



Why Most Equity Mutual Funds Underperform and How to Identify Those that Outperform

January 26, 2016

by C. Thomas Howard, PhD

Why do most active equity mutual funds underperform? I have researched this question over the last few years and have unearthed some surprising answers. It is neither because managers lack stock picking skill nor because of high fees, the two most often cited reasons. On the contrary, nearly 90% of active equity fund managers are superior stock pickers and, in addition, the funds most likely to outperform charge higher fees.

The real culprit for most underperformance is the structural decisions made by fund companies: asset bloat, closet indexing and over-diversification. Collectively, I refer to these as portfolio drag. These structural inefficiencies can be measured and ranked using a methodology dubbed the Portfolio Drag index (PDI). Once understood, it is fairly straightforward to avoid high portfolio-drag funds and reap the value add of skill.

There is a long line of research showing that stock picking skill exists among active equity mutual funds, generating returns that more than offset fees. Overall, these studies reveal a universe of investment teams who are very good at identifying profitable opportunities and portray an industry where superior skill is common. This contradicts a large body of literature that concludes, based on studies of aggregate active equity fund performance, that managers lack skill. I, along with others, argue that this underperformance is due to a variety of non-performance pressures and incentives that lead to building underperforming portfolios, rather than the lack of skill. The resulting impact of performance-destroying portfolio decisions – or, portfolio drag -- can be seen in a portfolio's structure.

In this study, the average skill among funds is found to be 3.81%, portfolio drag averages 2.71% and fees average 1.39%. As a result, 88% of funds display positive skill, with 79% of these large enough to cover fees, the latter being virtually the same as Berk and Green's estimate. Consequently, even though there is substantial skill among active equity funds, none of it is delivered to investors as a result of portfolio drag. This explains why, in spite of significant skill, the average alpha is -0.29% across all funds.

Given the important role of portfolio drag in determining performance, it would be useful to have a measure of the extent to which an individual fund is imposing such a drag. The PDI is introduced as this measure and is derived from measurements of size, closet indexing and conviction.

The resulting PDI is predictive of subsequent fund alpha and, in addition, future fund flows. Both alpha and flows decline precipitously and turn negative as PDI increases. A PDI of 40 or less is predictive of positive alpha and flows, 41 to 60 is predictive of near zero alpha and weak flows and 61-100 is predictive of both negative alpha and flows. Thus an effective way to identify funds with the best chance of subsequently outperforming and also generating positive flows is to focus on those with the lowest PDI, more specifically, those with a PDI of less than 40.

Disaggregating performance: Stock-picking skill and portfolio drag

A number of recent studies show that *truly* active funds – that is, not just calling themselves active but actually taking high-conviction positions -- outperform their closet-indexing counterparts. The three most important variables (asset bloat, closet indexing and conviction) identified in these studies are used for the purpose of disaggregating fund performance into skill and portfolio drag. Let's look at how each of those degrades portfolio performance.

Asset bloat

This study shows that fund performance declines with fund size as measured by AUM. This is likely the result of limiting aspects of managing an active equity portfolio. In pursuing a narrowly defined investment strategy, a fund manager ends up with a small number of “best idea” stocks (it will be shown shortly that this number is fewer than 20). Put another way, their highly developed strategy allows them to identify only a few stocks worthy of investment.

As the fund grows, it becomes increasingly difficult to effectively trade this small number of stocks. Eventually the fund reaches a size where it is not possible to limit the portfolio to best idea stocks, so a decision must be made either to limit the size of the fund or begin investing in other than best idea stocks. Unfortunately, the investment industry provides incentives to do the latter, most importantly because managers are compensated based on AUM. Thus begins the transformation from being truly active to a closet indexer.

Closet indexing

This study shows that fund performance declines as the R-squared to its benchmark increases. That is, the more closely a fund tracks its benchmark, turning itself into a closet indexer, the worse the performance. This is painfully obvious; in order to beat the benchmark, one must differ from the benchmark. It is truly strange that the industry has evolved to the point where funds are expected to closely track their benchmark while, at the same time, beat their benchmark!

The widespread requirement that funds maintain a high R-squared to a benchmark is the result of two important drivers.

First, investors suffer from myopic loss aversion (MLA), arguably the most important cognitive error identified by behavioral science. Individuals display 2-to-1 loss aversion; That is, they react twice as strongly to losses as they do to equivalent gains. In addition, investors focus on short-term performance even when they face a long investment horizon. Combining loss aversion with a short-

term focus leads to bad, myopic decisions by investors, resulting in poor long-term performance.

A fund that does not track its benchmark stirs up MLA in its investors as they emotionally react to short-term underperformance. This may prompt investors to leave the fund, thus turning investor emotion into business risk for the fund. One way to avoid those investor emotions, and the related business risk, is to closely track the fund's benchmark. That's why maintaining a high R-square is a common approach for catering to investor emotions.

Second, virtually every platform, broker-dealer and institutional consultant within the fund distribution system assigns a fund to a specific style box and demands that it stay there over time. In fact, style drift is often viewed as a more serious problem than underperformance. However, the assigned style box has little to do with a fund's strategy and, in turn, purchasing stocks in order to remain style consistent means not staying true to the fund's investment approach.

This study shows that funds with the greatest amount of style drift outperform those with the least drift by 3.00%. It also finds that a fund cannot outperform, on average, if it does not style drift.

Emotional catering and style drift avoidance encourage a fund to maintain a high benchmark R-squared.

Conviction

This study shows that individual stock alphas decline as a stock's relative portfolio weight rank declines. A fund's best idea or high-conviction stocks can be identified by ranking stocks based on their relative weight within the portfolio. The rank of these weights is predictive of future stock returns for up to a year ahead, which means the "shelf life" of fund holdings is at least 12 months.

Surprisingly, then, high conviction stocks can be identified by means of holdings and, in turn, these high conviction stocks subsequently outperform. The corollary is that low-ranked stocks reflect a lack of conviction on the part of the manager and, in turn, underperform.

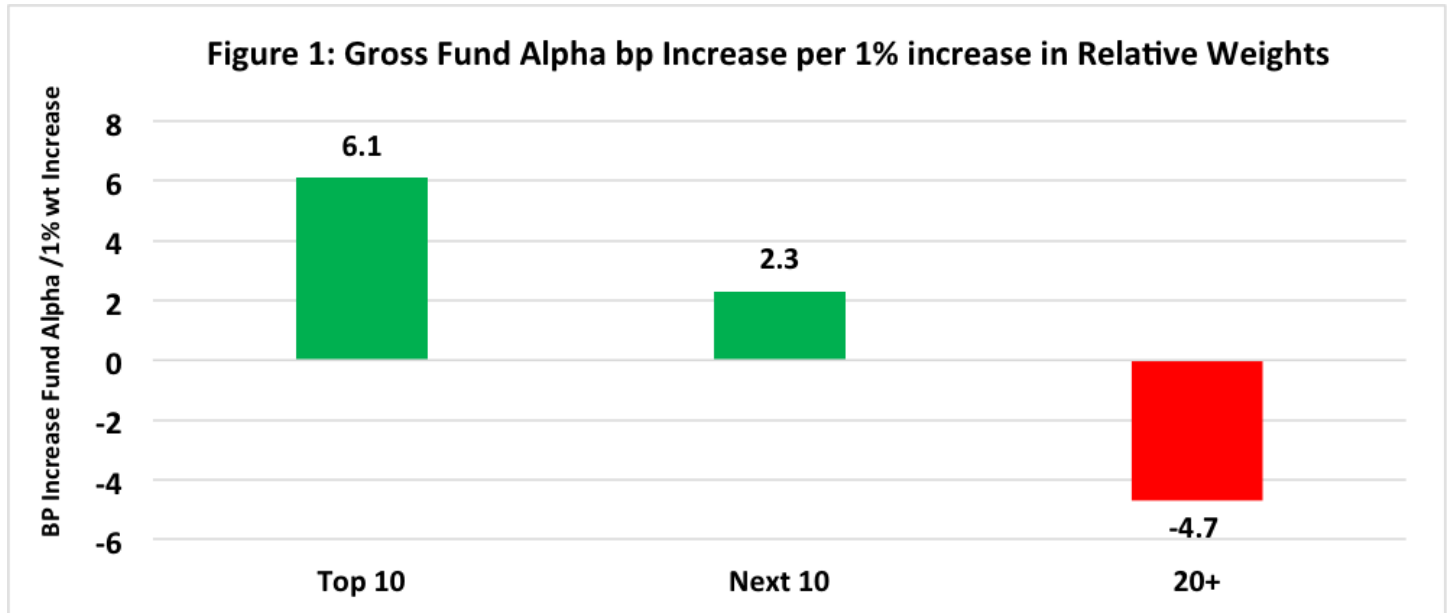
As reported in Figure 1, an increase in the top-20 stock weighting leads to an increase in a fund's subsequent alpha, with the gain to the top-10 stocks being nearly triple the next 10 stocks (6.1 versus 2.3 basis points annually). For example, increasing the top 10 weighting by 10% improves fund alpha by 61 basis points. However, increasing the weighting of stocks ranked lower than 20 hurts fund performance, as the impact on fund alpha is negative as shown in Figure 1.

The 4,000+ funds in the sample hold an average of 113 (median of 75) stocks, which translates into four- to five-times as many low-conviction as high-conviction stocks. Although there are legitimate reasons why some funds may choose to diversify more broadly, the results show fund managers heavily dilute performance by doing so.

One possible explanation is that investors and funds falsely believe that a large number of stocks are needed to achieve proper diversification in spite of the evidence to the contrary. Another possibility is that holding a small number of stocks exposes the manager to criticism if a stock dramatically

underperforms and thus has a significant impact on fund performance. As will be shown shortly, over- diversification may also be a byproduct of asset bloat and closet indexing.

Whatever the reason, investing in low-conviction instead of high-conviction stocks is a performance-destroying decision.



Based on ex-post gross fund alpha regressions on cumulative relative weights, estimated using a data set of 44 million stock-month equity fund holdings from January 2001 through September 2014.

A Common misconception: Stock picking is a zero-sum game

Before proceeding, let's address a pervasive misconception regarding stock-picking skill. A strongly held belief within the investment industry is that stock picking across active equity funds must be a zero-sum game. Such an assertion is true for the stock market as a whole, as stock picking must be a zero-sum game with as many losers as winners. But this does not have to be the case in every market segment.

The U.S. stock market has a current total market value exceeding \$38 trillion. Active U.S. equity mutual funds hold \$3.6 trillion, about 9% of all equities. So it is entirely possible for the average stock held by funds to outperform at the expense of the other 91% of the equity universe. Arguing that stock picking among equity funds must be a zero-sum game is akin to arguing it is impossible to drown in a lake of average depth of three feet; that lake may have pockets that are 20 feet or more deep. Both represent indefensible statements.

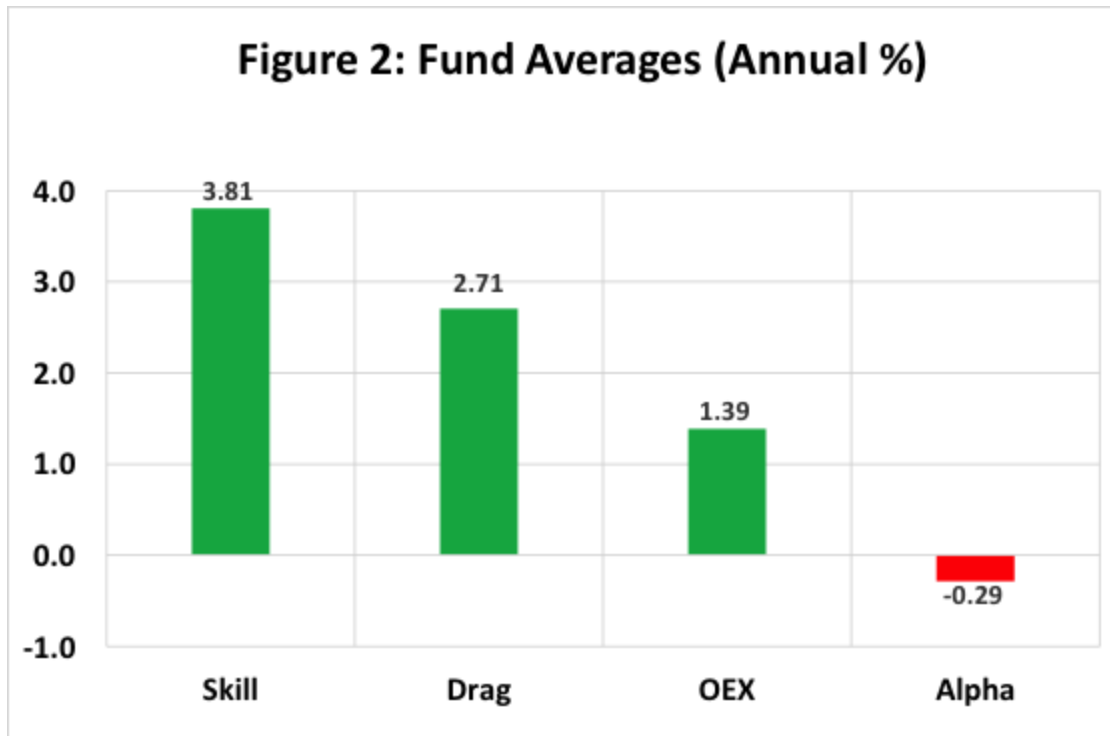
In particular, this study estimates that the average stock held by active equity mutual funds earns an alpha of 1.30%, confirming mutual funds do earn superior returns. Indeed, this must be the case in order for equity funds to cover their fees and, in turn, earn a near-zero collective alpha. Once the distorting effect of drag is removed, we uncover further evidence that funds are able to earn excess

returns at the expense of the rest of the market.

Estimating skill and drag

Individual fund stock-picking skill and portfolio drag are estimated using a multiple regression of ex-post gross fund alpha on ex-ante fund AUM, benchmark R-squared and top-10 relative weight. The survivor-bias-free sample includes all U.S. active equity mutual funds domiciled in the U.S. over the period from January 2001 through September 2014, resulting in over 250,000 fund-month observations.

Stock-picking skill and portfolio drag are reported in Figure 2. The overall average fund skill is 3.81%, where skill is the gross alpha the fund could earn if it did not face a portfolio drag, that is, no asset bloat, closet indexing or over-diversification. The average drag of 2.71% is double the explicit fees (OEX) of 1.39%. The average alpha is -0.29%, which means that, collectively, none of the substantial skill of active equity managers is delivered to investors, as fees and the much larger drag combine to wipe out all potential value added.



To demonstrate that skill is widespread, Figure 3 reports the percent of funds, with at least 24 fund-month observations, that outperformed based on various criteria. Strikingly, 88% displayed skill, meaning that virtually all funds in the sample were superior stock pickers. This further supports the growing body of research showing that skill is common among active equity funds.

Also strikingly, 79% of funds have enough skill to more than cover their fees. So explicit fees are not nearly as large a problem as the industry portrays. Finally, only 41% of funds generated a positive alpha. The reason, of course, is that the funds impose a drag sufficient to wipe out the skill benefit.

Figure 3: Percent of Funds Outperforming

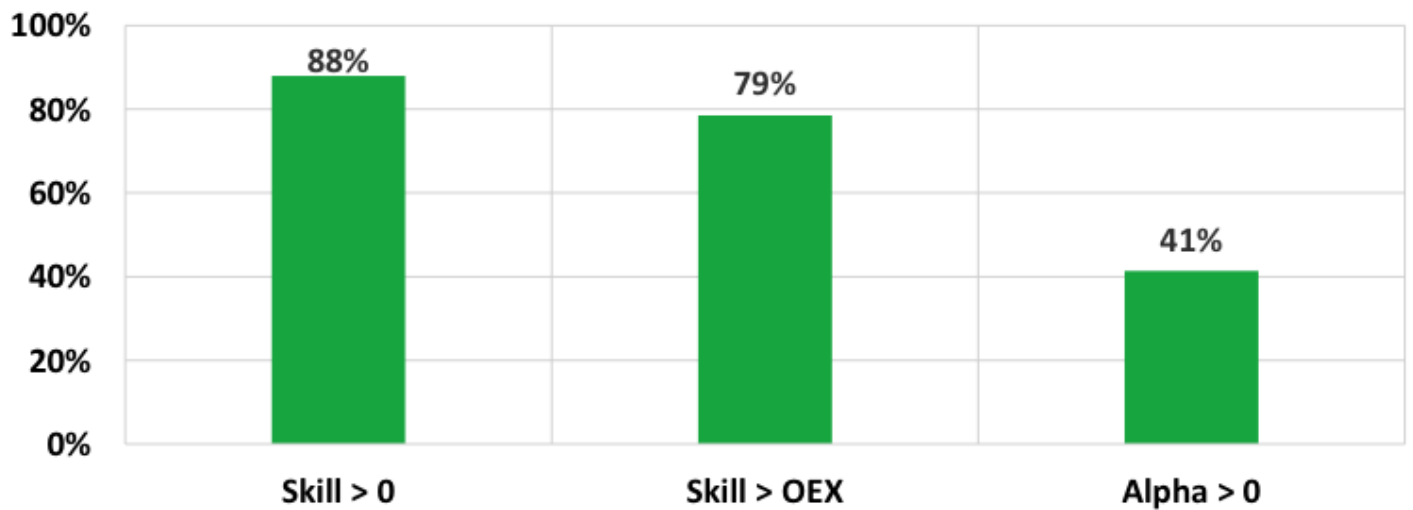


Table 1 reports the 2001-2014 annual averages for the three drag variables (AUM for asset bloat, R-squared for closet indexing and rw 1-10 for top-10 weighting), along with portfolio drag, alpha and fund flows. Over this period, both average AUM and R-squared increased while, on the other hand, top 10 weighting decreased, implying that asset bloat, closet indexing and over-diversification all grew worse over this 14-year period. Consequently, drag nearly doubled, expanding from 1.67 in 2001 to around 3.00 in 2014. At the same time, both alpha and fund flows decreased substantially.

This period provides a clear picture of how industry performance declines as drag increases. In the first half, when drag was relatively low, both alpha and flows were generally positive, but in the second half, when drag was higher, both alpha and flows turned negative. Sometime around 2007 the industry transitioned from being a collective value creator to a value destroyer. It is entirely possible that the market distress of 2008 was a major contributor to this change. Regardless, the state of the industry clearly deteriorated.

The portfolio drag index

The PDI is a simple measure of the extent to which funds impose a portfolio drag. PDI is calculated as a scaled value of a fund's portfolio drag and ranges from 0 to 100.

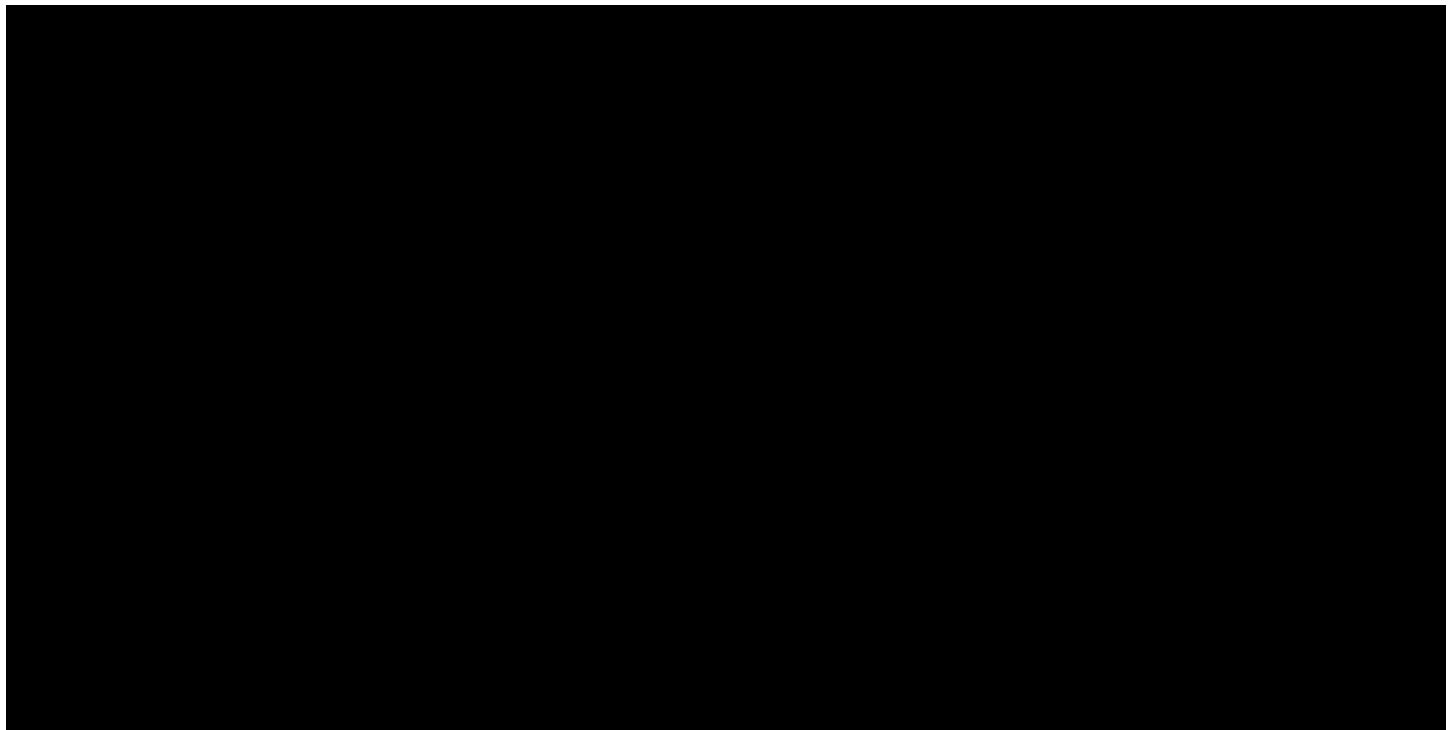
Table 2 reports average subsequent alpha, as well as annual fund flows for five PDI groups. Group 1 represents 4% of all funds and less than 1% of the industry AUM. In this group, PDI is 20 or less, average AUM is less than \$45 mil, R-squared is less than 75 and relative weight 1-10 exceeds 30%, producing a drag of 0.6%. Group 1 has the highest alpha at 1.73%, the highest OEX at 1.62% and the highest annual fund flows at 9.27%. Groups 1 and 2 are the only groups for which both alpha and fund flows are significantly positive, making them the top-performing groups, but disappointingly the smallest in terms of total AUM.

Fund Universe Statistics				Top PDI Average			Disaggregation			Performance	
PDI Group	PDI Range	Fund Count	Fund AUM	AUM	R-square	rw 1-10	Skill	OEX	Drag	Alpha	FF
1	0-20	4%	0.1%	45	75	30	3.92	1.62	0.58	1.73	9.27
2	21-40	23%	6%	659	83	19	4.25	1.58	1.40	1.28	4.65
3	41-60	30%	21%	1,187	92	15	4.00	1.46	2.34	0.19	2.03
4	61-80	27%	32%	1,655	96	11	3.55	1.34	3.19	-0.98	-0.62
5	81-100	16%	41%	9,664	98	6	3.95	1.15	3.82	-1.01	-1.36
Sum/Average		100%	100%	1,391	86	17	3.81	1.39	2.71	-0.29	-0.05

At the opposite end of the performance spectrum is Group 5, with a PDI exceeding 80, an average AUM exceeding \$1.7 bil, R-squared exceeding 96, top 10 weighting of less than 11% and a drag averaging 3.8%. This group displays the highest degree of asset bloat, closet indexing and over-diversification, resulting in the worst alpha at -1.0% and worst flows at -1.36%. Distressingly, this is the largest group, representing over 40% of industry AUM, which translates into an average annual outflow of \$15 billion during this time period. As was shown in Table 1, outflows have grown worse in recent years.

Groups 1 and 2 are the only ones for which both alpha and fund flows are positive. This implies that funds should have a PDI of 40 or less, grow no larger than \$700 AUM, have an R-squared less than 80 and top-10 weighting of no less than 20%, each based on the respective PDI 40 averages. These provide guidance on the extent to which a fund can asset bloat, closet index and over-diversify and still add value and generate positive flows. But of course, the less a fund does of each, the better.

Table 3 reports the top and bottom 10 funds, out of the total of nearly 1800 US active equity mutual funds, for which PDI was recently calculated. Note that those in the top 10 are less familiar while those in the bottom 10 are generally well known. It is the case that notoriety is frequently associated with large portfolio drag, implying future underperformance as well as outflows. Investing in a well-known fund is often detrimental to your wealth.



Higher fees of top-performing funds

The OEX and alpha columns in Table 2 challenge a widely held misconception. Many believe that the best funds going forward are those with the lowest fees. But Group 1 has the highest fees at 1.62% along with the highest alpha, while Group 5 has the lowest fees at 1.15% along with the lowest alpha.

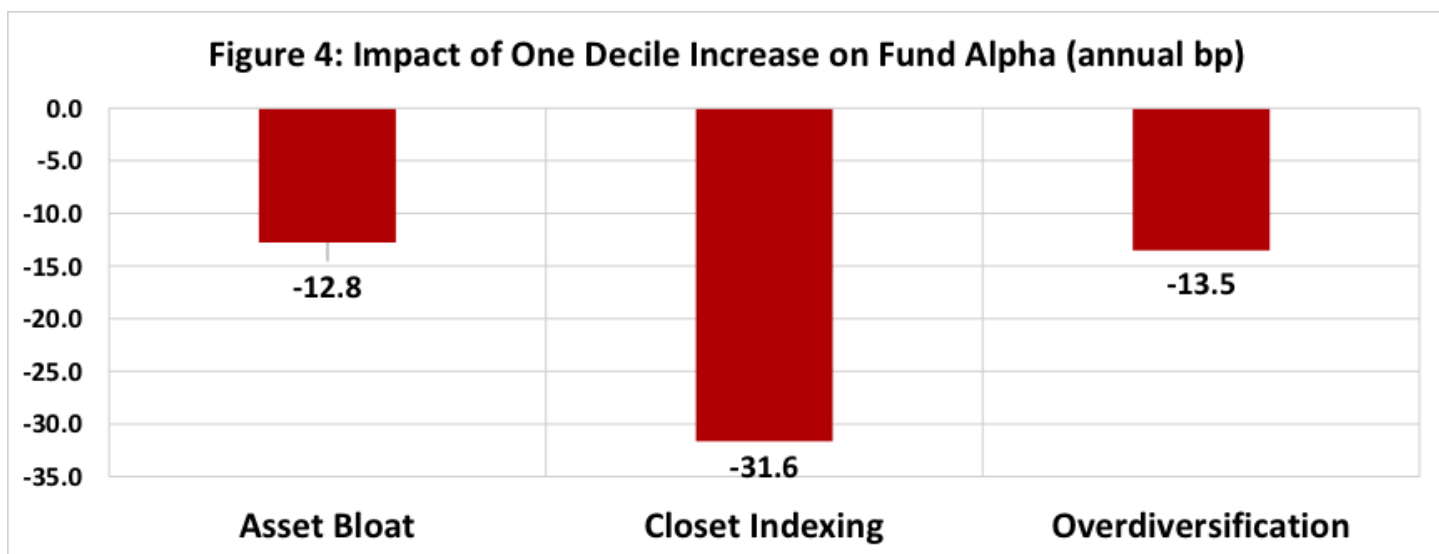
This is the opposite of the conventional wisdom.

The reason for this is that as fees fall, by moving down PDI groups, drag increases even faster. The underlying driver behind the results reported in Table 2 is an inverse relationship between explicit fees and implicit drag. Picking the fund with the lowest fees may very well lead to a large drag and thus poor future performance. Consequently, PDI should be a first level criteria in evaluating funds, with fees as a secondary criteria.

PDI components: A closer look

Let's examine the relationship between individual drag components and performance. The impact of each on fund performance is presented in Figure 4. The impacts of asset bloat and over-diversification are similar at -12.8 and -13.5, respectively. This means that every one decile increase in either variable results in a decline in annual fund return by about 13 basis points.

However, increased closet indexing is about 2.5 times more destructive, implying the most effective way to decrease PDI and reduce the accompanying negative impact on performance is to reduce benchmark tracking. The cross-variable correlations (not reported) are moderately low, confirming that the decision of which to change can be made fairly independent of one another.



Based on three subsequent gross alpha, single variable regressions.

Concluding remarks

I developed and tested a measure of stock-picking skill and portfolio drag for active equity mutual funds. Using a 2001-2014, 250,000 fund-month survivor-bias-free sample, the average fund skill is found to be 3.8%. Nearly 90% of funds displayed skill while nearly 80% had enough skill to cover their fees. This is the good news of this study.

The bad news is that funds make portfolio decisions that end up, collectively, wiping out all of the skill benefit intended for investors. By means of asset bloat, closet indexing and over-diversifying, funds hurt performance. The impact of these decisions is dubbed “portfolio drag.” The implicit portfolio drag is nearly two times explicit fund fees and consequently is the major reason why most funds underperform, in contrast to the often-stated reasons of a lack of skill and high fees.

The PDI is introduced as a simple combination of the three measures of asset bloat, closet indexing and over-diversification. It is shown that subsequent fund alpha and flows drop precipitously and turn negative as PDI increases (i.e. increasing portfolio drag). A PDI of 40 or less is predictive of positive alpha and flows, 41 to 60 is predictive of near-zero alpha and weak flows and 60-100 is predictive of both negative alpha and flows.

Making better decisions

PDI is a guide for making better decisions. Investors should purchase low (less than 40) PDI funds and when PDI rises above 40, sell and invest in other low PDI funds. There is no reason to stick with a fund whose PDI has risen since there is a large, continuous supply of low-PDI funds. Using such an approach, investors have the best opportunity to reap superior returns, the ultimate purpose of active-equity mutual-fund investing.

Thomas Howard, PhD is emeritus professor of finance at the Daniels College of Business, University of Denver, and CEO and director of research at AthenaInvest, Inc.