Mesirow

Out with the old and in with the new

A retrospective on Mesirow PrecisionAlpha® vs. the failure of traditional performance metrics

Executive summary

- No truer phrase has been uttered than "past performance is no guarantee of future results."
- The most widely used metric of trailing return is a poor predictor of performance at both the one-year and three-year horizon. The two broad equity category sets exhibited strong mean reversion tendencies, implying that the best on a trailing basis is the worst on a forward basis.
- Our results strongly confirm the tendency of traditional historical performance metrics to be negatively predictive of future relative performance (i.e., indicative of mean reversion), while supporting the historical efficacy of PrecisionAlpha as a positive predictor of future relative performance.
- **PrecisionAlpha, Mesirow's answer to the shortcomings of traditional performance metrics, is the most consistent predictor of performance across broad category sets for both one-year and three-year horizon periods in our study.**

Overview

"Past performance is no guarantee of future results" is a standard disclaimer that accompanies most financial marketing exhibits that are meant to highlight good past performance, while relegating the legal disclosures to footnotes. The implication is that potential investors can look past the pesky, legal disclosure requirements and assume that past will be prologue — what was good in the past will be good in the future as well.

Historical facts say otherwise, however. There exists an inconvenient statistical truth of reversion to the mean with respect to most backward-looking metrics of performance when used as a predictive tool for the future.

An active manager that outperformed its category peers over some historical period is more likely to underperform than outperform over a similar period of measurement in the future. Numerous studies have been done over past decades that examine whether "winning" managers can outperform a sample in the future.[1,2,3,4,5] Other studies have looked at the results of performance chasing, which is the investor behavior of replacing underperforming managers with other managers that previously outperformed with the hope that outperformance will continue.[6,7]

Keith Gustafson, CFA
Managing Director of Asset Allocation and Retirement Income

Christopher M. O'Neill, PhD, CFA, CFP®, FRM®, ChFC®
Managing Director, Chief Investment Officer and Director of Quantitative Research

An active manager that outperformed its category peers over some historical period is more likely to underperform than outperform over a similar period of measurement in the future.

Unfortunately, performance chasing typically leads to relative underperformance in the future and accompanying investor disillusionment. That has been identified as contributing to the increasing prevalence of passive investing that recently surpassed 50% of the total assets in the US equity market.[8]

In the search for ways to win this "loser's game" and achieve better predictability, investors have considered other metrics beyond simple trailing returns, such as those that incorporate risk adjustment, including Information Ratio, Sharpe Ratio or Treynor Ratio. Additionally, later generation metrics have become more sophisticated in statistical terms, utilizing returns-based style analysis to create custom benchmarks for metrics that include Selection Return, Timing Return and Information Ratio versions of each of these. Not many broad-based studies on predictive efficacy have been performed on these measures.

Other approaches to manager selection include non-standardized conditional analysis approaches such as up- and down-market analysis, as well as select regime periods, which can be difficult to test for effectiveness as predictors of alpha out of sample.[9,10,11]

Mesirow's quest to build a better mousetrap

Soon after the team's founding in 2006, Mesirow Fiduciary Solutions took a new, proprietary approach to building a better mousetrap for manager evaluation — an approach known today as PrecisionAlpha.

Knowing the inherent failings of traditional metrics, we created a proprietary evaluation framework that incorporates some elements of selection returns from returns-based style analysis, along with critical category-dependent statistical adjustments that account for elements of mean reversion. The final value is a performance forecast combining a sophisticated regression analysis of historical returns, and Bayesian statistical methods to enhance the precision of the alpha estimate.

We recently completed a study across numerous equity and fixed income categories that evaluated the historical predictive capability of most of these metrics that could be effectively evaluated along with Mesirow's PrecisionAlpha.

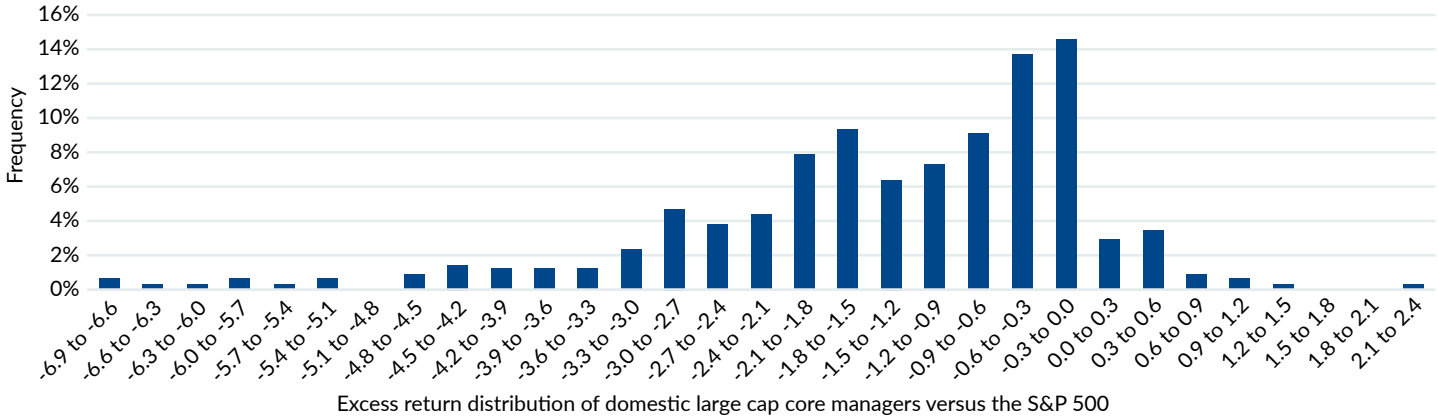
Our results strongly confirm the tendency of traditional historical performance metrics to be negatively predictive of future relative performance (i.e., indicative of mean reversion), while supporting the historical efficacy of PrecisionAlpha as a positive predictor of future relative performance.

Methodology

The results presented in this paper are based on five-year trailing periods of measurement. We evaluated all five-year trailing periods back to 2008 for domestic equity categories and back to 2012 for international equity and domestic fixed income. The periods selected were based on the availability of sufficient data population in each category for statistical efficacy in our quartile analysis. We utilized an overlapping data measurement framework in five-year rolling windows on an annual basis, such as 2008–2012, 2009–2013, 2010–2014, etc., such that there was no bias based on a specific starting and ending point as with non-overlapping analysis (e.g., 2008–2012, 2012–2016). We then evaluated out-of-sample future return horizon periods, specifically over the one- and three-year periods. For instance, 2013 would be the one-year out-of-sample evaluation period for 2008–2012 historical data that is sorted on the various metrics, while 2013–2015 would be the three-year out-of-sample evaluation period.

Analysis was performed on active managers grouped into quartiles over the evaluation period and then evaluated over the subsequent period in terms of return relative to the overall group average return. The quartile groupings differed depending upon the metric used, so that each metric analysis was independent. Metrics evaluated were Return, Sharpe Ratio, Information Ratio,[12] Selection Return, Timing Return, Selection Information Ratio, Timing Information Ratio and PrecisionAlpha. The same set of active managers was used in the measurement period and evaluation period, so it became the peer group. For instance, if the top quartile manager grouping on Sharpe Ratio in the measurement period is predictive of future outperformance, then it should produce a positive excess return relative to the overall peer group average return in the evaluation period.

FIGURE 1: **NEARLY 92% OF DOMESTIC LARGE CAP CORE MANAGERS FAILED TO BEAT THE S&P 500 OVER THE 10-YEAR PERIOD ENDED 12.31.2023**



Source: MPI Stylus and Mesirow.

Since we focused on an active peer group and predictability of outperformance from one period to the next, we wanted to focus on categories where there has been historical opportunity to beat a passive benchmark.[13,14,15] An active manager providing historical alpha relative to other active managers is not particularly useful for investors if such an active manager underperforms the relevant benchmark index on average over time. We required at least reasonable potential for alpha production because we did not want to skew the overall metric predictability results with low alpha probability categories over our result measurement period. This "potential alpha" criteria affected the categories that we selected to evaluate in the broad-based domestic equity set. We selected six domestic equity categories: mid core, mid growth, mid value, small core, small growth and small value.

Figure 1 shows an example of the excess return distributions that were considered for all manager categories, which in this case contributed to the decision to exclude large core, along with large growth and large value from our analysis in domestic equity. This exhibit displays the excess return distribution of domestic large cap core managers versus the S&P 500, where 91.6% of the managers failed to beat the index over this time period.[16]

Category choices in the broad-based international equity data set (that includes global equity and emerging markets) were solely a consideration of data population rather than alpha potential. International small/mid categories were excluded due to low data population, while global large core and emerging markets core categories had sufficient historical population and were included along with international large core, large growth and large value categories for a total of five manager categories.

The same was largely true for the broad-based fixed income data set where the category selection was mostly based upon historical data population considerations. Two categories, however, were excluded due to qualitative considerations because the category definitions were too broad based. For instance, short-term bonds as a category included government bond-oriented approaches, pure corporate strategies (including those that are mostly high yield), or other specialized approaches. The category grouping was too divergent for a meaningful peer comparison, while dividing up the broader category into sub-groups led to an insufficient population problem. Multisector bonds was a similar widely divergent category ill-suited to peer analysis in our estimation. The categories evaluated were five in total: bank loans (floating rate debt), high yield, intermediate government, core bond and core plus bond.

The population of managers naturally increases for each category as we move forward in time from the past to the present, but we do not want to skew the results towards the recent past by weighting based on observations. Therefore, each period of results is weighted equally within a category to avoid a time bias. Similarly, we equally weight each category within its applicable broad category set (domestic equity, international equity, or domestic fixed income) even though there are population differences from category to category within each broad category set. The reason for this are that we did not want overall average efficacy results skewed by a very high population category. For example, we did not want international large core managers to skew the results of the overall international equity category set simply because there are significantly more international large core managers than there are emerging markets core managers.

We wanted to evaluate how these historical metrics work in predicting future relative performance on average across categories and across time. By averaging and weighting in this manner, we eliminated substantial data noise that is endemic to individual categories in this type of statistical analysis. Therefore, we focused on average results for the metrics within the broad category sets of domestic equity, international equity (including global and emerging markets) and domestic fixed income.

Results: The top goes to the bottom and the bottom goes to the top for traditional measures, but not for PrecisionAlpha.

Our results show very poor predictability out-of-sample for the traditional performance metrics of trailing return (raw return or excess return results are identical when using a single category benchmark). There was also poor predictability for Sharpe Ratio and Information Ratio across all broad categories of domestic equity, international equity and fixed income over both a one-year horizon and the more meaningful three-year horizon.[17]

The results for the one-year horizon are shown in Figure 2, while the three-year results are shown in Figure 3 (numbers may not sum due to rounding). The best metric at the top, the bottom and for the spread in each category set is shown in **bold typeset**.

CHARTS AT A GLANCE

The top quartile represents the top performing managers — according to a variety of popular performance metrics — relative to the peer group, based on five-year rolling time frames during the sample period dating back at least 15 years. For this one-year time horizon chart, we then compared how those managers performed the next out-of-sample year.

One would expect the top managers to continue to outperform. However, the opposite occurs when using traditional performance metrics: the best managers tended to perform the worst in the subsequent one-year time period. Moreover, the "worst" performing managers tended to perform the best

This suggests that outperformance or underperformance doesn't persist on traditional performance metrics. Rather, the top and the bottom tend to revert to the mean.

One-year time horizon

FIGURE 2: **FIVE-YEAR TRAILING DATA TO PREDICT FORWARD ONE-YEAR RESULTS**

	Return	Sharpe	IR	Selection	Selection IR	Timing	Timing IR	PrecisionAlpha
Domestic Category Average								
Top quartile	-0.53	-0.57	-0.58	-0.16	-0.17	-0.37	-0.21	**0.03**
2nd quartile	-0.12	-0.12	-0.03	-0.02	0.11	-0.21	0.04	0.26
3rd quartile	0.03	0.27	0.19	0.07	0.04	0.14	0.15	-0.22
Bottom quartile	0.65	0.38	0.65	0.13	0.03	0.43	0.03	**-0.15**
Top-bottom spread	-1.18	-0.94	-1.23	-0.29	-0.20	-0.79	-0.23	**0.17**
International Category Average								
Top quartile	-0.39	-0.45	-0.48	-0.07	-0.16	-0.09	-0.33	**0.07**
2nd quartile	0.01	0.16	0.10	0.36	-0.08	0.10	-0.17	-0.05
3rd quartile	0.16	0.08	0.14	-0.31	0.09	-0.10	0.44	-0.01
Bottom quartile	0.24	0.22	0.25	0.03	0.15	0.11	0.07	**0.00**
Top-bottom spread	-0.63	-0.67	-0.73	-0.11	-0.32	-0.19	-0.40	**0.08**
Fixed Income Category Average								
Top quartile	0.02	0.04	0.08	**0.12**	0.09	0.02	-0.02	0.06
2nd quartile	0.07	-0.05	-0.06	0.01	0.03	-0.02	-0.07	-0.01
3rd quartile	-0.05	-0.03	0.00	-0.06	-0.01	-0.01	-0.02	0.05
Bottom quartile	-0.04	0.05	0.00	-0.06	**-0.11**	0.01	0.13	-0.09
Top-bottom spread	0.06	0.00	0.08	0.18	**0.20**	0.02	-0.15	0.15

Source: MPI Stylus and Mesirow.

These data represent the manager average excess return for the quartile relative to the peer group average in the evaluation period. The top-bottom spread represents the average persistence of the top quartile less the average persistence of the bottom quartile. Ideally, we would like to observe positive persistence at the top and negative persistence at the bottom. The spread captures these elements in one measure. It could also be interpreted as the return to a long/short strategy, were it possible to be long the top quartile bucket and short the bottom quartile bucket of managers.

Over a one-year horizon, the top quartile over the last five years produces substantially negative returns over the next year for Return, Sharpe Ratio and Information Ratio metrics within both domestic and international equity categories on average. Moreover, the bottom quartile produces substantially positive returns for the same metrics. The top goes to the bottom and the bottom goes to the top. Within the broad fixed income category set, there is weak persistence for the top quartile over the one-year horizon for Return and Sharpe Ratio, but only with Return for the bottom quartile.

The more statistically complex metrics of Selection Return, Selection Information Ratio, Timing Return and Timing Information Ratio are based on custom benchmarks for each manager determined from returns-based style analysis using high frequency daily data. These metrics basically partition excess return (positive or negative) generated by a manager into a component of timing variation around the benchmark and a selection component that better isolates security selection skill. Some historical research shows efficacy of selection returns as a predictive metric, while timing returns have usually shown to be a negative predictive component that is non-persistent.[18,19,20] While in the aggregate positive selection return may be outweighed by negative timing return, focusing on managers that are low tracking error with positive selection returns could be a potentially effective strategy.[21]

We confirm the inability of Timing Return and Timing Return Information Ratios to show positive performance persistence. We do show that there is some positive performance persistence at some horizons in some broad asset groupings for Selection Return and Selection Information Ratio, but it is inconsistent. The Selection Return produces the best persistence for the top quartile in fixed income, while the Selection Information Ratio produces the best negative persistence and top-bottom spread within the same category.

The PrecisionAlpha metric, which has underlying components of Selection Return embedded in its construction, is the most consistent predictor of performance across broad category sets at the one-year horizon, being the only effective metric in the two equity categories, while also positive in fixed income.

Figure 3 presents the more relevant three-year horizon out of sample forward returns for the trailing five-year periods of measurement. This horizon is a more typical horizon for measurement, albeit still somewhat short, that could result in manager evaluation and turnover relative to a one-year horizon outcome.

The PrecisionAlpha metric, which has underlying components of Selection Return embedded in its construction, is the most consistent predictor of performance across broad category sets at the one-year horizon, being the only effective metric in the two equity categories, while also positive in fixed income.

Three-year time horizon

FIGURE 3: FIVE-YEAR TRAILING DATA TO PREDICT FORWARD THREE-YEAR RESULTS

	Return	Sharpe	IR	Selection	Selection IR	Timing	Timing IR	PrecisionAlpha
Domestic Category Average								
Top quartile	-0.68	-0.52	-0.46	-0.22	-0.20	-0.43	-0.28	**0.11**
2nd quartile	0.06	-0.02	-0.12	-0.05	0.00	-0.05	-0.12	0.08
3rd quartile	0.21	0.24	0.16	0.09	0.10	-0.05	0.02	0.03
Bottom quartile	0.40	0.34	0.42	0.16	0.10	0.54	0.35	**-0.21**
Top-bottom spread	-1.08	-0.86	-0.88	-0.39	-0.30	-0.96	-0.63	**0.31**
International Category Average								
Top quartile	-0.06	-0.15	-0.07	**0.15**	0.13	-0.17	-0.39	0.09
2nd quartile	-0.07	0.12	0.06	0.10	0.02	0.00	-0.08	-0.02
3rd quartile	0.04	-0.09	-0.12	-0.23	-0.20	0.10	0.35	0.05
Bottom quartile	0.06	0.09	0.12	-0.04	0.04	0.05	0.11	**-0.12**
Top-bottom spread	-0.11	-0.24	-0.19	0.20	0.09	-0.22	-0.50	**0.21**
Fixed Income Category Average								
Top quartile	-0.01	0.01	0.10	0.04	0.00	-0.05	-0.03	**0.12**
2nd quartile	-0.03	-0.03	-0.08	-0.04	0.03	-0.09	-0.06	-0.04
3rd quartile	-0.02	-0.03	-0.08	-0.03	0.00	0.04	0.08	0.00
Bottom quartile	0.06	0.04	0.07	0.02	-0.04	0.08	0.02	**-0.07**
Top-bottom spread	-0.07	-0.03	0.03	0.02	0.04	-0.13	-0.05	**0.19**

Source: MPI Stylus and Mesirow.

The traditional metrics of Return and Sharpe Ratio fail to have positive performance persistence in the top quartile and negative persistence in the bottom quartile across domestic equity, international equity and domestic fixed income. In fact, they show strong reversion tendencies rather than no predictability in the equity categories. Information Ratio has some performance persistence in the top quartile of just the fixed income set, but reversion in the bottom quartile, leading to a weak top-bottom spread. The metric, however, is substantially inverse in the equity categories.

The Timing Return and Timing Information Ratio metrics are upside down on average across the three broad category sets. The Selection Return has mixed results with the best positive persistence in the broad international equity category set and a positive top-bottom spread in that set, as well as the broad fixed income category set, albeit weak. It is sharply inverse on average in the broad domestic equity category set. The same is largely true for the Selection Information Ratio, but with no positive persistence in the top quartile for the broad fixed income category set and weak negative persistence at the bottom.

The PrecisionAlpha metric, once again, is the most consistent predictor of performance across broad category sets, being the only measure that is positive across the three broad category sets with positive persistence at the top and negative persistence at the bottom. The efficacy of this metric appears to build with horizon, such that the annualized excess is higher at the top and lower at the bottom for the three-year horizon relative to the one-year horizon.

Conclusion

The most widely used metric of trailing return is a poor predictor of performance at both the one-year and three-year horizon, with strong mean reversion tendencies in the two broad equity category sets, meaning that the best on a trailing basis is the worst on a forward basis.

The risk-adjusted return metrics of Sharpe Ratio and Information Ratio do not fare much better. The Information Ratio showed some potential efficacy in fixed income at both horizons. However, in the case of these metrics overall, the disclaimer that *past performance is no guarantee*

of future results is meaningful. While they may be predictive, on average it is in the wrong direction. In fact, the historical failure of these traditional metrics in their ability to identify future manager alpha led to the creation of more customized and complex statistical measures.

We demonstrate that Timing Return and Timing Return Information Ratios generally have stronger mean reversion tendencies (rather than persistence) across categories than their Selection Return and Selection Information Ratio counterparts. This facet may be a fundamental contributor to the observed weakness of the overall return-related metrics (i.e., Return, Sharpe Ratio, Information Ratio) that subsume these sub-components of return generation in their total return measurements. We do show that there is some positive performance persistence at some horizons in some broad asset groupings for Selection Return and Selection Information Ratio, but it is inconsistent. The Selection Information Ratio produces the best persistence for the top quartile in fixed income, while the Selection Return produces the best negative persistence and top-bottom spread within the same category, both at the one-year horizon. At the three-year horizon, these metrics worked relatively well in the broad international equity category set, but curiously were very inverse in the broad domestic equity set.

The PrecisionAlpha metric, Mesirow's answer to the shortcomings of traditional performance metrics, is the most consistent predictor of performance across broad category sets for both one-year and three-year horizon periods in our study.

Of course, past performance is no guarantee of future results, but historical positive results are better than historical negative results.

Appendix | About PrecisionAlpha

PrecisonAlpha is a quantitative manager evaluation methodology designed to identify and measure persistent manager skill better than traditional performance metrics. It relies on a combination of returns-based style analysis (RBSA) methods, aimed at identifying Selection return, and two distinct Bayesian statistical methods, both to enhance the "signal to noise" of the alpha estimate and to infuse it directly with information about skill relative to peers. The PrecisionAlpha calculation process consists of three phases: initial categorization and style analysis, followed by two Bayesian statistical adjustments.

The first phase combines RBSA and initial fund categorization (from holdings-based analysis prior to the 2024 launch of Mesirow Category), which determines the benchmark set for RBSA and possible final peer groups. The RBSA analysis is a statistical precision-weighted average of single-period and rolling-period regressions that drives the active peer grouping process. The rolling-period RBSA captures dynamic and recent style changes more effectively than the single-period RBSA alone, and the combined measure captures how a manager has acted both long-term and across time. It also seeks to reduce the number of asset classes explaining manager returns, resulting in a more robust custom RBSA style benchmark (the manager "beta"), which in turn provides a more robust estimate of manager alpha.

The second phase of the PrecisionAlpha calculation incorporates beliefs about a manager's skill within a Bayesian framework. To enhance the precision of the first phase alpha estimate, the confidence level for a manager's skill is determined from the statistical confidence that its alpha is significantly different from zero, so no qualitative judgments are required. This confidence level is used in a significance-weighted, Bayesian average of skilled and unskilled alpha components. When we assume a manager has no real skill, its alpha estimate is based only on manager fees rather than historical manager performance.

The last calculation phase is a Bayesian adjustment based on the population of investment managers with similar strategies. When managers are grouped appropriately, academic research demonstrates a phenomenon known as "learning across funds." Conceptually, managers within a carefully constructed peer group can be viewed as individual draws or samples from a distribution of managers with similar characteristics. In a Bayesian framework, accounting for this phenomenon results in alpha estimates being pulled toward the mean. Integrating the additional information contained in peer group statistics, the statistical precision-weighted, Bayesian average of the second phase alpha and its peer group average adjusts the alpha estimate closer to its "true" (but unknowable) value, while reducing its statistical error in the process.

PrecisionAlpha is currently applied to analyze and rank approximately 70 peer groups. The granularity with which the process subdivides the universe enhances the likelihood that the managers within each group are suitable for "apples to apples" comparisons, and in turn enhances the integrity of both the peer groups and the absolute and risk-adjusted performance rankings that they generate.

About Mesirow

Mesirow is an independent, employee-owned financial services firm founded in 1937. Headquartered in Chicago, with locations around the world, we serve clients through a personal, custom approach to reaching financial goals and acting as a force for social good. With capabilities spanning Global Investment Management, Capital Markets & Investment Banking, and Advisory Services, we invest in what matters: our clients, our communities and our culture.

Mesirow Fiduciary Solutions creates institutional investment solutions for the retirement and health savings marketplace, supporting plan sponsors, financial advisors, recordkeepers and administrators, trust companies and insurers.

To learn more about our methodology and how it may be able to increase potential manager alpha for your plan going forward, please contact Mesirow Fiduciary Solutions at fiduciaryinquiries@mesirow.com or visit mesirow.com/fiduciarysolutions.

1. William N. Goetzmann, Roger G. Ibbotson, "Do Winners Repeat", The Journal of Portfolio Management Winter 1994, 20 (2) 9–18 DOI 10.3905/jpm.1994.9 | 2. Roger Ibbotson & Amita Patel, 2002. "Do Winners Repeat with Style?," Yale School of Management Working Papers ysm253, Yale School of Management, revised 01 Feb 2003 | 3. Mark M. Carhart, "On Persistence in Mutual Fund Performance", Journal of Finance Volume52, Issue1 March 1997 Pages 57–82 | 4. Keith Cuthbertson, Dirk Nitzsche, Niall O'Sullivan, "Mutual Fund Performance: Measurement and Evidence", Financial Markets, Institutions & Instruments, Volume19, Issue2 May 2010 Pages 95–187 | 5. By Kirsten Grind, Tom McGinty and Sarah Krouse, "The Morningstar Mirage: Investors everywhere think a 5-star rating from Morningstar means a mutual fund will be a top performer—it doesn't" Wall Street Journal Oct. 25, 2017 | 6. Mark M. Carhart, "On Persistence in Mutual Fund Performance", Journal of Finance Volume52, Issue1 March 1997 Pages 57–82 | 7. Amit Goyal and Sunil Wahal, "The Selection and Termination of Investment Management Firms by Plan Sponsors", The Journal of Finance, Vol. 63, No. 4 (Aug., 2008), pp. 1805–1847 | 8. James X. Xiong, Thomas M. Idzorek, and Roger G. Ibbotson, "The Diminishing Role of Active Mutual Funds: Flows and Returns", Journal of Investment Management, Quarter 4 2023 | 9. Kosowski, Robert. 2006. "Do mutual funds perform when it matters most to investors? US mutual fund performance and risk in recessions and expansions." Working paper, INSEAD | 10. Moskowitz, T. J. 2000. "Mutual fund performance: An empirical decomposition into stock picking talent, style, transactions costs, and expenses: Discussion."Journal of Finance, 55:1695–1703 | 11. Pástor, Ľuboš, M. Blair Vorsatz. 2020. "Mutual Fund Performance and Flows during the COVID-19 Crisis." The Review of Asset Pricing Studies, 10(4), 791–833 | 12. In this case the Information Ratio is total return divided by total risk, which is category agnostic and does not require switching the primary benchmark for each category. | 13. Anna von Reibnitz, 2017. "When Opportunity Knocks: Cross-Sectional Return Dispersion and Active Fund Performance." Critical Finance Review, 6: (2), 303–356 | 14. Wermers, Russ. 2000. "Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses." Journal of Finance, 55, 1655–1703 | 15. Wermers, Russ. 2021. "Active Investing and the Efficiency of Security Markets", Journal Of Investment Management,19(1), 5–24 | 16. We ran all of the domestic categories against both the S&P and Russell benchmark sets because of divergent size and style cut-offs and then averaged the results in our category selection process. On average, across the domestic large cap categories and utilizing both benchmark sets, 82.6% of managers underperformed over the last ten years after fees. | 17. A one-year horizon implies the ability and willingness to switch managers on an annual basis. Even if there are no transactions costs for participants, there still can be associated costs for plan sponsors or advisors. | 18. Keith Cuthbertson, Dirk Nitzsche, Niall O'Sullivan, "Mutual Fund Performance: Measurement and Evidence", Financial Markets, Institutions & Instruments, Volume19, Issue2 May 2010 Pages 95–187 | 19. W. Jiang, "A nonparametric test of market timing", Journal of Empirical Finance (2003) | 20. Connie Becker , Wayne Ferson , David H. Myers, Michael J. Schill. "Conditional market timing with benchmark investors", Journal of Financial Economics Volume 52, Issue 1, April 1999, Pages 119–148 | 21. Keith E. Gustafson, Patricia Halper, "Concentrated vs. Diversified Managers: Challenging What You Thought You Knew About 'High Conviction'", Alternative Investment Analyst Review, Q2 2016, Volume 5, Issue 1