stars, where 1 star is the worst rating and 5 stars is the best. Because of the rating system's similarity to the way we rate so many other products, and its high regard within the industry, the star rating system has become part of the accepted lexicon in mutual funds.

In fact, the rating system has become so popular that many people believe investment flows in and out of mutual funds are closely related to the Morningstar star rating. For example, a recent study in *The Wall Street Journal* reported that 97 percent of the money flowing into no-load equity mutual funds between January and August 1995 was invested into funds that were rated as 5 or 4 stars; funds with less than 3 stars actually suffered a net outflow during the same period.[1] Moreover, the heavy use of Morningstar ratings in mutual fund advertising suggests that mutual fund

companies believe that investors care about Morningstar ratings. Indeed, in some cases, the only mention of fund performance in the mutual fund advertisement is the Morningstar star rating.

In spite of its popularity, Morningstar makes no claim that its star rating system can predict future mutual fund performance. Rather, they regard the star ratings as "achievement scores" that investors should use to narrow down their search for the best mutual fund. While such advice is obviously sound, the simple fact is that many people do use the star ratings as signals of future performance. One only has to open up a newspaper to see many mutual funds advertising their Morningstar ratings in the belief that this will attract new investors.

The purpose of this article is threefold. First, we will introduce investors to the methods through which Morningstar calculates its star ratings and illustrate some of the limitations of these methods. We will show that while the Morningstar rating system has many attractive features, such as the ability to incorporate risk and different types of funds all within the simple framework of a star rating, the Morningstar system, like all rating systems, is not without its limitations. Second, we will examine some of the empirical evidence as to whether the Morningstar star ratings can actually predict future domestic equity fund performance. Third, we will examine how well the Morningstar summary star rating method predicts future performance as compared to alternative rating methods.

›››I. HOW THE MORNINGSTAR RATINGS ARE CALCULATED

This section explains the process by which Morningstar calculates its summary star ratings and, along the way, points out certain limitations that arise from the calculation method. Note that while Morningstar has other rating devices, this section only describes Morningstar's baseline star rating system, the most popular and well known of the Morningstar rating devices.[2] Also note that much more detailed information of the rating system can be found in Blume (1998) and Blake and Morey (2000).

Ia. Morningstar Return

To begin with, in order to calculate its summary star ratings, Morningstar determines the past 3-year, 5-year, and 10-year load-adjusted and risk-adjusted returns for each mutual fund. To calculate these returns, they first calculate what they call a Morningstar Return. To do this, they take the load-adjusted[3] excess return divided by the higher of two variables: the excess average return of the fund's broad asset class (more on this later) or the average 90-day U.S. T-bill rate:

(1)

Load-Adjusted Return on the Fund – T-bill Return

Higher of (Average Broad Asset Class Return – T-bill Return) or (T-bill Return)

Morningstar divides through by one of these two variables to prevent distortions caused by having low or negative average excess returns in the denominator of equation (1).[4]

Hence, as can be seen in equation (1), the Morningstar Return is essentially a load-adjusted relative return. Relative, in that, the return of the fund is compared to the other funds in its peer group (unless the broad asset class return is lower than the T-bill rate).

Ib. Broad Asset Classes

The term "broad asset class" deserves some clarification. To calculate its ratings, Morningstar classifies funds into one of four broad asset classes: domestic equity, international equity, municipal bond, and taxable bond. The determination of a fund's classification is conducted by Morningstar itself.

As can be seen in equation (1), the choice of the class by Morningstar can greatly affect the Morningstar Return, and as a result can greatly affect the overall rating. To illustrate, consider that before November 1996, Morningstar used four broad asset classes: equity, hybrid, municipal bond, and taxable bond. In this system, international equity funds were generally placed in the equity class. In November 1996, Morningstar developed a broad asset class breakdown in which the international equity funds had their own class. When this international equity class was developed, many international funds suddenly had considerably different star ratings because they were now compared to other international equity funds rather than all equity funds.

For example, one international equity fund was rated as a 3-star fund in October 1996; yet in January 1997, after the broad asset class reorganization, the same fund received a 5-star rating. Another international equity fund was rated as a 2-star fund in October 1996 and yet in January 1997 received a 4-star rating. Hence, the star rating is somewhat dependent upon the broad asset class used.

Ic. Morningstar Risk

After calculating the Morningstar Return, Morningstar then calculates a Morningstar Risk measure. This measure is calculated differently from traditional risk measures, such as standard deviation, which see greater-than and less-than expected returns as added volatility. Their rationale for this definition is that the greatest fear of most investors is losing money, which they define as underperforming the risk-free rate of return an investor can earn from the 90-day Treasury bill. Hence, their risk measure only focuses on downside risk. To calculate the Morningstar Risk, they plot the monthly returns in relation to T-bill returns. They add up the amounts by which the fund trails the T-bill return each month and then divide that total by the time horizon's total number of months. This number, the average monthly underperformance statistic, is then compared with those of other funds in the same broad asset class (i.e., domestic equity, international equity, municipal bond, and taxable bond) to assign the risk scores. The resultant Morningstar Risk score expresses how risky the fund is relative to the average fund in its category.[5]

To illustrate the calculation, Table 1 presents a hypothetical Morningstar Risk calculation for the 3-year time horizon. Table 1 indicates several interesting points about the Morningstar Risk measure. First, as with the Morningstar Return calculation, the Morningstar Risk is somewhat sensitive to the classification of the fund, as the average monthly underperformance is divided by the average monthly

Table 1: Understanding Morningstar Risk

MONTH	FUND RETURN (%)	T-BILL RETURN (%)	UNDERPERFORMANCE
1	2.0	0.5	—
2	–1.5	0.5	2.0
3	3.2	0.5	—
4	1.2	0.4	—
5	–4.0	0.6	4.6
6	2.1	0.5	—
7	0.7	0.5	—
8	2.3	0.5	—
9	–1.7	0.5	2.2
10	2.4	0.4	—
11	1.2	0.6	—
12	–3.9	0.5	4.4
13	2.0	0.5	—
14	–1.5	0.5	2.0
15	3.2	0.5	—
16	1.2	0.4	—
17	–4.0	0.6	4.6
18	2.1	0.5	—
19	0.7	0.5	—
20	2.3	0.5	—
21	–1.7	0.5	2.2
22	2.4	0.4	—
23	1.2	0.6	—
24	–3.9	0.5	4.4
25	2.0	0.5	—
26	–1.5	0.5	2.0
27	3.2	0.5	—
28	1.2	0.4	—
29	–4.0	0.6	4.6
30	2.1	0.5	—
31	0.7	0.5	—
32	2.3	0.5	—
33	–1.7	0.5	2.2
34	2.4	0.4	—
35	1.2	0.6	—
36	–3.9	0.5	4.4
TOTAL UNDERPERFORMANCE			**39.6**

$$\frac{\text{Total Underperformance}}{\text{Total Number of Months}} = \frac{39.6}{36} = 1.10, \text{ the Average Monthly Underperformance}$$

$$\frac{\text{Average Monthly Underperformance}}{\text{Average Monthly Underperformance}} = \frac{\text{3-Year Morningstar Risk}}{\text{of Investment Category}}$$

underperformance of the broad asset class. Hence, if a fund had been classified differently from what it was in an earlier time period, its Morningstar Risk could change. Second, the Morningstar Risk measure obviously penalizes fund managers for taking risks that could result in a fund underperforming the T-bill rate. Indeed, if a fund manager were concerned about the star rating for their fund falling, he or she might want to take investment positions to prevent the fund from suffering large losses (for example, hedged equity).

Id. Risk- and Load-Adjusted Return

Finally, to calculate the risk- and load-adjusted return, Morningstar then subtracts the Morningstar Risk number from the Morningstar Return number. Hence, for the 3-year time horizon period, they subtract the 3-year Morningstar Risk from the 3-year Morningstar Return, and the resultant number gives them the 3-year risk- and load-adjusted return. Assuming a fund has enough historical data, Morningstar calculates the 5-year and 10-year versions as well.

Ie. Morningstar Score, Age of the Fund, and Star Ratings

With the risk- and load-adjusted return measures in hand, Morningstar then calculates separate star ratings for each return period (3 years, 5 years, and 10 years) for each broad asset class. These period-specific star ratings are generated for only those funds within each class that have return history of at least 3 years, 5 years and 10 years, respectively. These ratings are assigned by a simple ranking of the risk- and load-adjusted returns. The funds having the highest 10 percent of risk- and load-adjusted returns within each class and time period are assigned 5 stars, the next 22.5 percent receive 4 stars, the next 35 percent receive 3 starts, the next 22.5 percent receive 2 stars, and the bottom 10 percent get 1 star. Once these period- and class-specific stars are determined, Morningstar then determines an overall star rating for each fund using an average of the period-specific star ratings for each fund. The way this average is calculated varies in an important way by the age of the fund being rated:

1. For **old** funds (funds with 10 or more years of historical data), the 3-year star rating (a number from 1 to 5) receives a 20-percent weighting, the

5-year a 30-percent weighting, and the 10-year a 50-percent weighting in the average. The resultant number, rounded to the nearest integer, is the number of overall stars for the fund.

2. For **middle-aged** funds (funds with at least 5 years, but less than 10 years of historical data), only the 3-year and 5-year star ratings are used in the average. The 3-year rating receives a 40-percent weighting and the 5-year rating receives a 60-percent weighting in the average. The average is again rounded to the nearest integer to determine the overall rating.

3. For **young** funds (funds with at least 3 years but less than 5 years of historical data), only the 3-year star rating is used. Hence, for these funds, the overall rating is equal to the 3-year rating.

4. For **very young** funds (funds with less than 3 years of historical return data), no overall star rating is calculated.

(Throughout the rest of this paper, unless otherwise indicated, when reference is made to an "old", "middle-aged", "young", or "very young" fund, it should be understood that the description is intended to reflect the definitions of these terms as indicated here.)

The central role of the age of the fund in the overall star calculation raises several interesting issues. The star-rating is quite ingenious in that it encompasses risk, loads, and short- and long-term performance in a single rating scheme. Nevertheless, because of the role that fund age plays in the calculations, it is possible that an old fund with strong long-term performance but relatively weak short-term performance can achieve a very high star rating, simply because of the heavy weight placed on past returns. Conversely, an old fund with poor long-term performance, yet strong short-term performance, may continue to receive low ratings for several years.

The weighting system could also affect the persistence of the ratings over time. Consider for the moment that as more and more mutual funds come onto the market, Morningstar will be rating more young funds as a percentage of the total funds it rates. Indeed, this point is illustrated in Table 2, which shows the number of funds evaluated by Morningstar and their respective ages from 1992 to 2000. For instance, in January 1997, Morningstar surveyed 7,857 funds of which 5,342

had less than 5 years' worth of returns (68 percent of the total funds surveyed). By comparison, in January 1993, Morningstar surveyed 2,532 funds, of which 958 had less than 5 years of returns (38 percent of the funds surveyed). Because of the way the weighting system is set up, the young funds are more likely to receive the highest and lowest performance rankings, even though older funds may have similar recent performance. This occurs because the overall ranking is an average for old funds, but not for young funds, and because of a survivorship bias, as demonstrated by Blume (1998). As the number of young funds has risen relative to the number of old funds, the group of 5-star and 1-star funds may therefore be dominated by younger funds. But as these young funds become old funds, they may drop (or rise) into the "middle of the pack" once their performance over different horizons is included in the overall rating. In recent research, Warshawsky, Mullen, and DeCarlantonio (2000) report that highly rated funds do not remain highly rated funds for very long after the time they are rated. Indeed, they find that less than half of all mutual funds rated 4 or 5 stars at the beginning of 1998 still held either of those high ratings at the end of 1998.

⟩⟩⟩ II. DO THE RATINGS HELP PREDICT FUTURE PERFORMANCE?

Since many investors use (or at least fund managers believe they use) the summary star ratings to help decide which fund to buy, an important question to ask is: Are the star ratings a signal of future performance? That is, does a 5-star fund signal that it will provide better future performance than a 3-star fund? Does a 1-star fund signify much lower future performance than a 4-star fund?

Recent research by Blake and Morey (2000) has investigated this question.[8] Our results indicate two robust findings about the predictive ability of the ratings system. First, low-rated funds (i.e., 1- or 2-star rated funds) generally fare significantly worse in the future than a 3-star-or-above rated fund. Second, and maybe more important, high-rated funds (i.e., 5- and 4-star funds) do not generally perform better in the future than do 3-star funds.

To get a better understanding of the Blake and Morey results, we focus here on one part of our research. In one section of our paper, we examined the 1993–1997 performance of all open diversified domestic equity funds rated by Morningstar in January 1993. These funds included all the aggressive growth, equity-income, growth, growth-income, and small company funds that an actual investor could have bought in January 1993 and were rated by Morningstar. Funds that were closed to new investors in January 1993 or were not diversified domestic equity funds were excluded from the sample. In addition, it should be noted here that this sample of funds consisted of 635 funds, which is just slightly less than 50 percent of the equity category funds rated by Morningstar in January 1993, as many hybrid and miscellaneous equity funds were not included as diversified domestic equity funds.

To measure the performance from 1993 to 1997, we examined each fund's Sharpe ratio, which is the excess mean monthly return[9] of the fund divided by its standard deviation. The idea behind the Sharpe ratio is that the standard deviation measures the amount of risk in the mutual fund's performance. Hence, the Sharpe ratio gives the return-per-unit risk of the fund. The higher the Sharpe ratio, the better the risk-adjusted performance of the fund.[10]

Another issue that should be briefly addressed before moving forward is that, because we are examining the 1993–1997 performance of funds that were rated by Morningstar in January 1993, a number of the funds drop out of our sample due to liquidations or mergers. If we were simply to drop these funds from the analysis, the data would be affected by what is known as a "survivorship bias," since we would only be examining the funds that survive to the end of the observation period. To deal with this issue, Blake and Morey assume that any time a fund merges or liquidates, the remaining monthly returns of the fund

become a weighted average of all the other funds of the same style (i.e., aggressive growth, equity-income, growth, growth-income, or small company). In this way, the survivorship bias is minimized.

The Sharpe ratio performance results are shown in Table 3. They reflect three different performance periods: 1 year (January 1993–December 1993); 3 years (January 1993–December 1995); and 5 years (January 1993–December 1997). The results show that, in general, there is not much difference in the Sharpe ratios of 5-star, 4-star, and 3-star funds. Indeed, when these numbers are tested for statistically significant differences, no difference is found among the 5-star, 4-star, and 3-star Sharpe ratios. The results also show that there is a considerable drop-off in Sharpe ratios between the 3-star funds and the lower-rated funds. This drop-off in performance was shown by Blake and Morey to be statistically significant, meaning that the difference in performance between high-rated and very low-rated funds was confirmed at a high degree of statistical confidence.

Before concluding this section, it should be noted that we found the results reported in Table 3 to be statistically significant across different sample periods, different styles of funds, different investment horizons, and, for the most part, across different ages of funds and ways of measuring performance.[11] Indeed, the only situation where the 5-star and 4-star funds were able to sufficiently beat the 3-star funds was when one more-sophisticated risk-adjusted return measure was used on the January 1993 sample.[12] In fact, among all the samples used in our study, the 1993 sample shows the most support (albeit not very strong support) for the idea that 5-star funds will do better than 3-star funds in the future.[13]

Table 2: Number of Funds Evaluated by Morningstar

| | **TIME EVALUATED BY MORNINGSTAR** | | | | | | | | |
	1/1992	**1/1993**	**1/1994**	**1/1995**	**1/1996**	**1/1997**	**1/1998**	**1/1999**	**1/2000**
NO. OF FUNDS EVALUATED	2,373	2,532	3,434	5,732	6,997	7,857	8,843	10,373	11,131
BY TYPE:									
OLD	502	532	608	770	942	1,240	1,582	1,750	1,843
	(21%)	(21%)	(18%)	(13%)	(13%)	(16%)	(18%)	(17%)	(17%)
MIDDLE-AGED	792	1,042	1,158	1,145	1,168	1,275	1,645	2,571	3,740
	(33%)	(41%)	(34%)	(20%)	(17%)	(16%)	(19%)	(25%)	(34%)
YOUNG	1,079	958	1,669	3,817	4,887	5,342	5,616	6,052	5,548
	(45%)	(38%)	(48%)	(67%)	(70%)	(68%)	(63%)	(58%)	(49%)
EQUITY	1,228	1,314	1,717	2,773	3,522	4,243	5,100	6,267	6,978
	(52%)	(52%)	(50%)	(48%)	(50%)	(54%)	(58%)	(60%)	(63%)
BOND	1,145	1,218	1,717	2,959	3,475	3,624	3,743	4,106	4,153
	(48%)	(48%)	(50%)	(52%)	(50%)	(46%)	(42%)	(40%)	(37%)

Source: Author's tabulation of Morningstar data.

›››› III. ALTERNATIVE PREDICTORS TO THE MORNINGSTAR STARS

In this section, we summarize research in Blake and Morey (2000) that attempts to answer the following question: Could an investor do as well by choosing funds based on alternative predictors rather than the Morningstar system? That is, can a rating scheme based on a different methodology predict fund performance as well as Morningstar star ratings? To answer this question, Christopher Blake and I examined a host of possible alternative predictors to see if they would do as well as the Morningstar stars in terms of predicting future performance. In the interest of space and clarity, I only describe the results relative to one of our alternative predictors.

To calculate this alternative predictor, we used each fund's 1990–1992 Sharpe ratio to rank-order all the funds. (Again, our sample consisted of the same 635 diversified domestic equity funds that were rated by Morningstar in January 1993.) The higher the Sharpe ratio, the higher the ranking of the fund. Then, to determine our alternative predictor, we allocated each fund "alternative stars" using the same proportional allocation of stars that Morningstar uses to generate its star ratings. That is, there were 635 diversified

domestic equity funds in our sample, of which Morningstar rated 54 as 5-star funds, 203 as 4-star funds, 282 as 3-star funds, 89 as 2-star funds, and only 7 as 1-star funds. For our alternative star ratings, we gave 5 alternative stars to the top 54 funds according to their 1990–1992 Sharpe ratios. The next 203 funds (according to their 1990–1992 Sharpe ratios) received 4 alternative stars, and so on, until we had the same distribution of alternative star ratings in the alternative predictor sample as we had in using the Morningstar ratings.

Using these alternative star ratings, we examined the 1993 (1-year), 1993–1995 (3-year), and 1993–1997 (5-year) load-adjusted mean Sharpe ratios. The results shown in Table 4 indicate that the alternative stars predict future performance poorly—and probably worse than the Morningstar method. In fact, the 3-alternative star funds do better in terms of performance than the 4- and 5-alternative star funds in each of the three time horizons. Even the funds with only 2 alternative stars perform similarly to the top-rated funds. Only in the case of the 1-star funds is there any evidence that future (poor) performance was predicted by the alternative stars.

These results suggest the possibility that even though the Morningstar ratings do not predict future high performance well, they may comprise a better method for doing so than the other alternative methods. However, a closer examination of the results in Table 4 does not bear this out. As mentioned in Section I, Morningstar rates funds using different information depending upon the age of the funds. Old funds are rated using the 3-, 5-, and 10-year risk-adjusted returns; middle-aged funds are rated using 3 and 5 years; and young funds are rated using 3 years of returns. Conversely, our alternative method uses just the past 3 years of returns to calculate the Sharpe ratios. Hence, for the old and middle-aged funds, Morningstar is using considerably more information. As a result, this may be why their method does better in terms of pre-dicting performance.

To test this notion, we separated out the 269 old funds from our sample of 635 funds. Using these funds, we then calculated their 1983–1992 Sharpe ratio. That is, instead of calculating the past 3-year Sharpe ratio, we calculated the past 10-year Sharpe ratio. We then rank-ordered the 269 funds by their 10-year Sharpe ratios and allocated alternative stars in a method similar to the one described above, and we tested to see if the alternative star method predicted better performance.

The results are shown in Tables 5 and 6. Table 5 contains the mean Sharpe ratios using the alternative stars as predictors, while Table 6 displays the mean Sharpe ratios using the Morningstar stars (note again that the funds in the table are the old funds only). The results show that when using the same amount of historical data on the funds, the alternative star rat-ings perform as well as, if not better than, the Morningstar ratings. The load-adjusted Sharpe ratios for 1993, 1993–1995, and 1993–1997 all show that the 5-alternative star funds outperform the 4-alternative star funds, the 4-alternative star funds outperform the 3-alternative star funds, and low alternatively rated funds do significantly worse than top alternatively rated funds.

Blake and Morey (2000) report that these results are statistically significant. When the same amount of information was used, alternative stars predicted future performance similarly to, if not better than, the Morningstar stars across different sample periods and time horizons. Moreover, other alternative predictors (based on measures other than the Sharpe ratio) also showed similar results.

Hence, alternative methods that utilize the same amount of historical data, but are in some cases less complex than the Morningstar system, can produce similar predictive performance.

▶▶▶ CONCLUSIONS

This paper has made three general points. First, we documented the method with which Morningstar determines its summary star ratings. We showed that the Morningstar rating system is ingenious in that it provides a simple and easy-to-understand star rating, incorporating risk-adjusted returns for different types of funds. However, investors should be aware that the star ratings are somewhat dependent upon the age and the broad asset class of the fund. Indeed, our overall analysis of the star rating system suggests that the star ratings are somewhat subjective since, in some cases, higher summary star ratings do not necessarily imply better past performance.

Second, we describe some of the results from Blake and Morey (2000) that illustrate the predictive abilities of the summary star ratings. The results document that while low ratings are solid indicators of poor future performance, high star ratings are not good indicators of superior future performance. Generally speaking, the performance of 5-star and 3-star funds is very similar after the time that they have been rated. These results were shown to be statistically significant across different samples, investment holding time horizons, and different performance measures.

Finally, we discuss some additional results from Blake and Morey (2000) regarding how well the Morningstar summary ratings predict future fund performance as compared to other, alternative predictors. We find that when alternative methods utilize the same amount of historical data, they predict future fund performance about the same as, if not better than, the Morningstar star system.

Table 3: 1993–1997 Performance of Funds Rated by Morningstar in 1993

Diversified Domestic Equity Funds*

SHARPE RATIOS TIME HORIZON WHEN PERFORMANCE IS MEASURED:	5-STAR FUNDS	4-STAR FUNDS	3-STAR FUNDS	2-STAR FUNDS	1-STAR FUNDS
1/1993–12/1993 (1-YEAR)	0.25	0.26	0.22	0.17	0.18
1/1993–12/1995 (3-YEAR)	0.26	0.25	0.24	0.18	0.13
1/1993–12/1997 (5-YEAR)	0.28	0.30	0.29	0.23	0.20

*635 funds were examined. See Blake and Morey (2000). Sharpe ratios are load-adjusted.

Table 4: 1993–1997 Alternative Star Rating Based on 1990–1992 Sharpe Ratios

Diversified Domestic Equity Funds*

SHARPE RATIOS TIME HORIZON WHEN PERFORMANCE IS MEASURED:	5-STAR FUNDS	4-STAR FUNDS	3-STAR FUNDS	2-STAR FUNDS	1-STAR FUNDS
1/1993–12/1993 (1-YEAR)	0.21	0.20	0.25	0.23	0.02
1/1993–12/1995 (3-YEAR)	0.23	0.23	0.25	0.23	0.06
1/1993–12/1997 (5-YEAR)	0.25	0.27	0.31	0.28	−0.01

*635 funds were examined. See Blake and Morey (2000). Sharpe ratios are load-adjusted.

Table 5: 1993–1997 Alternative Star Ratings Based on 1983–1992 Sharpe Ratios

Old Diversified Domestic Equity Funds Only*

SHARPE RATIOS TIME HORIZON WHEN PERFORMANCE IS MEASURED:	5-STAR FUNDS	4-STAR FUNDS	3-STAR FUNDS	2-STAR FUNDS	1-STAR FUNDS
1/1993–12/1993 (1-YEAR)	0.26	0.23	0.18	0.18	0.16
1/1993–12/1995 (3-YEAR)	0.28	0.26	0.21	0.18	−0.01
1/1993–12/1997 (5-YEAR)	0.33	0.31	0.27	0.21	0.15

*269 funds were examined. See Blake and Morey (2000). Sharpe ratios are load-adjusted.

Table 6: 1993–1997 Morningstar Star Ratings

Old Diversified Domestic Equity Funds Only*

SHARPE RATIOS TIME HORIZON WHEN PERFORMANCE IS MEASURED:	5-STAR FUNDS	4-STAR FUNDS	3-STAR FUNDS	2-STAR FUNDS	1-STAR FUNDS
1/1993–12/1993 (1-YEAR)	0.26	0.23	0.21	0.15	0.05
1/1993–12/1995 (3-YEAR)	0.28	0.25	0.22	0.14	0.16
1/1993–12/1997 (5-YEAR)	0.34	0.30	0.28	0.21	0.26

*269 funds were examined. See Blake and Morey (2000). Sharpe ratios are load-adjusted.

That said, a few final remarks should be made here. To begin with, we should take up the obvious question of why the Morningstar summary stars do not predict winners very well. The answer, we believe, is a combination of two factors. First, the Morningstar star rating method is somewhat subjective, so truly high-performing funds are sometimes not distinguished by high ratings. If they do not receive high ratings, yet do well in the future, it obviously weakens the predictive ability of high star rated funds. Second, and more important, the results of our paper (specifically, the ability of the stars to predict low-performing funds, yet the inability to predict superior-performing funds) are broadly consistent with much of the mutual fund performance persistence literature.[14] Researchers have found that while it is relatively easy to predict poor performance (regardless of how one measures performance), it is much more difficult to consistently predict superior performance. Why this occurs is somewhat a matter of debate. Some have suggested that when funds perform well, their assets increase so dramatically that it makes the funds unwieldy. Others have suggested that if markets are efficient, high fund performance is more a matter of luck than skill, and, over time, fund performance will tend to revert towards the mean.

Even if a high rating is not a very good indicator of future performance, another question that needs to be addressed here is: What should investors do if they want an accurate appraisal of past performance? After all, there are over 11,000 funds and some filtering mechanism is needed. To answer this question, we would offer advice similar to that distributed by Morningstar. That is, an investor can use the Morningstar summary star ratings, but must also look at other rating methods offered by Morningstar or others to complement the high summary star ratings. For example, the problem of classification of funds, as described in Section Ib, can be alleviated somewhat by examining the "Morningstar category rating" along with the star rating. This category rating system uses much the same methodology as the star rating system described above, including using a rating scale of 5 to 1 stars. However, it classifies funds into much smaller categories than the 4 broad asset classifications of the star rating. This narrower classification helps to prevent performance measurement irregularities resulting from too broad of a classification or changes in classification.

Another performance measurement problem of the star rating, which can be lessened by examining other rating systems, is that of the age of the fund. As mentioned in Section Ie, the star rating system aggregates the 3-year, 5-year, and 10-year returns together, which can result in young funds being rated higher or lower simply because of their age and not their recent performance. As an alternative, an investor could use the 3-year, 5-year, and 10-year Morningstar star ratings. These star ratings do not aggregate the 3-year, 5-year, and 10-year returns together. For example, the 5-year Morningstar rating uses the 5-year Morningstar Return and 5-year Morningstar Risk to arrive at a 5-year score for the fund. It then compares the 5-year score with the 5-year scores of other funds within the same broad asset classification. Like the star ratings, the ratings go from 5 stars to 1 star; however, all the funds are rated on the same amount of information.[15]

In conclusion, if a fund scores high ratings in all three ratings systems—the star rating, the category rating, and the time-dependent star rating—then an investor can be quite sure that the fund has been a high-performing fund. However, as this article has tried to document, high performance in the past is not at all a sign of future high performance.

ENDNOTES

[1] Damato, "Morningstar Edges Toward One-Year Ratings." *The Wall Street Journal*, April 5, 1996.

[2] In its publications, Morningstar sometimes refers to this rating as the risk-adjusted star rating.

[3] For more information on how Morningstar deals with loads, see Blume (1998) and Blake and Morey (2000).

[4] *Morningstar Principia Manual* (1998), p. 97.

[5] *Ibid.,* p. 98.

[6] Morey and Morey (1999) present a methodology that endogenously determines these weights.

[7] Blume (1998), in a study utilizing only 1996 data, provides some evidence that there is a relatively high percentage of young funds that are classified as 5-star or 1-star funds.

[8] Khorana and Nelling (1998) have also investigated this issue.

[9] Excess return implies the return of the fund minus the monthly T-bill return.

[10] Similar to Morningstar, the returns we used to calculate the Sharpe ratio are adjusted for loads.

[11] In terms of sample periods, we examined five other sample periods besides that of January 1993. These were January 1992, January 1994, January 1995, January 1996, and January 1997. In terms of other performance measures, we looked at two other risk-adjusted measures of performance; and when we could assume that the styles of the funds were similar, we also examined performance using simple mean monthly returns.

[12] This measure was the 4-index alpha. See Blake and Morey (2000) for more information.

[13] It should be noted here that an in-house Morningstar study also attempted to answer whether the stars predicted future mutual fund performance. Lallos (1997), which was published in a Morningstar newsletter, found that a statistically significant majority of funds receiving 4- and 5-star ratings in 1987 maintained those high ratings a decade later. The study notes: "By the standards of what it sets out to do—separating the long-term winners from the losers—the rating is actually quite successful." However, the study was not very forthcoming in terms of the methods used in the paper. Indeed, no mention of the survivorship bias or the actual methods used to conduct the tests are listed in the text.

[14] See Hendricks, Patel, and Zeckhauser (1993), among others.

[15] Note that the Morningstar category rating also has this quality. It only rates funds on their past three years of returns.

REFERENCES

Blake, Christopher and Morey, Matthew R., "Morningstar Ratings and Mutual Fund Performance," forthcoming in the *Journal of Financial and Quantitative Analysis*, September 2000.

Blume, Marshall, "An Anatomy of Morningstar Ratings," *Financial Analysts Journal*, March/April, 1998: 19-27.

Damato, Karen, "Morningstar Edges Toward One-Year Ratings," *The Wall Street Journal* (April 5, 1996): C1.

Hendricks, Darryll; Patel, Jayendu; and Zeckhauser, Richard, "Hot Hands in Mutual Funds: Short-Run Persistence of Relative Performance, 1974-1988," *Journal of Finance*, 1993: 48, 93-130.

Khorana, Ajay and Nelling, Edward, "The Determinants and Predictive Ability of Mutual Fund Ratings," *The Journal of Investing*, November 1998: 61-66.

Lallos, Laura, "The Star Rating, 10 Years Later," *Morningstar Commentary*, April 11, 1997: 1-2.

Morningstar Principia Manual, Morningstar Publications, 1998, Chicago.

Morey, Matthew R. and Morey, Richard C., "A Mutual Fund Performance Appraisal: A Multi-Horizon Perspective with Endogenous Benchmarking," *Omega: The International Journal of Management Science*, 1999: 27, 241-258.

Warshawsky, Mark; Mullen, Lisa; and DeCarlantonio, Mary, "The Persistence of Morningstar Ratings," TIAA-CREF Institute Working Paper, 2000.

▰ TIAA-CREF **INSTITUTE**