Dispatches

Rating the Star Rating
A recent study raises questions about the rating's worth.

Jeffrey Ptak

Does the Morningstar Rating for funds, more commonly known as the star rating, predict future performance?

We recently published a detailed study, "The Morningstar Rating for Funds: Analyzing the Performance of the Star Rating Globally," in which we sought to answer that question. What we found is that the star rating is predictive — to a point. It had moderate predictive ability for risk-adjusted returns, but less so for simple returns in most asset classes. The findings supported what we've long held, that the Morningstar Rating can be a helpful starting point, but should not be the final analysis. It is a handy tool, but not the only one in the tool kit.

Then, we shared a summary of the study on Morningstar.com, where it generated a flurry of comments from our website readers. These are generally sophisticated individual investors who enjoy getting into the weeds, and their questions and criticisms served as a reminder that this isn't simply an academic question. The star rating is still the best-known tool in the box, and investors want to know whether and how to use it.

Background
The star rating is a backward-looking, quantitative measure of a fund's past performance compared with its category peers. We've long stated that the star rating can be a good starting point for research for a few reasons:

▶ It is based on funds' trailing three-, five-, and 10-year returns versus category peers, not just the most recent year. Longer-term returns are less unstable and, thus, a better predictor of future performance than short-term results.

▶ It compares funds' net-of-fee returns. Our research has shown that fees are one of the best predictors of future success.

▶ It adjusts fund returns to account for risk (technically, "downside volatility"). This acknowledges that some funds are harder for investors to use and factors that into the rating accordingly.

In other words, the star rating distills some of the more telling indicators of fund success into one easy-to-use measure. This, in turn, can help investors cut a larger universe of investment options down to size, clearing the way for more-detailed research from there.

Study Approach
The time frame of our study was January 2003 to December 2015 (January 2008 to December 2015 for alternative funds) and included all rated funds in our database during that span. In addition, the study included obsolete funds (that is, those that have been merged or liquidated) and thus does not suffer from survivorship bias.

Using this data, we conducted two tests of the star rating's predictive power. The first was a Fama-MacBeth cross-sectional regression, which evaluated the relationship, if any, between a fund's star rating and its return in the next month. In particular, we sought to assess whether higher-rated funds were associated with superior forward one-month returns (relative to 3-star funds) than lower-rated funds after controlling for numerous variables, including asset class, Morningstar Category, fund expense, and risk.

The second test was an event-study procedure in which we took a snapshot of funds' star ratings at a given point in time, measured subsequent performance over varying time horizons, and then compared the returns of the ratings cohorts over those event horizons.

We ran twin tests for several reasons. We wanted to conduct a rigorous study that drew upon innovative research techniques that have been widely adopted and accepted in academic circles. However, recognizing that the Fama-MacBeth regression technique could be less intuitive to some investors, we also wanted an alternative that would be more accessible, hence our inclusion of the event-study procedure.

Key Findings

Using these two performance frameworks, we found that the star rating had moderate predictive power during the study period.

The Fama-MacBeth approach found that funds with higher star ratings had superior returns even after accounting for expenses and various risk exposures. Furthermore, these results held across all asset classes except alternatives:

▸ Among equity funds, the forward one-month returns of 5-star funds were 0.09% higher than those of 1-star funds, or 1.03% annualized.
▸ Similarly, 5-star fixed-income funds registered 0.09% higher forward one-month returns than 1-star funds (1.09% annualized), and 5-star allocation funds notched 0.15% higher returns than 1-star funds (1.75% per year).
▸ The premiums observed were highly statistically significant for fixed-income and allocation funds, but less so for equity funds.
▸ There wasn't a statistically significant difference between the performance of 5-star and 1-star alternative funds. These results could be explained by the smaller sample size (January 2008 to December 2015) and model misspecification (owing to unique return drivers among alternative funds).

The event-study approach led us to similar but less-convincing conclusions. The returns of the portfolios with 5- and 4-star ratings were higher than those of the other star rating cohorts, but not substantially so. Similar to the regression approach, the event study reveals a consistent (though weak) monotonic pattern throughout all periods, with the higher-rated portfolios delivering better average cumulative returns. Here's the table:

Average Cumulative Returns (%)

Time From Rating	★	★★	★★★	★★★★	★★★★★
1 Mo.	0.52	0.54	0.57	0.60	0.62
3 Mos.	1.70	1.72	1.79	1.87	1.92
6 Mos.	3.52	3.50	3.59	3.73	3.80
12 Mos.	6.90	6.89	7.03	7.35	7.51
36 Mos.	20.77	20.23	20.44	21.14	21.26
60 Mos.	28.56	28.30	28.87	29.65	29.22

Data from 01/01/2003–12/31/2015.
Source: Morningstar.

Taken together, our findings suggest that the star rating had moderate predictive ability for risk-adjusted returns in the short term.

It is also worth noting that 5-star funds were far likelier to survive the full event-study horizon, especially the 60-month horizon, than lower-rated funds. One could argue that the star rating confers the benefit of predicting survival.

Responses and Next Steps

We demand transparency when we're analyzing funds. Given that, it's only fair to hold ourselves to the same standard, which explains why we plan to regularly conduct studies like this one and make the results publicly available. It helps us to identify and prioritize potential future enhancements to the ratings. And it can benefit users who are given the opportunity to assess the strengths and deficiencies of our ratings for themselves and offer useful feedback.

Speaking of feedback, some readers of the study expressed disappointment that there wasn't greater separation in the subsequent performance of the ratings cohorts over longer time frames. That point is well taken, but it's worth noting that we conducted the event study in a very simple format mainly so that it was more understandable to most investors. Thus, it doesn't account for risk the way the Fama-MacBeth approach does. If it did, we'd likely see more separation. This will be the focus of additional analysis.

Moreover, while the event study isn't survivorship-biased, in that we incorporated the results of dead funds before they were wiped out, it somewhat understates the experience an investor would have had investing in the equal-weighted 1- and 2-star buckets. How so? It doesn't convey the amount of mortality in those cohorts. Lower-rated funds are merged and liquidated away more frequently, sometimes at considerable inconvenience and tax cost to investors. By contrast, 5-star funds live longer, affording investors greater opportunities to succeed without interruption and forestalling the need to make an additional investment decision.

One commenter suggested we add Analyst Ratings to the mix. This study did not encompass the Analyst Rating because that measure is still relatively new. However, it recently hit its five-year anniversary, and we plan to publish a similar paper on the efficacy of the Analyst Rating this year.

Other readers wondered whether the study would prompt us to re-examine the methodology of the star rating. For example, the star rating's efficacy varied by asset class. Higher-rated funds tended to outperform lower-rated funds by wider margins in allocation and fixed-income asset classes. By contrast, there was less separation among equity funds, and the star rating didn't appear to be predictive at all for alternative funds. This could imply that the star rating is a bit more effective in asset classes where there's a lower dispersion of returns and greater homogeneity among funds, though this is subject to further study and validation.

It also could suggest that the star rating has been more predictive in categories where performance differences are largely explained by cost advantages. Indeed, our research has found that expense ratios are one of the best tools for predicting future relative performance. To be sure, the Fama-MacBeth method found the star rating had predictive power *independent of expense differences*. But it's possible that the event-study results would have been better if the star rating methodology weighed expenses even more heavily. This, too, will be a focus of additional research. **IM**

Jeffrey Ptak, CFA, is head of manager research. He is a member of the editorial board of *Morningstar* magazine.