Examining the Morningstar Quantitative Rating for Funds

A new investment research tool.

Morningstar Quantitative Research
27 August 2018

Contents

Authors
Madison Sargis
Associate Director of Quantitative Research
+1 312 244-7352
madison.sargis@morningstar.com

Timothy Strauts
Director of Quantitative Research
+1 312 384-3994
timothy.strauts@morningstar.com

Nicholas Heise
Quantitative Research Intern

Katherine Olexa
Quantitative Research Intern

Executive Summary

Morningstar launched the Morningstar Quantitative Rating for funds to enlarge the universe of funds under coverage based on the research and data used by our analysts in their decision-making process for the Morningstar Analyst Rating for funds. With this new quantitative approach, we provide a forward-looking assessment for nearly every fund in the U.S. market through the combination of the two rating systems.

By analyzing the ratings—Quantitative or Analyst—for all funds, we find new insights into the Morningstar Category system, fund companies, and fund performance. Our key takeaways include:

- The Morningstar Quantitative Rating works, picking clear over- and underperforming funds, even in the absence of performance history.
- The Quantitative Rating system recommends higher ratings for passive strategies among sector funds and higher ratings for active strategies among illiquid bond funds.
- With the Analyst and Quantitative ratings, we now provide a comprehensive forward-looking set of ratings for a firm's entire lineup.
- By comparing a firm's asset-weighted and equal-weighted ratings, more-detailed insights can be obtained about the set of funds offered by a firm.

Introduction

Last June, Morningstar released the Morningstar Quantitative Ratings in Morningstar Direct℠ for the U.S. market. The latest rating system is based on the research and data used by our analysts and provides a forward-looking assessment for nearly every fund. Investors can now generate new insights across the entire fund industry and examine all funds globally through the same Morningstar lens.

Since the initial launch, we have rolled out the ratings to the EMEA and South Korea markets and have written about the machine-learning techniques driving the rating. Now, we figured it was time to revisit and evaluate the original ratings.

In the next section, we provide an abbreviated explanation for the Morningstar Quantitative Rating for funds methodology. Afterward, we provide five event studies on the Quantitative Rating in its first year of existence. For the purposes of the paper, the analysis includes all funds and exchange-traded funds in the U.S. that get an Analyst Rating or a Quantitative Rating. The only categories that are excluded are those for which Morningstar does not apply a Morningstar Rating™, or "star" rating, such as bear market and leveraged ETFs). Finally, we conclude with a few remarks on Morningstar ratings.

Abbreviated Methodology

The complete methodology is found in the Morningstar Quantitative Rating for funds Methodology document, which we link to in the References section. Below, we provide the overall structure of the algorithm but omit nuanced details for the sake of being concise. Our goal here is to arm readers with the bare bones of the methodology to understand origination of the Quantitative Rating for funds.

Pillar Rating Methodology

The five pillar ratings represent the foundation of the Analyst Rating. For the Quantitative Rating, the pillar ratings were estimated using a series of random forest models and rated on a scale of Positive, Neutral, and Negative.

Each pillar rating is estimated using a combination of two random forest models. First, a model is estimated that seeks to distinguish funds based on whether that fund's pillar rating would be rated Positive. Second, a different model is estimated that seeks to distinguish funds based on whether that fund's pillar rating would be rated Negative. Each model puts out probability scores that the fund would be Positive or Negative, respectfully. By combining these two probabilities, a more robust estimator is achieved.

The output for these pillar ratings will be on a scale of 0 to 1. The closer to 1 a fund's estimated pillar rating is, the more likely that the true pillar rating is Positive. Similarly, the closer to 0 a fund's estimated pillar rating is, the more likely that the true pillar rating is Negative.

The intuition underlying this method is subtle, yet important. First, the weighted summation captures information about a fund along two dimensions—the likelihood that a fund's pillar is Positive and the likelihood that a fund is not Negative. In practice, this has the result of classifying many Neutral pillars as decidedly not Positive and not Negative.

Calculating the Quantitative Rating

The final step in the Quantitative Rating involves predicting an overall rating on the scale of Negative, Neutral, Bronze, Silver, or Gold from our estimated pillar ratings. The Quantitative Rating applies the same set of pillar weightings to all uncovered funds. These weightings are determined empirically by examining the average pillar importance across Analyst Rating decisions historically through a multivariate linear regression. In this way, the Quantitative Rating pillar weightings represent the typical set of weightings used by the Analyst Rating. We use fixed percentile thresholds for final rating assignment. Exhibit 1 showcases these distribution breakpoints.

Exhibit 1 Rating Distribution Breakpoints

Morningstar Quantitative Rating™ for Funds	Breakpoints
Gold Q	> 95%
Silver Q	85%–95%
Bronze Q	70%–85%
Neutral Q	15%–70%
Negative Q	0%–15%

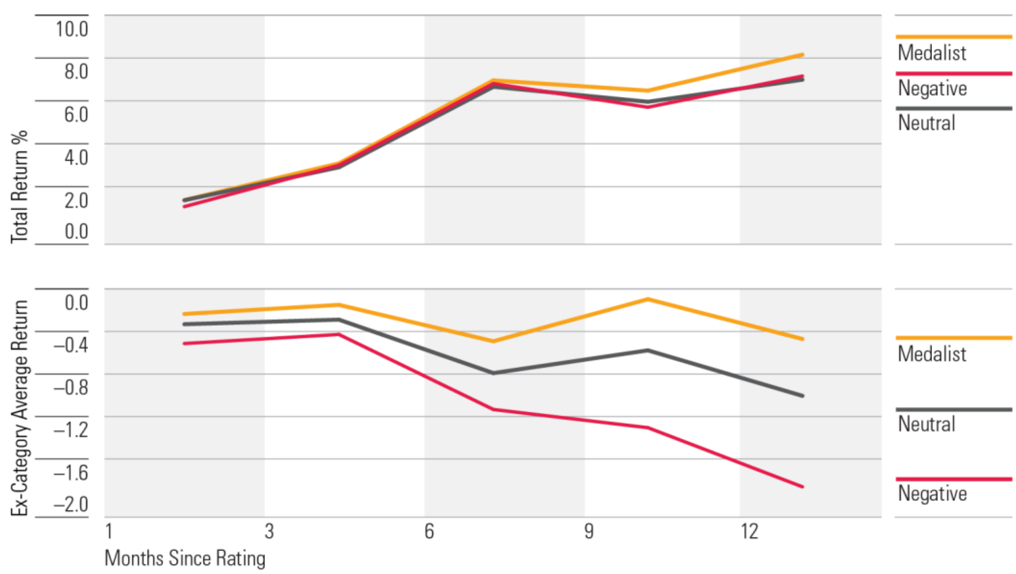
Source: Morningstar, Inc.

Key Takeaways

How Has the Morningstar Quantitative Rating Performed?

The data show that the Morningstar Quantitative Rating works. In Exhibit 2, we show the performance of the June 2017 ratings. Note that because of the small sample size of the recommended ratings, we combined the Gold, Silver, and Bronze ratings categories into the Medalist group. This group of recommended ratings clearly outperformed the group rated as Neutral and Negative, showing 12-month cumulative returns of 8.38% compared with returns over the same period from the Negative group of 7.44%. When we adjust for style and risk, we see an even more pronounced separation between funds rated as Negative, Neutral, or as a Medalist.

Exhibit 2 Morningstar Quantitative Rating Event Study of the Initial Ratings



Source: Morningstar, Inc. Data as of May 31, 2018

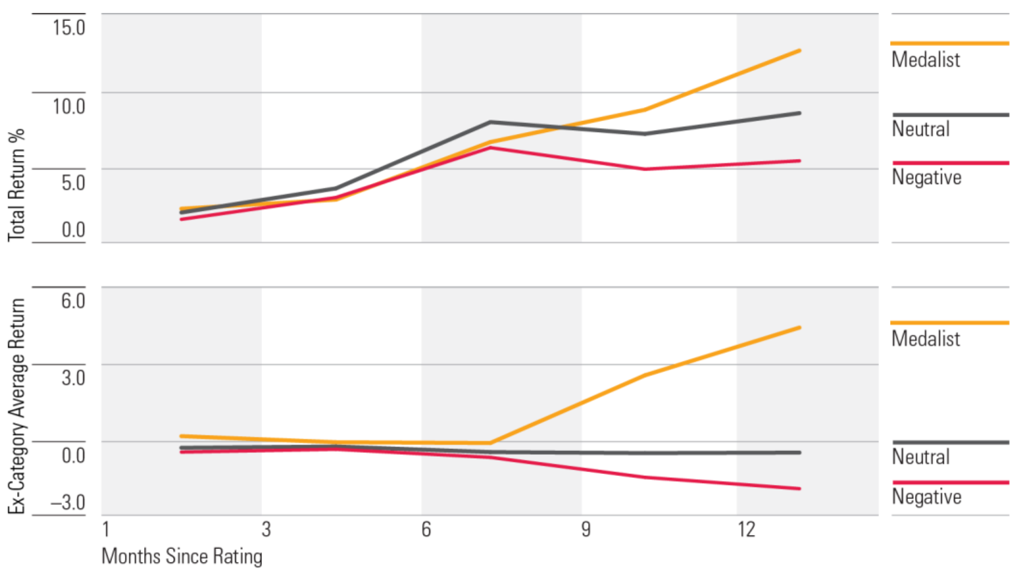
How Does This Help Investors?

Morningstar now provides a forward-looking assessment on every fund after just one month of history. Investors no longer need to wait for a fund to rack up a three-year track record to earn a star rating or accumulate enough assets to warrant attention. We provide an objective measure almost immediately upon launch.

To look at the success of the measure, we examine the performance of the initial June 2017 ratings on funds less than one year old. Exhibit 3 shows that even with essentially no performance information, the Quantitative Rating is still able to select above- and below-average funds.

In the absence of performance history, the ratings are driven almost entirely by the Price and Parent Pillars. This tends to skew the ratings toward Neutral. However, there continues to be a small subset of funds receiving Medalist or Negative ratings that prove to over or underperform, respectively. Obviously, the Quantitative Rating will give better recommendations when it has more data, but it is encouraging to see that the system still works well on average even in the absence of performance information.

Exhibit 3 Morningstar Quantitative Rating Event Study for New Funds



Source: Morningstar, Inc. Data as of May 31, 2018

In the year since the Morningstar Quantitative Rating launched, only 5.6% of share classes of new U.S. funds have been awarded Morningstar Medalist ratings. Bronze is the most frequent rating, making up approximately three fourths of the recommended ratings. To date, we have never issued a Morningstar Quantitative Rating of Gold to a new fund in the United States. Conversely, 17.0% of new funds are rated Negative after launch. This make sense, as the main differentiating factors for new funds within the quantitative rating system are the Parent and Price Pillars. If an expensive fund launched from a poorly rated fund company, the fund is likely to be rated Negative.

Category Ratings by Active and Passive Funds

At the Morningstar Category level, we can examine the difference between active and passive funds. The difference between the average active fund rating and the average passive fund rating in each category is shown in Exhibit 4. A value of 1.0 in this table would indicate that the average active fund in that category is one rating level above the average passive fund. This helps to demonstrate in which categories it is best to go active and in which to go passive. These differences between active and passive strategies are mostly driven by the Price and Performance Pillars because the People, Process, and Parent Pillars should cancel each other out between funds.

Exhibit 4 Average Active and Passive Rating by Category

Best Active		Best Passive	
Bank Loan	1.09	Consumer Defensive	–0.72
China Region	0.72	Consumer Cyclical	–0.65
World Small/Mid Stock	0.64	Communications	–0.63
Tactical Allocation	0.61	Latin America Stock	–0.55
Preferred Stock	0.54	Short Government	–0.54
High-Yield Bond	0.52	Equity Energy	–0.53
Multicurrency	0.46	Allocation 50%–70% Equity	–0.51
Diversified Pacific/Asia	0.41	Allocation 30%–50% Equity	–0.51
World Allocation	0.39	Inflation–Protected Bond	–0.45
Foreign Large Growth	0.38	Natural Resources	–0.33

Source: Morningstar, Inc. Data as of May 31, 2018

Bank loan, high-yield bond, and preferred stock make the list of top active categories. This shouldn't be surprising because these categories contain fixed-income securities that are generally illiquid, giving active managers an opportunity to beat an index. Several of the spots on the top active list are occupied by international funds, aligning with the commonly held view that active management is a better bet overseas because of the large amount of securities and unique political circumstances in each country.

The sector-specific categories of consumer defensive, communications, consumer cyclical, and natural resources all fall in the top 10 best passive list. Active managers do not have the same leg up when they are constrained to investing in securities of a particular sector. Much of the money in sector funds is earned with passive strategies.

Two fixed-income categories appear on the passive list: short government and inflation-protected bond. Funds in these categories typically have low returns, causing the higher fees of active management to have a greater drag effect, making it much more difficult to outperform. Additionally, the low-volatility nature of these categories will make it harder for any active manager to differentiate their performance.

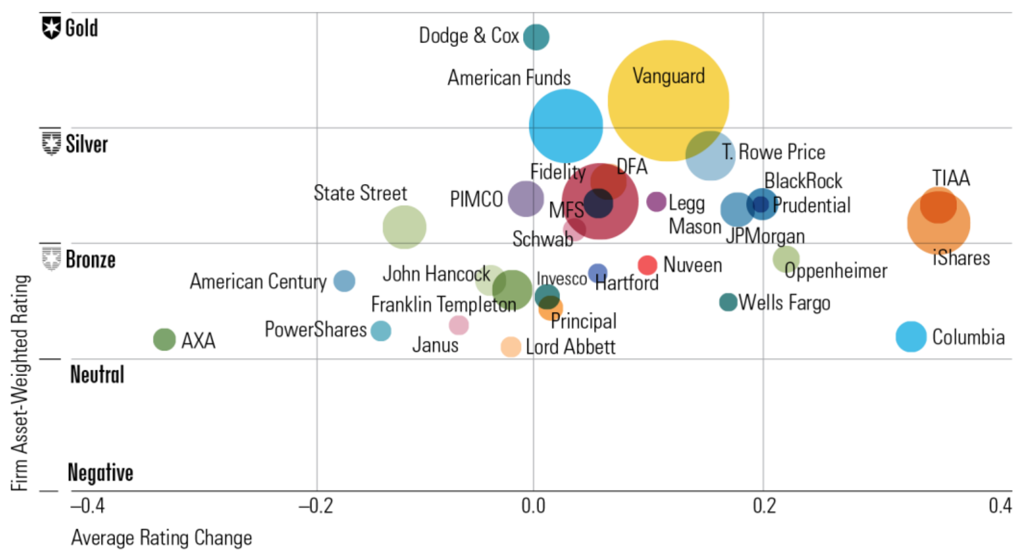
A Rating System for the Entire Firm

In Exhibit 5, we display the current asset-weighted average along the y-axis and one-year change in rating along the x-axis. Fund companies are ranked from 5 to 1, with each level representing Gold, Silver, Bronze, Neutral, and Negative, respectively. This allows us to see both where each firm's ranking sits today and which direction it has moved over the past year. In addition, each firm's bubble represents its size with respect to assets under management.

Ideally, a firm has an average Gold rating and 0 change. This is right where we find Dodge & Cox. It has maintained five Gold-rated funds and one Bronze-rated fund for the past 12 months. Note, its one Bronze-rated fund makes up a small portion of its assets.

The worst spot is a Negative change in rating and low average rating. Of the largest 30 firms, this is where we find AXA. AXA suffered the largest ratings change for the trailing 12 months, decreasing by 0.3 points. In the past year, its U.S. fund line up is entirely covered by the quantitative system and 71 of its share classes were downgraded.

Exhibit 5 One-Year Change in Morningstar Analyst Rating and Morningstar Quantitative Rating



Source: Morningstar, Inc. Data as of May 31, 2018

Ratings: Equal-Weighted vs. Asset-Weighted Rating

We look deeper into the relationship between a firm's rating and its funds in Exhibit 6 by comparing the firm's equal-weighted rating to its asset-weighted rating. Doing so helps us answer two questions: For the typical fund launched by the firm, what is the average rating? For the typical dollar invested in the firm, what's the average rating of the fund it is invested in?

Typically, we'd expect assets to concentrated in higher-rated funds, so a firm's asset-weighted rating would be higher than its equal-weighted rating. This is what we find for a firm like Fidelity. The average rating of its assets is 3.5—halfway between a Bronze and Silver rating—while its average fund rating is 3.1, or approximately Bronze-rated.

Out of the largest fund companies, we find PIMCO and BlackRock with some of the largest dispersions between their asset ratings and their fund ratings. The average dollar is invested in a fund rated 1.09 and 0.99 notches, respectively, higher than the average fund issued by each firm. This suggests that the firm has many smaller, lower-rated funds. Among BlackRock funds, only 1.8% are rated Negative, but 59.5% are classified as Neutral. For PIMCO, just 2.5% of offered funds are given Negative ratings, but 57.8% are Neutral.

Exhibit 6 Equal-Weighted and Asset-Weighted Ratings by Firm

Firm	Asset-Weighted	Equal-Weighted	Difference
PIMCO	3.61	2.52	1.09
Prudential	3.49	2.41	1.08
BlackRock	3.52	2.53	0.99
State Street	3.28	2.33	0.95
Legg Mason	3.52	2.62	0.90
Oppenheimer	2.78	1.94	0.83
DFA	3.72	2.97	0.75
MFS	3.60	2.93	0.67
Vanguard	4.35	3.71	0.65
Schwab	3.17	2.53	0.64
Invesco	2.65	2.08	0.57
JPMorgan	3.49	2.94	0.55
iShares	3.33	2.78	0.55
Hartford Investments	2.85	2.32	0.53
Fidelity	3.51	3.02	0.49
American Funds	4.19	3.70	0.49
Franklin Templeton	2.78	2.34	0.44
TIAA	3.47	3.06	0.41
Nuveen	2.87	2.51	0.36
T. Rowe Price	3.77	3.41	0.35
Wells Fargo	2.60	2.26	0.34
Dodge & Cox	5.00	4.67	0.33
Columbia	2.33	2.01	0.32
Principal	2.57	2.25	0.31
John Hancock	2.84	2.61	0.23
American Century	2.71	2.49	0.23
Lord Abbett	2.31	2.19	0.11
Janus	2.53	2.55	-0.02
PowerShares	2.34	2.38	-0.04
AXA	2.14	2.36	-0.22

Source: Morningstar, Inc. Data as of May 31, 2018

Conclusion

By combining the Morningstar Analyst Rating for funds with the Morningstar Quantitative Rating for funds, Morningstar now provides the investor with a complete suite of forward-looking ratings based on analyst-driven research. Investors can generate new insights across the entire fund industry and dissect all funds globally through the same Morningstar lens.

We will continue to monitor funds through these two ratings systems and will provide updated insights as we see fit. We expect that, over time, we will enhance the Quantitative Rating to improve performance. We will note methodological changes in this document as they are made.

References

Morningstar Analyst Rating for Funds Methodology Document. 2011.
http://hkbeta.morningstar.com/Productdata/Methodology/analyst_rating_methodology.pdf

Morningstar Quantitative Ratings for Funds Input Data Methodology. 2017.
http://corporate1.morningstar.com/ResearchLibrary/article/813699/morningstar-quantitative-rating-for-funds-input-data-methodology/

Morningstar Quantitative Ratings for Funds Methodology. 2017.
http://corporate1.morningstar.com/ResearchLibrary/article/813568/morningstar-quantitative-rating-for-funds-methodology/

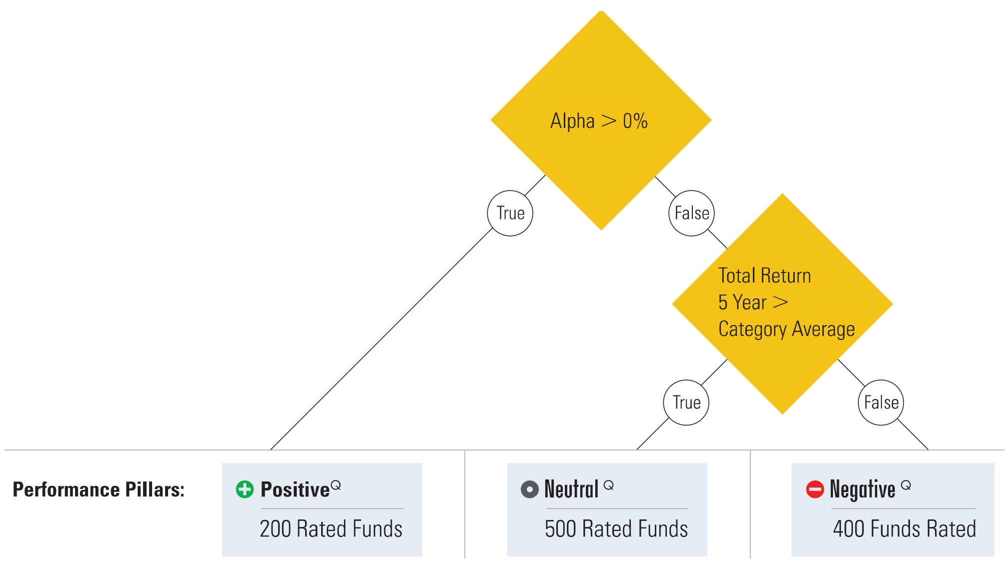
Appendix A: Random Forest

A random forest is an ensemble model, meaning its end prediction is formed based on the combination of the predictions of several submodels. In the case of a random forest, these submodels are typically regression or classification trees (hence the "forest" in "random forest"). To understand the random forest model, we must first understand how these trees are fit.

Regression Trees

A regression tree is a model based on the idea of splitting data into separate buckets based on your input variables. A visualization of a typical regression tree is shown in Exhibit 7. The tree is fit from the top down, splitting the data further into a more complex structure as you go. The end nodes contain groupings of records from your input data. Each grouping contains records that are similar to each other based on the splits that have been made in the tree.

Exhibit 7 Sample Regression Tree With Dummy Data



Source: Morningstar, Inc.

How Are Splits Determined?

The tree is composed of nodes that then are split until they reach terminal nodes that no longer split. Each split represents a division of our data based on a particular input variable, such as alpha, or total return five-year versus the category average (Exhibit 7). The algorithm determines where to make these splits by attempting to split our data using all possible split points for all of the input variables, and it chooses the split variable and split point to maximize the difference between the variance of the unsplit data and the sum of the variances of the two groups of split data as shown in the following function.

$$VarDiff = \frac{\sum(y-\bar{y}_{presplit})^2}{N_{presplit}} - \left[\frac{\sum(y-\bar{y}_{left})^2}{N_{left}} + \frac{\sum(y-\bar{y}_{right})^2}{N_{right}}\right]$$

Intuitively, we want the split that maximizes the function because the maximizing split is the one that reduces the heterogeneity of our output variable the most. That is, the companies that are grouped on each side of the split are more similar to each other than the pre-split grouping.

A regression or classification tree will generally continue splitting until a set of user-defined conditions has been met. One of these conditions is the significance of the split. That is, if the split does not reduce heterogeneity beyond a user-defined threshold, then it will not be made. Another condition commonly used is to place a floor on the number of records in each end node. These conditions can be made more or less constrictive in order to tailor the model's bias-variance trade-off.

How Are the End-Node Values Assigned?

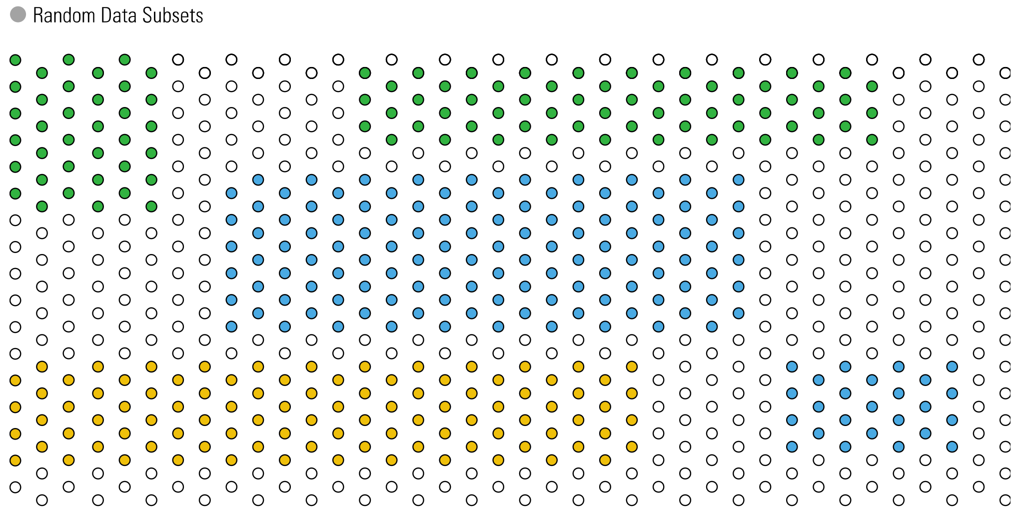
Each tree, once fully split, can be used to generate predictions on new data. If a new record is run through the tree, it will inevitably fall into one of the terminal nodes. The prediction for this record then becomes the arithmetic mean of the output variable for all of the training set records that fell into that terminal node.

Aggregating the Trees

Now that we understand how trees are fit and how they can generate predictions, we can move further in our understanding of random forests. To arrive at an end prediction from a random forest, we first fit N trees (where N can be whatever number is desired—in practice, 100 to 500 are common values), and we run our input variables through each of the N trees to arrive at N individual predictions. From there, we take the simple arithmetic mean of the N predictions to arrive at the random forest's prediction.

A logical question at this point is: Why would the N trees we fit generate different predictions if we give them the same data? The answer is: They wouldn't. That's why we give each tree a different and random subset of our data for fitting purposes (this is the "random" part of "random forest"). Think of your data as represented in Exhibit 7.

Exhibit 8 Sample Random Forest Data Representation



Source: Morningstar, Inc.

A random forest will choose random chunks of your data, including random cross-sectional records as well as random input variables, as represented by the highlighted sections in Exhibit 8, each time it attempts to make a new split. While Exhibit 8 shows three random subsets, the actual random forest model would choose N random subsets of your data, which may overlap, and variables selected may not be adjacent. The purpose of this is to provide each of your trees with a differentiated data set and thus a differentiated view of the world.

Ensemble models use a "wisdom of crowds" type of approach to prediction. The theory behind this approach is that many "weak learners," which are only slightly better than random at predicting your output variable, can be aggregated to form a "strong learner" so long as the weak learners are not perfectly correlated. Mathematically, combining differentiated, better-than-random, weak learners will always result in a strong learner or a better overall prediction than any of your weak learners individually. The archetypal example of this technique is when a group of individuals is asked to estimate the number of jelly beans in a large jar. Typically, the average of a large group of guesses is more accurate than a large percentage of the individual guesses.

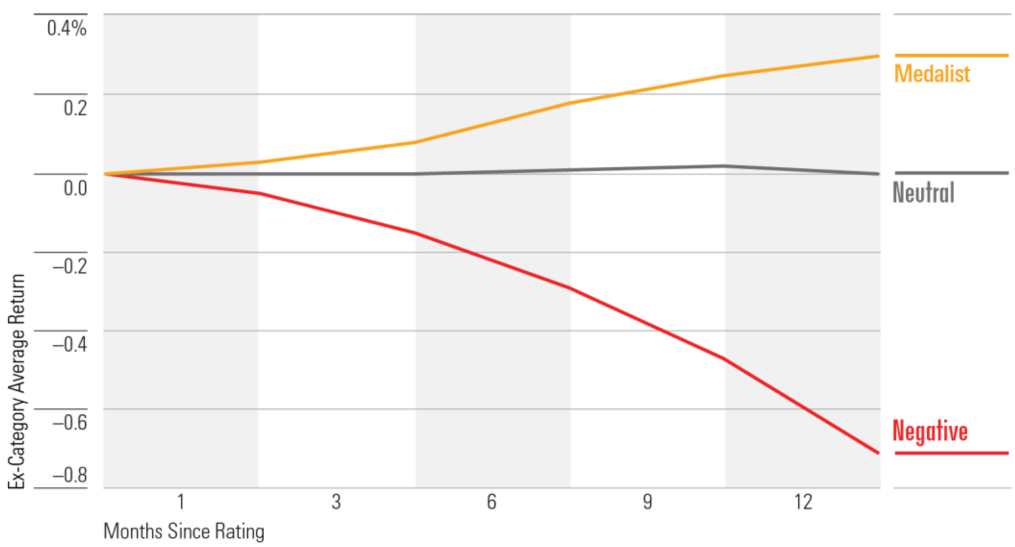
Random forests can also be used for classification tasks. They are largely the same as described in this appendix except for the following changes: Slightly different rules are used for the splitting of nodes in the individual tree models (Gini coefficient or information gain), and the predictor variable is a binary 0 or 1 rather than a continuous variable. This means that the end predictions of a random forest for classification purposes can be interpreted as a probability of being a member of the class designated as "1" in your data.

Appendix B: Additional Performance Studies

Event Study

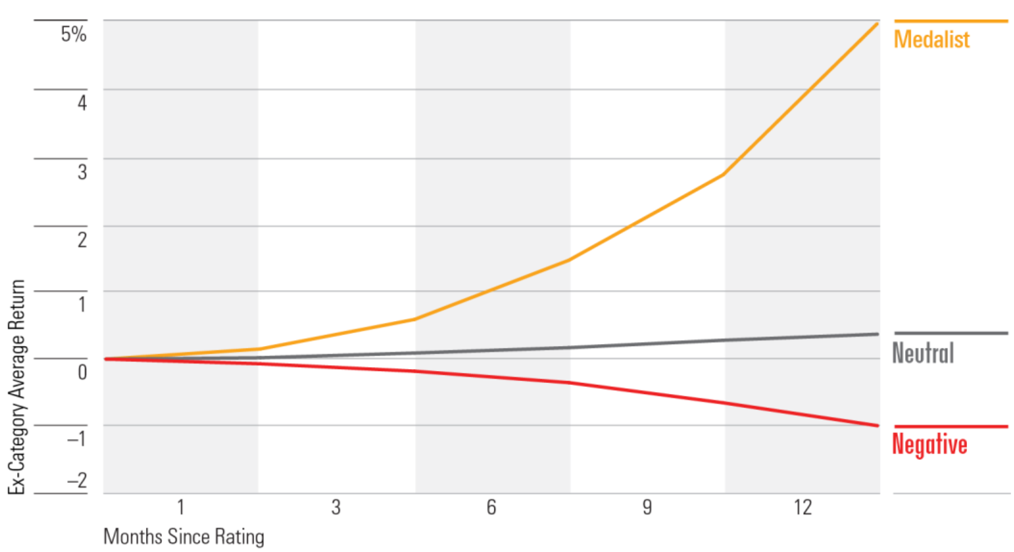
In addition to the event study of the June 2017 Morningstar Quantitative Ratings, we conducted event studies on all subsequent issued ratings. The analysis was repeated for all funds and the subset of new funds. Below shows the average ex-category average return for each subgroup of ratings.

Exhibit 9 Morningstar Quantitative Rating Event Study for All U.S. Funds



Source: Morningstar, Inc.

Exhibit 10 Morningstar Quantitative Rating Event Study for New U.S. Funds



Source: Morningstar, Inc.

T-Tests

The p-values from the t-tests on the above event studies are displayed below. In all cases, the difference in mean performance of the Morningstar Quantitative Rating is significant at the 5% level, as displayed by the small p-values below.

Exhibit 11 Event Study Excess Return t-test p-values for All Issued Morningstar Quantitative Ratings

	All Funds			New Funds		
	Medalist	Neutral	Negative	Medalist	Neutral	Negative
12 Months						
Medalist	—	—	—	—	—	—
Neutral	2.07E-09	—	—	4.86E-04	—	—
Negative	6.53E-42	2.96E-27	—	2.34E-05	1.41E-08	—
9 Months						
Medalist	—	—	—	—	—	—
Neutral	9.80E-19	—	—	9.80E-07	—	—
Negative	2.89E-78	5.93E-48	—	3.46E-10	2.63E-13	—
6 Months						
Medalist	—	—	—	—	—	—
Neutral	2.53E-26	—	—	2.48E-06	—	—
Negative	6.63E-87	3.02E-46	—	3.43E-10	3.33E-12	—
3 Months						
Medalist	—	—	—	—	—	—
Neutral	9.13E-21	—	—	4.07E-05	—	—
Negative	1.79E-64	2.69E-33	—	1.98E-09	5.59E-10	—
1 Month						
Medalist	—	—	—	—	—	—
Neutral	4.13E-08	—	—	1.75E-02	—	—
Negative	6.34E-26	1.69E-14	—	8.88E-05	3.04E-05	—

Source: Morningstar, Inc.

About Morningstar® Quantitative Research

Morningstar Quantitative Research is dedicated to developing innovative statistical models and data points, including the Morningstar Quantitative Rating, the Quantitative Equity Ratings and the Global Risk Model.

For More Information

+1 312 244-7541
lee.davidson@morningstar.com

22 West Washington Street
Chicago, IL 60602 USA