Don’t be Misled by Morningstar’s Box Score Results
October 20, 2009
by Robert Huebscher

Indexing does not necessarily work best in asset classes that are often perceived to be less efficient, such as small-cap stocks. And if you chose active management three years ago, you would have had a one-in-three chance of picking a fund that outperformed its benchmark in the period since, after adjusting for risk.

Those insights are among the key findings of two research studies published in the last week. Although they are compatible, it turns out that the findings of the first study invalidate some findings of the second.

The first study, “When Indexing Works and When It Doesn’t in U.S. Equities: The Purity Hypothesis,” is by William Thatcher of St. Louis-based Hammond Associates, an institutional consulting firm, and was published in the Journal of Investing. The second is by Morningstar’s index group, and is available here.

The Purity Hypothesis

Conventional wisdom is that active management has an advantage over indexing in small-cap stocks, because they are a less efficient market than large-cap stocks, where indexed strategies prevail.

That belief prevails today, although it was originally dispelled by William Bernstein in 1999 (see When Indexing Fails). Bernstein identified periods when indexing outperformed active management for large-cap stocks, but he said that over the long term small-cap indexing should do better than large-cap indexing, because of the small cap premium and cost advantages. Those cost advantages stem from higher turnover among small-cap managers and larger spreads among small cap-stocks, which combine to create higher “impact costs.” For example, when a fund sells a stock it depresses the market price of that stock.

Thatcher extends Bernstein’s work by showing that indexed strategies work best when their style is performing best. Using Morningstar data from 1998 to 2007, based on S&P indices, he showed that style box index performance is highly correlated to whether that index outperforms active managers.

For example, over the ten-year period he studied the top-performing style box was mid-cap core, and in this style box index funds outperformed active managers 78% of the time. Thatcher’s results are statistically significant at the 0.05% level.

The Purity Hypothesis explains this correlation, according to Thatcher. Indices are always more pure than active funds; by definition, they contain only those stocks whose characteristics meet the definition of the style box. Within a style box, such as mid-cap core, an active fund, unlike the index, will hold stocks that fall outside the definition for that box. Those “impurities” detract from active performance in bull markets, giving the fund a measurable headwind. When a style box is performing...
poorly, impurities are a tailwind to the active fund relative to the index.

Industry consultant Ron Surz has written extensively on this topic (see here, for example). According to Surz, the underlying cause of the Purity Hypothesis is “classification bias,” which is the unavoidable result of index providers assigning active funds to an index, even when those funds hold securities that do not meet the provider’s formal requirements for that index. Surz has similarly documented a tendency among active managers to underperform their index when the index is in favor and outperform when their index is out of favor.

The Purity Hypothesis, Thatcher showed, works for Russell, MSCI, and Morningstar indices in addition to S&P indices. Unlike the other three indices, Morningstar’s are unique in that they are pure – each stock is categorized in one, and only one, index. Other index providers allow a stock to be in more than one index. (Surz offers his own set of pure style indices, but Thatcher did not analyze these in his study.)

As a result, for Morningstar indices the Purity Hypothesis works over one-year intervals. If a Morningstar index outperforms over a one-year period, active managers tied to that style box are likely to underperform the index. According to Thatcher, a possible implication for advisors is that if they forecast underperformance for a Morningstar index over a one-year period, they could benefit from choosing active management for that index – although Surz contends that such active manager outperformance will be a mirage created by classification bias, and Thatcher believes that consistent success in this type of short-term strategy would be very difficult to attain.

Thatcher’s methodology did not incorporate two considerations – survivorship and risk.

Thatcher’s data includes only those active funds that existed over the entire ten-year period, and omits those which failed or otherwise failed to survive the period. The absence of those funds artificially elevates the aggregate performance of the remaining funds, and Thatcher cites a study estimating this bias to be 66 basis points over the ten years. Assuming this survivorship bias is the same across style boxes, then the Purity Hypothesis still holds; the omission merely indicates that the diagonal line in the graph above is shifted upward to properly reflect the performance of active managers.

Thatcher also measured the absolute performance of active funds without adjusting for risk. Over- or underperformance of funds can also be attributable to the extra risk associated with, for example, small-cap stocks. Since Thatcher compared performance only within a style box, the variation in risk among active funds may be less. But whether the Purity Hypothesis truly holds for risk-adjusted performance will be left for future researchers to answer.

An active-versus-passive scorecard

Morningstar’s study, which measured the performance of active versus passive funds over the five years ending June 30, 2009, made the proper adjustment for risk. The study was an update to Morningstar’s previous Box Score Reports (BSRs) which showed how active and passive strategies fared across Morningstar’s nine style boxes for U.S. equities.
In this latest installment, Morningstar noted that approximately half of active funds outperformed their index over the last three years, using an equal-weighted (or headcount) methodology. More significantly, though, after adjusting for risk using the Fama-French three-factor model, 37% of funds beat their index. Using Jensen’s alpha, which adjusts for market risk but not for the small-cap and value effects, 41% of funds outperformed their index.

Since this is a headcount and not an asset-weighted analysis, it does not shed much light on the active-passive debate, but active proponents should be reassured to know that, over a three-year period, they have a better than one-in-three chance of picking a fund that adds alpha after proper risk adjustments. Of course, three years is an awfully short time period in the context of long-term retirement-oriented investing, and Morningstar’s results say nothing about the likelihood of successfully picking active managers over longer time periods.

Asset-weighted performance exceeded equal-weighted performance across all nine style boxes over the last three years by a margin of 67 basis points per year on average. Consistently, the larger the fund, the better the performance.

Morningstar also noted that cash balances in funds have declined from 5% at the beginning of the decade to 3% now, indicating that active managers, on average, did not increase cash balances in anticipation of or in reaction to the most recent bear market.

Morningstar’s analysis adjusts for survivorship by including the performance of closed funds for the time until their closure.

So far, so good.

Where Morningstar’s results are misleading, however, is in its presentation of the BSR:

These results show the percentage of active funds outperforming their index (in the upper left of each box) in each of the nine style boxes over 6-month (shown above), one-year, 3-year and 5-year periods. The index returns are in the lower right of each box.

For example, over the last six months, most active growth funds underperformed their indices (17% of large-cap growth, 41% of mid-cap growth, and 39% of small-cap growth). Over that same six-month period, large-, mid-, and small-cap growth were the best-performing styles, registering returns of 16.0%, 13.3%, and 12.6% respectively, versus 4.3% for the overall market.

The opposite was true for active value managers; they outperformed their indices while their index returns lagged the overall market.
It is wrong to conclude that active growth managers failed during this period, just as it is to conclude the value managers succeeded; such conclusions are merely artifacts of ignoring the Purity Hypothesis and classification bias.

In fairness, Morningstar acknowledged this bias in their prefatory comments in the BSR. Hopefully, careful readers will heed that warning before making any investment decisions based on the BSR.

Morningstar is not alone in this respect. Any peer group ranking that relies on arbitrary classification rules to assign funds to indices will generate similarly misleading conclusions.