Received: 9 November 2018 | Revised: 16 November 2019 | Accepted: 18 June 2020

DOI: 10.1002/ijfe.1981

RESEARCH ARTICLE

WILEY

Uncovering investment management performance using SPIVA data

Imran Hussain Shah[1] | **Hans Matthias Wanovits**[2] | **Richard Hatfield**[1]

[1]Department of Economics, University of Bath, Bath, UK

[2]Business School, Imperial College Business School, London, UK

Correspondence
Imran Hussain Shah, Department of Economics, University of Bath, Bath BA2 7AY, UK
Email: i.h.shah@bath.ac.uk

Abstract

Which performs better, passive or active funds management, a question that both fund managers and academics fiercely debate. Why does fund size matter? These are a number of typical questions that puzzle practitioners and academics alike. To date, the data has been shown to be somewhat problematic. This paper exploits the SPIVA and passive fund datasets with several novel methods in order to build a foundation for unbiased fund performance analysis and comparison. For this, we address a number of questions including: passive versus active management, fund size, time horizon and fund style on performance. We find that in general, passive funds outperform active funds due to lower management costs, larger funds tend to perform better and funds with longer (3+ years) records of accomplishment tend to perform better. Short termism tends to have a significant detrimental effect on performance. We introduce Dynamic Generalized Method of Moments to show that competition has a significant effect on fund performance. Furthermore, this demonstrates that SPIVA data has a significant dynamic panel time series that was largely ignored by prior research. This integrated dataset and associated methods that we illustrate here, provide both academic researchers and industry analysts alike with an environment to investigate and potentially draw conclusions about the fund factors that affect performance without the inherent limitations of the original sources.

KEYWORDS

active and passive funds, alpha, biases, mutual funds, SPIVA

1 | INTRODUCTION

The industry contentious debates on relative merits of active versus passive funds, large versus small funds has many protagonists. We contribute to these debates utilising several innovative methods with SPIVA (Standard and Poor's Index Versus Active) that shows the relative merits of active and passive fund management, size, time horizon and style. There are many issues in the data collection and its impact on the analysis in other approaches, our methods with the SPIVA data navigate around the biases limiting their impact on the analysis. Furthermore, these methods provide the opportunities for broadening future research utilising SPIVA dataset effectively.

Our contribution is to exploit SPIVA dataset, in an unbiased manner, to analyse the factors that affect fund performance. We address the debate on active versus passive fund management; how size, longevity and style factors affect fund performance. Utilising several contextually novel methods such as Generalized Method of Moments (GMM), fixed and random effects with cross sectional and panel analysis to analyse the dataset. This provides a foundation for future research to exploit this dataset with robust tools and techniques. To demonstration our methods, we utilise these empirical methods to show the contrasting performance of different fund timeframe, style and other factors. We show how to address the issues of active versus passive fund management. We consider the effects that fund size in the context of management when considering returns and performance. Finally, we explore the time horizon and fund style to show new interpretations on fund performance.

Many studies do explore the relative benefits of active or passive portfolio management in several ways, examples include Blake, Caulfield, Ioannidis, and Tonks (2017); Bogle (2002); Brown and Goetzmann (1995); Holmes (2007); Malkiel (1995); Sharpe (1991). Some existing literature considers the importance of fees in fund performance, hence their importance for investors' decision making. Carhart (1997), Indro, Jiang, Hu, and Lee (1999) and Kinnel (2016) found inverse association between fund's fee and its returns and that larger funds can reduce costs through exploiting returns to scale. There are many factors that affect fund performance and its success or failure. We investigate a some of these factors in the context of removing biases from the data.

Overall, our objective is to demonstrate that the use of SPIVA dataset can follow best practice by avoiding potential biases. These biases could impact the findings of previous studies concerning investigations into: why passive funds outperform active funds or why fund size, NAV,[1] matters regarding performance, what style factors really matter in fund performance. We address these biases with a range of integrated methods for this type of analysis. Furthermore, we uniquely address the dynamic relationship between performance, return index and macroeconomy to obtain more meaningful analysis than prior research.

We illustrate our contribution with two propositions to illustrate our empirical methods to demonstrate unbiased analysis of SPIVA data. The first proposition; do passive funds outperform active funds and if so why? The second, does fund size matter regarding outperformance or underperformance[2] considering fund style? Previously, using SPIVA scorecards without correcting for all of the biases could possibly lead to somewhat problematic results, particularly when considering active and passive fund comparison with style factors. Our methods applied to the SPIVA data supplemented with other data provides a basis for investigating some largely overlooked questions in the active and mutual fund literature.

Blanchett and Israelsen (2007) identify several biases to the SPIVA dataset that could materially affect analysis, these being: Survivorship bias, Equal weighted performance measurement bias, Incorrect comparison to index/style bias, single index comparison bias, Index funds, ETF and others inclusion bias, Fund fees bias, Indices and index funds comparison bias and finally, different timeframes comparability bias. Our methods, that we illustrate here, address these biases to the extent that they have little impact. Little material impact on the results and analysis.

Although these biases exist in the data and have largely were identified by prior research, the still industry regards SPIVA scorecard as one of the most accurate fund management reporting datasets generally available. It provides the industry with a basis to benchmark relative fund performance amongst their peers in the industry. Furthermore, it is a major input to financial media analysis and reporting (Foley, 2016; Schroders, 2014). Therefore, if our methods improve the quality of the analysis then this materially improves both public information and provides a basis for further exploration.

The methods we employ include System GMM, a technique not used with SPIVA data. Using System GMM allows us to demonstrate the dynamic nature of some of the panel data with some interesting and thought-provoking results. We run a series of tests comparing the performance without transformation/correction and with combining rules and methods to confirm our assertions and to demonstrate that the comprehensive dataset addresses and methods we propose here work consistently and reliably. Furthermore, we provide guidance on how to prevent spurious evaluation results when using SPIVA data.

Our tools and techniques promote reuse by other researchers where future updates are unlikely to necessitate modification of the core empirical analysis. Furthermore, the SPIVA scorecards methodology is available for Australia, Europe and other geographies. With limited modification, our tools could be used on these other datasets. SPIVA scorecards provide the opportunity to explore the active versus passive debate in investment management and our tools enable unbiased analysis with a high degree of accuracy (Standard & Poor's, 2016a).

The paper proceeds as follows: Section 2 reports a brief review of the origins of the active versus passive management datable and investigate the role of the SPIVA scorecard in the debate, Section 3 discusses the biases and how this SPIVA scorecard (research) addresses them, Section 4 empirically evaluates the Large versus Small fund proposition, Section 4 evaluates the Passive

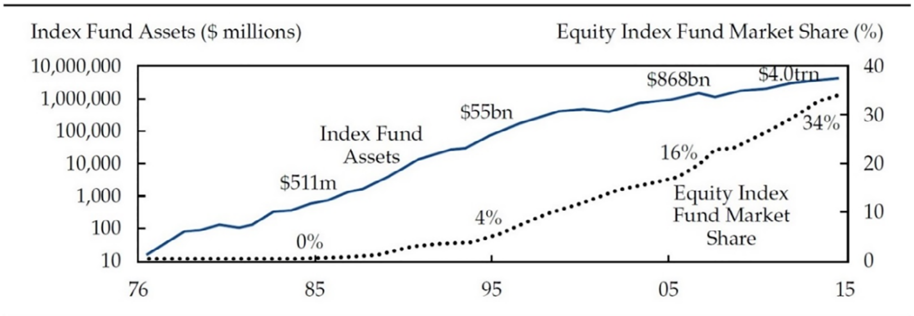

FIGURE 1 Growth in passively invested assets and market share in the USA [Colour figure can be viewed at wileyonlinelibrary.com]

versus Active proposition, this is followed by a discussion in Section 5 and concludes in Section 6.

2 | LITERATURE REVIEW

2.1 | Active versus passive management

With the world is witnessing a rapidly growing investment management industry and the need to justify the costs makes the debate over active fund management being beneficial over passive funds poignant to both investor and manager alike. This is particularly brought into sharp focus since the 2008 financial crisis and the historically low returns from "safe" assets. There are highly divergent opinions on how to invest money and allocate assets either actively or passively. This divergence combined with the rapid growth of assets being managed leads debate over active versus passive Investment Management in a new direction. Figure 1 illustrates this point showing significant flows over the last 40 years into passively managed funds in the United States (USA) The passive fund market share is more than one third of the USA equity markets. This clearly reflects the fact that investors are recognising the long-term benefits of passive investment over active therefore to reallocating their assets accordingly.

The running debate of active versus passive investment management is not fully explored in the current literature, particularly using the SPIVA dataset. Core is that active investment management seeks to outperform the market by identifying, though research and analysis, assets that perform better than the market, thus "beating" the Efficient Market Hypothesis (EMH) (Fama, 1965 & Fama, 1970). However, research is costly and assets with high returns tend to be more volatile. Managers of this school of thought analyse the market in order to identify and purchase investments that are undervalued, effectively trying to exploit market inefficiencies with price discrepancies. Therefore, one of the core tenors for active management is that the price does not reflect fully the information that is available to the market, thus managers may exploit an asymmetry and gain an advantage over the market thus contradicting the EMH. Following Fama, we define three strengths of market efficiency, (weak, semi-strong, and strong). Weak is of little interest, more that the semi-strong information set includes all past and current information, however, future cannot necessarily be accurately forecasted by the market, hence price might not reflect the future, as the information set is deficient.

Alternatively, passive portfolio management bases its decisions on the factors affecting an independent measure, normally an index. It will use the same assets with the same weighting and normally the same time periods for review as the benchmark index or measure (Fuller, Han, & Tung, 2010). However, passive management should not be confused with a "buy and hold" approach. Managers of passive funds need to act on corporate actions such as dividend payments, secondary issues, or mergers and acquisitions to match the methodology of the index they are tracking. The term "Indexation" is thus arguably the better description of this management style. In some cases, investment managers may follow sampling approach where mathematical models refine the index or measure to assess the market risk and performance constraints to identify possible risky denominators in passive investing.

A significant part of the literature uses performance comparisons of a specific set of funds versus arbitrarily selected benchmarks over multiple time periods. Researchers (such as Carosa, 2005; Malkiel, 1995; Wermers, 2000) generally report the mean performance of funds and their respective benchmarks concluding when active or passive funds are more preferable considering the market conditions. The nature of these studies tends to draw on different factors, models and benchmarks that make comparison suspect. Thus drawing compelling, consistent and reliable conclusions that stand scrutiny somewhat difficult, if not impossible (Holmes, 2007).

Starting with Malkiel (1995) and the argument that passive funds net returns outperform those of active funds. Malkiel finds that is the gross return advantage gained by active fund management is eliminated by the costs (fees) imposed either directly or indirectly on the investor. In most cases the payoff from active management does not offset the costs, time and effort of managing actively. Largely, costs are fixed (i.e., not proportionate to the returns) amplifying the effects of any downside and attenuating the upside. As in Kinnel (2016) the scale of fund fees are a strong predictor of funds likelihood of outperforming indicating some economies of scale possibly could rule. This represents potentially a significant barrier to entry and may consolidate the managers of such services. Furthermore, Indro et al. (1999), finds that active management requires significant analysis and research utilising specialist, often expensive, staff and resources. Therefore, some economies of scale and scope must rule with large funds, that is high Net Asset Values (NAV). Large funds achieve economies by lowering transactional costs, synergies in the research and efficient resource utilisation/sharing, consequentially, a reduction in fees that can be passed onto the investor. This implies that one should observe larger active funds potentially outperforming rather than smaller ones.

For passives, size might not matter as there is no such need for such research and analysis. Their actions are limited to a set of prescribed portfolio adjustment timing to follow the index. Sharpe (1991) explains that in a market only populated by active fund managers, by construction, 50% of the active managers will outperform the benchmark index, before fees. However, real world imperfections such as the existence of fees and the fact that there are other market participants, result in deviations from this hypothetical split.

This paints a dire view of active management "value" to the investor. Carosa (2005) counters with the benefits of active management far outweigh the costs with their ability to exploit opportunities and to protect investments in "at risk" situations. Earlier, Sharpe (1991) could explain some of this by the hypothesis that if the market only populated by active fund managers, by construction, 50% of the active managers will outperform the benchmark index, before fees. However, markets are rarely "pure" with the existence of fees and other market participants, results in significant deviations from this hypothetical "pure" market view.

A somewhat less extreme view to Carosa is occasionally, active management has an edge under certain circumstances. Others present a balanced picture giving active the occasional edge and may give better returns if they can control costs (Blanchett & Israelsen, 2007; Holmes, 2007). However, Bogle (2016), Malkiel (1995) and Jensen (1968) concur with the proposition that, in the long run, passive management outperforms active management. To note that there are several studies in aspects of fund management such as the persistence in returns of active management (Brown & Goetzmann, 1995; Carhart, 1997), and the drivers of active fund performance (Nihar & Murty, 2010).

A recent exploration into active fund management, strategies and behaviours by Fasano and Boido (2017). identifies that behavioural biases can significantly handicap investment strategies. This becomes evident in times of significant events such as the China crisis (2015) drawing several similarities to that of investment manager's behaviour in the Brazilian, Russian, Indian crises previously. They find that strong behavioural traits such as over-confidence (confirmation bias), framing and priming compounded with other psychological behvaiours, significantly weakens active management strategies, hence returns.

Still, a compelling argument either way is not self-evident (Holmes, 2007). The plethora of comparisons funds to a variety of benchmarks over a multitude of time periods that purport to address the active versus passive debate just adds to the confusion. Holmes concludes that there are no consistent, repeatable, reliable and transparent methods to draw any meaningful conclusions. We aim to assist with clarification of some of the evidence within active and passive management debate.

2.2 | The SPIVA scorecard and investment management

As previously mentioned, active versus passive studies lack comparability due to differing inputs. SPIVA scorecards provide a possible route to aiding comparisons between fund types. At their core, SPIVA scorecards compare active and passive performance with a level of consistency not found in other datasets. This scorecard specification allows investors to compare the aggregate returns of active managers to broad market indices. This enables them to decide if the potential benefits of active management outweigh the costs.

The investment management industry considers SPIVA scorecards to be the most accurate aggregate performance specification. Rigorous compilation and frequent updates have continued to enhance its reputation and usefulness to asset managers for performance assessment and benchmarking. Additionally, media, commercial literature and specialist commentators often cite scorecards as an accurate data source (Foley, 2016; Schroders, 2014). With the current popularity of such scorecards, we aim to investigate whether the SPIVA USA scorecards are constructed with the same effort as indicated in existing body of literature or if the dataset may have some dubious potential biases limiting the usefulness. We already note that the value of many studies pre-2000 is somewhat limited by the inherent biases though inconsistent data collection and compilation (e.g., Carosa, 2005; Jensen, 1968; Malkiel, 1995).

There are various pitfalls that can severely impact the validity of results in active versus passive studies and therefore it is equally true to SPIVA scorecard (Blanchett & Israelsen, 2007). However, despite their role in the active versus passive debate, the methodologies leading to the results presented in SPIVA scorecards have never been rigorously researched and investigated. This research contributes by categorising potential biases, how to avoid them and the potential impact on the findings of previous studies. Core is investigating whether SPIVA scorecards follow best practice in managing these biases or not. Our research attempts to address these issues/biases with an empirical approach using SPIVA data, correcting for the biases to produces reliable, repeatable results and demonstrate the methodology.

3 | THE SPIVA SCORECARD METHODOLOGY AND DATA ISSUES

The SPIVA scorecard dataset evaluates the performance of actively managed funds compared to benchmark indices. Initially started in 2003 the SPIVA dataset receives regular updates half year updates of the data and continues to expand methodology (Bogleheads, 2017). Some of these updates are in response to developing issues in both collection and evaluation, however they add to rather than change for methodology. Its foundation is the survivor-bias mutual fund returns from the free monthly CRSP database(USA). One of the core characteristics of SPIVA data is that the data updates are consistent changing the methodology, permitting long timeseries analysis. This is a benefit over the more traditional snapshot nature of many other studies.

In studying the performance of active funds, that is returns versus benchmark, researchers need to be very careful of any systematic biases into the data. Avoiding such biases can be complex as they can arise during all stages of the process including performance measurement and performance comparison. Here, the paper summarises the biases this research addresses. The principle source of bias identification and characterisation for this research comes from the work of Blanchett and Israelsen (2007). Furthermore, consideration of Carosa (2005) work in this area assists with illustrating the biases, however Blanchett and Israelsen is the principle source that we summarise here.

3.1 | Survivorship

This occurs when lapsed funds are excluded from the estimate of aggregate fund returns. This has the effect of over estimating the returns creating an upward bias by excluding poorly performing funds prior to closure. The main reason for fund closure is investor capital withdrawal after poor returns. This effectively cuts the lower part of the distribution of fund performance. Carhart, Carpenter, Lynch, and Musto (2002) quantified the size of survivorship bias of approximately 7 bps higher over 1 year, and 85 bps over 15 years.

Popular techniques amongst researchers includes "Reachback"[3] that excludes funds that have terminated prior to the date causing this bias. Although the Centre for Research in Security Prices (CRSP) seeks to address this bias, critics such as Elton, Gruber, and Blake (2001) indicate omission of data though discretionary reporting requirement creates a similar effect.[4] This is particularly evident in 1960's–80's, hence, SPIVA excludes this problematic period concentrating on post 2000 data. This research follows the same regime.

3.2 | Equal weighted performance measurement

This bias occurs when studies only report the fund manager performance as an arithmetic average rather than weighting the average by factors that affect both performance and decision making. In a market dominated by a few large firms with market power then the small firms are put at a considerable disadvantage. Large firms may negotiate favourable terms whilst attracting investors by their mere presence and perceived reputation, a key element in investment decision making. This creates a heavily skewed to right and a left long right tail distribution where wealth concentration is in few large companies. Small firms are disadvantaged by such things as price taking on investing terms, research costs and client turnover rates, fear of mass withdrawal making them less stable. A self-reinforcing cycle when combined with perceived reputation. Economies of scale furthermore puts small firms under greater cost pressures dragging own their performance, hence the average of the market. This would exaggerate outperformance over average of larger firms further exacerbating the large firm bias. SPIVA resolves this by reporting both arithmetic average and a weighted average on Investment Net Asset Value (NAV) a proxy for fund size. Such weightings largely address the large firm bias.

3.3 | Incompatibility between fund and index

Most funds specialise their investment activities enabling them to conduct much more thorough research and gain significant market knowledge. Active funds returns are normally compared to a relevant and pertinent index to assess their over (or under) performance in relation to the passive market. This eliminates claims of fund manager outperformance that are more market driven rather than their own investing activities. Selecting a suitable index or indices can be problematic on two counts. First is the composition of the index may include asset classes that are not relevant to the fund being compared. Second, it is common for indices to occasionally change the composition that might make them less reliable benchmarks for the fund. A change in composition may make a comparison less reliable especially where funds have limited specialisations. S&P address this bias by matching and pairing Lipper fund database classifications with the CRSP database returns (Standard & Poor's, 2016a). This provides a near perfect cross match between the classifications and S&P indices. Furthermore, the Lipper database classifies fund holdings rather than what the fund espouses, eliminating another potential bias.

3.4 | Single index benchmark comparison

Indices, by their very nature, may have substantially different methodologies that affect the index movements compared to the underlying market (Fuller et al., 2010). Although the index and fund may be comparable, the choice of index may affect the over or under performance comparison potentially leading to potentially spurious results. The index is a proxy for the real market and ideally, funds would be compared to multiple comparable indices. However, this introduces a weighting problem to construct the measure and likely to be subjective by its very nature. As in Fuller et al. (2010), all major indices are reasonable proxies for the market and fund comparisons, if they are consistent, are valid to use in SPIVA analysis.

3.5 | Indices and index funds comparison

Active funds are normally benchmarked against a comparable index. Indices provide absolute returns, whereas the argument presented by Blanchett and Israelsen (2007) that equivalent passive funds would represent a better comparator. This ignores the variations deriving from fees, valuation errors and the specific characteristics of the fund. Hence, there is a trade-off between absoluteness of the index and the proxy accuracy of the index. Naturally, S&P, the compiler of the SPIVA, would tend toward indices of their own making rather than any other benchmark index. The SPIVA is consistent in is application, although some accuracy may be sacrificed, the differences are likely to be marginal and inconsequential.

3.6 | ETF and other fund inclusions

Inclusion of funds that use methodologies that link composition decisions to something other than investment manager's research and evaluation (such as index linked funds) may be problematic. This introduces a potential bias the average performance includes these funds. SPIVA excludes all leveraged and indexed derived funds to eliminate this bias.

3.7 | Fund fees

Fund managers tend to present gross returns rather than inclusive of fee returns with fee rates varying considerably across the funds and investors. However, investors are only interested in net position as it tends to be more representative of the performance of funds.[5] Some fund fees are fixed (NAV or other basis) regardless of returns skewing the risk to the investor.

SPIVA does not directly address this bias, however this research compares 2015 SPIVA and the impact of fees on mutual funds (Standard & Poor's, 2016b) enabling calculation of net returns. For multiple share classes, SPIVA reports only the fee structure for the highest NAV class in the fund, introducing some bias. This bias may not be material to the average investor as they are likely to place most of their investment in the highest NAV class.

3.8 | Differing timeframes

Certain fund management strategies may materially benefit or be of detriment at different timeframes depending on market volatility or if the market is in either an up or down cycle (Philips, 2008). This is particularly important when a study timeframe is short or the underlying events in comparing study timeframes are significantly different. For any study to be reliable, then consideration of environmental and timeframe factors that affect the comparison of time periods. SPIVA limits time frame effects by regular by half yearly updates.

3.9 | Eliminating biases and empirical testing

Confirming robustness of both the SPIVA dataset and the methods we propose, in the face of such biases, necessitates several empirical tests. In summary, some biases can be eliminated, firstly by time, such as Survivorship by excluding data prior to 2000 in this specific setup. Secondly, by type, namely ETF and other funds bias, then exclusion of all leveraged or indexed derived funds seeks to eliminate this bias. Thirdly, by report returns, namely fund fees bias by only using net returns (derived by alternative sources) and eliminating gross returns. The differing timeframes bias is addressed by ensuring frequent (half yearly) updates on performance. Others, as indicated above, are of little consequence to the results and analysis. The principle biases that this research seeks to address are Equally Weighted Performance bias and the fund performance relative to benchmark (indices and index fund comparison and single index comparisons) which cause the most difficulty in evaluating fund performance.

4 | ANALYTICAL FRAMEWORK

4.1 | Large versus average fund investment

4.1.1 | Equal and asset weighted alphas

The first empirical testing is to investigate SPIVA's reporting of fund outperformance within category over its benchmarks considering the Equally Weighted bias (Section 3.2) and its effects on reporting accuracy. Generally, equally weighted and asset weighted (the net asset value weighted) returns' benchmark is α (alpha[6]). To determine if there is a benefit from investing in funds using fund size as a method of selection then we need to consider the difference in performance of namely equally weighted and asset weighted performance measures though the α. Therefore, we can understand the effect of the equally weighted performance bias in the context of the combined dataset. For instance, when a large fund performs well, then equally weighted will report a small alpha, (many funds equally contribute to the benchmark), whereas when asset weighted α, the effect may dominate over the smaller funds. This shows the bias factor when utilising equally weighting fund performance.

Table 1 compares the benchmark S&P Composite 1,500 and average funds (All Domestic) alphas to compute the outperformance for 2015.[7] All funds underperform then benchmark in either equally or asset weighted over all time periods. However, equally weighted All Domestic measurements are significantly less than the asset weighted, implying that there is a bias as we have discussed. This implies that the effects of small fund under performance has a significant effect on the equally weighted benchmark. However, this effect is attenuated by asset weighting the performance, hence improving the measure. This is reflected in the last row, where difference is some 30–70% less, depending on timeframe.

From this table, we show that size matters regardless of duration, with large benefiting the most. Intuitively, this would align with one's own thinking and that of prior research that “economies of scale and scope” play a significant part in performance (Carhart, 1997; Indro et al., 1999). Furthermore, large funds tend to have the resources and facilities to access a wide range of information, can have more broadly based specialist analysis and perform more extensive monitoring. One could easily conclude that large active funds should have a marked advantage over other sizes and styles. However, passive funds tend to be comparable returns to that of the best performing large active funds. As passives require much less intervention, then one would suspect that large passive funds would be the outperform the funds management market. The role of active versus passive is a subject of a later analysis in this article.

Next, we consider over time, this bias in the context of the effects of fund style, competition and capitalisation are needed to complete the analysis. Over time, there is a significant turnover of funds and change to their fund style (Papadamou & Sirpiopoulos, 2004). For example, during 2015, 4% of funds were terminated and 10% modified their investment style. Over 10 years, 40% of funds were terminated and very few maintained their investment style. As SPIVA scorecard takes account of these changes by dynamic adjustment in every time period and not perceived by many as a problem. However, to be thorough, we conduct further investigation by comparing specific funds over time. Our investigation is in two parts, a cross sectional view and then a panel regression with the results in the next three tables.

4.1.2 | Cross sectional analysis: Econometrics analysis, data and variables

The first part is the investigation of the differences between equal and weighted alphas on a cross sectional basis, regardless of the SPIVA scorecard. We utilise SPIVA US semi-annual from 2010[8] to 2016 (available on the

TABLE 1 Comparison of equal and asset weighted alphas

	Equally weighted				Asset weighted			
Duration (years)	**1**	**3**	**5**	**10**	**1**	**3**	**5**	**10**
S&P Comp 1500	0.99	14.90	12.39	7.41	0.99	14.90	12.39	7.41
All domestic	−1.93	12.12	9.49	6.20	0.03	13.22	10.47	6.67
Alpha (2)–(1)	−2.92	−2.78	−2.90	−1.21	−0.96	−1.68	−1.92	−0.74

Note: All are per annum percentages for the duration in years for equally weighted and asset weighted funds. The S&P 1500 is a benchmark index consisting of the largest 1,500 US stocks. The All Domestic category shows the equal/asset weighted performance of ALL actively managed US funds. The last row is the difference (2)–(1) indicating the effect of the weighting either equally or by asset.

Standard and Poor's, 2017). We then split the data into two equivalent balanced panel datasets, one equally weighted, the other asset weighted. Each panel has 12 time dimensions and 51 cross sections, totalling 612 observations. Each cross-section represents the alphas for a specific style, size, and timeframe.[9] Benchmark indices derive from Thomson Reuters DataStream services.

Analysis is by two cross sectional regressions using equivalently balanced panel datasets with the respective alphas as the dependent variable. The regression has control/explanatory variables including benchmark index performance. Additionally, dummy variables specifically to test the significance of some characteristics including fund time horizon, fund style and competition (see Table B1 in the Appendix B for dummies specification and description of the variables). Refer to Appendix B.1 for details on the exact description of the variables and model specification. The regression model:

$$\boldsymbol{\alpha_{i,w}} = \boldsymbol{\beta_0} + \sum_{s=\{3.5.10\}} \boldsymbol{\beta_s Y_{i.s}} + \sum_{j=\{g.v.a\}} \boldsymbol{\beta_j S_{i.j}} + \sum_{k=\{mi.sm.mu\}} \boldsymbol{\beta_k C_{i.k}} + \sum_{m=\{c.r,e\}} \boldsymbol{\beta_m F_{i.m}} + \boldsymbol{\varepsilon_i}, \quad (1)$$

where $\boldsymbol{i}$ is the specific fund category, $\boldsymbol{\alpha_{i,\,w}}$ is the alpha for the fund by its weighting type (where w indicates equally or asset weighted), $\boldsymbol{\beta_0}$ is the intercept, $\boldsymbol{Y_{i,\,s}}$is the dummy for time horizons 2, 5, 10 years, $\boldsymbol{S_{i,\,j}}$ is the fund style factors, $\boldsymbol{j} = \{\boldsymbol{g}, \boldsymbol{v}, \boldsymbol{a}\}$ is growth, value and aggregate respectively, $\boldsymbol{C_{i,\,k}}$ is market capitalisation, $\boldsymbol{k} = \{\boldsymbol{lg}, \boldsymbol{mi}, \boldsymbol{sm}\}$ for large cap, mid cap and small cap funds and, $\boldsymbol{F_{i,\,m}}$ factors that affect fund performance or measurement $\boldsymbol{m} = \{\boldsymbol{c}, \boldsymbol{r}, \boldsymbol{e}\}$ for competition (SPIVA), return, and a dummy if $\boldsymbol{\alpha_{i,\,w}}$ is equally weighted. Each cross section is time period making up the panels.

Running five regressions, the first with only the time horizon dummies and intercept the rest includes several control variables. The first regression follows:

$$\boldsymbol{\alpha_{i,w}} = \boldsymbol{\beta_0} + \sum_{s=\{3,5,10\}} \boldsymbol{\beta_s Y_{i,s}} + \boldsymbol{\varepsilon_i}, \quad (2)$$

Table 2 are the results from comparison equally and asset weighted fund performance using a number of factors that may or may not influence alpha. Conduct of this experiment is by using 17 comparison categories of SPIVA score cards, controlling for selected characteristics and using the comparison between equally and asset weighted alphas.

The first regression (1) looks principally at fund time horizon. The results indicate that funds underperform their benchmark by 2.52% (intercept) in the first year and outperform benchmark in subsequent years ranging from 0.6% at 3 years to 1.2% per annum at 10 years ($\boldsymbol{\beta_s}$ where $\boldsymbol{s} = \{\mathbf{3}, \mathbf{5}, \mathbf{10}\}$). The coefficients are positively correlated with the time horizon indicating that it is increasingly difficult for fund managers to achieve higher returns in later years. A word of caution each fund reports at a different time where market conditions may change. This may lead to a spurious conclusion if not careful in the fund reporting timeframes.

Another consideration is that if market performance is positive then fund performance will likely be positive. This raises the question; will an active fund outstrip the proportionate increase in market performance or not? Some studies indicate that active funds tend to outstrip the market (Goldman, 2010; Vayanos & Woolley, 2016) however, the cost of managing those funds may impose a burden and that reduces net performance (Barnes, 2003; French, 2008; Sharpe, 1991). This research shows that, depending on time horizon, both effects are evident.

Next is to consider several other factors that might affect fund alpha including fund size, style, index returns, and competition. We report these combinations in columns (2)–(5) of Table 2. First, in (2) we look at the effect of competition, equally weighting (discussion above) and index return (passives). Competition is only significant at the 10%, whereas both equal asset and index return, for all time horizons, are significant. Furthermore, this result

TABLE 2 Cross-sectional results fund performance with equally and asset weighted $\alpha_{i,\,w}$

Dependent variable: $\alpha_{i,\,w}$	(1)	(2)	(3)	(4)	(5)
Intercept	−2.517***	−3.085***	−3.085***	−3.346***	−3.359***
_3Y	0.574***	0.571***	0.570***	0.570***	0.570***
_5Y	0.926***	1.002***	1.009***	1.010***	1.017***
_10Y	1.199***	1.339***	1.349***	1.360***	1.365***
Large			0.447***		0.459***
Mid			0.036		0.059*
Small			−0.520***		−0.503***
Aggregate				0.127	0.196
Style-value				0.692***	0.684***
Style_growth				−0.008	−0.014
Competition		0.001*	0.001	0.002**	0.001
Equal_asset		0.445***	0.445***	0.445***	0.445***
Index return		0.020***	0.022***	0.021***	0.024**
R^2	0.06	0.09	0.13	0.12	0.16
Adjusted R^2	0.06	0.09	0.12	0.11	0.15
F-test p-value	.000	.00	.00	.00	.00
Observations	1,394	1,394	1,394	1,394	1,394

Note: ***, ** and * indicates 1, 5 and 10% significance with White adjusted standard errors. (1) Represents regression coefficients for time horizons only, (2) adds the competition, equal/asset with and index return with time, (3) shows variables in (2) with size, (4) indicates all variables except size and (5) includes all variables. Adjustments for heteroskedasticity is by reporting White adjusted standard errors.

is reinforced by regressions (3)–(5) where competition is insignificant (except in [4]) and time horizon, equal asset and index return remain highly significant and there is little change to their coefficients. This indicates that with long time horizons, and asset weighting tend are more likely to generate excess returns. However, competition only seems to have significance when fund size is not considered (regressions [2] and [4]) and with fund size (regressions [3] and [5]) being significant and consistent implies that competition has little impact on alpha. In regression (5) large funds, that focus on value, with long time horizons and asset weighting have a significant advantage over small funds. Furthermore, we find that, counter to the view that efficiency in investment firms is driven by competition, competition plays no role in fund performance.

Following our earlier discussion on fund size, we observe that smaller funds tend to perform worse (−0.503) than larger funds (0.459) indicating that size really matters, confirming our earlier predictions and intuition. Furthermore, the dummy, equal asset (β_e) is both positive and significant confirms that when using equally weighted, a bias causes the overstatement of smaller fund performance and the understatement of larger funds. This supports the view that asset weighted alpha is reliable and contains little bias particularly compared to equally weighted. Interestingly, competition has negligible effect on fund performance. The expectation is that increasing competition should boost fund performance (i.e., a positive coefficient), however, we find no evidence supporting that view.

As discussed in Section 3.2, SPIVA scorecards report outperformance of each fund category to its relevant benchmark on an equal and asset-weighted basis. We report the results in Table B1 in the Appendix B on the benefits of investing into large funds compared to the average fund for completeness.

If we consider net returns from high Net Assets funds are likely to be greater than equivalent smaller funds imply that there are significant investment advantages with larger funds. Those funds may be able to exploit economies of scale, wider breath of coverage, access to information and with a wider investor base, less likely to be exposed to individual investor's withdrawal risk.

Competition between large funds implies that most of the benefits from cost savings and competitive advantage would be passed onto the investor in the form of higher returns. This suggests that smaller funds are at a significant disadvantage in both attracting and retaining investors. The SPIVA data does not directly identify Net Assets, therefore would seem to be impossible to answer this question. This analysis suggests that this impossibility can be answered by exploiting the difference between equal and asset value weighting to indicate fund size, thus infer that the hypothesis size matters has support of the data from the SPIVA database. It is evident that those funds with large Net Assets outperform generally outperform those with lesser assets. The exact nature of what drives large fund outperformance over smaller similar funds is a subject for later research.

4.1.3 | Panel analysis: Econometric analysis and data description

The next question is does index return and competition have a positive or negative effect on alpha. We use a reduced form of the above cross-sectional regression (2) in the form of a time series in dependent and independent variables to test these influences. A balanced panel data regression model is used to capture the impact of above variables on alpha using semi-annually data from mid-year 2010 to year-end 2016 for 51 cross-sections. Therefore, alpha will vary over time within fund category. This makes fund performance measurement over time accounting for competition and index return. Fund size is captured by the proxy of equally and asset weighted alphas. The specification of the reduced form is:

$$\boldsymbol{\alpha_{i,t,w}} = \boldsymbol{\beta_{i,0}} + \boldsymbol{\beta_c F_{i,c,t}} + \boldsymbol{\beta_r F_{i,r,t}} + \boldsymbol{\beta_g GDP_t} + \boldsymbol{\varepsilon_{i,t}}, \tag{3}$$

where $\boldsymbol{\alpha_{i,\ t,\ w}}$ is the alpha for fund category $\boldsymbol{i}$, with weighting either equal or weighted Assets and at time $\boldsymbol{t}$. Fund factors, $\boldsymbol{F_{i,\ c,\ t}}$ and $\boldsymbol{F_{i,\ r,\ t}}$ are for the competition and S&P benchmark index returns respectively. The fund category $\boldsymbol{i}$ is within one of the 51 cross sections of funds. $\boldsymbol{GDP_t}$ is the growth rate of real GDP.

The we report in Table 3 the regression results for pooled, fixed and random effects. The coefficients for competition, $\boldsymbol{F_{i,c,t}}$ are small and insignificant in both pooled and random effects, thus is not a factor in fund performance within a category. This contradicts prior research by Nihar and Murty (2010). They find strong association between competition and fund performance. In contrast, the fixed effects competition coefficient is negative and significant indicating that increasing competition has a negative effect on fund performance. Intuitively, small market saturation effect may cause funds to try and differentiate themselves.

Several studies report the importance of fees on fund performance and, as a consequence, investors' decision making (Carhart, 1997; French, 2008; Khorana, Servaes, & Tufano, 2005; Malkiel, 2003). They identify that there is a negative relationship between a fund's fees and

TABLE 3 Regressions results for pooled, fixed effects and random effects

	Equal-weighted alpha			Asset-weighted alpha		
	Pooled	**FE**	**RE**	**Pooled**	**FE**	**RE**
Intercept	−2.954***	−1.452***	−2.977***	−2.241***	−1.150***	−2.256***
Competition	0.0001	−0.004**	0.0001	0.0001	−0.003**	0.0001
Index return	0.020**	0.044***	0.008***	0.004	0.021***	0.005
GDP growth	0.196	0.246**	0.196	0.170	0.209*	0.171
R^2	0.02	0.21	0.02	0.01	0.16	0.01
Adjusted R^2	0.01	0.14	0.02	0.00	0.08	0.00
F-test *p*-value	0.02	0.00	0.01	0.40	0.00	0.38
Log-likelihood	−1,213.60	−1,144.62		−1,206.60	−1,155.91	
Hausman test		63.80***			35.71***	
LR test	366.50***			331.77***		
Breusch-Pagan LM		3,423.43***			25.47.02***	
Pesaran LM		42.54***			25.19 ***	
Bias-corrected LM		40.54***			22.87***	
Pesaran CD		44.40***			24.37***	
Observations	612	612	612	612	612	612

Note: ***, ** and * indicate the level of significance at the 1, 5 and 10% respectively. Panel level heteroskedasticity test is the likelihood-ratio (LR) test. The standard tests for unit root with the results reported in the Appendix A, Table A1 as are all the other tests.

its returns. That is fees will reduce gross returns or potentially, if fixed fees, then make net returns negative. As with Indro et al. (1999), we find larger funds can exploit both economies of scale and scope to both maximise their profits and returns to investors. As a result, large funds have a significant marketing and retention advantage over an equivalent smaller fund. One aspect of smaller funds is liquidity risk. A less diverse investor profile coupled with the impact of withdrawals substantially increases the liquidity risk to smaller funds. Evidence by Ben-Rephael (2017) that smaller funds tend to have a shorter life through insolvency, particularly in downturns. Larger funds with proportionately lower turnover of investors and the depth of the portfolios, would needs to retain less proportionate liquidity. Hence, large funds can deploy more funds to productive investment and not incur a penalty for holding liquid assets to mitigate liquidity risk.

We use the Poi and Wiggins (2001) to test panel-level heteroskedasticity in pooled regression. Table 3 (LR test) indicates that there is pooled heteroskedasticity invalidating the pooled regression model. Using the Hausman test to determine if the data supports fixed or random effects, we find that fixed effects is the more appropriate for these specifications. We apply cross-section dependency (CD) (Breusch & Pagan, 1980), Lagrange Multiplier (LM) and Pesaran (2004) Scaled LM and CD, biaed-corrected scaled LM (Baltagi, Feng, & Kao, 2012) to ensure results from Fixed effects are consistent and valid. We find that they are not, hence we look to use GMM to estimate the equations.

This introduces another problem, endogeneity,[10] which could invalidate any results We turn to the literature to resolve this matter, particularly, Ketokivi and McIntosh, (2017)[11] that provides a strategy GMM to address endogeneity in the panel data. Furthermore, we utilise Arellano and Bond (1991) for a one step GMM and the Arellano and Bover (1995) two step GMM forward orthogonal deviations (FOD) transformation to minimise the possibility endogeneity in panel analysis. We find that these two approaches provide us with the certainty that endogeneity is not relevant to this study. We confirm our results by using AR(1) and AR(2) tests to check for autocorrelation using Arellano and Bond (1991) methods. Our final test is Sargan-Hansen J-statistic test (Hansen, 1982) to check the validity of instruments. The Sargan-Hansen test identifies that the over identification restrictions are valid and the AR(2) confirms that there is no autocorrelation. This is consistent with the findings of Schultz and Tan (2010) that demonstrates GMM standard errors are robust and without heteroskedasticity and autocorrelation.

We use of instrumental variables to create an endogenous structure into what now can be termed a dynamic panel analysis. One of the facets of using GMM is the use of instrumental variables with lags in differences and levels (Arellano & Bond, 1991) commonly referred to as System GMM (Arellano & Bover, 1995). This permits the use of small period panel data, therefore needing a small number of instruments. We ensure the validity of the results using the Hansen/Sargan test (J-statistic) for over-identifying restrictions and to a test of the departure of serial correlation of the residuals. We should observe little difference to evidence that the coefficients are robust. The General Dynamic GMM model specification is;

$$\boldsymbol{\alpha}_{\{i,t,w\}} = \boldsymbol{\gamma}_0 \boldsymbol{\alpha}_{\{(i,t,w)-1\}} + \boldsymbol{\gamma}_1 \boldsymbol{F}_{i,c,t} + \boldsymbol{\gamma}_2 \boldsymbol{F}_{i,r,t} + \boldsymbol{\gamma}_3 \boldsymbol{GDP}_t + \boldsymbol{v}_{i,t}, \tag{4}$$

where $\boldsymbol{\alpha}_{\{(i,t,w)-1\}}$ is a lag operator of the alpha.

Table 4 reports the results from our different GMM transformation methods. There is little difference between table columns 1 & 2 and 3 & 4 (GMM one step, GMM two step) in both significance and magnitude thus confirming our coefficients and results are robust. Note that AR(1) is and AR(2) is not significant at 5% level indicating no evidence of second order autocorrelation.

This confirms that Tables 1–3, namely, index returns and GDP growth are material to fund performance. However, this result also implies that competition has a positive significant impact on fund performance. The dynamic nature of System GMM using instruments lagged alpha ($t - 1$), indicates there is some persistence in Carhart alpha. In contrast to the random effects and pooled in Table 3 where we note the competition insignificance implies that a fixed effects provides a more complete result. In system GMM, competition being positive, as one would expect, is contrary to the results in Table 3 fixed effects. This makes the dynamic specification a more appealing that any static nature of Random, Pooled or fixed effects.

Our results indicate that competition is not affected by the asset weighting method, implying that fund size has no differential effect on performance. However, GPD and index returns are materially different when considering weighting method. With the asset weighting coefficients being somewhat less than equal weighting in both cases indicate that large funds play a significant positive role in performance when considering competition and index returns. We use the Sargan-Hansen test to show that no evidences of incorrect specification indicating overidentifying restrictions are valid and confirming our findings. To sum up, our test statistics hint at a proper specification.

TABLE 4 Regressions results for system Generalized Method of Moments (GMM)

	Equal-weighted alpha		Asset-weighted alpha	
	GMM (1S)	**GMM (2S)**	**GMM (1S)**	**GMM (2S)**
Alpha (−1)	0.322***	0.322***	0.298***	0.297***
Competition	0.013***	0.013***	0.014***	0.014***
Index return	0.072***	0.072***	0.047***	0.046***
GDP growth	0.300***	0.299***	0.285***	0.282***
J-statistic	50.27	50.15	50.79	50.67
J-statistic *p*-value	.34	.35	.33	.33
AR(1) *p*-value	.02		.00	
AR(2) *p*-value	.12		.06	
Instrument rank	51	51	51	51
Observations	510	510	510	510

Note: ***, ** and * indicate the level of significance at the 1, 5 and 10% respectively. GMM (1S) is the one-step GMM and GMM (FOD) is the two-step GMM. J-statistic is the Sargan-Hansen overidentification test.

4.2 | Passive and active fund performance

4.2.1 | Data description and variables

The prior discussion on fees in active funds draws us to the performance of passive and active funds. Expectation is that lower fees for passive funds over that of active funds, with the potential the lack of fee burden increasing net returns making up for the inability to exploit opportunities as they arise. To what degree do active or passive funds outperform each other? We know that size has an impact on active funds, does this change the dynamic with passives?

Over time, active versus passive remains a topic of much controversy both in industry and academia. On the one hand, active managers portray that they take advantage of upswings and protect on downswings, protecting the investor. According to the claim of passive managers, in the long run, it is hard to beat the market (EMH) and whilst active managers may take tactical advantage of market conditions. However, those conditions are rarely long lived and the advantage gained may be far outweighed by the cost of monitoring and managing. Academic research also reflects this juxtaposed opinion. For example, Philips (2008) asserts that an active fund manager is abler to exploit the market by positioning the portfolio aggressively in a climbing market and defensively in a declining market, something that passive funds cannot do. Still, there are significant transactional and research costs with such manipulations complicated by just plainly choosing the wrong strategy. Generally, active fund defensive measures tend to be more effective over the equivalent passive fund in the short run implying a falling market whereas Demos (2010) identifies that active funds outperform passives in a climbing market. These contradictions are doing not permit drawing such simple conclusions that one style outperforms another (Steverman, 2011). Example is that mutual passives tend to have a long horizon performing well during crises periods bond performance improves.

These rather inconclusive outcomes from both industry and academia suggest that incorporating market returns into the model in addition to GDP growth, inflation and outperformance are necessary to evaluate passive and active fund performance. The correlation between corporate profits to GDP growth is evident in many markets. Such increases may not be mirrored the equity market (Ritter, 2005). Furthermore, high GDP growth may also be associated with increasing inflation expectations that have an attenuating effect on GDP growth (Fama, 1981). Combined with outperformance (or underperformance), GDP growth and inflation are all important factors in the market performance, hence fund performance, therefore this research includes these variables in measuring fund outperformance (MSCI, 2010).

Passive funds follow some form of index, that is, their composition and decision making over investments is exogenous to the fund. A reasonable proxy for any passive fund is the associated category of indices. Hence indices provide a benchmark performance indicator for passive funds when comparing them with active funds equivalents. By accounting for GDP growth and inflation factors and measuring their effects on the relative performance differential, provides a basis for determining which outperforms the other (see the Appendix B1 further details about the relationship between the variables).

We run regressions where fund style is the dependent variable using the same methods as above, namely, pooled, fixed effects and random effect with the addition of GMM:

$$\boldsymbol{OP_{i,t}} = \boldsymbol{\beta_0} + \boldsymbol{\beta_r F_{i,r,t}} + \boldsymbol{\beta_g GDP_{t-1}} + \boldsymbol{\beta_\pi \pi_t} + \boldsymbol{\beta_{op} OP_{i,t-1}} + \boldsymbol{\varepsilon_{i,t}}, \quad (5)$$

where, $\boldsymbol{OP_{i,\ t}}$ is the number of funds that outperform the benchmark (namely the index or passive funds) in style category $\boldsymbol{i}$ at time $\boldsymbol{t}$. $\boldsymbol{F_{i,\ r,\ t}}$ is the return of benchmark index as before. GDP and inflation $\boldsymbol{\pi_t}$, $\boldsymbol{OP_{i,\ t-1}}$ is one period lagged as is the previous outperformance with style category. The panels cover the time-period 2000–2016 inclusive, with 14 cross sections (representing the different fund style/size categories), taken from the SPIVA US Scorecard as before. Data on US GDP growth and inflation was collected from the Federal Reserve Economic Data (FRED) database. The dataset has 238 observations. As before, we exclude data prior to 2000 to ensure consistency. Unfortunately, the 17 periods, introduces a dimensionality problem, hence our concentration is on dynamic fund behaviour. We perform standard unit root tests on the variables that we report in Table A2, Appendix A and reject that we have unit roots. The cross sections denoted by subscript $\boldsymbol{i}$ cover nine combinations of large, mid, and small-cap funds with the S&P indices styles of Core, Growth, and Value.[12] These nine combinations combined with an aggregate outperformance percentage for each fund-size category, and an overall average, giving 14 cross sections overall.[13] We use panel analysis techniques to determine the driving factors for outperformance that differentiate between passive and active funds.

4.2.2 | Econometrics analysis

We already identify that dynamic models have greater success in uncovering the driving factors. To complement our portfolio of techniques with autoregressive distributed lag (ARDL) analysis to confirm by contrasting results with GMM in the face of limited observations. We analyse the co-integrated lagged values using methods of Pesaran and Smith (1995) and Pesaran and Shin (2002). We use several post-estimation tests including the Bruesh-Pagan test for cross-section dependence in residuals, AR(2), Jarque-Bera (JB) for normality and Sargan test of over-identification restrictions to confirm result reliability. We include pooled, fixed and random effects by way of comparison and for completeness.

Table 5 reports the results that determine the factors affecting passive and active fund management. Note that the Hansen/Sagan J-Statistics indicate that the model is correctly specified for both GMM transformations. The AR(2) insignificance indicates that there is no second order autocorrelation. The Bruesch-Pagan test indicates that there is a cross dependency in both Fixed and Random Effect analysis therefore, necessitating the use of GMM and ARDL methods. Normality test indicates that Pooled, fixed and Random effects estimations are not normal confirms this approach.

However, fixed and random effects have both, their signs and magnitude similar to those of GMM and ARDL. This implies that a positive movement to the index has about 24–35% detrimental effect on the proportion of funds that maintain outperformance of the benchmark. This confirms Malkiel (2003) work. We confirm the view that active management relies on fund managers being able to outpace the market. However, active management costs and, in many cases, may not exceed the returns from passives in the long run. Further, there may be short term gains that might be outweighed market efficiency in price adjustments. With zero lags nothing is, significant Using one lag tends to give better results implying that SPIVA dataset has dynamic relationships rather than static making Pooled, Random and Fixed effects redundant.

Note that GDP and its lags in the ARDL model implies that shocks to GDP tend to take time to work though the economy and onto returns. Outperformance in both GMM and ARDL support the old investment adage, past performance is no predictor of future performance [positively]. We find quite the reverse; prior past good performance tends to lead to poor performance! One could conjecture overconfidence has a role to play.

As to the passive and active fund debate these findings concur with Philips (2008) in this regard. From Table 5, the negative index returns (i.e., the proxy for passive funds) to outperformance indicates that active funds perform poorly compared to passives, in a rising or bull market. This implies that active fund managers are not necessary as good as the market at exploiting the gains from a rising market. However, in a falling or bear market, active fund managers can take “rear guard” actions that mitigate losses and protect the fund. These interventions are not available to the passive fund and decline is inevitable. Chan and Cheng (2003) suggest that actively managed mutual funds underperform with respect to the market portfolio in average return and we concur this that view.

Finally, we find that GDP growth has a positive effect on fund performance reflected in an increase in the number of outperforming funds. This, supports the view that

TABLE 5 Regression results active and passive funds

Variables	Pooled	FE	RE	GMM (1S)	GMM (2S)	ARDL
Intercept	4.566***	4.749***	4.566***			5.226***
Index return	−0.293***	−0.296***	−0.293***	−0.368***	−0.327***	−0.241***
GDP_growth	0.028**	0.029**	0.028**	0.027**	0.032***	−0.102***
Inflation	−0.023	−0.027	−0.023	−0.019***	−0.022***	−0.187***
GDP growth (−1)						−0.026*
GDP growth (−2)						0.060***
Outperformance(−1)	−0.119*	−0.163**	−0.119*	−0.213***	−0.204***	−0.135**
R^2	0.06	0.11	0.06			0.19
Adjusted R^2	0.04	0.03	0.04			
Hausman test			10.22**			
JB test *p*-value	.00	.00	.00	.86	.02	.30
J-statistic				12.11	12.94	
J-statistic *p*-value				.28	.23	
AR(1) *p*-value				.01		
AR(2) *p*-value				.26		
Breusch-Pagan LM		3,122.03***				
Instrument rank				14	14	
Observations	223	223	223	210	210	210

Note: ***, ** and * indicate the growth of significance at the 1, 5 and 10% respectively. GMM (1S) is the one-step GMM, GMM (2S) is the two-step GMM and JB is the Jarque-Bera test for normality. J-statistic is the Sargan-Hansen overidentification test.
Abbreviations: ARDL, autoregressive distributed lag; GMM, Generalized Method of Moments.

not all GDP growth converts to index returns hence, passive funds. GDP growth has a positive effect on active funds supporting the results of Jiranyakul (2009). As expected and confirming Monjazeb and Ramazanpour (2013), Table 5 identifies that inflation has a detrimental effect on returns.

5 | DISCUSSION

5.1 | Biases

Although S&P attempt to limit the effects and in some cases eliminate these biases their remains much discussion SPIVA data's biases, the causes and the effects (see Section 3). These biases cause many issues in using SPIVA datasets for both academic and commercial analysis. We address many of the biases by the methods of selecting data from the SPIVA scorecards dataset as we have discussed above. Some of the more difficult biases, for example: fund weighting performance measurement, indices and index fund comparisons, gross and net of fund fees (net returns) and single index benchmark comparisons we address with these methods.

By selecting data post 2000, we eliminate survivorship bias This does not necessarily hold for other databases where similar to SPIVA data is available. Although this does impose some restrictions on the number of observations, with care, researchers can work round this, as we demonstrate in our analysis.

The fund weighted performance measurement bias causes many issues in comparing funds to their respective alphas as we discuss in Section 2. Our analysis demonstrates that using asset value as a weighting (available in the SPIVA dataset) rather than equally weighted largely addresses the effects of this bias. This approach provides the level of accuracy and consistency needed to analyse funds management using SPIVA data. Furthermore, we show that by using the difference in the alphas and the coefficients that smaller funds have some significant disadvantages over larger funds. This confirms that economies of scale and scope derives from fund size implying that information sharing, fund manager resources and transactional costs can be mitigated when funds are of sufficient size.

Our finding from the analysis of benchmark returns, alphas and fund characteristics are consistent with those of Fuller et al. (2010), that is single index poses little effect on analysis and, if the index is categorised according to the fund, then the results can be relied upon. One factor in the passive active debate is returns. Active funds "claim" that they have better gross returns than the equivalent passive fund. However, using net returns eliminates the fund fees bias and the perceived advantage of active management under a growing market. Active management comes into its own when there is a declining market, particularly in the short term.

5.2 | Empirical results

We conclusively show that asset weighted alpha gives a clearer, more accurate picture of fund performance accounting for fund style. Mitigating this equally weighting bias though asset weighting achieves a "price" (alpha) that compares favourably to Fama (1970) EMH. We evidence that using equally weighted as a benchmark overstates the outperformance and would lead to spurious conclusions about fund size.

Regarding fund size, average and larger funds tend to perform better over time than smaller funds when the benchmark uses asset weighting. The implication is that larger funds are more efficient that the smaller equivalents. Normative analysis indicates that economies of scale and scope in both the investor and investment sides are significantly beneficial to the larger funds, particularly when they are actively managed. Normatively, the range and diversity of investors makes less likely a "run" by investors, therefore the necessity to hold liquidity is limited. Likewise, the range, specialisation, depth of the investment portfolios limits their exposure to a sectoral or geographical downturn.

Within a category, we show that large funds have a significant effect on alpha. By comparing category alphas with coefficients for the fund size dummies indicate that size matters, and it matters a great deal. The extent is of the same magnitude however opposite signs. Smaller funds tend to turn a greater frequency to larger funds, particularly in downturns.

Lastly, we contribute to the active passive fund performance debate. A passive fund proxy is an underlying index, hence a benchmark. Active funds intention is to beat the market, however Malkiel (1995) and Jones and Wermers (2011) finds that active funds underperform the benchmark and we confirm those results. Although active may have a short-term advantage, their costs of operation are not fully covered by any long term advantage. Hence, we conclude that passives outperform active funds in the long run co. This brings into question the whole reasoning for active fund management, we will leave that for future work.

Which introduces the next question, where do active funds outperform? We concur with Philips (2008) finding that, using SPIVA data, active fund outperformance declines in a bull market. However, seemingly somewhat contradictory, is that GDP growth has a positive effect on outperformance. This is explained that the mechanism between GDP, firm profits and index returns is not necessarily a straight 1:1 ratio channel. Consequently, indices and passive funds may not benefit from GDP growth in the same way. Where active funds become into their own, is in a bear market. They able to act to protect their assets and exploit opportunities not open to passive funds. However, the effect is not so great in the long run. Generally, passive funds outperform active funds in the long run though the attenuation from fees deriving active management.

Of interest is that if the category has many passive funds linked to similar indices then changes in the composition of those indices is likely to trigger mass changes in the passive portfolio of funds. If the movement and volume of similar assets is coincident, then this is likely to have a marked effect on both the assets prices being disposed of, negatively and positive effect on assets being purchased (Sushko and Turner (2018). This creates arbitrage opportunities for active traders (Da & Shive, 2018; Leippold, Su, & Ziegler, 2016). This opens another opportunity for further research, particularly with the debate on fund size and performance.

Finally, all our results are confirmed with several robustness checks ensuring that we can draw the results for this discussion.

6 | CONCLUSION

This research extends the use and value of SPIVA scorecards by demonstrating the application of novel methods to correct for several significant biases and conduct robust analysis. We address some of the most significant biases such as equal weighted asset and the indices based biases to show robust results and analysis.

Moving to our analysis, we show that smaller funds perform less well than larger funds across all categories. Bar the economies of scale and scope, there is less turnover of large funds and we infer that fund liquidity and diversity of both investment and investors plays a significant role in survivability.

Regarding the passive active debate, we find that passive funds generally outperform active funds in the long run, any advantages of active funds are wiped out by the fees. Only under a bear market does active funds demonstrate an advantage. It does bring into question the value that active fund management brings to investors.

The effects of index performance on fund category outperformance (that this passive vs. active funds) has contradictory results to intuitive effects of GDP on company, hence market performance. We concur with other researchers on this effect, however, we demonstrably show the effects of this the contradiction on the difference between active and passive funds. By using Dynamic time series panel analysis though GMM we identify that competition is important and that GDP growth tends to take time to filter though to fund performance.

The adage, "past performance is not indicator of future performance" is something that the fund management industry tries to refute particularly when performance reflect favourably. We provide evidence that prior good performance tends to result in poor performance implying a cyclical relationship. Many investors will invest when a fund is high (funds tend to market their success) and we would predict that they will lose as a result. One could hypotheses that fund managers tend to "sit on their laurels" when good performance occurs and "buck their ideas up" when they have poor performance to survive. This leads us to a potentially fruitful area of future research, namely, the interaction of fund size and the active passive debate. This could greatly be enhanced by studying the mass movement across multiple passive funds in a category using our methods and research findings using SPIVA data.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from Mendeley database. See link below: https://data.mendeley.com/datasets/pmtp6zn3ny/2

ORCID

Imran Hussain Shah https://orcid.org/0000-0002-3662-4128

ENDNOTES

[1] Net asset value.

[2] For consistency, we use the terms "Outperformance" and "Underperformance" with a specific industry meaning, that is Underperformance is the proportion of funds that performed worse, that is, have lower returns, than their respective benchmark. For example, a 70% underperformance indicates that 70% of funds in that category performed worse than their benchmark index within a specific size/style category. It does not indicate the loss or attenuation of returns, just that they underperformed the benchmark. Likewise, Outperformance, is proportion/percentage of funds that exceeded index performance. It is unlikely that a significant amount of funds performs at par with index performance, therefore the total proportion of underperforming, plus performing, plus outperforming must equal unity.

[3] Reachback simply take all funds that are alive today and then look at their returns over the previous X years. This clearly excludes any funds that were alive during the X years but ceased to exist, therefore leading to survivor bias.

[4] Funds are less likely to report when performing poorly in a discretionary environment.

[5] This is one of the perennial arguments between active and passive funds, the cost of the research and analysis does not necessarily improve the performance of an active fund over similar passive fund. Fees for passive funds tend to be much less than those for equivalent active funds.

[6] Alpha is the return of fund category (equally weighted, asset-weighted forms) minus return of benchmark for the relevant category. Alpha deals with the size of the shortfall. For example, we would clearly prefer a fund that outperforms its benchmark returns by say 10% than just 1%.

[7] 2015 is reasonably representative of a normal year, rather than the periods such as immediately prior to or post the financial crisis.

[8] The reason for using this time period is that the mid-year 2010 scorecard was the earliest one that is publicly available.

[9] For example, one cross section would be the alpha of large-cap value funds over a 1-year horizon compared to the relevant S&P benchmark index.

[10] In economic terms, endogeneity can be interpreted as the act of the past on the present, both on the model (dependent variable) and on the independent variables, or as the causality relationship between regressors and explained variable along the time.

[11] Readers may refer to Guide and Ketokivi (2015), Zaefarian, Kadile, Henneberg, and Leischnig (2017) further support with marketing research panel data (a major component of SPIVA dataset) from Subhan, Pervaiz, and Ghasem (2018).

[12] These are three styles of S&P indices.

[13] These are maximum cross-sections available based on structure of S&P indices.

REFERENCES

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, *58*(2), 277–297.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error component models. *Journal of Econometrics*, *68*(1), 29–51.

Baltagi, B. H., Feng, Q., & Kao, C. (2012). A Lagrange multiplier test for cross-sectional dependence in a fixed effects panel data model. *Journal of Econometrics*, *170*, 164–177.

Barnes, J. (2003). Active vs. passive investing. Retrieved from https://www.cfainstitute.org/en/research/cfa-magazine/2003/active-vs-passive-investing.

Ben-Rephael, A. (2017). Flight-to-liquidity, market uncertainty, and the actions of mutual fund investors. *Journal of Financial Intermediation*, *31*, 30–44.

Blake, D., Caulfield, T., Ioannidis, C., & Tonks, I. (2017). New evidence on mutual fund performance: A comparison of alternative bootstrap methods. *Journal of Financial and Quantitative Analysis*, *52*(3), 1279–1299.

Blanchett, D. M., & Israelsen, C. L. (2007). Spotlighting common methodological biases in active vs. passive studies. *Journal of Financial Planning*, *20*(11), 64–72.

Bogle, J. C. (2002). An index fund fundamentalist. *The Journal of Portfolio Management*, *28*(3), 31–38.

Bogle, J. C. (2016). The index mutual fund: 40 years of growth, change, and challenge. *Financial Analysts Journal*, *71*(1), 9–13.

Bogleheads, (2017). SPIVA scorecards [online]. Retrieved from https://www.bogleheads.org/wiki/SPIVA_scorecards

Breusch, T., & Pagan, A. (1980). The Lagrange multiplier test and its application to model specification in econometrics. *Review of Economic Studies*, *47*, 239–253.

Brown, J. S., & Goetzmann, N. W. (1995). Performance persistence. *The Journal of Finance*, *50*(2), 679–698.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, *52*(1), 57–82.

Carhart, M. M., Carpenter, J. N., Lynch, A. W., & Musto, D. K. (2002). Mutual fund survivorship. *Review of Financial Studies*, *15*(5), 1439–1463.

Carosa, C. (2005). Passive investing: The emperor exposed. *Journal of Financial Planning*, *18*(10), 54–60.

Chan, Y. C., & Cheng, L. T. W. (2003). Asset allocation and selectivity of Asian mutual funds during financial crisis. *Review of Quantitative Finance and Accounting*, *21*, 233–250.

Da, Z., & Shive, S. (2018). Exchange traded funds and asset return correlations. *European Financial Management*, *24*(1), 136–168.

Demos, T. (2010). Active funds found to outperform in bull markets. Financial Times. Retrieved from https://www.ft.com/content/f50b364a-83a6-11df-b6d5-00144feabdc0

Elton, E., Gruber, M., & Blake, C. (2001). A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases. *The Journal of Finance*, *56*(6), 2415–2430.

Fama, E. F. (1965). The behaviour of stock-market prices. *The Journal of Business*, *38*(1), 34–105.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417.

Fama, E. F. (1981). Stock returns, real activity, inflation and money. *American Economic Review*, *71*(4), 545–565.

Fasano, A., & Boido, C. (2017). Concentration and behavioral biases in the active management of bric funds. *Ekonomika*, *96*(1), 58–73.

Foley, S. (2016). What the hell am I paying my EM fund manager for? Financial Times. Retrieved from https://www.ft.com/content/40ad8862-78b2-11e6-97ae-647294649b28

French, K. (2008). Presidential address: The cost of active investing. *The Journal of Finance*, *63*(4), 1537–1573.

Fuller, R., Han, B., & Tung, Y. (2010). Thinking about indices and "passive" versus active management. *The Journal of Portfolio Management*, *36*(4), 35–47.

Goldman S. (2010). Re-thinking the active vs. passive debate. Asset Management. Retrieved from http://s3.amazonaws.com/zanran_storage/www2.goldmansachs.com/ContentPages/630009710.pdf

Guide, V. D. R., & Ketokivi, M. (2015). Notes from the editors: Redefining some methodological criteria for the journal. *Journal of Operations Management*, *37*, v–viii.

Hansen, L. P. (1982). Large sample properties of the generalized methods of moments. *Econometrica*, *50*, 1029–1054.

Holmes, M. (2007). Improved study finds index management usually outperforms active management. *Journal of Financial Planning*, *20*(1), 48–58.

Indro, D., Jiang, C., Hu, M., & Lee, W. (1999). Mutual fund performance: Does fund size matter? *Financial Analysts Journal*, *55*(3), 74–87.

Jensen, M. (1968). The performance of mutual funds in the period 1945-1964. *The Journal of Finance*, *23*(2), 389–416.

Jiranyakul, K. (2009). Economic forces and the Thai stock market, 1993- 2007. *NIDA Economic Review*, *4*(2), 1–12.

Jones, R. C., & Wermers, R. (2011). Active management in mostly efficient markets. *Financial Analysts Journal*, *67*(6), 29–45.

Khorana, A., Servaes, H., & Tufano, P. (2005). Explaining the size of the mutual fund industry around the world. *Journal of Financial Economics*, *78*(1), 145–185.

Ketokivi, M., & McIntosh, C. N. (2017). Addressing the endogeneity dilemma in operations management research: Theoretical, empirical, and pragmatic considerations. *Journal of Operations Management*, *52*(2017), 1–14.

Kinnel, R. (2016). Predictive power of fees: Why mutual fund fees are so important. *Morningstar*, *24*(8), 1–3.

Leippold, M., Su, L. & Ziegler, A. (2016). How index futures and ETFs affect stock return correlations. University of Zurich. https://doi.org/10.2139/ssrn.2620955.

Malkiel, B. G. (1995). Returns from investing in equity mutual funds 1971 to 1991. *The Journal of Finance*, *50*(2), 549–572.

Malkiel, B. G. (2003). Passive investment strategies and efficient markets. *European Financial Management*, *9*(1), 1–10.

Monjazeb, M., & Ramazanpour, E. (2013). The effect of economic factors on the efficiency of mutual funds in Iran. *Technical Journal of Engineering and Applied Sciences*, *3*(15), 1707–1711.

MSCI. (2010). *Is there a link between GDP growth and equity returns*. New York, NY: MSCI Barra Inc.

Nihar, K. L., & Murty, K. S. (2010). Mutual fund performance—A review. *Drishtikon: A Management Journal*, *1*(2), 89–102.

Papadamou, S., & Sirpiopoulos, C. (2004). American equity mutual funds in European markets: Hot hands phenomenon and style analysis. *International Journal of Finance and Economics*, *9*, 85–97.

Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels. *CESifo*. (Working Paper No. 1229).

Pesaran, M. H., & Shin, Y. (2002). Long run structural modelling. *Econometric Reviews*, *21*(1), 49–87.

Pesaran, M. H., & Smith, R. (1995). Estimating long-run relationship from dynamic heterogeneous panels. *Journal of Econometrics*, *68*, 79–113.

Philips, C.B. (2008). The active-passive debate: Bear market performance. The Vanguard Group, 1–11. Retrieved from https://global.vanguard.com/.

Poi, B. and Wiggins, V. (2001) STATA: How do I test for panel-level heteroskedasticity and autocorrelation? Retrieved from http://www.stata.com/support/faqs/statistics/panel-level-heteroskedasticity-and-autocorrelation/.

Ritter, J. R. (2005). Economic growth and equity returns. *Pacific-Basin Finance Journal*, *13*(5), 489–503.

Schroders (2014). New report: Investors worldwide turn to equity markets for growth in 2014. Schroders Global Investor Study. Retrieved from https://www.schroders.com/hu/sysglobalassets/digital/press/investors_worldwide_turn_to_equity_markets_for_growth_in_2014.pdf

Schultz, E. L., & Tan, D. T. (2010). Endogeneity and the corporate governance-performance relation. *Australian Journal of Management*, *35*(2), 145–163.

Sharpe, W. (1991). The arithmetic of active management. *Financial Analysts Journal*, *47*(1), 7–9.

Standard & Poor's (2016a). U.S. Scorecard–Year End 2015. S&P Dow Jones Indices LLC. New York. Retrieved from https://us.spindices.com/documents/spiva/spiva-us-yearend-2015.pdf

Standard & Poor's (2016b). SPIVA® Institutional Scorecard–How much do fees affect the active versus passive debate? S&P Dow Jones Indices LLC. New York. Retrieved from https://www.spindices.com/documents/spiva/research-spiva-institutional-scorecard-how-much-do-fees-affect-the-active-versus-passive-debate-year-end-2016.pdf

Standard & Poor's (2017). SPIVA® Institutional Scorecard–How much do fees affect the active versus passive debate? S&P Dow Jones Indices LLC. New York. Retrieved from https://us.spindices.com/documents/spiva/spiva-us-year-end-2017.pdf

Steverman, B. (2011). Gloomy bond investors clash with upbeat stock managers. Bloomberg Businessweek. Retrieved from https://www.bloomberg.com/news/articles/2010-07-05/gloomy-bond-investors-clash-with-upbeat-stock-managers

Subhan, U., Pervaiz, A., & Ghasem, Z. (2018). Dealing with endogeniety: The generalized method of moment (GMM) for panel data. *Industrial Marketing Management*, *71*, 69–78.

Sushko, V., & Turner, G. (2018). The implications of passive investing for securities markets. *BIS Quarterly Review*, 113–131.

Vayanos, D. & Woolley, P. (2016). The curse of the benchmarks. Paul Woolley Centre Financial Markets Group. (Discussion Papers, No. 747).

Wermers, R. (2000). Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses. *Journal of Finance*, *55*, 1655–1695.

Zaefarian, G., Kadile, V., Henneberg, S. C., & Leischnig, A. (2017). Endogeneity bias in marketing research: Problems, causes and remedies. *Industrial Marketing Management*, *65*, 39–46.

How to cite this article: Shah IH, Wanovits HM, Hatfield R. Uncovering investment management performance using SPIVA data. *Int J Fin Econ*. 2021;26:3676–3695. https://doi.org/10.1002/ijfe.1981

APPENDIX A: Unit root tests

TABLE A1 Unit root test results

Variables	Level	LLC	IPS	ADF	PP
Alpha equal-weighted	Intercept	−12.09***	−9.89***	−234.79***	−196.53***
Index return	Intercept	−3.64***	−2.82***	−128.70***	−181.06***
Alpha equal-weighted	Intercept	−10.93***	−6.69***	−217.29***	−190.15***
Competition	Intercept	−3.515***	−1.133	−124.34***	−227.27***

Note: *** and **indicate the level of significance at the 1 and 5% respectively.

TABLE A2 Unit root test results

Variables	Level	LLC	IPS	ADF	PP
Outperformance	Intercept	−12.43***	−10.15***	−130.82***	−199.11***
Index return	Intercept	−9.89***	−7.20***	−98.19***	−209.88***
GDP growth	Intercept	−3.76***	−4.13***	−41.71***	−59.48***
Inflation	Intercept	−6.39***	−2.91***	−46.76***	−85.56***

Note: *** indicates the level of significance at the 1%.

APPENDIX B: Data description and model specification

Large versus average fund investment

TABLE B1 Definition of dummy variables

Grouping	Variable	Type	Description
Timeframes	_Y1, _Y3, _Y5, _Y10	Dummy (0, 1)	The SPIVA scorecards compare aggregate fund returns to relevant benchmarks on a 1, 3, 5 and 10 year (annualised) basis. We create four dummy variables (_1Y, _3Y, _5Y, and _10Y), for each time horizon. This supports the analysis of benchmarking performance by category across multiple years
Styles	Core, growth, value, and aggregate	Dummy (0, 1)	Representing the fund styles in the SPIVA scorecard namely, Core funds, growth funds, value funds, and aggregate measures.
Size	Large, mid, small, multi	Dummy (0, 1)	One each for the four types of funds presented in the SPIVA scorecard, which are large-cap, mid-cap, small-cap, and multi-cap funds
	Competition	SPIVA	SPIVA proxy for competition, the SPIVA scorecards report the number of funds in each style-size box at the start of the period. We took these numbers from the corresponding SPIVA scorecards to get a proxy for competition in each class
	Index return	Index	The performance of the relevant index over the horizon investigated for each observation
	Equal_asset	Dummy(0,1)	To represent when the alpha derives from equally weighted rather than asset weighted.

Passive and active fund performance:

The rationality for the use of independent variables in our econometric analysis and their collection are outlined below:

Constant: As indicated by Sharpe (1991), if the market only consists of active funds, by construction, 50% will outperform and 50% will underperform the benchmark index before management fees. Post-fees, more than 50% will underperform. Although this assumption does not match actual financial markets, the concept of the market only being as good as its agents still holds, and therefore a constant is needed in the regression.

Relevant Benchmark return: As the dataset includes two strong bull and bear markets the relevant benchmark return for each fund category over time is included in the regression. It is generally believed that active management performs better in falling (bear) markets than in rising (bull) due to its ability to position the portfolio more defensively, to hold cash, or even to short-sell securities (Philips, 2008). However, research by FundQuest (Demos, 2010) suggests that active funds outperform in rising markets compared to passive benchmarks. While these two pieces of research produced conflicting findings, they both indicate that market returns are a valid regressor as they have an impact on performance.

GDP growth: Including GDP growth in the model controlled for any index variation based on GDP, but also allowed GDP to influence the number of outperformed funds through other channels.