Executive summary. Since the origin of modern portfolio theory and indexing as an investment strategy, empirical evidence has supported the notion that a low-cost index fund is difficult to beat consistently over time. Yet, despite both the theory and the evidence, most mutual fund performance ratings have given index funds an “average” rating. This paper addresses two questions surrounding mutual fund rating systems. First, we examine why index funds tend to receive an average rating on the basis of relative quantitative metrics. Second, we analyze whether a given performance rating offers actionable information: Specifically, we look at whether higher-rated funds can be expected to outperform lower-rated funds in the future. Ultimately, we conclude that investors should expect an average rating for index funds when relative quantitative metrics are used. This is because the natural distribution of the actively managed fund universe around a benchmark dictates that an appropriately constructed and managed index fund should fall somewhere near the center of that distribution. We also find that a given rating offers little information about expected future relative performance; in fact, our analysis reveals that higher-rated funds are no more likely to outperform a given benchmark than lower-rated funds, and that the value of indexing stems in large part from low operating costs and the zero-sum game.
The theory of indexing as an investment strategy is powerful in its simplicity and effectiveness (Sharpe, 1991; Philips, 2010). Yet, despite both the theory and the evidence, most mutual fund performance ratings score index funds as average (as of December 2009, 54% of all stock and bond mutual funds had a 3-star rating on a 5-star scale, according to Morningstar, Inc.). Indeed, it’s not uncommon for clients to question why an average-rated index fund should be given preference as a portfolio option over potentially higher-rated actively managed funds. Such questions provided the catalyst for this paper’s study, having initially surfaced with respect to Morningstar’s first Target-Date Fund Series Rating and Research Reports (Morningstar Research, 2009). In its initial rating of the target-date fund universe, The Morningstar Target-Date Fund Series Rating and Research Reports rated Vanguard’s Target Retirement Funds first out of 20 competing products, yet the Morningstar Rating accorded just 3 out of 5 stars to each of the Vanguard Target Retirement and underlying component funds. Investors logically ask: Why the discrepancy between Morningstar’s rating systems?

The simple answer is that while the Morningstar Rating system focuses on a purely historical, quantitative performance evaluation, the Morningstar Target-Date Fund Series Rating and Research Reports system includes, in addition, broader, qualitative metrics (the management team, the parent organization, and pricing, to name a few). So although a review of prior performance alone rates the index funds as “average,” when one takes into account the criteria in the broader evaluation, the Target Date Fund ratings for Vanguard’s funds improved significantly.

A deeper, more involved, answer addresses why an index portfolio would be rated average by any performance rating system—a rating that seems to be in direct contrast to both the theoretical expectation and empirical evidence supporting the success of indexing as an investment strategy. We focus on this question in the first part of this analysis. We then examine whether a higher or lower rating offers actionable results. In other words, does investing in higher-rated funds (or avoiding lower-rated funds) lead to outperformance? For this analysis, we referred to Morningstar’s rating system, since it is the most widely used rating system in the financial services industry and has the most readily available and reliable data. Finally, we look at costs as a potentially more meaningful metric for selecting investments.

Indexing as an ‘average’ investment strategy

In their quest to outperform a given benchmark, active fund managers typically incur significant costs, which must then be overcome to deliver that outperformance to the fund’s shareholders. In addition, managers face the cold reality that outperformance is a zero-sum game: For every buyer of a security, there must be a seller; that is, for every belief that a security will outperform, there is a counter view that it will underperform. The net result is that for any given period, the returns of active managers form a distribution around the return of the benchmark (represented by the medium blue curve in Figure 1). However, the constant drag of transaction, management, and other costs serves to push a majority of portfolios to the losing side of the benchmark (represented by the brown curve...
It therefore seems curious that when index funds are evaluated using common rating systems, the funds typically are rated as falling in the middle of the pack. However, this seemingly counterintuitive logic can be addressed with the understanding that because all active managers combine to form a distribution around a benchmark (see Figure 1), the benchmark should be rated as average, simply because it falls near the middle of the distribution at all times. The funds that outperform the benchmark will be highly rated, while those that underperform the benchmark will be given a low rating. The benchmark, by definition, must then be rated as average.

By extension, a portion of the fund universe should outperform an index fund, just as a portion of the fund universe should underperform an index fund. For example, Figure 2, on page 4, shows the three-year annualized range of excess returns for funds in each of the nine Morningstar style boxes overlaid with the annualized excess returns of every index fund that operates in each of the style boxes. The cluster of light-brown dashes in the middle of the large-capitalization core distribution represents 229 unique index funds.

While index funds may differ in their expenses and implementation efficiency, the performance distribution of index funds for a given style box has been much tighter than that of actively managed funds. Most important, the index funds are all close

---

Notes on risk: Past performance is no guarantee of future results. All investments, including a portfolio’s current and future holdings, are subject to risk. Investments in bond funds are subject to interest rate, credit, and inflation risk. Investors in any bond fund should anticipate fluctuations in price, especially for longer-term issues and in environments of rising interest rates. Diversification does not ensure a profit or protect against a loss in a declining market.

Investments in Target Retirement Funds are subject to the risks of their underlying funds. The year in the fund name refers to the approximate year (the target date) when an investor in the fund would retire and leave the workforce. The fund will gradually shift its emphasis from more aggressive investments to more conservative ones based on its target date. An investment in the Target Retirement Fund is not guaranteed at any time, including on or after the target date. The performance of an index is not an exact representation of any particular investment, as you cannot invest directly in an index.

---

4 A recent Morningstar report (2009; see also Mamudi, 2009) found that less than 40% of actively managed funds beat their respective Morningstar indexes after adjusting for risk, style, and size biases over the previous three, five, and ten years.
to the x-axis, representing returns very similar to the benchmark. Because of this, we would expect most index funds to be rated as average, because they represent the very benchmark that active managers strive to beat.

In addition to the distributional effect of the fund universe, the methodology of many rating systems further ensures that index funds will receive a middle-of-the-pack rating. Again looking at Morningstar, the star methodology weights three-year performance more heavily than five- or ten-year performance (and if longer-term performance is unavailable, then ratings are based entirely on three-year performance). This is important, because the methodology inherently rewards short-term results at the expense of longer-term performance. With the focus on shorter-term performance, we would expect to see significant volatility with respect to funds’ ratings, primarily because a three-year performance window is narrow enough to permit the portfolio decisions of active managers to outweigh any potential cost disadvantages. It is over longer periods that costs have a greater influence on the distribution of relative performance. As a result, as the evaluation period extends, we would expect index funds to be rated more favorably as costs and the zero-sum game overshadow near-term performance.
Star rating and performance predictability

A natural result of the performance distribution is that investors would rather invest in winning funds than losing funds. And it’s during the selection process for these winning funds that investors often turn to rating systems. Such systems rate the available funds based on one or more performance metrics that categorize fund results as ranging from poor to exceptional. Of course, while we used Morningstar’s system for this paper’s study, Morningstar (2008) clearly states that “the star rating isn’t a complete solution but rather an aid that helps you to narrow the field and improve your chances for success.” That said, the natural use of such a tool is to build a portfolio of highly rated funds with the expectation that such a process will ultimately lead to outperformance relative to a given benchmark.

The question, therefore, is whether such rating systems provide any tangible performance information to investors going forward. This question is not new, and the predictive power of the Morningstar Rating system has been explored before—see, for example, Huebscher (2009), Morey (2005), Morey and Gottesman (2006), and Antypas et al. (2009). Each of these studies evaluated whether any information could be gleaned from performance following a given star rating. While some of the studies found that higher-rated funds do outperform lower-rated funds, others found that this could not be proven to any degree of significance, and still others found no actionable information. One common theme in most of the studies is the difficulty active managers face in simply outperforming a benchmark over time, regardless of their prior performance.

Our analysis—results of which are shown in Figure 3—looked at excess returns versus a relevant style benchmark (nine styles covering large-, mid-, and small-capitalization growth, blend, and value mutual funds) over the three-year period following a given rating. We used a three-year period for two primary reasons: (1) Morningstar requires at least three years of performance data to generate a rating and (2) investment committees typically use a three-year window to evaluate the performance of their portfolio managers. We used style benchmarks instead of the broad market because evaluating performance relative to the broad market would not account for style biases and/or risk-factor bets (Philips and Kinniry, 2009).

5 In the case of Morningstar, funds are rated on trailing risk-adjusted performance whereby it is assumed that, all else being equal, investors prefer higher returns to lower returns and lower risk to higher risk. Morningstar changed its methodology in June 2002 to account for this utility function in addition to market-risk factors such as size and style biases of managers (Morningstar, 2002). Before June 30, 2002, Morningstar rated funds’ risk-adjusted excess returns versus broad benchmarks such as the U.S. stock market, the U.S. bond market, or the international stock market. However, such a methodology does not account for additional broad risk factors such as growth/value or large/small.

6 Morningstar also regularly evaluates the performance of its rating system. For example, in its December 2008 evaluation, Morningstar found that following an initial rating in December 2003, on average, 5-star funds beat 4-star funds, 4 beat 3, and so on, whether on the bases of annual returns, ensuing star ratings, or batting averages (a measure of what percentage of funds beat their peer-group averages) over the five years ended December 2008. For additional details, see Morningstar (2008).
Figure 3 shows that, on average, 39% of funds with 5-star ratings outperformed their style benchmarks for the 36 months following the rating, while 46% of funds with 1-star ratings outperformed their style benchmarks for that period. The figure also shows the average 36-month excess returns (versus the funds’ style benchmarks) over time, based on the median fund in each rating bucket. Here the top-rated funds are shown to have actually generated the lowest excess returns across time, while the lowest-rated funds generated the highest excess returns. Also of interest, the average excess returns across most buckets were significantly negative. Clearly, regardless of whether we look at the likelihood of outperforming or the magnitude of excess returns, investors, on average, have not benefited from basing their investment decisions solely on historical quantitative performance metrics.

Figure 4 expands the averages to show the probability that for any month from June 30, 1992, through August 31, 2009, a randomly selected fund from a given bucket will generate positive excess returns versus its style benchmark over the 36-month period after receiving its star rating. For example, the left-most points intersecting the y-axis represent the percentage of funds with a given star rating on June 30, 1992, that subsequently generated positive excess returns versus their style benchmarks in the next 36-month period (ended June 30, 1995). Over that time period, 51% of 5-star-rated funds on June 30, 1992, generated positive excess returns versus their style benchmarks. For the same period, 35% of 4-star funds, 39% of 3-star funds, 50% of 2-star funds, and 50% of 1-star funds generated positive excess returns. This can be interpreted to mean that an investor essentially had a 50% chance or less of picking a fund that beat its benchmark, regardless of the initial rating.
Several additional points are worth mentioning regarding Figure 4. First, there was no systematic outperformance by funds rated 4 or 5 stars or underperformance by funds rated 1 or 2 stars. In fact, there was tremendous volatility with respect to leadership in any given period. Second, higher ratings in no way ensured that an investor would increase his or her odds of outperforming a style benchmark in subsequent years. In fact, more often than not, all five of the buckets saw probabilities of less than 50%, meaning that an investor had less than a 50–50 shot of picking a fund that would outperform regardless of its rating at the time of the selection. Only ratings that occurred from 1997 through early 2000 led to outperformance. And even then, an investor was at an apparent disadvantage by selecting a 5-star fund instead of a 1-star fund, since 1-star funds posted the highest probability of outperformance relative to their style benchmarks.7

Finally, if we look at the results before and after the methodology change, we find no significant differences. That said, due to the methodology changes, we would expect less-dramatic shifts in average results after June 2002 because the influences of the primary risk factors have been largely removed from the ratings.

In addition to analyzing the probability that an investor would pick a winning fund, we also looked at the median excess returns of the funds in each bucket. Here the reasoning is that probabilities treat all funds equally—yet, in fact, outperforming by 0.01% is not equivalent to underperforming by –1.00%. Figure 5 plots median excess returns generated by funds in each bucket using the same methodology as in Figure 4. The figure can be interpreted to mean that at any point in time, 50% of the funds generated an excess return greater than the median, and 50%

7 Much of this cycle of outperformance can be explained by the shift in market dynamics that occurred from the late 1990s to the early 2000s, coupled with Morningstar’s pre-2002 rating methodology. See Philips and Kinniry (2009) for a discussion of the market dynamics, and footnote 6 and Morningstar (2002) of this paper for details on the Morningstar methodology change.
Figure 6. Percentage of intermediate-term bond funds with positive excess returns versus U.S. bond market for 36 months following Morningstar Rating: January 31, 1992, through August 31, 2009

We also replicated our analysis for the universe of intermediate-term diversified bond funds (both active and passive—see Figure 6), with results differing somewhat from those of equity funds. Broadly, we found that bond fund investors may be able to glean more information regarding future relative performance than stock fund investors. For example, Figure 6 shows that, in general, the probability of beating the bond market decreased as the star rating decreased. We also found (not shown here) that median excess returns followed a similar pattern, with 5-star funds generally outperforming 4-star funds, 4-star funds outperforming 3-star funds, and so on. Of course, while there is greater differentiation across rating buckets, it is important to point out that, as we saw in equity funds, the probability of the average intermediate-term bond fund outperforming the benchmark (here represented by the aggregate bond market) has been poor across ratings.

The dynamic we observe here can be explained to a great degree by the impact of costs. Historically, the range of returns across bond portfolios has been much smaller than the range of returns across stock portfolios. As a result, costs have tended to affect the distribution of bond returns to a greater extent than the distribution of stock returns. In addition, because excess returns for bond managers are almost completely dependent on duration positioning, and because yield changes are notoriously difficult to predict, it can be extremely hard for active bond managers to consistently outperform enough to overcome high fees. Therefore, the higher the fees, the greater the likelihood that bond funds will underperform. As a result, performance-based rating systems have tended to reward lower-cost bond funds and punish higher-cost bond funds to a much greater extent than equity funds.
generated an excess return less than the median. If the median excess return is less than 0%, then, intuitively, more than 50% of the funds underperformed the benchmark and vice versa. Figure 5 again demonstrates that little performance information can be gleaned from one rating versus another (the median 5-star funds’ excess return was not consistently higher than the median 1-star funds’ excess return). And, recalling Figure 3, we again point out that, on average, the subsequent excess returns were negative, regardless of the initial star rating.

One important implication of the observed lack of performance persistence is that funds are likely to jump from one ranking to another over time. This is demonstrated in Figure 7, which shows the likelihood that a stock fund maintained its rating for at least 12 months. We found that most funds had less than a 50% chance of earning the same rating just 12 months following the initial rating. Only 3-star funds had a greater than 50% chance of maintaining their rating, albeit by a slim margin. And, of note, 5-star funds showed the lowest probability of maintaining their rating, further confirming that sustainable outperformance is difficult. This means that investors who focus on investing only in highly rated funds may find themselves continuously buying and selling funds as ratings change. Such turnover could lead to higher costs and lower returns as investors are continuously chasing yesterday’s winner.

The role of costs

To this point, we have demonstrated the inherent challenge to investors of any rating system that focuses on quantitative metrics as the sole factor in performance evaluation. Instead, investors may look to costs as perhaps a more reliable indicator of relative subsequent performance. Figure 8, on page 10, shows the relationship between expense ratios and excess returns over the ten years ended December 31, 2008. Specifically, the figure shows the 10-year annualized excess returns of each fund relative to its style benchmark. To demonstrate the impact of costs, we show a fund’s excess return relative to its expense ratio. The red line in each style box represents a trend line that plots the overall relationship between expenses and excess returns for the funds in that style box. This analysis makes clear that higher costs have historically tended to lead to lower relative returns. For investors, the clear implication has been that focusing on low-cost funds has potentially increased the probability of outperforming higher-cost portfolios.

Conclusion

This analysis has demonstrated why investors should not be surprised by an “average” rating for index funds when performance is based on short-term quantitative results relative to a benchmark. The natural distribution of the actively managed fund universe around a benchmark dictates that an appropriately constructed and managed index fund will fall somewhere near the center of that distribution. We have also demonstrated the difficulty in predicting mutual fund performance based on these relative performance ratings. As this analysis has shown, quantitatively based rating systems do a tremendous job of explaining past performance, but generally offer little insight into future performance.

---

8 For an evaluation of alpha persistence in small-cap funds, see Davis et al. (2007).
9 Morningstar (2008) found that when a cost factor is included in an analysis of subsequent performance, the predictive power of the rating system is improved.
It should also be noted that investors looking to use a given star rating as the sole criterion for selecting funds are picking funds focused on near-term quantitative metrics. This potentially leaves them all the more exposed to the risk that the funds they choose will underperform. By focusing only on the highest star ratings, they may overlook other, more qualitative aspects, such as the fund manager, the parent team, and cost—aspects that, in combination, may yield better overall results to investors in the long run.

A direct implication of the lack of persistence in relative fund performance, combined with the power of costs, is that indexing is a powerful strategy for producing consistent, competitive results. Indeed, if there is no surefire way to pick a consistently winning fund, and, as we have shown here, an investor is likely to pay more in expenses for an actively managed fund than an index fund, indexing would seem to be a more prudent strategy.

References


For more information about Vanguard funds, visit www.vanguard.com, or call 800-662-2739, to obtain a prospectus. Investment objectives, risks, charges, expenses, and other important information about a fund are contained in the prospectus; read and consider it carefully before investing.

CFA® is a trademark owned by CFA Institute.

The funds or securities referred to herein are not sponsored, endorsed, or promoted by MSCI, and MSCI bears no liability with respect to any such funds or securities. For any such funds or securities, the prospectus or the Statement of Additional Information contains a more detailed description of the limited relationship MSCI has with The Vanguard Group and any related funds.

Standard & Poor’s®, S&P®, S&P 500®, Standard & Poor’s 500, and 500 are trademarks of The McGraw-Hill Companies, Inc., and have been licensed for use by The Vanguard Group, Inc. Vanguard mutual funds are not sponsored, endorsed, sold, or promoted by Standard & Poor’s, and Standard & Poor’s makes no representation regarding the advisability of investing in the funds.

Russell is a trademark of The Frank Russell Company.