The Morningstar Rating™ for Funds

Analyzing the Performance of the Star Rating Globally

Morningstar Quantitative Research
10 November 2016

Jeffrey Ptak, CFA
Head of Manager Research
+1 312-384-4928
jeffrey.ptak@morningstar.com

Lee Davidson, CFA
Head of Quantitative Research
+1 312-244-7541
lee.davidson@morningstar.com

Christopher Douglas
Director
Manager Research Ratings, APAC
+1 312-384-3769
christopher.douglas@morningstar.com

Alex Zhao, CFA
Analyst,
Equity Research
+1 312-244-7559
alex.zhao@morningstar.com

Introduction

The Morningstar Rating™ for funds, or "star rating," was introduced in 1985 to help investors and advisors better understand and assess the crowded mutual fund landscape.

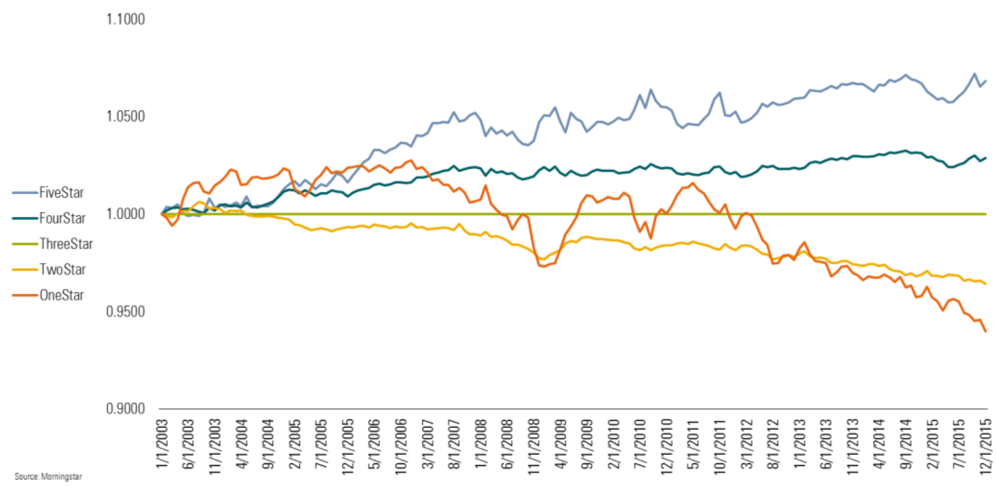
We designed it as a grade on past performance, much as a university would assign grades to students to document who demonstrated better effort and ability in their coursework. We wanted a standard that went beyond the short-term performance that dominated fund marketing at the time. Accordingly, we focused the star rating on longer-term performance and, following the lead of academic finance experts, incorporated risk and all sales charges into our rating. We believe that the star rating provides a good scorecard, giving unbiased documentation of which funds have served investors best. We have consistently said the star rating is the start of a selection process, not its conclusion. Just as a university wouldn't claim that its "A" students will achieve greater success than its "B" or "C" students, we haven't promoted the star rating as an infallible predictor of future success, even though we do believe that holding managers to the standard of delivering better long-term risk- and cost-adjusted returns is beneficial. Nevertheless, we think it's worthwhile to measure the efficacy of the star rating and are committed to reviewing it regularly because fund investors around the world use it as part of their fund selection process and, in some cases, to forecast fund returns.

Executive Summary

In this study, we analyzed the global performance of the star rating in terms of its ability to predict risk-adjusted fund returns. We employed two common approaches to measure the predictive power of these ratings as investment signals: 1) Fama-MacBeth cross-sectional regressions and 2) an event study. The timeframe is January 2003 through December 2015.

The results suggest that the star rating had some moderate predictive power during our sample period. The Fama-MacBeth cross-sectional regression shows that funds with higher star ratings had superior returns in the cross section even after accounting for expenses and various risk exposures. Furthermore, these results held across all asset classes except alternatives.

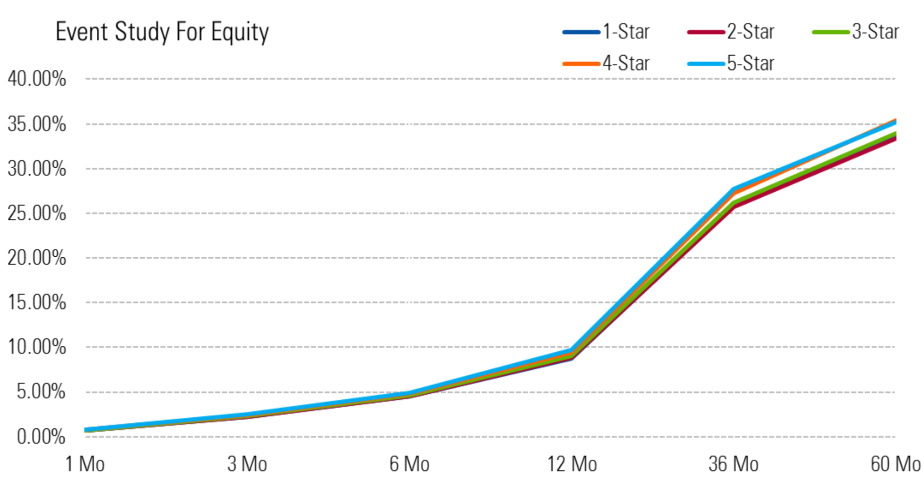
Exhibit 1 Aggregate Return Premiums by Morningstar Rating for Funds (Equities)



Source: Morningstar, Inc.

The event study pointed in the same direction as the regression findings, though the significance indicated was much lower. An event study is meant to showcase the typical experience for the typical investor according to the star rating for various holding periods. Our findings in the event study suggest that 5-star funds outperformed 1-star funds by approximately 0.25 percentage point annualized on average across multiple holding periods and market cycles. This result held clearly in different magnitudes across all asset classes except alternatives. It should be noted that 5-star funds were also more likely to survive a full event-study horizon than lower-rated funds that more often were merged or liquidated, a key distinction given the potential inconvenience and cost of having to switch funds.

Exhibit 2 Average Return by Morningstar Rating for Funds (Equities) Over Different Horizons



Source: Morningstar, Inc.

Taken together, our results suggest that the star rating had moderate predictive ability for risk-adjusted returns but less so for simple returns in most asset classes, as evidenced by the Fama-MacBeth and event study results, respectively. Furthermore, this predictive power appeared to sustain itself even as the holding period lengthened. In this paper, we discuss the data and our interpretations for these results in more detail. The study is divided into the following sections:

Section 1 – Overview
Section 2 – Methodology and Explanation of Process
Section 3 – Regression Results
Section 4 – Event Study Results
Appendix 1 – Net Expense Ratio Equivalent Global Calculation
Appendix 2 – Literature Review
References

Section 1: Overview

The Morningstar Rating for funds, or “star rating,” was introduced in 1985 and quickly became a leading marker for fund appraisal. Initially only available for U.S. mutual funds, it now covers a range of investment types from exchange-traded funds to variable annuities. Morningstar now publishes star ratings on more than 217,000 managed investments across 72 countries.

The methodology is straightforward. Funds must first have a minimum three-year track record. Performance is then assessed after fees on a risk-adjusted basis. The Morningstar Rating rewards longer-term performance, low volatility of returns, and low fees. The formula is based on past performance, so it is a backward-looking indicator. Investors seeking Morningstar's view on a fund's future prospects should refer to the Morningstar Analyst Rating and associated commentary.

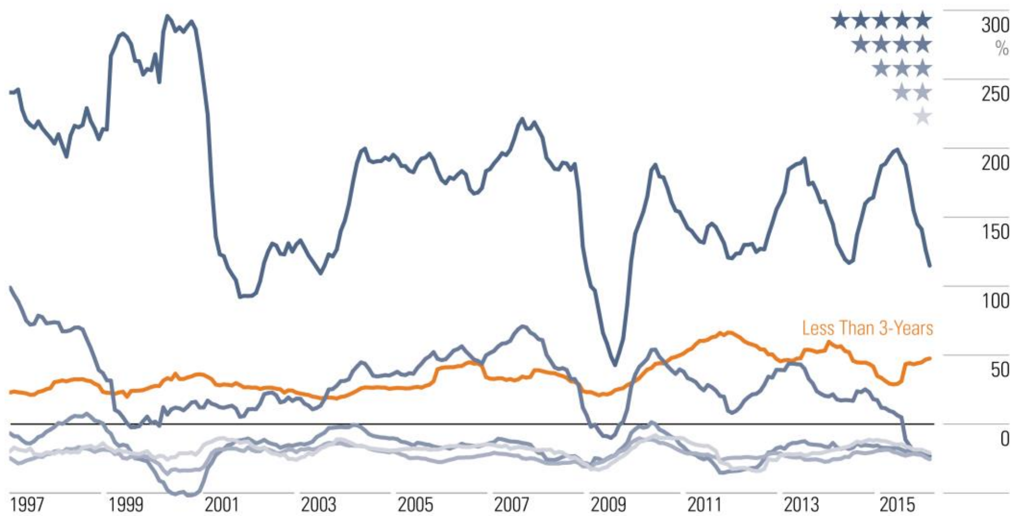
Historical Effect of the Morningstar Rating for Funds
Before evaluating the predictive power of the star rating, it is important to look back. We believe the introduction of the star rating had a significant positive effect on the mutual fund industry and, as a predictor of future returns, was much more predictive than what it replaced.

In 1985, investors had very little information available to assess a fund, so most could draw conclusions from only short-term performance, which was heavily marketed by fund companies. Information on sales loads was ignored, risk was not part of the conversation, and there was little focus on long-term returns. These are all facets that were embedded in the star rating's formulas since day one.

The star rating, as well as written commentary from Morningstar’s fund analysts, helped steer investors away from funds with excessive loads and high fees toward those that have proved to deliver over the medium to long term and with fewer swings in performance.

Investors incorporated the star rating into their decision-making process soon after its initial launch. Overwhelmingly, we find that investors tend to allocate money toward 5-star funds and away from 1-star funds going back to at least 1997. In Exhibit 3, we show rolling one-year flows into each of the five rating cohorts, which clearly shows that investors preferred 5-star funds throughout this period. The pattern has held up convincingly during the subsequent two decades. Given that investors began to allocate to higher-rated funds at the expense of lower-rated funds, the star rating effectively contributed to an investing culture that gives incentives for low fees, no loads, low risk, and solid long-term performance.

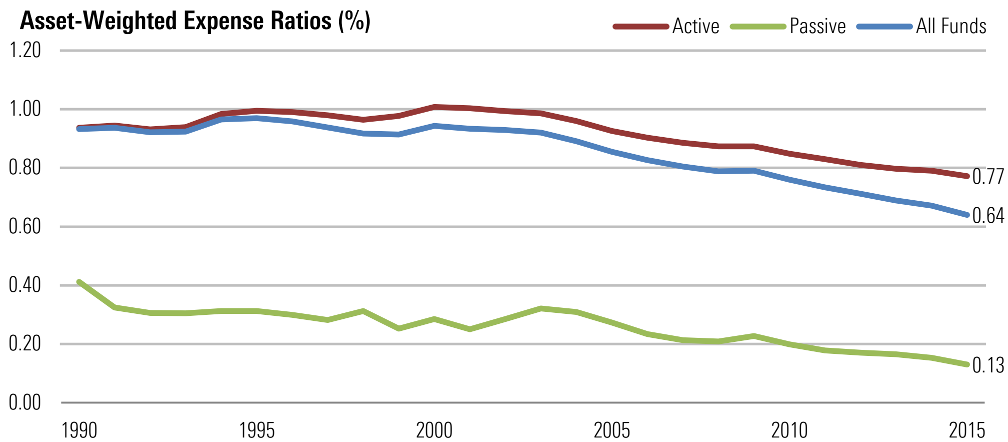
Exhibit 3 Investors Tend to Buy 5-star Funds



Source: Morningstar, Inc.

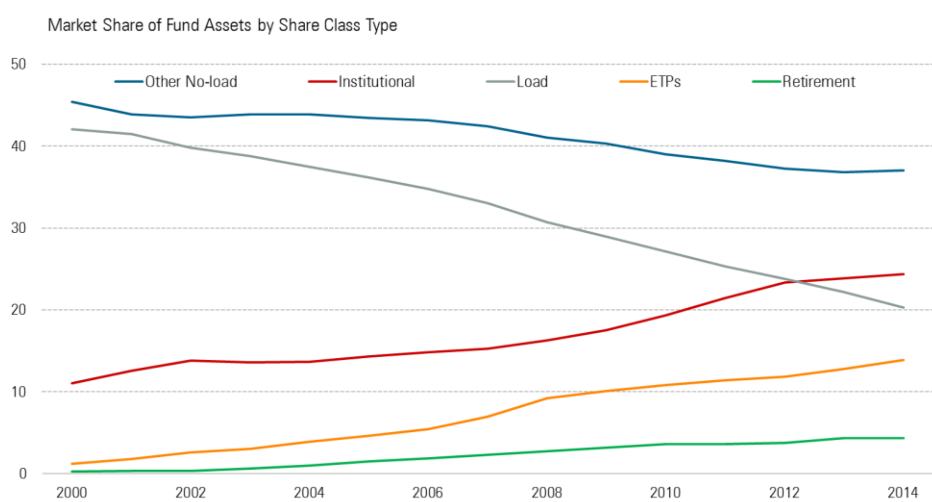
A fund company looking to grow its assets under management must first recognize that investors preferred the types of products that the star rating rewarded and then build their lineups accordingly. Over time, it has been widely observed that fees have all come down (Exhibit 4). Load funds have also seen their share of AUM decline (Exhibit 5). Certainly, other factors besides the star rating were at work during this time, such as the rise of passive investing and ETFs. But we believe the star rating was one of the first summary statistics that investors could use to identify preferred funds and over time has contributed to the transparency investors enjoy today.

Exhibit 4 Expenses Have Come Down for Active and Passive Funds



Source: Morningstar, Inc.

Exhibit 5 Load Shares Are on Their Way to Extinction as Institutional Shares and ETPs Have Gained Ground



Source: Morningstar, Inc.

Examining the Predictive Power Today

The star rating was originally designed as a simple analysis of a fund, but this paper explores whether it is right to use it as an investment signal for a fund. Our results suggest that it has been a useful predictor of future fund returns or, at the very least, has not caused harm for investors in most asset classes over most horizons. It has led clients to funds with lower fees, a focus on risk, and a longer-term horizon. These are all positive things, but the large influence of past performance in the calculation resulted in some periods of poor returns owing to market sell-offs and inflection points that the rating is not designed to predict.

This paper analyzes data since January 2003, a few months after we made significant changes[1] to the Morningstar Rating methodology, so it is not relevant to assess the data prior to this date. Morningstar has published a broad range of articles on the star rating over many years. This paper targets broad audiences, including academia, so rather than assess the star rating against attributes like batting averages or other well-known measures of success as we have in the past, we used two approaches to test performance: 1) Fama-MacBeth cross-sectional regressions and 2) an event study in which funds were grouped by star rating. We used a similar process to analyze the Morningstar Rating for stocks in previous papers.

The Fama-MacBeth cross-sectional regression aims to discover whether higher-rated funds have superior one-month forward returns in the cross section. As part of this framework, we controlled for expenses and a number of widely recognized risk factors to better isolate the contribution from Morningstar's ratings on performance beyond these drivers of returns. By "controlled for," we mean that we have attempted to remove the effect of these characteristics on the result. Therefore, it is difficult to attribute the return differences observed between 5- and 1-star funds to differences in expenses or risk factors.

The set of risk factors we used varies by the broad asset categories. Specifically, we controlled for market, size, value, and momentum risk for equity funds; duration and credit (default) risk for fixed-income funds; market, size, value, duration, and credit risk for allocation funds; and market and commodity exposure for alternatives funds. We controlled for expenses in all asset classes.

We recognize that although Fama-MacBeth cross-sectional regressions are a rigorous way to evaluate the efficacy of our ratings, they may not be intuitive to all readers nor simulate a real-world scenario. Therefore, in our second approach, we tested the efficacy of the Morningstar Ratings by constructing an event study. This view of our ratings provides a picture for what the typical investor could expect to experience over a variety of holding periods. The event study uses simple and transparent construction

1. In 2002, Morningstar enhanced the star rating with new peer groups and a new measure of risk-adjusted return. The peer groups for the rating were changed to the smaller category groups instead of the broad asset classes. As a result, for this study we analyzed the Morningstar Rating from January 2003 onward.

rules without any steps taken to isolate the influence of risk factors or expenses on returns. In this sense, it is simple and intuitive but also ignores well-known return drivers.

Using these two performance frameworks, we found relatively strong results that suggest the star rating had predictive power during the sample period. The Fama-MacBeth cross-sectional regression approach strongly supports the hypothesis that funds with higher star ratings had superior returns in the cross section even after accounting for expenses and various risk exposures. Furthermore, these results held across all asset classes except alternatives. Among equity funds, moving from 1 star to 5 stars was correlated with 0.09-percentage-point higher returns per month on average (1.03 percentage points annualized). Similarly, 5-star fixed-income funds registered 0.09-percentage-point higher returns per month more than 1-star funds (1.09 percentage points annualized), and allocation funds achieved 0.15-percentage-point higher returns (1.75 percentage points annualized). Alternatives funds, on the other hand, saw much less discrimination in performance during the shorter time period of study (2008-14), with a 0.04-percentage-point per month difference (0.48 percentage point annualized) estimated between 5- and 1-star funds.

The event study approach led us to similar but less convincing conclusions. The event study aims to showcase the typical investor's experience of investing according to the various levels of the star rating for different holding periods. The study is constructed by sorting funds into groups according to their star ratings each month, equally weighting them, and then tracking each group's subsequent performance during several time periods: one month, three months, six months, 12 months, 36 months, and 60 months. We formed these pairings each month for the entire sample period and then averaged the returns over each subsequent time period for each rating group. We reconstituted each rating group each month to account for any funds that became obsolete. This analysis covers the period January 2003 to December 2015 for global funds and January 2008 to December 2015 for alternatives funds.

Among equity funds, we found that the average three-year forward cumulative return was 27.74% for 5-star funds and 25.78% for 1-star funds (annualized outperformance of 0.64 percentage point). Similarly, fixed-income funds registered average three-year forward cumulative returns of 13.71% for 5-star funds compared with 12.66% for 1-star funds (annualized outperformance of 0.35 percentage point). Allocation funds have an annualized return gap of 0.59 percentage point between 5- and 1-star funds for the same typical three-year period. Finally, the star rating again did not fare as well in the alternatives asset class, with average three-year forward cumulative returns of 19.15% for 5-star funds and 18.64% for 1-star funds (0.17 percentage point annualized).

It is important to remember that the event study results do not control for the risks of the funds and must be interpreted carefully. This method of analysis cannot account for any correlation between the star rating assignments and the level of factor risk exposure or fees. For this reason, we caution against strong conclusions based on these data alone.

While we do not control for risk in the event study, we do attempt to insulate the study from any survivorship bias. The event study includes funds that may have not persisted for the entire holding period. Obsolete funds may contribute to the study up until the time that they are no longer active.

Taken together, however, our findings suggest that the star rating had moderate predictive ability for risk-adjusted returns in the short term. Over the long term, the event study indicates that the star rating has appeared to do more good than harm to investors during the period studied. At a one-month horizon, higher-rated funds exhibited strong persistence in performance that cannot be explained by their fees or risk exposures; this performance persistence correlated very nicely with the star ratings. The event study results corroborate these results over the one-month horizon. Longer-term event study results further suggest that, on average, investors who have bought higher-rated funds have tended to earn higher returns than those who have purchased lower-rated funds, though the magnitude of this outperformance is low.

Given that the star ratings are based on past performance, we think that these findings are consistent with a "momentumlike" interpretation of the star rating in the short run. Like traditional momentum, the star rating is a backward-looking performance indicator (albeit risk-adjusted), and as such, it makes sense that the relevance of this information should decay as time goes on. One possible conclusion from our findings suggests that sorting on past risk-adjusted returns–of which the star rating is but one example–offers a different dimension of return behavior than traditional momentum. The event study suggests that the performance gap between higher-rated and lower-rated funds persists as the period lengthens. We believe that this relationship is likely caused by the star rating's high correlation to fees and that over time these differences accumulate in an economically significant way.

Through the years we have repeatedly stated that the Morningstar Rating for funds is best used as the starting point in fund research, and this paper does not alter this view. We acknowledge, however, that it is important to understand what, if any, predictive power the Morningstar Rating possesses. What can investors expect by selecting a fund with a 4- or 5-star rating? Are there particular asset classes or regions where investors can have greater confidence in the rating? This paper seeks to uncover answers to these questions and more.

While this paper presents global results, regional results, such as U.S.-domiciled funds, are available upon request. Going forward, we will update this paper regularly and, as always, invite any feedback and all discussion.

Section 2: Methodology and Explanation of Process

Morningstar Rating for Funds

The Morningstar Rating for funds (or "star rating") is a quantitative assessment of a fund's past performance—both return and risk—as measured from 1 to 5 stars. It uses focused comparison groups to better measure fund manager skill on an after-fees basis. The Morningstar Rating is intended for use as the first step in the fund evaluation process. A high rating alone is not a sufficient basis for investment decisions.

The Morningstar Rating has two key characteristics:

- The peer group for each fund's rating is its Morningstar Category.
- Ratings are based on funds' risk-adjusted returns versus category peers.

Morningstar Category

The original Morningstar Rating was introduced in 1985 and was often used to help investors and advisors choose one or a few funds from among the many available within broadly defined asset classes. Over time, though, increasing emphasis was placed on the importance of funds as portfolio components rather than "stand-alone" investments. In this context, it was important that funds within a particular rating group be valid substitutes for one another in the construction of a diversified portfolio. For this reason, beginning in 2002, Morningstar began assigning ratings based on comparisons of all funds within a category rather than all funds in a broad asset class.

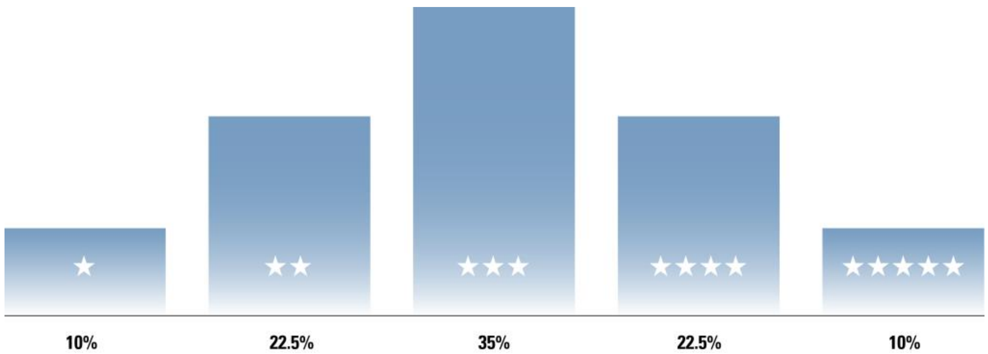
Morningstar Risk-Adjusted Returns

The Morningstar Rating is based on "expected utility theory," which recognizes that investors are 1) more concerned about a possible poor outcome than an unexpectedly good outcome and 2) willing to give up some portion of their expected return in exchange for greater certainty of return. The rating uses the Morningstar Risk-Adjusted Return measure, which accounts for all variations in a fund's monthly performance, with more emphasis on downward variations. It rewards consistent performance and reduces the possibility of strong short-term performance masking the inherent risk of a fund. MRAR is similar to the Sharpe ratio but goes a step further by penalizing downside risk, making it a more intuitive measure.

How Does It Work?

The Morningstar Rating for funds methodology rates funds based on Morningstar Risk-Adjusted Return, which accounts for the effects of all sales charges, loads, or redemption fees. Funds are ranked by their MRAR scores, and stars are assigned using the scale in Exhibit 6:

Exhibit 6 Morningstar Ratings



Source: Morningstar, Inc.

Funds are rated for up to three periods—the trailing three, five, and 10 years. For a fund that does not change categories during the evaluation period, the overall rating is calculated using the weightings in Exhibit 7.

Exhibit 7 Morningstar Ratings

Age of Fund	Overall Rating
At least three years, but less than five	100% three-year rating
At least five years, but less than ten	60% five-year rating 40% ten-year rating
At least ten years	50% ten-year rating 30% five-year rating 20% three-year rating

Source: Morningstar, Inc.

A link to the full methodology can be found here.

Data

Our sample encompasses funds given the Morningstar Rating for funds. The sample relies entirely on Morningstar fund data sources, begins in January 2003, and ends in December 2015. Monthly fund counts ranged from 7,669 to 44,071, having mostly grown during the period. Our sample includes multiple broad asset classes: balanced funds (fund counts ranged from 946 to 12,098 funds), equity (3,361 to 21,636 funds), and fixed income (1,681 to 9,727 funds). Given the small universe of alternatives funds, we restricted our analysis of that subgroup to the period of January 2008 to December 2015, when fund counts ranged from 643 to 1,224. In December 2015, our sample spanned a total of 43,698 funds.

Our sample was geographically diverse throughout the period tested. By the end of the period, funds domiciled in Europe accounted for 43.7% of the total universe, North and South America 33.0% (U.S. funds represented 18% of the total universe), Asia-Pacific (ex Australia and New Zealand) 13.3%, Australia and New Zealand funds 6.9%, and Middle East and Africa funds 3.1%. Of all developed countries, Japan was least represented in the sample owing to a lack of reliable data on Japanese funds.

We restricted our analysis to the oldest share classes of individual funds to avoid overweighting funds with multiple share classes. We recognize that there are other ways to select representative share classes, such as choosing "retail" or "institutional" shares. However, it was not feasible to select share classes in this manner because of regional differences in share-class topology. Furthermore, we believe the primary differences between share-class selection approaches resolve to cost. Given that we included fees in all of our regressions, we believe our analysis controls for any performance disparities related to fee differences (stemming from share-class differences or otherwise).

We further restricted our analysis to funds that possessed monthly returns and funds that were categorized into a Morningstar broad asset class. This represents a trivial criterion, as funds must possess both historically to receive a star rating.

Our sample does not suffer from survivorship bias. Morningstar's global fund databases retain a full history of obsolete funds, and our sample includes these funds. Moreover, our evaluation technique dynamically incorporates monthly changes in fund-universe composition, providing a more holistic and realistic picture of historical performance. Each monthly snapshot captures any funds that were subsequently merged or liquidated.

Regression Coefficients

The control variables in our regressions are important to understand. The net expense ratio data point is described in detail in Appendix 1. We also include fund betas estimated from time series regressions dependent on asset class.

Equity funds: We ran rolling three-year regressions of a fund's return onto the region-appropriate Fama-French-Carhart factors: market risk, or RMRF; value, or HML; size, or SMB; and momentum, or UMD. All returns, which we sourced from the Kenneth R. French Data Library, are in U.S. dollars, include dividends and capital gains, and were not continuously compounded. We selected region-appropriate factors based on each fund's Morningstar Category, which is based in turn on the fund's portfolio holdings data. For example, funds that invest in the stocks of U.S. large-cap firms were classified into the U.S. large-blend category. The Fama-French-Carhart factors were calculated for each of the following regions: global, global ex U.S., Europe, Japan, Asia-Pacific ex Japan, and North America. Each regional set of factors contained the following:

RMRF (also known as "excess return on the market") is the excess return of the region-specific market portfolio, calculated as each region's market-cap-weighted portfolio return minus the regional risk-free rate (that is, the one-month U.S. Treasury bill).

SMB ("small minus big," or small-cap minus large-cap) and HML ("high (ratio of book value to price) minus low," or value minus growth) portfolio returns represent factor portfolios designed to proxy a common risk in equity returns arising from cross-sectional differences in market capitalization and valuation. To construct the SMB and HML factors, stocks in a region are sorted into two market-cap and three book/market equity, or B/M, groups at the end of June each year. "Big" stocks are those in the top 90% of June market cap for the region, and "small" stocks are those in the bottom 10%. The B/M breakpoints for a region are the 30th and 70th percentiles of B/M for that region's "big" stocks.

SMB is the equally weighted average of the returns on the three "small" stock portfolios for the region (small value, small core, and small growth) minus the average of the returns on the three "big" stock portfolios (large value, large core, large growth).

HML is the equally weighted average of the returns for the two high B/M portfolios for a region (small value and large value) minus the average of the returns for the two low B/M portfolios (small growth, large growth).

UMD ("up minus down") is a factor portfolio designed to proxy an observed return pattern of momentum in equities where recent winners keep winning and recent losers keep losing. The 2x3 sorts on size and lagged momentum to construct UMD are similar to those used for value/growth, but the size-momentum portfolios are formed monthly (instead of annually). For portfolios formed at the end of month t–1, the lagged momentum return is a stock's cumulative return for month t–12 to month t–2. The momentum breakpoints for a region are the 30th and 70th percentiles of the lagged momentum returns of the "big" stocks of the region. UMD is the equally weighted average of the returns for the two winner portfolios (large and small) for a region minus the average of the returns for the two loser portfolios (large and small).

The regression rolls monthly, providing a set of factor betas, alpha, and R-squared values each month estimated from the prior 36 months' experience. The equity asset-class regression took the form:

$$r_{i,t} = \alpha_i + \beta_i^{rmrf} RMRF_t + \beta_i^{hml} HML_t + \beta_i^{smb} SMB_t + \beta_i^{umd} UMD_t + e_{i,t}$$

Subsequently, we used the estimated time series of factor alphas and betas above $\left(\alpha_i, \beta_i^{rmrf}, \beta_i^{hml}, \beta_i^{smb}, \beta_i^{umd}\right)$ as our explanatory variables in our cross-sectional regressions. For simplicity, we refer to these as Alpha (α_i), Market Beta $\left(\beta_i^{rmrf}\right)$, Value Beta $\left(\beta_i^{hml}\right)$, Size Beta $\left(\beta_i^{smb}\right)$, and Momentum Beta $\left(\beta_i^{umd}\right)$.

Fixed-Income and Balanced Funds

We ran rolling three-year regressions of a fund's return onto the region-appropriate Fama-French factors–RMRF, HML, and SMB–as well as term premium and default premium factor series, which we computed in a manner consistent with that set forth in Chen, Roll, & Ross (1986) and Fama-French (1993).

TERM (term premium) is a factor portfolio designed to proxy a common risk in bond returns arising from unexpected changes in interest rates. The portfolio return is calculated by going long the Barclays Capital U.S. Treasury 10-20 Year TR USD Index and shorting the Barclays Capital U.S. Treasury Bill 1-3 Month TR USD Index.

DEF (default) is a factor portfolio designed to proxy a common risk in bond returns arising from shifts in economic conditions that could change the likelihood of default. The portfolio return is calculated by going long the Barclays Capital U.S. Corporate High Yield TR USD Index and shorting the Barclays Capital U.S. Government TR USD Index.

The regression rolls monthly, providing a set of factor betas, alpha, and R-squared values each month estimated from the prior 36 months' experience. The fixed-income and balanced asset-class returns-based style-analysis regression took the form:

$$r_{i,t} = \alpha + \beta_i^{rmrf} RMRF_t + \beta_i^{hml} HML_t + \beta_i^{smb} SMB_t + \beta_i^{term} TERM_t + \beta_i^{def} DEF_t + e_{i,t}$$

Subsequently, we used the estimated time series of factor betas above $\left(\alpha_i, \beta_i^{rmrf}, \beta_i^{hml}, \beta_i^{smb}, \beta_i^{term}, \beta_i^{def}\right)$ as our explanatory variables in our cross-sectional regressions. For simplicity, we refer to these as Alpha (α_i), Market Beta $\left(\beta_i^{rmrf}\right)$, Value Beta $\left(\beta_i^{hml}\right)$, Size Beta $\left(\beta_i^{smb}\right)$, Duration Beta (β_i^{term}), and Credit Beta $\left(\beta_i^{def}\right)$.

Alternatives Funds

We ran rolling three-year regressions of a fund's return onto the region-appropriate RMRF factor and a dummy variable for inclusion in the commodity asset class.

The regression rolls monthly, providing a set of factor betas, alpha, and R-squared values each month estimated from the prior 36 months' experience. The alternatives asset class returns-based style-analysis regression took the form:

$$r_{i,t} = \alpha + \beta_i^{rmrf} RMRF_t + e_{i,t}$$

Subsequently, we used the estimated time series of factor betas above $(\alpha_i, \beta_i^{rmrf})$ as our explanatory variables in our cross-sectional regressions. For simplicity, we refer to these as Alpha (α_i) and Market Beta (β_i^{rmrf}).

To estimate a fund's beta to the factors above, we required 36 months of return history. Historically, there are a few examples of star-rated funds that do not have recorded net expense ratios or equivalent fee-related data. Rather than eliminate those funds from our sample, we chose to impute the broad asset-class median cross-sectionally by date onto these funds. No other imputation was performed or necessary. Finally, we winsorized all control variables at the 99% level cross-sectionally by date to mitigate the effect of potential outliers.

Regression Methodology

To evaluate the star ratings' efficacy, we employed a Fama-MacBeth regression of forward one-month fund returns on different independent variables for each asset class. A Fama-MacBeth regression is essentially a two-stage least-squares procedure. In the first stage, we estimated a fund's betas to various factors from one to 36 months ago (t-1 to t-36). Then, in the second stage, we used those betas as control variables in a cross-sectional regression at time t. We repeated the cross-sectional regressions each month, incorporating changes to the composition of the fund universe.

We ran each Fama-MacBeth regression separately by broad asset class to take into account different control variables. In the equity asset class, we controlled for fees, market exposure, and style tilts (for example, value/growth, large cap/small cap, and momentum). In the fixed-income asset class, we controlled for fees, credit exposure, and duration exposure. In the balanced asset class, we controlled for fees, equity market exposure, style tilts (for example, value/growth, large cap/small cap, and momentum), credit exposure, and duration exposure. In the alternatives asset class, we controlled for fees and equity market exposure and included a flag for commodity funds to capture the unique behavior of this sub-asset class.

We defined dummy variables for each of the star rating levels that take the value 1 when a fund is rated a specific level for month t and 0 otherwise. Star ratings can take on values of 1 star, 2 stars, 3 stars, 4 stars, and 5 stars. Because our universe consists entirely of star-rated funds, we had to choose one rating level to eliminate as an independent variable. We chose to leave out the 3-star level. Therefore, λ_t^{1-Star} can be interpreted as the average percent return of a 1-star-rated fund above/below a 3-star-rated fund after the control variables are taken into account. In the same way, other rating dummy variables can be interpreted as percent premiums above a typical 3-star-rated fund.

Morningstar typically publishes star ratings three business days after month-end. It could be argued, therefore, that any results obtained by using star ratings as of time t to predict returns at time t+1 will not represent an investable or actionable insight. We recognize this and sought to insulate our study from this criticism in addition to any potential look-ahead bias. Therefore, we used star ratings as of time t-1 to test the efficacy of the rating system. As a result, the rating information would have been available for nearly a month, allowing plenty of time for investors to act on this information.

For alternatives funds, we included a dummy variable in our regressions that represented a fund's assignment to the commodity asset class. The commodity dummy variable took the value 1 when a fund was assigned to the commodity asset class and 0 otherwise. Because commodity funds tend to be only loosely correlated to equity markets, a significant portion of the systematic risk associated with commodities markets would be unaccounted for without the inclusion of some control variable. We opted for a dummy variable in this case because it easily allows us to control for the average return in this asset class each month without needing to specify an appropriate global commodity benchmark.

One potential critique of our methodology is that it is too short-term in nature and does not reflect the long-term investor experience because we use one-month forward returns as opposed to longer periods. We do recognize this as a drawback but believe that the shorter horizon has some benefits. First, the timing is standard in academic papers using the Fama-MacBeth technique and essentially equates to the same assumptions as monthly rebalancing. As ratings change from one month to the next, our methodology captures and incorporates this information immediately. Furthermore, given the relatively short time frame for this analysis (12 years), we needed to find a balance in our choice of period between having a decent length horizon and having a decent number of nonoverlapping periods in order to draw some reliable statistical inferences. If we had more data, we could justify using a longer horizon.

Cross-Sectional Regression:

$$r_{i,\,t+1} = \gamma + \Omega_t Stars_{i,t-1} + \phi_t Z_{i,t} + \lambda_t Expense_{i,t} + \varepsilon_{i,t+1}$$

Where $r_{i,\,t+1}$ is defined as the return for fund i for time t+1, $Stars_i$ is a vector of star rating dummy variables for fund i at time t-1, Z_i is a vector of returns-based style analysis variables obtained from t-36 months ago to time t, and $Fees_i$ is the equivalent all-in expense for fund i at time t.

The vector of returns-based style analysis variables, Z_i, was estimated uniquely by fund from a regression on the prior 36 months of returns. The regression rolled monthly, providing a set of factor betas, alpha, and R-squared values each month estimated from the prior 36 months' experience. The factors included in the regression change depending on the broad asset class considered.

The equity asset class returns-based style-analysis regression took the form:

$$r_{i,t} = \alpha + \beta_i^{rmrf} RMRF_t + \beta_i^{hml} HML_t + \beta_i^{smb} SMB_t + \beta_i^{umd} UMD_t + e_{i,t}$$

The fixed-income and balanced asset class returns-based style-analysis regression took the form:

$$r_{i,t} = \alpha + \beta_i^{rmrf} RMRF_t + \beta_i^{hml} HML_t + \beta_i^{smb} SMB_t + \beta_i^{term} TERM_t + \beta_i^{def} DEF_t + e_{i,t}$$

The alternatives asset class returns-based style-analysis regression took the form:

$$r_{i,t} = \alpha + \beta_i^{rmrf} RMRF_t + e_{i,t}$$

The contents of the vectors -$Stars_i$, Z_i, X_i were as follows:

$\boldsymbol{Stars_i}$	$\boldsymbol{Z_i}$	$\boldsymbol{Expense_i}$
1 star (yes/no)	Market Beta	Net Expense Ratio equivalent
2 stars (yes/no)	Value Beta (not for alts)	
4 stars (yes/no)	Size Beta (not for alts)	
5 stars (yes/no)	Momentum Beta (equity only)	
	Credit Beta (fixed income and balanced only)	
	Term Beta (fixed income and balanced only)	
	Commodity (alts only - yes/no)	

Obtaining Final Estimates

Each month we ran cross-sectional regressions, as specified above. As a result, we were left with several vectors of coefficients on each date estimated from each model. For example, we had a matrix $\vec{B}$ that collected the time series of estimated coefficients from t=1 to t=T for each vector:

$$\vec{B} = \begin{bmatrix} \Omega_1 & \phi_1 & \lambda_1 \\ \vdots & \vdots & \vdots \\ \Omega_T & \phi_T & \lambda_T \end{bmatrix}$$

Then, the final estimates of the coefficient vectors Ω, ϕ, λ were averages across time:

$$\hat{\Omega} = \frac{1}{T}\sum_{t=1}^{T} \widehat{\Omega_t}$$
$$\hat{\phi} = \frac{1}{T}\sum_{t=1}^{T} \widehat{\phi_t}$$
$$\hat{\lambda} = \frac{1}{T}\sum_{t=1}^{T} \widehat{\lambda_t}$$

Standard errors were assumed to be uncorrelated over time in the standard way. Robust standard errors using the Shanken (1992) correction to account for the error-in-variables problem were also calculated in an earlier version of the paper but found to be similar. In the literature, error-in-variables is known to be a problem when using regressors that had themselves been estimated rather than known with certainty. Since we are using betas estimated from time-series regressions, our reported standard errors may suffer from this problem. However, correcting for this problem typically results in small changes to Fama-MacBeth standard errors in the literature we have read and examples we have seen. Regardless, reporting both versions of the standard errors has been identified as an improvement we can make in the next iteration of this paper. But the reader should be cautioned that reported standard errors may be slightly larger after accounting for the Shanken (1992) correction.

$$\sigma(\hat{\Omega}) = \frac{1}{T} var(\Omega_t) = \frac{1}{T^2} \sum_{t=1}^{T} \left(\widehat{\Omega_t} - \Omega\right)^2$$
$$\sigma(\hat{\phi}) = \frac{1}{T} var(\phi_t) = \frac{1}{T^2} \sum_{t=1}^{T} \left(\widehat{\phi_t} - \phi\right)^2$$
$$\sigma(\hat{\lambda}) = \frac{1}{T} var(\lambda_t) = \frac{1}{T^2} \sum_{t=1}^{T} \left(\widehat{\lambda_t} - \lambda\right)^2$$

Methodology for Event Study Analysis

To complement the Fama-MacBeth approach to evaluating the star ratings, we also conducted an event study analysis. While not as rigorous, we hoped that the event study approach would be more straightforward and closer to the typical investor experience. For the purposes of this analysis, we grouped funds by their one-month lagged star rating and looked at the average return by star rating cohort in the next one month, three months, six months, 12 months, 36 months, and 60 months. We reconstituted each rating group monthly in order to account for any funds that became obsolete. We also did this for the forward return calculation. The returns presented here are, therefore, equivalent to a monthly rebalanced portfolio of funds using the star ratings from one month ago, three months ago, six months ago, and so forth.

As in the regression analysis, investors would have one full month in which to respond to the star rating news. Like our previous analysis, we split our analysis up by global broad category group. Unlike the previous analysis, average portfolio returns were calculated using local fund returns and were not U.S. dollar-converted. Given the heterogeneity of each star rating bucket in terms of regional exposures and currency types, we believe the most comparable benchmark is the 3-star rating cohort. Therefore, while our analysis shows that 4- and 5-star funds tend to outperform 3-star funds, it does not necessarily follow that 4- and 5-star funds would outperform a chosen benchmark. The star rating is designed to sort funds on the basis of relative outperformance versus other funds, which is why we chose to test it in this manner.

Section 3: Regression Results

As described in Section 1, for each one-month period and each asset category (equity, bond, balanced, and alternatives funds) there exists a corresponding regression coefficient for the dummy variables associated with 1, 2, 4, and 5 stars (3 stars is the benchmark). We interpret such coefficients as performance premiums/discounts that are purely attributable to star ratings after controlling for other independent variables, like expenses and asset-class-specific variables (for example, market beta, size, style, momentum, and credit). We then examined the time series of star rating premiums to test the hypothesis that these coefficients are nonzero.

Exhibit 8 displays the average coefficients of the regression on Morningstar Ratings for each of the four asset class groups over time. In the fixed-income and allocation asset classes, the t-test results strongly support the hypothesis that the star ratings had significant forecasting capability after controlling for other factors, such as market beta, size, style, momentum, and credit. Among equity funds, we observe the correct directionality among the coefficients (for example, higher-rated funds have higher returns), but these results were not as significant or conclusive. For alternatives funds, the star rating had very little bearing on future performance as each of the rating cohorts did not appear to be statistically separable.

Exhibit 8 Aggregate Regression Results

Statistic	Asset Class	Category	One Star	Two Star	Four Star	Five Star	Market Beta	Value	Size	Momen tum	Expense	Intercept
Mean (%)	Equity	Aggregate	-0.04	-0.02	0.02	0.05	0.20	0.10	0.39	0.49	-0.04	0.56
SE (%)	Equity	Aggregate	0.03	0.01	0.01	0.03	0.23	0.16	0.19	0.54	0.08	0.25
t-stat	Equity	Aggregate	-1.18	-2.26**	1.76	1.75	0.86	0.61	1.95*	0.91	-0.46	2.25**

Statistic	Asset Class	Category	One Star	Two Star	Four Star	Five Star	Credit	Term	Expense	Intercept
Mean (%)	Fixed Income	Aggregate	-0.05	-0.02	0.02	0.04	1.31	-0.03	-0.01	0.23
SE (%)	Fixed Income	Aggregate	0.02	0.01	0.01	0.01	0.44	0.16	0.05	0.06
t-stat	Fixed Income	Aggregate	-2.51**	-2.92***	2.51**	3.02***	3.00***	-0.19	-0.01	3.75***

Source: Morningstar, Inc. * significant at $p<0.1$; ** significant at $p<0.05$; *** significant at $p<0.01$

Exhibit 8 Aggregate Regression Results (Continued)

Statistic	Asset Class	Category	One Star	Two Star	Four Star	Five Star	Market Beta	Value	Size	Credit	Term	Expense	Intercept
Mean (%)	Balanced	Aggregate	-0.06	-0.02	0.04	0.08	0.26	0.35	0.18	0.36	-0.38	-0.05	0.43
SE (%)	Balanced	Aggregate	0.03	0.01	0.01	0.03	0.18	0.20	0.23	0.30	0.23	0.05	0.16
t-stat	Balanced	Aggregate	-2.14**	-1.87*	3.18***	3.22***	1.39	1.78	0.79	1.21	-1.68	-0.90	2.61**

Statistic	Asset Class	Category	One Star	Two Star	Four Star	Five Star	Market Beta	Commo dity	Expense	Intercept
Mean (%)	Alternative	Aggregate	0.08	0.07	0.05	0.12	0.01	-0.45	0.14	0.17
SE (%)	Alternative	Aggregate	0.15	0.05	0.04	0.08	0.28	0.40	0.11	0.14
t-stat	Alternative	Aggregate	0.50	1.16	1.19	1.48	0.00	-1.06	1.15	1.17

Source: Morningstar, Inc. * significant at p<0.1; ** significant at p<0.05; *** significant at p<0.01

For the ease of interpreting the regression results, we annualize the star rating coefficients and display them in Exhibit 9. For instance, on average, the 5-star equity fund cohort outperformed the 3-star cohort by an annualized 0.55 percentage point, after adjusting for factor exposures. Similarly, 1-star equity funds trailed 3-star funds by 0.48 percentage point per year. Among fixed-income funds, the 5-star cohort outperformed 3-star funds by 0.47 percentage point per year, whereas 1-star funds lagged 3-star funds by 0.61 percentage point.

Exhibit 9 Annualized Star Rating Coefficients

Statistic	Asset Class	Category	One Star	Two Star	Four Star	Five Star
Mean (%)	Equity	Aggregate	-0.48	-0.29	0.23	0.55
Mean (%)	Fixed Income	Aggregate	-0.61	-0.28	0.20	0.47
Mean (%)	Balanced	Aggregate	-0.73	-0.28	0.44	1.02
Mean (%)	Alternative	Aggregate	0.96	0.82	0.57	1.45

Source: Morningstar, Inc.

The statistical significance of ratings' predictive power is also consistent with the difference between the distributions of the higher-rating premiums and those of the lower-rating premiums. To highlight the difference, Exhibit 10 shows the percentage of monthly star rating premiums above 0%. It is clear that a greater portion of the monthly 4- and 5-star rating premiums are above zero, whereas more 1- and 2-star rating premiums are below zero.

Exhibit 10 Percent of Positive Premiums by Asset Class and Morningstar Rating

Premia	Equity	Bond	Balanced	Alternative
1 star	49.7%	35.3%	35.9%	44.8%
2 star	38.2%	33.3%	38.5%	53.1%
4 star	58.6%	61.5%	62.8%	53.1%
5 star	55.4%	65.4%	66.0%	61.5%

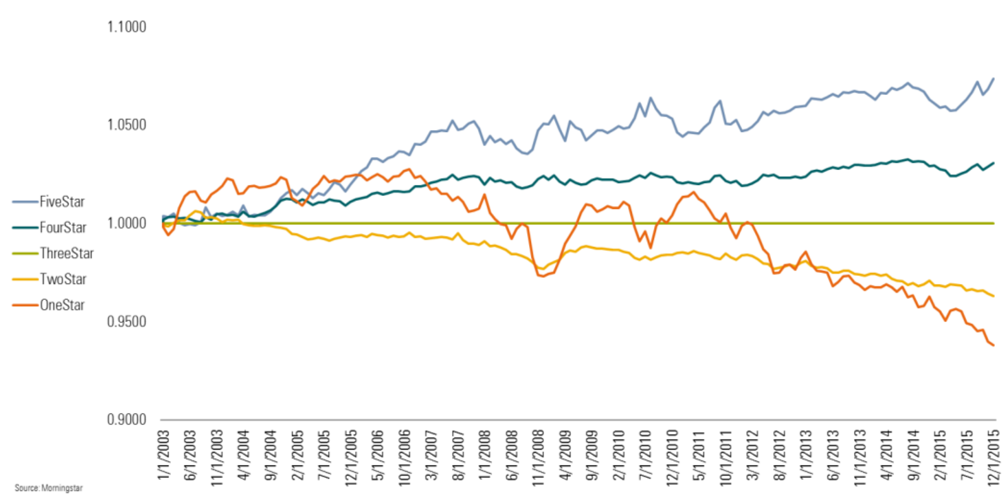
Source: Morningstar, Inc.

Cumulative Return of $1 in Premium Series

In this section, we examine the time series of the star rating premiums. We first display the growth of $1 invested in the premium series for each asset category over time. Then, we collect the calendar-year returns of the premium series.

Exhibit 11 shows the cumulative returns of $1 invested in the star rating premium series of equity funds over time. After controlling for common factors, the average returns of equity funds exhibit a strictly monotonic relationship with the star rating, with increases in star rating level directly associated with increases in average monthly returns. Higher-rated funds (4- and 5-star cohorts) delivered consistently strong outperformance against the 3-star cohort, whereas the 1-star cohort underperformed 3-star funds for the period.

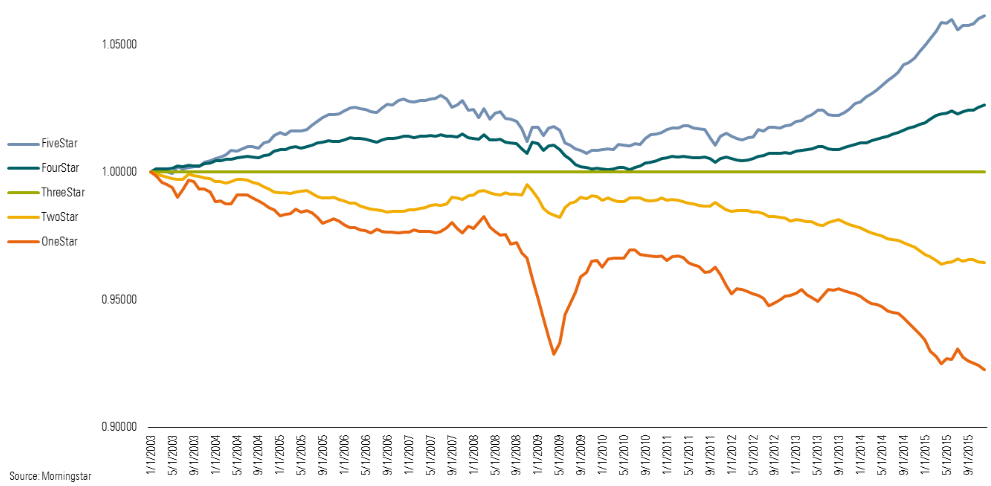
Exhibit 11 Aggregate Return Premiums by Morningstar Rating for Equity Funds



Source: Morningstar, Inc.

Exhibit 12 displays the time series results for fixed-income funds. Similar to equities, higher-rated funds tended to have higher average returns after accounting for their duration and credit exposures. During the entire period, 4- and 5-star cohorts posted strong performance, whereas the 1-star cohort underperformed all other cohorts.

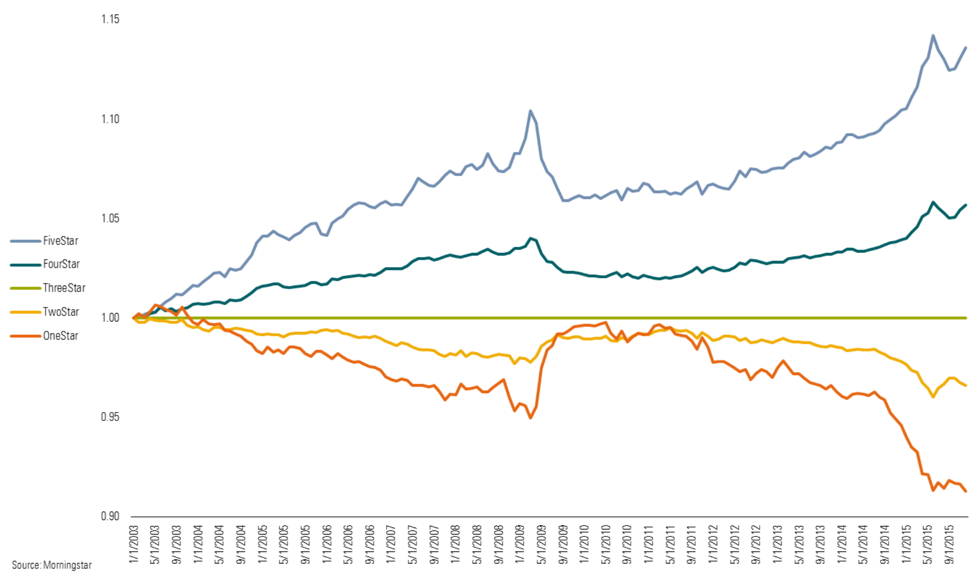
Exhibit 12 Aggregate Return Premiums by Morningstar Rating for Fixed-Income Funds



Source: Morningstar, Inc.

Exhibit 13 shows the familiar monotonic pattern between performance and star ratings for balanced funds, highlighting the dominance of 4- and 5-star cohorts over the other groups. This long-term trend experienced a short yet sharp reversal in the six months between March and September 2009, when U.S. equity markets staged the first rally after massive losses during the global financial crisis.

Exhibit 13 Aggregate Return Premiums by Morningstar Rating for Balanced Funds

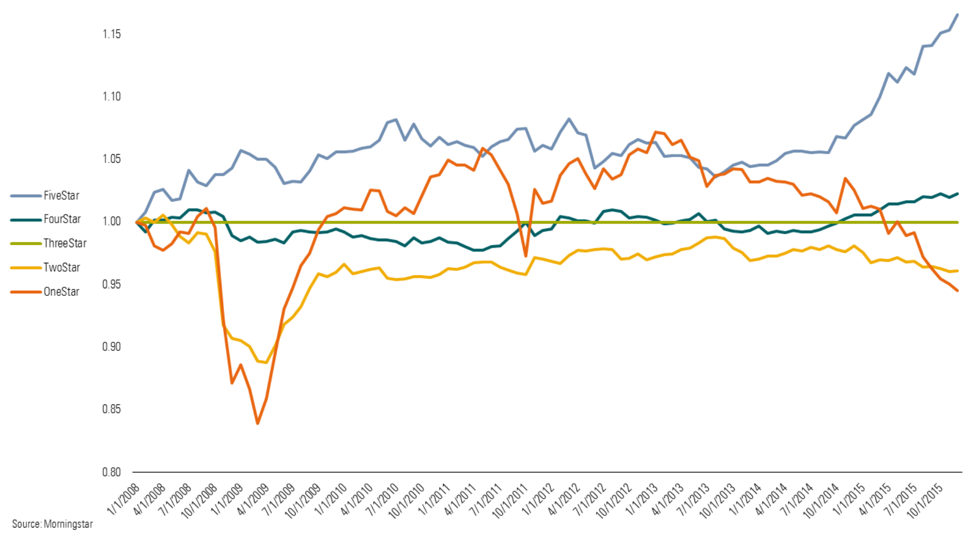


Source: Morningstar, Inc.

Exhibit 14 shows the premium series for alternatives funds since January 2008. After controlling for market beta and a dummy variable indicating commodity exposure, the premium returns do not exhibit the monotonic patterns that we saw in other asset groups. The 2-star cohort consistently underperformed the 3-star cohort, and the 5-star cohort exhibited clear dominance over 2-, 3-, and 4-star cohorts. However, 4-star funds delivered virtually no outperformance over 3-star funds, and the premium return of the 1-star cohort exhibited a whipsawed pattern and outperformed the 5-star cohort for several periods.

Overall, we do not draw a meaningful or convincing conclusion from Exhibit 14 that would suggest the star ratings have predictive power during this period for alternatives. Part of the problem could be that the model is improperly specified by leaving out potentially important return drivers. We would like to further analyze factor exposures of alternatives. Specifically, we want to investigate the fund-specific drivers of returns that are unique to alternatives relative to other asset classes. Currently, we are only crudely evaluating systematic risk in this asset class by controlling for equity market beta and the average return in commodities. It is likely that there are other ways to improve on this model, which we would like to explore in later updates of this paper. Further, the depth and breadth of alternatives fund data that we were able to analyze in this paper was quite small compared with the equity, fixed-income, and balanced asset classes. Therefore, it is not surprising that the results from this evaluation differed from those of other asset classes.

Exhibit 14 Aggregate Return Premiums by Morningstar Rating for Alternatives Funds



Source: Morningstar, Inc.

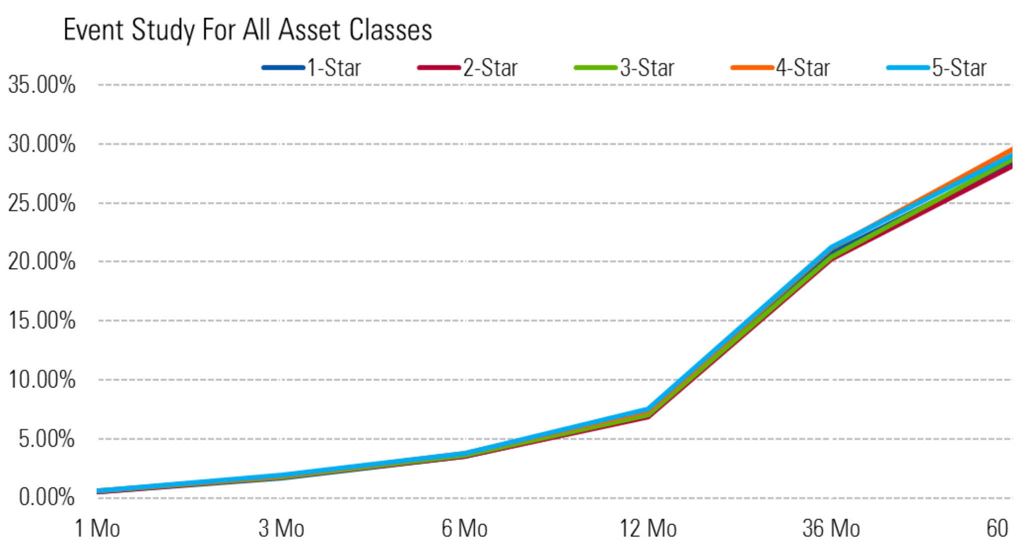
Section 4 – Event Study Results

In this section, we examine the event study grouped by Morningstar Rating. This analysis follows the method described on page 17.

Event Study for All Asset Classes

Exhibit 15 displays the performance of the event study for all asset classes. In Exhibit 16, we present a table of the cumulative returns for each star rating cohort. The returns of the 5- and 4-star rating portfolios have been higher than those of the other star rating cohorts, but not substantially so. Similar to the premium series, the event study showcases a consistent (though weak) monotonic pattern throughout all periods in the analysis, with the higher-rated portfolios delivering better average returns.

Exhibit 15 Event Study for All Asset Classes



Source: Morningstar, Inc.

Exhibit 16 Average Cumulative Return by Time Period – All Asset Classes

MonthsFromRating	1-Star	2-Star	3-Star	4-Star	5-Star
1 Mo	0.52%	0.54%	0.57%	0.60%	0.62%
3 Mo	1.70%	1.72%	1.79%	1.87%	1.92%
6 Mo	3.52%	3.50%	3.59%	3.73%	3.80%
12 Mo	6.90%	6.89%	7.03%	7.35%	7.51%
36 Mo	20.77%	20.23%	20.44%	21.14%	21.26%
60 Mo	28.56%	28.30%	28.87%	29.65%	29.22%

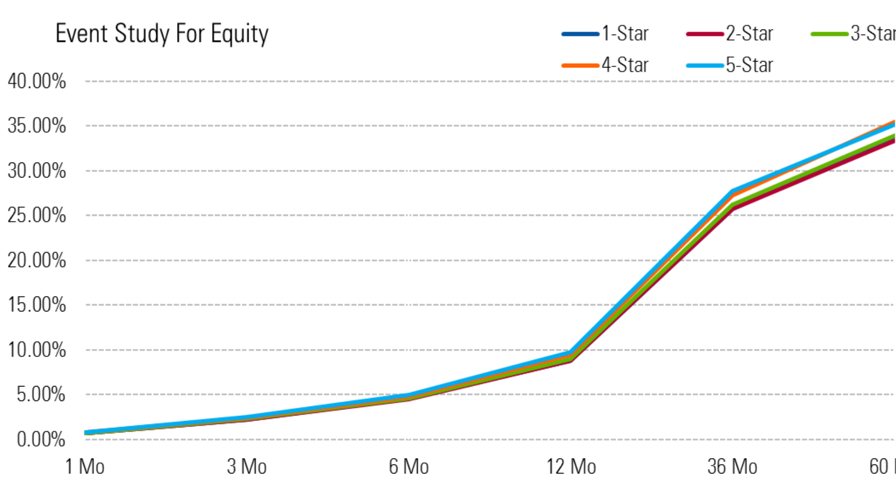
Source: Morningstar, Inc.

It is additionally 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. 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 or the need to make an additional investment decision to select a replacement fund in the event the investor's fund is slated for merger or liquidation. In this way, the star rating arguably confers the benefit of predicting survival.

Event Study for Equity

Exhibit 17 displays the performance of the event study for the equity asset class. In Exhibit 18, we present a table of the cumulative returns for each star rating cohort. The returns of the 5- and 4-star rating portfolios were higher than those of the other star rating cohorts, but not substantially so. Similar to the premium series, the event study showcases a consistent (though weak) monotonic pattern throughout all periods in the analysis, with the higher-rated portfolios having delivered better average returns.

Exhibit 17 Event Study for Equity



Source: Morningstar, Inc.

Exhibit 18 Average Cumulative Return by Period – Equity

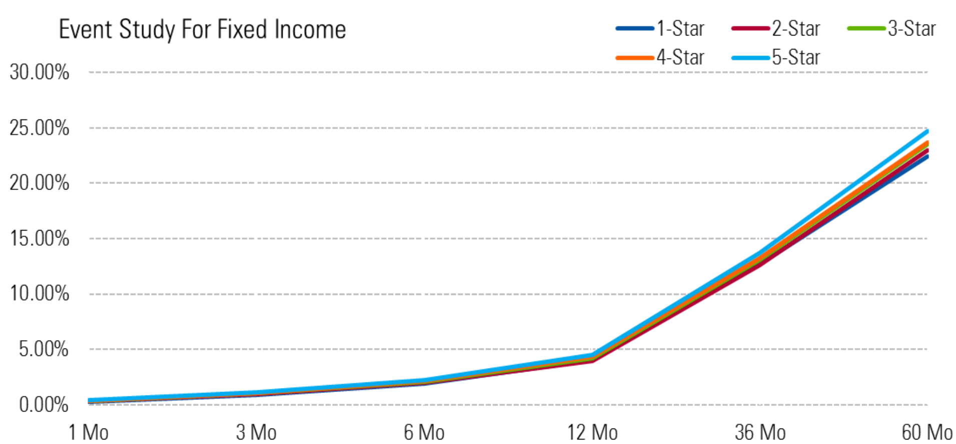
MonthsFromRating	1-Star	2-Star	3-Star	4-Star	5-Star
1 Mo	0.66%	0.68%	0.70%	0.74%	0.78%
3 Mo	2.19%	2.20%	2.27%	2.37%	2.46%
6 Mo	4.55%	4.52%	4.61%	4.79%	4.93%
12 Mo	8.78%	8.82%	9.02%	9.44%	9.72%
36 Mo	25.78%	25.74%	26.18%	27.28%	27.74%
60 Mo	33.86%	33.36%	33.92%	35.43%	35.17%

Source: Morningstar, Inc.

Event Study for Fixed Income

Exhibit 19 displays the performance of the event study for the fixed-income asset class. In Exhibit 20, we present a table of the cumulative returns for each star rating cohort. The returns of the 5- and 4-star rating portfolios have been higher than those of the other star rating cohorts, but not substantially so. Similar to the premium series, the event study showcases a consistent (though weak) monotonic pattern throughout all periods in the analysis, with the higher-rated portfolios delivering better average returns.

Exhibit 19 Event Study for Fixed Income



Source: Morningstar, Inc.

Exhibit 20 Average Cumulative Return by Period – Fixed Income

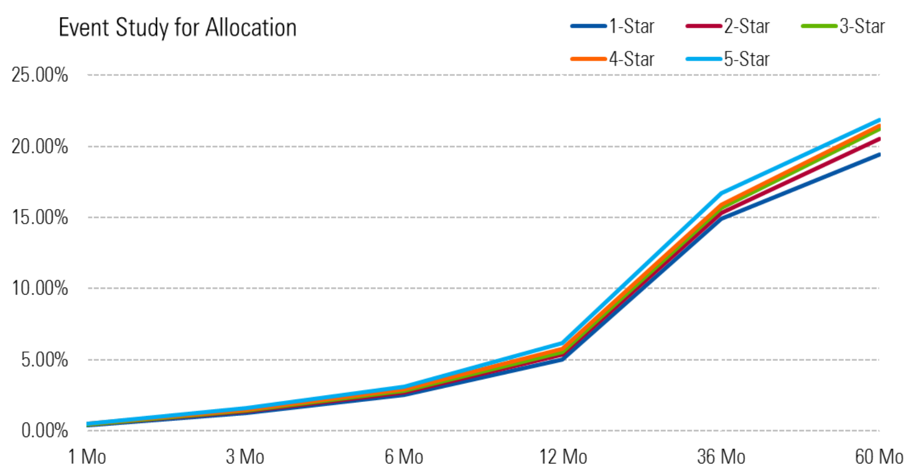
MonthsFromRating	1-Star	2-Star	3-Star	4-Star	5-Star
1 Mo	0.28%	0.32%	0.35%	0.36%	0.39%
3 Mo	0.90%	0.96%	1.04%	1.08%	1.14%
6 Mo	1.93%	1.95%	2.07%	2.13%	2.21%
12 Mo	4.01%	3.97%	4.18%	4.32%	4.47%
36 Mo	12.66%	12.63%	13.07%	13.24%	13.71%
60 Mo	22.39%	22.95%	23.52%	23.72%	24.71%

Source: Morningstar, Inc.

Event Study for Allocation

Exhibit 21 displays the performance of the event study for the allocation asset class. In Exhibit 22, we present a table of the cumulative returns for each star rating cohort. The returns of the 5- and 4-star rating portfolios have been higher than those of the other star rating cohorts, but not substantially so. Similar to the premium series, the event study showcases a consistent (though weak) monotonic pattern throughout all time periods in the analysis, with the higher-rated portfolios delivering better average returns.

Exhibit 21 Event Study for Allocation



Source: Morningstar, Inc.

Exhibit 22 Average Cumulative Return by Time Period – Allocation

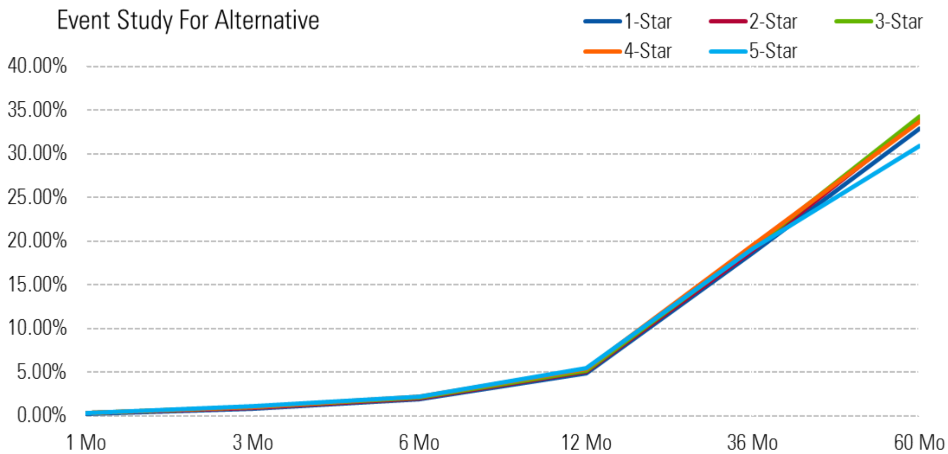
MonthsFromRating	1-Star	2-Star	3-Star	4-Star	5-Star
1 Mo	0.38%	0.43%	0.45%	0.48%	0.52%
3 Mo	1.24%	1.35%	1.41%	1.48%	1.59%
6 Mo	2.54%	2.71%	2.81%	2.93%	3.12%
12 Mo	5.03%	5.35%	5.53%	5.77%	6.20%
36 Mo	14.89%	15.32%	15.68%	15.88%	16.68%
60 Mo	19.44%	20.54%	21.22%	21.44%	21.88%

Source: Morningstar, Inc.

Event Study for Alternatives

Exhibit 23 displays the performance of the event study for the alternatives asset class. In Exhibit 24, we present a table of the cumulative returns for each star rating cohort. For horizons less than three years, the higher-rated portfolios tend to perform better. However, the return of the 1-star rating funds has been higher than that of the other star rating cohorts for the five-year horizon.

Exhibit 23 Event Study for Alternatives



Source: Morningstar, Inc.

Exhibit 24 Average Cumulative Return by Time Period – Alternatives

MonthsFromRating	1-Star	2-Star	3-Star	4-Star	5-Star
1 Mo	0.23%	0.28%	0.30%	0.32%	0.34%
3 Mo	0.81%	0.92%	0.97%	1.01%	1.06%
6 Mo	1.92%	2.02%	2.09%	2.14%	2.20%
12 Mo	4.84%	5.04%	5.17%	5.32%	5.47%
36 Mo	18.64%	18.88%	19.24%	19.47%	19.15%
60 Mo	32.84%	33.93%	34.23%	33.70%	30.85%

Source: Morningstar, Inc.

Appendix 1: Net Expense Ratio Equivalent Global Calculation

Different regions have different reporting requirements for mutual fund expenses. For example, in the United States, net expense ratio is the most commonly used data point that encompasses all fees levied on the investor during the past year, including performance-based fees. In the United Kingdom and Europe, ongoing charge is the most commonly used data point to express fees levied on investors in the past year. Unlike net expense ratio, ongoing charge does not include performance-based fees. Therefore, to harmonize net expense ratios of U.S., U.K., and Europe-domiciled funds, we added performance fees back into ongoing charge.

For funds of funds, we also included acquired fund expenses.

For all domiciles in our purview, we do our best to harmonize fee-reporting differences across geographies using the following mapping procedure.

Annual Report Net Expense Ratio Equivalent

To get an annual report net expense ratio equivalent data point, we combine various annual report net expense ratio equivalent data points.

$$NetExpenseRatio = \begin{cases} NetExpenseRatio & \text{Domicile = USA} \\ IndirectCostRatio(orMER) & \text{Domicile = AUS} \\ ManagementExpenseRatio & \text{Domicile = CAN or NZL} \\ OngoingCharge + PerformanceFee(orNER) & \text{Region = UK, EU} \\ JPAf - TaxTotalExpenseRatio & \text{Domicile = JAP} \\ FoF.NetExp_i = FoF.exp_i + AcquiredFundExpense & \text{FoF = Yes, Acq Fund Exp} \neq \text{NA} \\ FoF.NetExp_i = FoF.exp_i + \sum_{i=1}^{N} w_i exp_i & \text{FoF = Yes and Region = US} \\ NetExpenseRatio & \text{Otherwise} \end{cases}$$

Appendix 2: Literature Review

The Morningstar Rating for funds is a five-star ranking system based on past performance of mutual funds; the calculation is risk-adjusted and accounts for fees. The purpose of this paper is to examine the predictive power of this ranking system. Generally speaking, strong predictive power purports that past performance of mutual funds, either good or bad as indicated by ratings, will persist into the future.

There is no shortage of academic studies on the general topic of mutual fund performance, with some dating back to the 1960s. The common framework of these studies has been to first separate out performance attributable to common factors and assess either the residual (alpha) or the coefficients associated with regressors are nonzero. Statistical tests between in-sample and out-of-sample performance have been applied frequently to evaluate the future success of past winners. Through the years, the research approach has evolved from a single-factor model, that is, the capital asset pricing model in Sharpe (1966) and Treynor (1965), to multifactor models. The well-known Fama-French three-factor model was found in Fama & French (1992). Carhart (1997) built on this framework by adding a momentum factor first revealed by Jegadeesh & Titman (1993). Although their paper sought to explain returns of stocks instead of mutual funds, Chen, Roll, & Ross (1986) did find two other economic factors—term and credit—that are significant in explaining asset prices. Fund expenses are often found to be significant in explaining mutual fund performance by Sharpe (1966), Carhart (1997), Kinnel (2010), and Fama & French (2010). Our framework includes all of these factors. We adopt the regression approach from Fama & MacBeth (1973) to easily calculate standard errors that correct for correlation across assets. Furthermore, this method makes it easy for us to construct models with independent variables as time series.

Specifically, a number of academic papers have studied the persistence of mutual fund performance. Elton, Gruber, & Blake (1996) argued that survivorship bias was present in many of the studies before the 1970s, for example, Sharpe (1966). However, researchers have since implemented measures to adjust for this bias. Our sample universe is free of survivorship bias as we include both active and obsolete funds and their complete performance history.

More recent studies offer contradicting opinions around whether superior fund performance can persist and, if so, over what time frame. By analyzing quarterly returns, Hendricks, Patel, & Zeckhauser (1993) found evidence of short-term persistence of mutual fund performance, referred to as the “hot hands” effect, during the following one or two years. Grinblatt & Titman (1992) identified positive persistence in performance during a five-year span or longer. Elton et al. (1996) confirmed persistence in risk-adjusted performance in both the short and long run. Goetzmann & Brown (1995) revealed that relative risk-adjusted performance persists among funds lagging benchmark returns and suggested that the persistence effect is attributable to investment strategies not captured by traditional categories that are based on style and size. Building on their earlier work in Goetzmann & Ibbotson (1994), Ibbotson & Patel (2002) found that past performance and rankings do predict future returns, even after adjusting for fund styles. Elton, Gruber, & Blake (2011) used Morningstar's holdings data and found that alphas estimated by holdings-based betas have stronger predictive power of future alphas than beta estimated from time series regression on common factors.

On the other side, Carhart (1997) found that the one-year momentum effect of fund performance can be largely explained by common factors and costs and that longer-term persistence is not statistically robust. Fama & French (2010) used bootstrap simulations in conjunction with multifactor models to infer that few active managers possess skills sufficient enough to cover their investment costs after controlling for factors and that the positive effect of the few skilled managers on the aggregate result is overshadowed by the bad performance of the managers without skill.

Most of the highly cited studies mentioned above focus on funds domiciled in the U.S. for reasons such as data availability, limited effect of survivorship bias, and standardized and regulated reporting practices. This paper covers a global data set.

Finally, there are a few past studies conducted by Morningstar and other researchers that focused specifically on the predictability of the Morningstar Rating system. Blake & Morey (2000) concluded that the predictabilities of ratings are more pronounced for lower-rated funds and found weak statistical evidence of outperformance of 5-star funds over 4- and 3-star funds. Kinnel (2010) shared the conclusion that the 1-star rating predicts bad performance and the likely consequence of mergers or liquidation. He also found that expense ratios are the most powerful predictor of performance. Rekenthaler (2014) documented the phenomena that top-performing funds are noticeably more likely to stay better-rated in the future. Our paper revisits this topic with a more rigorous approach and highlights a number of findings that would be of interest for scholars and practitioners in this debate.

References

Blake, C.R., & Morey, M.R. 2000. "Morningstar Ratings and Mutual Fund Performance." J. *Financial and Quantitative Analysis*, Vol. 35, No. 3, P. 451.

Brown, S.J., & Goetzmann, W.N.1995. "Performance Persistence." *J. Finance*, Vol. 50, P. 679.

Carhart, M.M. 1997. "On Persistence in Mutual Fund Performance." *J. Finance*, Vol. 52, No. 1, P. 57.

Chen, N., Roll, R., & Ross, S.A. 1986. "Economic Forces and the Stock Market." *J. Business*, Vol. 59, P. 383.

Elton, E.J., Gruber, M.J., & Blake, C.R. 1996. "Survivor Bias and Mutual Fund Performance." *Rev. Financial Studies*, Vol. 9, No. 4, P. 1097.

Elton, E.J., Gruber, M.J., & Blake, C.R. 2011. "Holdings Data, Security Returns, and the Selection of Superior Mutual Funds." *J. Financial and Quantitative Analysis*, Vol. 46, No. 2, P. 341.

Elton, E.J., Gruber, M.J., Das, S., & Blake, C.R. 1996. "The Persistence of Risk-Adjusted Mutual Fund Performance." *J. Business*, Vol. 69, No. 2, P. 133.

Fama, E.F., & French, K.R. 1992. "The Cross-Section of Expected Stock Returns." *J. Finance*, Vol. 47, No. 2, P. 427.

Fama, E.F., & French, K.R. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *J. Financial Economics*, Vol. 33, No. 1, P. 3.

Fama, E.F., & French, K.R. 2010. "Luck Versus Skill in the Cross-Section of Mutual Fund Returns." *J. Finance*, Vol. 65, No. 5, P.1915.

Fama, E.F., & MacBeth, J.D. 1973. "Risk, Return, and Equilibrium: Empirical Tests." J. Political Economy, Vol. 81, No. 3, P. 607.

Goetzmann, W.N., & Ibbotson, R.G. 1994. "Do Winners Repeat?"
J. Portfolio Management, Vol. 20, No. 2, P. 9.

Grinblatt, M., & Titman, S. 1992. "The Persistence of Mutual Fund Performance."
J. Finance, Vol. 47, No. 5, P. 1977.

Hendricks, D., Patel, J., & Zeckhauser, R. 1993. "Hot Hands in Mutual Funds: Short-Run Persistence of Performance 1974-1988." *J. Finance*, Vol 48, No. 1, P. 93.

Ibbotson, R.G., & Patel, A.K. 2002. "Do Winners Repeat With Style? Summary of Findings."
Ibbotson Associates.

Jegadeesh, N., & Titman, S. 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *J. Finance*, Vol. 48, No. 1, P. 65.

Kinnel, R. 2010. "How Expense Ratios and Star Ratings Predict Success,"
Morningstar, Aug. 9, 2010. http://news.morningstar.com/articlenet/article.aspx?id=347327

Rekenthaler, J. 2014. "The Fund Winner Curse Is an Optical Illusion." Morningstar, Oct. 8, 2014.
http://news.morningstar.com/articlenet/article.aspx?id=667868

Shanken, J., 1992. On the estimation of beta-pricing models. Review of Financial Studies 5. 1-33.

Sharpe, W. 1966. "Mutual Fund Performance." *J. Business*, Vol. 39, P. 119.

Treynor, J.L. 1965. "How to Rate the Performance of Mutual Funds." *Harvard Business Rev.*,
Vol. 43, P. 63.

MORNINGSTAR®

22 West Washington Street
Chicago, IL 60602 USA