



The Predictive Power of Morningstar's New Rating System

See our related [article](#), which examines whether advisors and their HNW/UHNW clients hold funds with higher ratings. Also, see our related [article](#) which shows Morningstar's analysis of the predictive value of their rating system.

A recent [article](#) in Investment News ("Morningstar's Stars Shine After Overhaul," September 10, 2007) cited a research study which showed Morningstar's ratings "actually are meaningful in predicting future performance of a fund — at least for a while." Several of our readers asked us to look at this research in more detail, with the goal of evaluating whether Morningstar's ratings convey useful predictive information for financial advisors and wealth managers.

Background

Morningstar dramatically revised its rating methodology, and this new system became effective June 30, 2002. The primary change in the new system is that the universe of mutual funds is segregated into 48 categories. For US equity funds, these categories correspond to the nine style boxes (small, mid, and large cap v. growth, value, and blend). Previously, there were just four broad asset class categories. The rationale behind the new system is to facilitate the comparison of funds against more relevant peer groups. Two other changes were made to Morningstar's methodology – to more accurately measure downside risk and to classify funds with multiple share classes more meaningfully.

Now, with several years of data available, researchers are beginning to examine the effectiveness of the new rating system.

The [study](#) cited in the Investment News article is by far the most comprehensive on this topic. It was done by two professors at Pace University, Matthew Morey and Aron Gottesman. Morey was the author or co-author of studies on the previous Morningstar rating system, and we spoke with him in conjunction with preparing this article. Morey and Gottesman looked at the fund ratings as of June 30, 2002, and the risk-adjusted performance of these funds over the subsequent three year period (through June 30, 2005). Their conclusion was that there is "widespread support for the notion that the new Morningstar rating system can predict future performance, at least within the first three years out-of-sample." Specifically, they found that 5-star funds outperformed 4-star funds, 4-star funds outperformed 3-star funds, etc.



Morey and Gottesman looked at 1,902 actively traded US equity funds (they excluded index funds, ETFs, and other funds not rated by Morningstar). They present their results with and without adjusting for sales loads. When adjusting for sales loads, they amortized the load evenly over the three year time period. They also adjust for survivorship with three different methodologies. Survivorship is a significant issue with this study, since approximately 20% of the funds at the beginning of the study did not exist at the end (due to mergers or terminations). The first method of adjustment was to assume that the terminated fund took on the average characteristics of the remaining funds. The second method of adjustment was to use the returns of terminated funds, as long as there were at least 12 months of data available for these funds. The third, and most sophisticated, method was to replace a terminated fund with a similar fund, based on the fund's category, rating, load, turnover, expense ratio, and fund size.

Morey and Gottesman measured the alpha (relative to a benchmark consisting of all NYSE, AMEX, and NASDAQ stocks) and Sharpe ratio of funds in each rating category. Below are the results of their study, using the third (most sophisticated) adjustment for survivorship, as well as adjusting for loads:

	5 star funds	4 star funds	3 star funds	2 star funds	1 star funds
Sharpe Ratio	0.1742	0.1576	0.1449	0.1300	0.1032
Jensen Alpha	-0.0232	-0.1192	-0.1719	-0.2435	-0.4170
4-index Alpha	-0.1505	-0.2318	-0.2598	-0.3132	-0.4820
Conditional Alpha	0.0275	-0.0569	-0.1168	-0.2017	-0.3879

For an explanation of alpha and Sharpe ratio, see the accompanying [article](#) and the note below. The alphas in this table are monthly values.

The above table shows the Sharpe Ratio declining monotonically from 5-star to 1-star ratings, and the alphas (measured in each of the three different ways) similarly declining from 5-star to 1-star funds. These findings (which are consistent with the results without adjusting for load and with the other two methods of survivorship bias adjustment) led Morey and Gottesman to their conclusion regarding the predictability of the Morningstar ratings.



Other Studies

There have been two other studies that have looked at the new Morningstar rating system for US securities.

The [first](#) concluded that the new Morningstar rating system “reduced the predictive performance of the rating system as a whole.” However, because it looked at only one year of data (from 2003) and did not look at risk-adjusted returns, we do not attach any significance to its findings.

The second² looked at US growth funds during 2003, and found that, of the 91 funds that were rated 5-star at the beginning of 2003, only 39 were rated 5-star at the end of 2003. We do not attach much significance to these findings, for a number of reasons. First, it used a short time period and a very small sample size. Second, it does not attempt to measure ratings as a predictive factor of risk-adjusted returns. Third, it unfairly penalizes 5-star funds that are ‘on the bubble’ between 5- and 4-star status, especially since a 5-star fund can only move down. However, we would like to see more studies of this nature with a larger sample size and longer time period. If there is little or no persistence in the status of ratings, then the value of ratings as a predictive tool is compromised.

Our Analysis

We believe that Morey and Gottesman used reasonable methodology. However, we are concerned about several aspects of their findings:

- (1) In the above tables the alphas (with the exception of the Conditional Alpha for 5-star funds) are all negative, implying that investors would be better off in index funds, regardless of the rating of the fund they purchased. This is may be more of a commentary on the universe of actively traded mutual funds than on the predictive power of Morningstar’s ratings. However, if the rating system cannot identify funds that beat index funds, it cannot be of much value to investors or their advisors.
- (2) The results are indicative of the performance of a portfolio of all funds with a given rating. In order to achieve these results, an advisor would need to purchase all the funds in a given rating category (e.g., for 5-star funds, purchase all 192 funds with this rating). No advisor can afford the administrative cost of holding so many funds to capture an extra 1% or 2% per year. Because of the big range of performance within each category, an advisor selecting a handful of funds within each category has little chance of capturing the purported benefit. An advisor needs to know the probability that a randomly chosen fund with a given rating will outperform



(on a risk-adjusted basis) a randomly chosen fund with a different rating. Morey and Gottesman do not provide this data. Fortunately, Morningstar provided us with this analysis, which appears in our related [article](#).

- (3) A number of studies^{3,4,5} have shown that mutual funds are often mis-categorized and therefore the selection of peer groups is widely flawed. In some cases, fund companies may actually benefit from mis-categorization, by allowing their fund to be compared to a peer group with poorer performance characteristics. Such biases in the underlying data could easily overwhelm Morey and Gottesman's findings, rendering them useless.
- (4) Morey and Gottesman state that fund expense ratios do not explain their results, citing the fact that expense ratios for 5-star funds are slightly higher than those of 4-star and 3-star funds. However, the expense ratios of 1-star funds are significantly higher. While expense ratios may not say much about higher rated funds, they certainly say a lot about why the 1-star funds are poor performers.
- (5) The time period of this study (June 2002-June 2005) was mostly an up market. The change to Morningstar's methodology was to more accurately measure downside risk in funds. Given this, we believe that more data is required to evaluate the effectiveness of the new methodology, and that the data must encompass a market cycle that includes a down market.
- (6) The adjustments Morey and Gottesman use for sales loads can have some undesirable effects. For example, if a fund with a 4% front-end load experienced significant positive performance at the beginning of the study, the methodology used would credit the fund with virtually all of this performance. Yet, because of the front-end load, only 96% of the assets would benefit from the performance. We would prefer to see a study that looked exclusively at load-waived and no-load share classes, which are the relevant universe for advisors.

Conclusion

For the above reasons, we believe the Morey and Gottesman study does not provide any actionable results for advisors. We believe "the jury is still out" as to whether the new Morningstar rating system offers predictive value, in terms of risk-adjusted performance. We would like to see the methodology used by Morey and Gottesman extended to adjust for the problems in mis-categorization of funds, and applied to a universe of funds without loads, appropriate for advisors. We would like to see the probabilities of randomly chosen funds with



one rating outperforming funds with a different rating. Most importantly, we believe that more time is required to study the behavior of funds. Although we do not wish for a down market, it will take at least one such market cycle to truly test the predictive value of the new Morningstar ratings methodology.

Notes

1. Jensen alpha is the conventional representation of risk-adjusted return, and shows the return that would be expected over and above the return of the associated index or benchmark security. A Jensen alpha of $-.0232$ implies a monthly return of $.0232\%$ less than that of the benchmark or index. The 4-index alpha applies an adjustment to the Jensen alpha that corresponds to the Fama-French 3 factor model, in that it takes into account risk associated with the size, style, and capitalization of the funds. It also uses a momentum effect. The conditional alpha applies an adjustment to the Jensen alpha that allows for the fact that the risk associated with the benchmark or index may change over time.
2. Gerrans, Paul, 2006. "Morningstar ratings and future performance", *Accounting and Finance* 46 pp. 605–628.
3. DiBartolomeo, Dan and Erik Witkowski. "Mutual Fund Misclassification: Evidence Based On Style Analysis," *Financial Analyst Journal*, 1997, v53(5,Sep/Oct), 32-43.
4. Brown, Stephen J. and William N. Goetzmann. "Mutual Fund Styles," *Journal of Financial Economics*, 1997, v43(3,Mar), 373-399.
5. Kim, Moon, Ravi Shukla and Michael Thomas. "Mutual Fund Objective Misclassification," *Journal of Economics and Business*, 2000, v52(4,Jul/Aug), 309-324.

www.advisorperspectives.com

For a free subscription to the Advisor Perspectives newsletter, visit:
<http://www.advisorperspectives.com/subscribers/subscribe.php>