ORIGINAL ARTICLE

Puzzle solved? A comprehensive analysis of hedge fund-like mutual funds according to the value-added paradigm

Nathaniel Light[1] · Ivan Stetsyuk[2]

Abstract

Mutual funds that claim hedge fund strategies (HMFs) have experienced large increases in assets under management, yet academic studies are nearly unanimous in their negative appraisal of HMF performance. This paper examines whether this inconsistency can be resolved through the 'value added' paradigm. We demonstrate instead that pockets of value production and skill coexist alongside a largely wasteful HMF space as a whole. We document a highly fractured market structure in which the top 10 firms control 48% assets, the bottom half only 4%, and nearly 40% of firms disappear within a brief window. We explore the implications for tests of performance under such conditions.

Introduction

Hedge fund-like mutual funds, referred to alternatively either as 'hedged mutual funds' (HMFs) or as 'liquid alternatives,' were all the rage in the early years after the financial crisis of 2008–2009. Supported by academic evidence such as the findings of Agarwal et al. (2009) that HMFs outperform regular mutual funds by as much as 4.8% annually, and spurred on by arguments that their greater flexibility enabled returns both stronger and less correlated with the market than those of conventional mutual funds, HMFs saw massive inflows and increases in assets under management (AUM) in the years subsequent to the crisis. For example, inflows into these funds in the year 2009 amounted to $121 billion, a *full quarter* of inflows into all mutual funds, and the trend continued into 2010.[1] Yet subsequent academic studies about HMFs paint a portrait much less flattering than those of Agarwal et al. (2009). These studies mainly find evidence of underperformance. What, then, accounts for the seeming discrepancy between large inflows into HMFs and the negative appraisal from academic research? This paper applies recent theoretical and empirical advances in the mutual fund literature in an attempt to resolve this issue.

More specifically, we examine HMFs in the context of the 'value added paradigm,' which stresses that the proper skill measure for individual funds centers around their gross alpha scaled by assets, and that the value proposition for a group of mutual funds must weight any analysis by the AUMs of the respective funds, thereby recognizing the greater importance of funds with high AUM. To our knowledge, this is the first paper to apply the value-added approach to HMFs, a specific subsection of the mutual fund industry. Notably, we find a highly embryonic market structure for HMFs in which a rapid expansion of funds and AUM turns to a gradual decline in which nearly 40% of funds disappear within a brief window, and in which the bottom half of funds control only 4% of assets as of 2018. These figures highlight the limitations insofar as broad inferences about HMF performance for investors in the prior studies that rely almost

✉ Ivan Stetsyuk
   stetsyuk.ivan@uqam.ca

   Nathaniel Light
   tub45479@temple.edu

[1] St. John Fisher College, 235 Salerno Hall, 3690 East Avenue, Rochester, NY 14618, USA

[2] École des sciences de la gestion, Université du Québec á Montréal, 315, rue Sainte-Catherine Est, Montréal, Québec H2X 3X2, Canada

exclusively on equal-weighted portfolios in their empirical design. When we evaluate the performance of HMFs in the value-added approach, we get mixed results that actually appear worse than what the theories predict. HMFs produce no value in aggregate, suggesting that investors essentially pay high expenses for pre-expense returns equivalent to those on benchmarks. On the other hand, we conduct a bootstrap analysis of gross alphas and we also test for persistence in value added, the results for which suggest persistent value generation among the top 10% of funds. While these results overall contradict value-added predictions of positive value added, we also find a 25% drop in AUM in the years 2015–2018, which may suggest the HMF space is gradually moving towards its proper efficiency point in size after an initial over-expansion.

We use the term 'value added paradigm' to refer to the contributions of Berk and Green (2004) and Berk and van Binsbergen (2015). These papers substantially recast how observers should measure skill among mutual fund managers, and they also delineate an equilibrium condition in the mutual fund industry in terms of gross returns, net returns, and sizes of funds. Berk and Green (2004) lay out the theoretical side, in which they argue that persistent alphas in mutual fund performance represent market inefficiencies in the same manner as do stock market anomalies, insofar as they imply investors leave money on the table. In equilibrium, the paper explains, investors reward outperforming funds with more AUM until these funds no longer achieve superior results for their investors, at which point their outperformance after fees (net alpha) will equal zero and their performance before fees (gross alpha) will equal the amount of their fees. This conclusion follows from the assumption that a manager's ability to beat the market decreases with the size of the AUM. In other words, persistent net alphas for a particular fund or strategy should disappear once investors recognize the success by substantially increasing AUM.

According to the above argument, Berk and Green (2004) argue further, the proper measure of mutual fund skill should be gross alpha scaled by AUM, which they term 'value added', rather than net alpha. This preference overturns the prior consensus of mutual fund research in which net alpha represents the 'gold standard' of fund performance and of the indicator of skill. For purposes of this paper, it is important to note, we use the above definition of 'value added', whereas other papers may state offhandedly that a fund 'adds value' or demonstrates 'value added' in any instance of net alpha. On the empirical side, Berk and van Binsbergen (2015) demonstrate that fund-level value added persists and that the mutual fund industry as a whole produces large amounts of value added confirming a great degree of skill among mutual fund managers and justifying the flows into active management. This result counters the conclusions of several prior studies such as Jensen (1968)

and Fama and French (2010), which find no evidence of skill in the mutual fund industry based on their analyses of net alpha. Finally, Berk and van Binsbergen (2015) also introduce tradable Vanguard funds as the proper benchmark of performance in place of more widespread factor models, since they argue that value added should be measured relative to returns available to investors rather than to factors which investors are often unable to replicate in practice.

What, then, suggests that studying HMFs in the context of the 'value added' paradigm can improve our understanding beyond that attained by the prior studies on HMFs? While the vast majority of earlier studies base their results on equal-weighted calendar portfolios, Berk and van Binsbergen (2015) find that the vast bulk of value added occurs within a small number of funds that control disproportionately large AUM. The inflows into the HMF space, then, may be justified if only a few dominant funds outperform their peers. To the extent that weak performance from smaller HMFs hurts the returns on the equal-weighted portfolios in other papers, these papers only offer results that allow a more limited set of inferences than does a study that examines value added. More specifically, ignoring the value added of HMFs may overlook the underlying skill apparent on a value-weighted basis. On a more general level, as argued by Berk and Green (2004) and Berk and van Binsbergen (2015), we should expect mutual funds to reach an equilibrium in which their net alphas equal zero while at the same time they produce large value added.

Therefore, this paper examines whether the value added of HMFs can overturn a negative evaluation of their performance in the same manner that the value added results of Berk and van Binsbergen (2015) argue against a prior negative evaluation of mutual funds in general. We perform several tests: We produce point estimates of value added for several different sub-categories of HMFs; We conduct bootstrap analyses of their net and gross alphas; In addition, we implement the value added persistency test of Berk and van Binsbergen (2015). We find results that, while interesting, are in part contrary to the above stated equilibrium predictions. For example, our estimates for value added differ across our four main sub-categories. For two of the four, the average value added is positive at the 5% significance level. Yet the value added is either negative or close to zero for the other two categories and insignificantly different from zero for all categories aggregated. This differential performance across categories may reflect genuine differences in performance of the underlying strategies, though they may also hinge on the few top funds by AUM within each category.

On the other hand, we conduct bootstrap analysis of gross alphas and show evidence that some HMFs achieve positive gross alphas, indicating some underlying skill. Similarly, when we test for persistence in value added, the results offer more encouragement: The top decile of funds sorted

on value added produces persistently positive value added in subsequent periods, and in most specifications the top decile funds beat the median decile and the bottom decile. Finally, the results for bootstrapped net alphas offer a point of clarity: The numbers unambiguously suggest a complete absence of positive net alpha funds. This result casts doubt on the ability of mutual fund investors to achieve superior results by chasing positive publicity for funds or strategies.

Furthermore, our collection of summary data reveals some important aspects of the structure of the HMF space. In addition to the previously reported large inflows and creation of many new funds in the years after the crisis, we also observe a sharp decline in the overall AUM as well as in the number of funds in the years 2017 and 2018. More specifically, the aggregate AUM of funds in our sample drops by approximately 24.2% from December 2015 to December 2018. Furthermore, we find a surprisingly high attrition rate, with a full 41% of funds exiting the sample by its end, and an average age for funds of only 5.60 years. Finally, we also find that only a few firms dominate the rest in terms of AUM. As of December 2018, the four main categories in our HMF sample consist of 201 funds, of which the top five, ten, and twenty control 31.3%, 47.6%, and 64.8%, respectively, whereas the bottom 50% of funds amount to a paltry 4.2% of overall AUM. These figures buttress a claim that an equal-weighted portfolio might provide results distinctly different from the value-weighted emphasis in the value-added approach. In fact, the rapid expansion of the HMF space seemingly accompanies a deterioration in the performance of small funds relative to large ones.

As stated earlier, the theory of Berk and Green (2004) predicts that positive net alphas for a fund or strategy induce inflows until the point at which net alphas equal zero. We note, however, that the rapid expansion and subsequent drop in the number of funds accord with a slightly more cynical view in which success in the mutual fund space invites unskilled copycats. This view finds some currency among both academics and practitioners. For example, Jones and Mo (2021) point out that any characteristic associated with mutual fund success may attract unskilled mimicry so that the characteristic is corrupted as a predictor, even if the alphas of individual funds remain constant. Similarly, Cooper et al. (2005) shows how funds can attract inflows through mere cosmetic name changes that cater to 'hot' investment ideas. Some industry observers offer yet harsher assessments. "Investors often chase new products with no track record based on unsubstantiated marketing hype," according to one New York-based liquid alternatives specialist.[2] On a related note, Badrinath and Gubellini (2011)

observe higher flows of funds into bear market and equity market neutral funds after down market states, and Jiang and Yüksel (2019) link mutual fund flows to investor sentiment.

Thus, our findings provide a mixed picture for the HMF space and the extent to which its growth fits into the theories of Berk and Green (2004). On the one hand, evidence for positive net alphas is absent, suggesting that whatever the merits of the HMF strategies investors can no longer benefit from them. This is consistent with the theory (assuming net alphas equal zero, that is, rather than some significantly negative amount). Furthermore, the value added of the HMF space is statistically indistinguishable from zero, a result which calls the rationale for the funds into question, and a large number of smaller funds enter the space and then exit it by the end of the sample period. On the other hand, the persistence in value added among the top funds sorted along this measure points to a degree of underlying skill among HMF funds. One final test reveals a mainly inefficient HMF market for asset management with only pockets of efficiency. Thus, the overall picture is consistent with a melee of genuine skill and a multitude of mimicry in which eventually the space moves from a state of confusion towards one that weeds out the weaker players.

Our research belongs to several strands of the HMF literature. First and most obviously, several prior papers, most of them quite recent, examine the performance of HMFs and other related categories of alternative mutual funds. These papers offer conflicting assessments. On the one hand, Agarwal et al. (2009) show that HMFs outperform traditional mutual funds (TMFs) by as much as 4.8% annually though they trail hedge funds themselves. Similarly, McCarthy (2013) show that long/short HMFs offer similar returns to the investors as hedge funds (HFs) and they can serve as reasonable substitutes. On the other hand, Kanuri (2016) find that while HMFs help investors diversify risk, they underperform most asset categories. Kooli and Stetsyuk (2020) show that an average long/short HMF underperforms an average actively managed long/short HF by $2.52 million per year. Moreover, Kanuri and McLeod (2014) show that such types of HMFs as Long/Short and Equity Market Neutral fail to hedge during crisis but instead destroy value.

The remainder of the paper is structured as follows. Section "Benchmarks and estimation models" describes benchmarks and estimation models. Section "Data" describes the data, our sample, and value-added measure. In Section "Results" we explore HMF managers' skill across various models and compare the results with the literature. Section "Conclusions" concludes.

---

[2] https://www.etfstrategy.com/the-future-of-liquid-alts-goldmans-answer-replication-10338/.

Benchmarks and estimation models

Four-factor alphas

We begin our analysis using the risk-based approach that measures the outperformance of the fund against the three factors from Fama and French (1995) augmented with the fourth momentum factor implemented in Carhart (1997), hereafter four-factor model. We identify the alpha of fund i as the intercept in the time series regression

$$R_{it} - R_{ft} = \alpha_i + \beta_i^{\text{mkt}}\text{MKT}_t + \beta_i^{\text{sml}}\text{SML}_t + \beta_i^{\text{hml}}\text{HML}_t + \beta_i^{\text{umd}}\text{UMD}_t + \varepsilon_{it}. \tag{1}$$

In this regression, R_{it} is the return on fund i for month t; R_{ft} is the risk-free rate that equals the 1-month U.S. Treasury bill rate return; MKT_t, SML_t, HML_t, and UMD_t are respectively market, size, book-to-market, and momentum factors. We only consider the funds that have more than 24 months of returns.

Such approach has been applied in mutual fund literature (e.g., Fama and French 2010; Berk and van Binsbergen 2015), in hedge fund literature (e.g., Fung and Hsieh 2005), and in hedged mutual fund literature (e.g., Agarwal et al. 2009). Four-factor specification should be reported as an adjustment for risk. However, it cannot be interpreted as an alternative investment opportunity as explained below.

While factor portfolios are widely used as benchmarks in mutual fund literature, they can be misleading proxies for investment opportunities available to fund managers. Berk and van Binsbergen (2015) point out that several of the factors were discovered only after the inception dates of many mutual funds as is evident in the CRSP Mutual Fund database. For example, active managers investing in size-based strategy were not benchmarked properly against the alternative investments since they were limited to diversified index funds or stocks. Only after the size factor became known to the investors, they started rewarding mutual fund managers for outperforming respective size-based benchmarks. Even after the factor portfolios had been discovered, the costs associated with their implementation were hard to estimate precisely. For instance, implementing momentum strategy is costly but neither indirect expenses nor direct expenses (such as trading costs linked to high level of turnover) are documented properly. Not surprisingly, only several funds sell directly momentum strategies to investors. The literature on performance evaluation acknowledges that due to problems inherent in risk factor models, alphas estimated in such models are imprecise.

Accordingly, Berk and van Binsbergen (2015) argue that due to limitations of the factor models, they cannot be used as reliable benchmarks and propose to use a set of Vanguard index funds as benchmark portfolios. They claim that a set of Vanguard funds can be used as an alternative investment opportunity set because these funds are well diversified, easily tradable, and readily available to investors at a low cost. We follow their approach and choose the same 11 tradable Vanguard index funds, which we list below in the paper. Due to the fact that Mid-Cap Index, Small-Cap Growth Index, and Small-Cap Value Index funds were only incepted on May 21, 1998, our sample uses June 1998 as the first month-year observation.

In Vanguard indices model, we identify the alpha of fund i as the intercept in the time series regression

$$R_{it} - R_{ft} = \alpha_i + \beta_i f_t + \varepsilon_{it}. \tag{2}$$

In this regression, R_{it} is the return on fund i for month t; R_{ft} is the risk-free rate that equals the 1-month U.S. Treasury bill rate return; f_t is a vector of factors (returns on 11 tradable Vanguard index funds).

The value added as a measure of skill

Berk and Green (2004) reason that in a competitive environment informed investors should drive any abnormal fund performance to zero. While the net alpha is a measure of the abnormal return, it cannot be used to measure the skill of the manager. Pástor and Stambaugh (2012) lends credence to this argument by showing that equity mutual funds have provided investors with net returns below those generated by passive benchmarks due to decreasing returns to scale. A negative net alpha indicates that investors are channeling too much capital to an actively managed fund. When a fund generates positive net alpha, it indicates that investors are providing insufficient capital to eliminate abnormal returns. Likewise, recent mutual fund literature increasingly criticizes the choice of the gross alpha as a measure of skill (e.g., Berk and van Binsbergen 2015). The gross alpha is shown to be a return measure, not a value measure, and therefore cannot be used to establish the value contribution of the fund.

We measure benchmark performance in two different ways. First, we calculate the return of a portfolio of equivalent riskiness constructed from the following four-factor portfolios

$$R_{it}^B = \hat{\beta}_i^{\text{mkt}}\text{MKT}_t + \hat{\beta}_i^{\text{sml}}\text{SML}_t + \hat{\beta}_i^{\text{hml}}\text{HML}_t + \hat{\beta}_i^{\text{umd}}\text{UMD}_t,$$

where $\hat{\beta}_i^{\text{mkt}}$, $\hat{\beta}_i^{\text{sml}}$, $\hat{\beta}_i^{\text{hml}}$, and $\hat{\beta}_i^{\text{umd}}$ are the estimated coefficients on the regression of the excess gross returns $R_{it}^g - R_{ft}$ on respective risk factors: MKT, SML, HML, and UMD.

Second, we compute the benchmark return, R_{it}^B, constructed from 11 Vanguard benchmark portfolios as

$$R_{it}^B = \sum_{j=1}^{n} \hat{\beta}_i^j R_t^j,$$

where R_t^j, $j = 1, \ldots, n$, are excess returns on n Vanguard benchmark portfolios, and $\hat{\beta}_i^j$ are the estimated coefficients from the regression of the excess gross returns for fund i, R_{it}^g, on R_t^j. We call the resulting mean value—the value added from Vanguard indices model. As in the analysis of net and gross alphas, we limit our analysis to the funds that have more than 24 months of returns.

We then closely follow Berk and van Binsbergen (2015) methodology who multiply the benchmark adjusted gross return, $R_{it}^g - R_{it}^B$, by the inflation-adjusted fund i size, $q_{i,t-i}$, to compute the realized value added V_{it} between periods $t-1$ and t:

$$V_{it} = q_{i,t-1}(R_{it}^g - R_{it}^B), \tag{3}$$

where R_{it}^g is the excess gross return on fund i at time t and R_{it}^B is the benchmark return. For a fund that exists for T_i periods, the average value added is calculated as follows

$$S_i = \frac{1}{T_i} \sum_{t=1}^{T_i} q_{i,t-1}(R_{it}^g - R_{it}^B), \tag{4}$$

Data

Our data comes from the Center for Research in Security Prices (CRSP) mutual fund survivor-bias-free database. Our sample relies mainly on CRSP's classification of 'Hedged Mutual Funds,' corresponding to the CRSP objective code 'EDYH,' though we also apply keyword searches to follow the methodology of similar papers such as Agarwal et al. (2009) and Huang and Wang (2013). Because the CRSP categories have changed somewhat across the years, we briefly describe below the various categories, their history, and how the changes to the categories result in our sample differing somewhat from those of previous papers that also examine HMFs. Afterwards we discuss our own search criteria in more detail.

The CRSP objective code for 'hedged' mutual funds currently corresponds 1-to-1 with the five Lipper Class categories of 'Absolute Return' (ABR), 'Long-Short Equity' (LSE), 'Equity Market Neutral' (EMN), 'Extended US Large-Cap Core' (ELCC), and 'Equity Leverage' (DL). These categories first enter the CRSP database in 2011, 2006, 2006, 2008, and 2008, respectively. Neither Agarwal et al. (2009) nor Huang and Wang (2013) include Absolute

Return funds in their sample, possibly because the samples for these papers predate the category's initiation in 2011.

Whereas the strategies of Equity Market Neutral and Long-Short Equity funds are relatively well understood, the other three categories require a brief introduction. Absolute Return funds follow investing strategies that hope to deliver positive returns (usually within a reasonable time-frame, say, 3 years) irrespective of economic and market conditions. There are two senses in which Absolute Return funds stand apart from the other categories and require a justification for inclusion. First, despite CRSP's designation of this category as equity-based, such funds frequently combine equity and debt instruments in their portfolios. Clifford et al. (2013) examines these funds in detail and finds them approximately evenly split between those above and below a 70% equity exposure. Despite our paper's focus on equity-based funds, we decline to exclude any Absolute Return funds for too high debt exposure because of Clifford et al. (2013)'s finding that both types of Absolute Return funds share quite similar loadings on the Carhart factors as well as similar alphas from the Carhart model. Second, again as Clifford et al. (2013) note, their marketing tends to emphasize their dependability rather than alpha *per se*, for example targeting a particular level of risk or average return above a benchmark such as inflation or the T-Bill rate. Nevertheless, we decide on inclusion because like the other categories they have seen rapid increases in AUM while at the same time the literature finds little evidence for alpha delivery, and so the central question as to whether the value added paradigm can justify the flows still is in effect.

The Lipper Class category of Extended US Large-Cap Core corresponds to what is more commonly known as 130/30 funds. These funds will balance a leveraged long equity exposure of, say, 130% of their asset base, with a short exposure of 30% such that their overall market exposure will equal that of a fund invested 100% in equities without leverage. While 130/30 is the most common type, a few funds are 120/20, 150/50, and even 170/70. These funds resemble Long-Short Equity funds in that they aim to beat the market by simultaneously buying underpriced securities while shorting overpriced ones. Lo and Patel (2008) provide a useful description of their development and rationale.

Finally, the Equity Leverage category describes funds that aim for returns that are equal over some short-term time horizon, usually a day, to a target multiple of some benchmark. For example, the Rydex NASDAQ-100 2X Strategy fund describes itself thusly on its website: "Seeks to provide investment results that match, before fees and expenses, 200% of the daily performance of the NASDAQ-100 Index. . . intended for investors who expect the NASDAQ-100 Index to go up and want accelerated investment gains when it does so. However, there is an increased risk of accelerated losses if the market declines." Typically, such funds engage in heavy use of derivative contracts as part of their

strategies. Within our sample we find funds whose target multiples range from 125 to 300% of their benchmark. Since these funds explicitly follow a benchmark rather than pursue active management strategies, we exclude them from our sample, though we include them in the summary data tables and highlight some interesting preliminary observations. After the summary tables, however, this category is absent from our study.

Let us now turn to the construction of our sample. Similar to Agarwal et al. (2009) and Huang and Wang (2013), we begin by backdating the Lipper Class entries for all funds that later are classified as among the five Lipper Class categories of 'EYDH,' since as we discuss above these classifications begin only after the start of our sample period. We also re-assign funds whose names include 'Long/Short', 'Market Neutral', and '130' or '120' to the respective categories for any funds categorized otherwise. Next we conduct a keyword search among the fund names. We apply the same list of terms as Agarwal et al. (2009), except that we add '130' and '120' to identify funds in the ELCC category: Market Neutral, Arbitrage, Hybrid, Hedge, Merger, Distressed, Alternative, 130, and 120. Next we drop all funds indicated as an ETF or index fund, all those whose names indicate 'Fixed Income', 'Bond', or 'Credit', all those classified as Dedicated Short Bias, those that deal primarily in currencies or commodities, and those funds geared towards a particular retirement 'target' date. We keep only those funds whose modal CRSP Objective Code either begins with the letters 'ED,' indicating domestic equity, or else 'O', which stands for 'other' and is the classification of some funds that follow miscellaneous hedge fund related strategies.

For each month we aggregate multiple share classes into a single observation on a value-weighted basis. More specifically, we use the variable crsp_cl_grp as the identifier of funds, we weight each of return, expense ratio, and turnover based on the weights of all observations that share each one and AUM, and we sum the AUM from all classes to derive the AUM of the fund. Afterwards, we delete any monthly observation with a missing value for any of expense ratio, return, or AUM. To match HMFs with reliable observations of various fund-specific variables, our data starts from December 1998. The results should not be sensitive to the period selection because few HMFs existed prior to 1999.

We also impose screens common in the mutual fund literature. Following Berk and van Binsbergen (2015) we drop all observations prior to a fund reaching $5 Million in AUM (in constant January 2000 dollars), and to account for the incubation bias, we follow Sherrill et al. (2017) and remove fund observations prior to their first offer date as reported in CRSP. For our analysis, we require a minimum of 24 monthly observations.

Furthermore, because our persistence tests require us to track funds across time, we take special care to insure consistency in our classifications as well as consistency in identifying distinct funds. Therefore, we must specifically deal with some issues with the CRSP data. To the extent that CRSP fund classifications change across time, we either keep or reject the fund's entire history depending on whether its modal classification corresponds to one of the desired categories, and we likewise define such funds for their entire history according to their modal classification. Finally, we notice from visual inspections that CRSP sometimes assigns different entries across time for the variable crsp_cl_grp, which is meant to uniquely identify funds. For this problem we design a workaround in which we group all such funds that share a common class of fund (directly or indirectly) as recorded by the CRSP variable crsp_fundno. Finally, for the sake of simplicity, any fund that passes through all the above screens without belonging to any of the five categories is assigned to the Absolute Return category.

Table 1 shows the summary statistics for the AUM and number of funds in December of each year for these subcategories of HMFs : Absolute Return (ABR); Long/Short Equity (LSE); Equity Market Neutral (EMN); Extended U.S. Large-Cap Core (ELCC); Equity Leverage (DL); and all categories combined (ALL). It also reports mean AUM and median AUM. All AUM figures appear in millions of U.S. dollars adjusted for inflation and expressed in January 2000 dollars. The table reveals two dramatic trends in the data: First, the number of funds and total AUM rises dramatically in the years from the end of the financial crisis to 2014 ; second, the number of funds and AUM decline in the years from 2015 to 2018. These trends are mostly driven by the following types of HMFs: ABR, LSE, and EMN which control 42%, 31%, and 14% of total AUM, respectively. The remaining ELCC and DL funds together control only 13% of total AUM.

On the way up, the total AUM rises from $19.4 billion in 2008 to $95.7 billion in 2014, while at the same time the number of funds rises almost fivefold between 2008 and 2014, going from 58 to 284, respectively. The subsequent decline in these figures, while less pronounced than the initial rise, is still noticeable: from 2015 to 2018, the AUM drops almost 25% from $95.7 billion to $72.6 billion; likewise, the number of funds falls approximately 15% in just the 2 years from 2016 to 2018. The greatest decline occurs among the ELCC funds, which drop from a peak of 12 to only 7 by the end of the sample. The mean AUM reaches $228 million in 2005 and subsequently climbs slowly to $299 million in December 2018. The mean AUM always greatly exceeds the median AUM, suggesting the presence of a few very large HMFs. For example, as we mention earlier, by December 2018 the non-DL categories contain 201 funds, of which the top five, ten, and 20 control 31.3%, 47.6%, and 64.8% of the total AUM, whereas the bottom 50% of the funds control only around 4.2%.

Table 1 The table shows the summary statistics for the count (N) and assets under management (AUM) of the following categories of hedged mutual funds at the end of the year: Absolute Return (ABR); Long/Short Equity (LSE); Equity Market Neutral (EMN); Extended U.S. Large-Cap Core (ELCC); Equity Leverage (DL); all categories combined (ALL)

Year	N						AUM						AUM All	
	ABR	LSE	EMN	ELCC	DL	All	ABR	LSE	EMN	ELCC	DL	All	Mean	Median
1999	0	1	2	0	0	3	0	13	56	0	0	69	23	13
2000	1	6	5	0	3	15	1116	893	147	0	1046	3203	214	121
2001	1	13	5	0	19	38	1013	1111	126	0	1474	3724	98	26
2002	3	16	6	0	22	47	1272	1275	232	0	1413	4192	89	24
2003	4	21	7	0	25	57	2224	2051	497	0	2195	6967	122	39
2004	4	22	7	0	26	59	3460	2871	1740	0	3074	11145	189	59
2005	4	23	8	0	28	63	3944	4355	3013	0	3056	14368	228	63
2006	6	23	8	1	7	45	4793	6645	4280	163	1043	16924	376	128
2007	8	25	11	3	11	58	5314	7778	5352	1980	1211	21635	373	130
2008	10	29	12	6	28	85	3989	6577	5560	2512	761	19398	228	47
2009	15	28	14	11	41	109	5404	8040	7527	4675	1343	26988	248	47
2010	24	37	16	11	42	130	9972	9615	13155	5364	1180	39287	302	35
2011	33	46	18	11	41	149	13954	10156	10441	5641	871	41064	276	36
2012	45	48	21	12	41	167	18088	10429	12054	5968	854	47393	284	40
2013	70	59	24	11	40	204	31217	15226	12540	7488	1780	68251	335	52
2014	89	73	26	12	41	241	44728	22883	16316	9295	2518	95741	397	61
2015	107	87	27	10	41	272	51189	23674	9378	9522	1893	95655	352	63
2016	105	99	28	9	43	284	42402	22939	9646	8102	1935	85024	299	57
2017	101	88	26	8	42	265	36788	27517	9814	7811	2742	84672	320	67
2018	91	79	24	7	42	243	30591	22763	10504	6652	2051	72561	299	56

AUM are reported in millions of U.S. dollars adjusted for inflation and expressed in January 2000 dollars

As benchmarks, we choose the monthly returns on a set of Vanguard index funds given by the CRSP mutual fund database. If the fund has multiple share classes, we select the one with the lowest fees. Following Berk and van Binsbergen (2015) we select the following equity index funds: the S&P 500 Index Fund (VFINX), Extended Market Index Fund (VEXMX), Small-Cap Index Fund (NAESX), European Stock Index Fund (VEURX), Pacific Stock Index Fund (VPACX), Value Index Fund (VVIAX), Balanced Index Fund (VBINX), Emerging Markets Stock Index Fund (VEIEX), Mid-Cap Index Fund (VIMSX), Small-Cap Growth Fund (VISGX), and Small-Cap Value Index (VISVX). Table 2 provides the detailed information on the selected Vanguard index funds and shows that all these funds were traded when HMFs from our sample were available to investors.

Results

Summary data for performance and fund characteristics

Our next three tables examine summary performance statistics. The first of these looks at calendar-time portfolios both value-weighted (VW) and equal-weighted (EW), while the latter two consider cross-sectional averages calculated according in slightly different manners. The results for these three tables partially overlap, but also reveal some notable differences.

Table 3 displays results for calendar-time value-weighted and equal-weighted portfolios in Panels A and B, respectively. We construct the portfolios by combining all funds into a single time series in which the data for each month is weighted either according to the fund's prior-month AUM or else equally. Discussing first the value-weighted panel, the turnover for the various sub-categories is relatively high, perhaps unsurprisingly for hedge fund-like funds that should pursue active management. The figures range from 279% for the Equity Market Neutral Category to 130% for the ELCC funds (we discuss separately the 'DL' category, see below). The expense ratios also are relatively high, ranging from 1.59 to 1.11%. The betas from the Carhart Model generally

Table 2 The table shows a set of tradable Vanguard index funds used to calculate Vanguard benchmarks

Fund name	Ticker	Category	Inception date
S&P 500 Index	VFINX	Large Blend	08/31/1976
Extended Market Index	VEXMX	Mid Blend	12/21/1987
Small-Cap Index	NAESX	Small Blend	10/03/1960
European Stock Index	VEURX	Europe Stock	06/18/1990
Pacific Stock Index	VPACX	International/Global Stock	06/18/1990
Value Index	VIVAX	Diversified Pacific/Asia Stock	11/02/1992
Balanced Index	VBINX	Moderate Allocation	11/09/1992
Emerging Markets Stock Index	VEIEX	Diversified Emerging Markets	05/04/1994
Mid-Cap Index	VIMSX	Mid Blend	05/21/1998
Small-Cap Growth Index	VISGX	Small Growth	05/21/1998
Small-Cap Value Index	VISVX	Small Value	05/21/1998

Fund names, tickers, asset classes, and inception dates are taken from the Vanguard website. The returns are used for the Investor class shares until Admiral class shares with lower fees are introduced for the respective funds

Table 3 The table shows several statistics for value-weighted (VW) and equal-weighted (EW) calendar-time portfolios in Panel A and Panel B, respectively

	Turnover	Expense Ratio	Market Beta	Net Alpha	T-Stat (Net Alpha)	Gross Alpha	T-Stat (Gross Alpha)	Value Added	T-Stat (Value Added)
Panel A. Value-weighted calendar-time portfolios									
ABR	2.35	1.58%	0.19	− 0.21%	− 0.24	1.38%	1.58	− 3.11	− 0.43
LSE	1.80	1.59%	0.35	− 2.94%	− 3.09	− 1.36%	− 1.43	− 8.99	− 1.36
EMN	2.79	1.21%	0.05	0.43%	0.81	1.64%	3.07	6.96	2.68
ELCC	1.30	1.11%	1.04	0.27%	0.37	1.37%	1.87	7.40	2.17
DL	4.36	1.69%	2.17	− 6.85%	− 3.73	− 5.15%	− 2.80	− 4.93	− 2.37
All except DL	2.15	1.44%	0.28	− 1.09%	− 1.85	0.35%	0.60	0.03	0.00
Panel B. Equal-weighted calendar-time portfolios									
ABR	2.62	1.72%	0.29	− 0.41%	− 0.51	1.31%	1.65	− 5.89	− 1.00
LSE	2.40	1.83%	0.51	− 1.99%	− 2.88	− 0.16%	− 0.23	− 2.61	− 0.69
EMN	5.25	1.44%	0.05	0.21%	0.39	1.65%	3.04	6.86	2.81
ELCC	1.61	1.33%	1.00	− 1.59%	− 2.34	− 0.26%	− 0.39	0.23	0.08
DL	7.55	1.77%	1.97	− 6.18%	− 3.55	− 4.40%	− 2.53	− 5.83	− 3.70
All except DL	2.82	1.69%	0.39	− 1.05%	− 1.87	0.64%	1.15	− 1.13	− 0.11

Betas, net alphas, and gross alphas are calculated according to the Carhart four-factor model, in which annualized intercepts ($12 \times \alpha$) and respective t-statistics are estimated on VW net and gross returns on the portfolios of HMFs in Panel A and on EW net and gross returns on the portfolios of HMFs in Panel B. Average Value Added is expressed per month and is calculated by regressing calendar-time VW portfolios in Panel A (calendar-time EW portfolios in Panel B) against the Vanguard benchmarks, then calculating the value added for each month according to equation separately and then averaging these amounts, as specified in Eq. 4. Value Added is expressed in millions of U.S. dollars adjusted for inflation and expressed in January 2000 dollars

comport with the claims as to how the strategies operate. For example, the beta for the ELCC category, which aims to balance a leveraged equity position with a corresponding short position while maintaining average market exposure, does in fact come quite close to 1. Similarly, the beta for the Equity Market Neutral category is close to 0. Betas for the Long-Short and Absolute Return categories fall in the middle at 0.35 and 0.19, respectively. With respect to the equal-weighted figures, the expense ratios are quite similar to those of their value-weighted counterparts, yet on the other hand Turnover figures for EMN and DL are substantially higher.

The summary performance statistics from Table 3 reveal only lukewarm performance, the most generous interpretation of which would identify zero net alphas and slightly positive gross alphas for some of the sub-categories. The two alpha variables and their T-Statistics come from Carhart model regressions, while for the Value-Added figure we follow Berk and van Binsbergen (2015) and use the Vanguard

Table 4 This table shows equal-weighted averages of several variables across individual funds

	Market Beta	SMB Beta	HML Beta	UMD Beta	Net Alpha	Gross Alpha	Annualized Return	Average Age	Percent Funds Surviving
ABR	0.26	− 0.02	− 0.02	0.00	− 2.43%	− 0.76%	0.70%	4.93	64.09%
LSE	0.55	0.02	− 0.02	0.00	− 3.14%	− 1.43%	1.87%	5.85	57.25%
EMN	0.09	− 0.01	0.01	0.04	− 1.11%	0.41%	0.52%	6.77	55.81%
ELCC	1.03	− 0.06	− 0.06	− 0.01	− 3.30%	− 1.91%	8.75%	6.23	38.89%
ALL	0.40	0.00	− 0.02	0.00	− 2.60%	− 0.95%	1.57%	5.60	58.94%

The four betas are calculated according to the Carhart Four-Factor model. The net and gross alphas represent the respective monthly intercepts multiplied by 12. Annualized Return is calculated on a geometric basis based on net returns. Average Age is reported in years and is calculated as the number of months the fund is present in the sample divided by 12. Percent Funds Surviving represents the percentage of the funds in the sample still present in December 2018

portfolios as the benchmarks. In the case of the value-weighted results, for three of the four main sub-categories (excluding DL), the net alphas are indistinguishable from zero. For the same three sub-categories, the gross alphas are either significant or nearly significant, and range from 1.37 to 1.64% on an annualized basis. Only the Long-Short category deviates from this pattern, with its net alpha coming in at − 2.94% and its gross alpha at − 1.36%, the former of which is statistically significant even at the 1% level. The gross and net alphas for the equal-weighted portfolios are roughly similar. In The figures for Value Added, expressed in Millions of US dollars in January 2000 constant dollars, roughly follow this pattern. The Equity Market Neutral and ELCC categories have statistically positive Value-Added figures, suggesting some measure of skill, though the other two categories come in with negative Value Added but statistically insignificantly so. When we aggregate all four of these fund types into a single time series, the net alpha is statistically significantly negative, whereas the gross alpha and Value Added are close to zero. It is unclear why the Long-Short category performs so much worse than the other categories do, especially given that the ELCC funds essentially follow a form of Long-Short in which they aim for a market exposure corresponding to a market beta of 1; It is possible that something internal to the Long-Short strategies breaks down during our sample period, or else the result may simply reflect weak performance of a particularly large fund or funds that influence the aggregate results.

As we have mentioned earlier, we ignore in the rest of our study the Leveraged Equity ('DL') category because such funds merely amplify the returns of some benchmark through the application of leverage, yet we briefly highlight here some interesting results that to our knowledge no other study has examined. As expected, the Carhart model market beta comes in well above 1 at 2.17, and the turnover reaches 436% yearly. The VW expense ratio for these funds is the highest of any category in our study, 1.69%. The more interesting results are that the value-weighted

net alpha is − 6.85% and the gross alpha is − 3.73%, both statistically significant. Essentially, then, traders in these funds pay for a highly leveraged bet, losing slightly on average. We leave for other researchers whether they could do so more effectively on their own with home-made portfolios of puts, calls, or futures contracts.

Our next two tables cover cross-sectional averages calculated in two slightly different manners. First, Table 4 shows 'pure' cross-sectional averages in which each fund counts as a single observation. Afterwards, Table 5 follows Agarwal et al. (2009) and gives panel means from statistics calculated annually at year-end on a rolling 24-month basis (with a minimum of 18 observations for inclusion). Turning first to Table 4, it presents some interesting contrasts with the calendar-time results in Table 3. More specifically, these cross-sectional performance results appear substantially worse in terms of net and gross alpha. For example, the net alpha for the ELCC category is − 3.30% in the cross-section, while it is 0.27% in the earlier value-weighted result in Table 3. Similarly, the net alphas for other three categories are − 2.43%, − 3.14%, and − 1.11%. As a group, the funds earn a negative net alpha of − 2.60%. The gross alphas are also worse than the gross alphas in Table 3, all negative except for the Equity Market Neutral category, whose gross alpha is slightly positive at 0.41%. In addition, we calculate the average annualized return, net of fees, for each category, and the results look quite modest. Three of the four categories earn average returns between 0 and 2%, while the ELCC funds earn 8.75%. However, given that the funds in these three categories carry low betas with respect to the Market factor, it bears noting that they can achieve positive alpha (in theory) even with quite low returns. In addition, the table displays the average beta exposures to the other three factors of the Carhart model, all of which look small for all fund categories.

The last two columns in Table 4 deserve special mention. First, we see that the funds in our sample are quite young,

Table 5 Reported alphas represent panel means of alphas calculated for each fund on a year-end annual rolling basis using the prior 24 months return history (with a minimum of 18 months required)

	Net Alpha	T-Stat (Net Alpha)	Gross Alpha	T-Stat (Gross Alpha)	Value Added	T-Stat (Value Added)
ABR	− 1.62%	− 10.50	0.03%	0.18	− 0.09	− 0.70
LSE	− 2.32%	− 11.31	− 0.60%	− 2.95	− 0.15	− 1.64
EMN	− 0.74%	− 2.58	0.70%	2.41	0.00	− 0.01
ELCC	− 2.53%	− 6.33	− 1.20%	− 3.04	0.75	1.97
ALL	− 1.83%	− 15.81	− 0.21%	− 1.76	− 0.05	− 0.75

Alphas are calculated with respect to the Carhart model. Value Added represents panel means of monthly value added calculated for each fund on a year-end annual rolling basis using the prior 24 months return history (with a minimum of 18 months required). Value Added is based on the Vanguard Index model and is expressed in millions of U.S. dollars adjusted for inflation and expressed in January 2000 dollars

averaging only 5.60 years in age. In addition, we see that only approximately 59% of funds overall still exist by the end of the sample period, and only 39% for the ELCC funds. These numbers suggest an almost embryonic nature to the HMF space in which many firms attempt to enter the market and then quickly exit.

A well-known caveat applies with respect to cross-sectional averages such as those in Table 4. It can be problematic to compare funds with different history lengths or those that span different time periods during which their performances come under the influence of distinct macroeconomic conditions. One way to address this issue is to apply the panel approach of Agarwal et al. (2009) who, as mentioned above, calculate performance statistics on an annual rolling basis based on the prior 24 months. Table 5 implements this approach with respect to net alpha, gross alpha, and value added. The alphas reflect the Carhart model; value added, the Vanguard model. The resultant alphas occupy a middle-ground between the calendar-time alphas of Table 3 and the pure cross-sectional alphas of Table 4. That is, they are more negative than the calendar-time EW alphas, but less negative than the 'pure' cross-sectional ones. From the latter comparison, we infer that some number of short-lived, poorly performing funds drags down the cross-sectional figures while exerting a lesser influence on the panel means calculated in the manner of Agarwal et al. (2009). On the other hand, the value-added figures in Table 5 roughly mirror those of the VW calendar-time portfolios, with only one of five categories (EMN) changing with respect to its statistical significance. Finally, we should note that the performance averages of the various categories remain somewhat incomparable to the extent that the categories themselves span entirely distinct periods, as is true with the ELCC category and its later entry into the data.

On the surface, the prior three tables present a puzzle insofar as the cross-sectional average alphas produce results so much worse than those of the equal-weighted calendar portfolios. This issue warrants further investigation. At least two possible explanations present themselves. The first is the 'reverse survivorship bias' first documented by Linnainmaa (2013), to which we will return later. The second potential explanation, which we now pursue, may lay in the rapid expansion in the number of funds in the second half of our sample period. That is, if funds generally perform worse in the later years than they do in the earlier ones, while at the same time more funds exist in this later period, then the cross-sectional averages will take a greater hit than will the calendar-time portfolios, for which the earlier period equally impacts performance estimates as does the later period. Furthermore, inferior performance in recent years stands as a prominent finding in current literature such as Bollen et al. (2021) for hedge funds and Barras et al. (2010) for mutual funds.

Therefore, Table 6 splits the sample into two periods: Period 1, from inception through 2009; Period 2, from 2010 to 2018. We select the cut-off mark of 2009-2010 to coincide with the publication of Agarwal et al. (2009) in 2009, which in essence serves as the 'discovery' of the prior HMF outperformance 'anomaly,' so to speak. Note that this cutoff date also approximately partitions the sample into periods before and after the rapid expansion of HMF funds and aggregate AUM, as documented in Table 1, and which as described in the introduction followed on the heels of a wave of positive publicity. For each period, the table presents equal-weighted as well as value-weighted net alphas and gross alphas, all formed from calendar-time portfolios. In other words, for each of the five fund categories the table sets out four sets of results formed by splitting first on period and then again on VW versus EW.

The table reveals two prominent patterns. First, the EW results suffer a substantially worse dropoff from Period 1 to Period 2 than do their VW counterparts. In raw numbers, for example, the EW gross alpha drops by approximately 1.74% for 'All' funds, whereas the equivalent VW figure is only approximately 0.19% (the net alpha figures are 1.57% and 0.17%, respectively). In connection with the rapid expansion of the number of funds in Period 2, these results raise the prospect that the phenomenon of managers and fund families

Table 6 The table shows equal-weighted as well as value-weighted gross alphas, net alphas, and their *t*-statistics from calendar-time portfolios

	Equal-Weighted Gross Alpha	*T*-Stat	Value-Weighted Gross Alpha	*T*-Stat	Equal-Weighted Net Alpha	*T*-Stat	Value-Weighted Net Alpha	*T*-Stat
Panel A. Period 1								
ABR	2.06%	1.70	1.65%	1.17	0.26%	0.22	0.10%	0.07
LSE	− 0.05%	− 0.04	− 2.11%	− 1.35	− 2.00%	− 1.49	− 3.81%	− 2.43
EMN	2.18%	2.38	1.71%	1.98	0.73%	0.79	0.46%	0.53
ELCC	3.22%	2.23	6.06%	6.38	1.92%	1.31	4.99%	5.25
All	1.01%	1.03	− 0.03%	− 0.03	− 0.76%	− 0.78	− 1.48%	− 1.46
Panel B. Period 2								
ABR	− 0.39%	− 0.74	0.40%	0.61	− 2.03%	− 3.83	− 1.23%	− 1.89
LSE	− 0.97%	− 2.31	− 0.20%	− 0.17	− 2.65%	− 6.34	− 1.64%	− 1.38
EMN	0.60%	1.16	0.44%	0.89	− 0.82%	− 1.59	− 0.70%	− 1.43
ELCC	− 1.61%	− 2.04	− 0.75%	− 0.92	− 2.93%	− 3.70	− 1.86%	− 2.31
All	− 0.73%	− 1.83	− 0.22%	− 0.43	− 2.33%	− 5.85	− 1.65%	− 3.17

Panel A and Panel B respectively split the sample into two periods: Period 1, from inception through 2009; Period 2, from 2010 to 2018

attempting to cash in on 'hot' styles of mutual funds, as discussed in Cooper et al. (2005), may occur more commonly among smaller funds than among larger ones.

The other noteworthy feature of the table is that in Period 2 the EW returns clearly underperform the VW ones, whereas no such tendency holds in Period 1. More specifically, in Period 2 the EW alpha is inferior to the VW alpha for four of the five categories in the gross case and in all five of the categories for the net alphas. With respect to the gross alphas in Period 2, the average underperformance for the EW portfolios is 0.55% across the five categories. On the other hand, In Period 1 there is no such underperformance; it is noteworthy, though, that in Period 1 both EW and VW portfolios achieve mainly positive gross alphas (with the exception of long-short funds) but insignificant net alphas, indicating that the opportunity for investors to benefit from HMF outperformance perhaps had already passed by the time of publication for Agarwal et al. (2009). The imbalanced results in Period 2 offer a contrast to the calendar-time numbers for the entire sample period in Table 3, where minimal difference was observed between EW and VW portfolios over the entire sample period of the study.

On a minor side note, the ELCC category's abnormally strong alpha in Period 1 relative to Period 2 and to the entire sample period can be explained by the late start for the ELCC funds (only a single fund observation prior to 2007) so that the vast majority of ELCC observations occur in Period 2.

Earlier, we mentioned the 'reverse survivorship bias' of Linnainmaa (2013) as a potential explanation of the differential figures for the calendar-time portfolios and the cross-sectional averages. This phenomenon occurs whenever unlucky bad performance induces fund closures while fortuitously successful funds stay in operation. Indeed, one of the three

estimation procedures in Linnainmaa (2013) is simply the difference in performance between the cross-sectional average and the equal-weighted calendar-time portfolio. The original paper's various estimates for this bias range from approximately 0.60% to 1%. In the case of HMFs, however, the abnormally high failure rates of the fund may intensify the effect so that it becomes particularly severe. Given this context, studies of HMFs that rely on cross-sectional averages may arrive at erroneous inferences, and the same danger applies to studies of other active management areas experiencing harsh attrition.

With respect to testing procedures when evaluating rapidly expanding sectors of the asset management industry, such as the HMF space during our sample period, the 'reverse survivorship bias' creates yet another source of caution. The value-added paradigm of Berk and Green (2004) already stresses the importance of value-weighted results over equal-weighted ones, highlighting the potential for false inferences when examining only equal-weighted results. To this pitfall we should now add another, namely, that cross-sectional results may suffer from the reverse survivorship bias, particularly so when a large number of funds enter and then quickly exit the field.

The overall findings from this section, then, paints a picture in which the number of funds and AUM rapidly increases and then partially reverses course, and in which many new funds enter the scene and then subsequently perish. In addition, a substantial gap exists between the calendar-time and cross-sectional performance measures, suggesting that such measures may lead to erroneous conclusions in isolation. These problems may particularly intrude in cases such as this where AUM is highly concentrated and, as mentioned earlier, the bottom half of funds account for only approximately 4% of AUM. Nevertheless, it remains the case

that even the most charitable interpretation of the calendar-time performance measures must admit mainly disappointing results, with none of the categories earning significantly positive net alpha, where three of the four categories show gross alphas approximately equal to the expense ratio, but where the overall results with all four funds reveal a gross alpha insignificantly different from zero.

In terms of the predictions of the value-added paradigm, this section's results partially disappoint. In table 3 only two of the four categories carry significantly positive value added for the entire sample period when using value-weighted returns, though admittedly this is better than the result for the equal-weighted returns for which only a single category is positive and significant. On the other hand, in the second half of our sample period there is evidence that value-weighted returns outperform equal-weighted ones.

Value added

Although the summary performance statistics provide some useful insights, the real test of whether HMFs provide a useful service relates to their 'value added' as defined earlier according to equation 4. Indeed, one of the central research questions in this paper pertains to whether a positive figure for value added of HMFs can reconcile the poor results from earlier literature based on equal-weighted portfolios with the strong inflows of money into HMFs. Such a result would run parallel to the finding of Berk and van Binsbergen (2015) of positive value added for the mutual fund industry as a whole, regardless of any results for net alphas or other statistics that fail to weight for AUM. Table 7 provides monthly value added of HMFs by fund type. Following Berk and van Binsbergen (2015), we calculate the value-added statistic for each category in the *ex post* manner that weights funds according to the number of monthly observations. This approach more closely resembles the performance of the overall category than does the alternative *ex ante* method that weights each fund equally.

To some extent the results here point in the same direction as do the earlier ones. With the Vanguard benchmarks, Equity Market Neutral and ELCC funds produce positive value added at the 1% and 5% significance levels, respectively. Specifically, Equity Market Neutral and ELCC HMFs generate $348.7 thousand and $831.8 thousand of monthly value added, respectively. The Absolute Return, Long-Short, and 'ALL' funds display values that are statistically insignificantly different from zero. Note that value added of zero corresponds to gross alpha of zero, implying that investors pay simply for funds to deliver a return no better than the benchmark return. On the surface, then, these results should foster discouragement among investors in HMFs. The results appear worse when we use the Carhart Model as the benchmark. Only the Equity Market Neutral category retains

Table 7 The table shows monthly average value added of HMFs by fund type

	Vanguard funds	Four-factor
ABR	− 0.0295	− 0.0501
Standard errors	0.1079	0.1262
LSE	− 0.0794	− 0.2401***
Standard errors	0.0942	0.0970
EMN	0.3487***	0.2661**
Standard errors	0.1175	0.1222
ELCC	0.8318**	0.0786
Standard errors	0.3722	0.3885
All	0.0576	− 0.0747
Standard errors	0.0628	0.0685

Value Added is calculated according to equation 4 in which excess gross returns relative to the indicated benchmarks are multiplied by AUM in the prior month. Value Added is reported in millions of U.S. dollars adjusted for inflation and expressed in January 2000 dollars

*, **, and ***Indicate statistical significance at the 10%, 5%, and 1% level, respectively

significantly positive value added, and the Long-Short funds actually destroy value at the 1% significance level. All other estimates are statistically insignificant for the Carhart model, including that for the value added of All HMFs. The difference in estimates between the Vanguard and Carhart benchmarks invites varying interpretations. Benchmarking against Vanguard indices measures value added against an investment opportunity that was tradable and marketed at the time of the investment. Benchmarking against the factor portfolio measures risk-adjusted value added that neither takes into account transaction costs nor accounts for the managerial skill required to uncover the factor strategies that might not be available to most investors in the initial periods as discussed by Berk and van Binsbergen (2015).

Bootstrap simulations

The analysis of the previous section shows that HMF managers fail to produce value added. However, it remains possible that some managers possess superior skill that allows them to outperform benchmarks, in which case the disappointing earlier results would stem from underperforming managers whose influence on the average results might negate the success of the skilled managers. To test for the existence of skilled managers, we use bootstrap simulations in a manner that closely follows the methodology proposed by Kosowski et al. (2006) and applied by Fama and French (2010) on the universe of all equity mutual funds. Accordingly, in addition to four-factor model, we report results for CAPM and three-factor models. As recommended by Kosowski et al. (2006), we report *p*-values only for the resultant t-statistics.

Table 8 The table shows $t(\alpha)$ (based on net returns) bootstrap results for the following categories of hedged mutual funds at the end of the year: Absolute Return (ABR); Long/ Short Equity (LSE); Equity Market Neutral (EMN); Extended U.S. Large-Cap Core (ELCC); all categories combined (All Except DL)

		t-stat: net fund returns				
		80th	85th	90th	95th	Max
ABR	CAPM	1	1	1	1	0.19
	FF3	1	1	1	1	0.13
	Four-factor	1	1	1	1	0.01
LSE	CAPM	1	1	1	1	0.92
	FF3	1	1	1	1	0.96
	Four-factor	1	1	1	1	0.86
EMN	CAPM	0.96	0.81	0.73	0.70	0.12
	FF3	0.95	0.80	0.25	0.62	0.10
	Four-factor	0.78	0.54	0.33	0.39	0.29
ELCC	CAPM	1	1	1	1	0.97
	FF3	1	1	1	1	1
	Four-factor	1	1	1	1	0.99
All	CAPM	1	1	1	1	0.13
	FF3	1	1	1	1	0.19
	Four-factor	1	1	1	1	0.04

The table reports p-values of the empirically observed percentiles as the fraction of realizations in the bootstrapped distributions of $t(\alpha)$ that are above the empirical percentiles under the assumption of a world in which true α is zero. The p-values are reported for the CAPM, three-factor, and four-factor models

We test for the existence of nonzero true α in 30 different cross-sections of α estimates - for the five types of HMFs, for the CAPM, three-factor, and four factor models, and for the net and gross excess returns. In all cases we construct 1000 bootstrapped samples of excess returns for each fund and run regressions to find α estimates and their respective t-statistics, $t(\alpha)$. We then analyze the cross-section of the $t(\alpha)$ and find their percentiles. Because we focus on top performing managers, we test for the existence of skill in the top 80th, 85th, 90th, and 95th percentiles. In addition, we bootstrap the maximum of $t(\alpha)$. After obtaining the bootstrapped distributions of each percentile and the maximum under the assumption of the absence of skill, we compute p-values of the empirically observed percentiles as the fraction of realizations in the bootstrapped distributions that are above the empirical percentiles. To test for nonzero true α in actual fund returns, we set the true α of each fund to zero and then use bootstrap simulations on returns. After recording the empirical values for the respective percentiles of the t-statistics, we run 1000 simulations with bootstrapped standard errors. The reported p-values correspond to the proportion of times that the simulated t-statistics, under the null hypothesis of zero alpha, exceed those obtained from the actual distribution.

Table 8 shows that as in the case of net alphas, the vast majority of p-values are one, suggesting no benefit to investors of allocating money to these funds. The sole exceptions to this pattern are the 'Maximum' for All funds and for Absolute Return funds. Zero net alpha agrees with the theories of Berk and Green (2004) that net alphas should

converge to zero as investors allocate funds towards skilled managers, yet nevertheless such a result in no way excludes the possibility of skilled management. Indeed, it would be consistent with the theory that some gross alphas would be positive, indicating skill, but that a competitive landscape among investors result in an absence of positive net alphas. The results for the gross alpha are consistent with this narrative. To varying degrees, many of the p-values suggest some measure of underlying skill. Table 9 reports the bootstrap p-values of the selected percentiles of $t(\alpha)$ for gross alphas and the p-values of their maxima. The managers of EMN HMFs clearly demonstrate skill at every analyzed top percentiles with the p-values significant at 1% cutoff and at maximum with the p-values significant at 10% cutoff. The Absolute Return and All categories also demonstrate skill through significant p-values, but this is mostly untrue for the Long-Short and ELCC funds. Finally, Table 10 shows the results with the Vanguard funds as the benchmarks .They are quite similar to those with the factor-based models. For the net returns, almost all results are close to 1 except for the 'Maximum' for ABR and 'All', while several figures among the gross return numbers suggest underlying skill.

Overall, the results of the bootstrap simulation complement the results based on the value-added figures described in prior tables. Namely, while the value-added figures suggest that the HMF categories on average mainly fail to extract value from the markets, the bootstrap simulations indicate that a subset of superior funds possesses skill.

Table 9 The table shows $t(\alpha)$ (based on gross returns) bootstrap results for the following categories of hedged mutual funds at the end of the year: Absolute Return (ABR); Long/Short Equity (LSE); Equity Market Neutral (EMN); Extended U.S. Large-Cap Core (ELCC); all categories combined (All Except DL)

		t-stat: gross fund returns				
		80th	85th	90th	95th	Max
ABR	CAPM	0.24	0.06	0.01	0	0.07
	FF3	0.57	0.18	0.01	0	0.09
	Four-factor	0.47	0.08	0.04	0	0.01
LSE	CAPM	1	1	1	1	0.01
	FF3	1	1	0.98	0.99	0.01
	Four-factor	1	1	1	1	0.03
EMN	CAPM	0.00	0.00	0.00	0.00	0.07
	FF3	0.00	0.00	0.00	0.00	0.05
	Four-factor	0.00	0.00	0.00	0.00	0.05
ELCC	CAPM	0.93	1	0.7	0	0.24
	FF3	0.98	0.84	0.71	0.39	0.32
	Four-factor	0.99	0.91	0.92	0.65	0.40
All Except DL	CAPM	0.42	0.35	0.02	0	0.1
	FF3	0.51	0.32	0.01	0.01	0.08
	Four-factor	0.86	0.32	0.02	0	0.01

The table reports *p*-values of the empirically observed percentiles as the fraction of realizations in the boot-strapped distributions of $t(\alpha)$ that are above the empirical percentiles under the assumption of a world in which true α is zero. The *p*-values are reported for the CAPM, three-factor, and four-factor models

Table 10 The table shows bootstrapped $t(\alpha)$ for gross and net alphas with respect to the Vanguard benchmarks following categories of hedged mutual funds: Absolute Return (ABR); Long/Short Equity (LSE); Equity Market Neutral (EMN); Extended U.S. Large-Cap Core (ELCC); all categories combined

	T-stat: net fund returns				
	80th	85th	90th	95th	Max
ABR	1.00	1.00	1.00	1.00	0.01
LSE	1.00	1.00	1.00	1.00	0.86
EMN	0.78	0.54	0.33	0.39	0.29
ELCC	1.00	1.00	1.00	1.00	0.99
All Except DL	1.00	1.00	1.00	1.00	0.04
	T-stat: gross fund returns				
	80th	85th	90th	95th	Max
ABR	0.47	0.08	0.04	0.00	0.01
LSE	1.00	1.00	1.00	1.00	0.03
EMN	0.00	0.00	0.00	0.00	0.05
ELCC	0.99	0.91	0.92	0.65	0.40
All Except DL	0.86	0.32	0.02	0.00	0.01

The table reports *p*-values of the empirically observed percentiles as the fraction of realizations in the bootstrapped distributions of $t(\alpha)$ that are above the empirical percentiles under the null assumption that α equals zero

Persistence of value added

If some funds possess skill, despite the aggregate failure of HMF funds to produce value added, then this skill may persist across periods. In order to investigate this issue, we follow the procedure for testing the persistence of value added of Berk and van Binsbergen (2015). We first

examine how funds perform during the 'sorting period,' in which we assign funds into deciles based on their value added, and we then examine their performance afterwards in the subsequent period, the 'horizon window.' More specifically, within each sorting window we rank funds according to their 'skill ratio:'

Table 11 This table shows the performance of top decile funds following the procedure for testing the persistence of value added of Berk and van Binsbergen (2015)

Top decile	Window = 3 (2007–2018)		Window = 4 (2007–2018)		Window = 5 (2009–2018)	
	Value added	p-value	Value added	p-value	Value added	p-value
ABR	0.1190	0.2145	− 0.0408	0.5899	0.1609	0.1313
LSE	0.2201	0.0852	0.2930	0.2150	0.9901	0.0000
EMN	1.0441	0.0001	1.3062	0.0365	3.1114	0.0014
ALL	0.7648	0.0009	1.5635	0.0000	2.0228	0.0001
Top beats bottom	Percent	p-value	Percent	p-value	Percent	p-value
ABR	0.6389	0.0013	0.6250	0.0297	0.5417	0.1577
LSE	0.4706	0.7258	0.5764	0.0275	0.6300	0.0033
EMN	0.5664	0.0470	0.5694	0.0399	0.5667	0.0602
ALL	0.5903	0.0121	0.6667	0.0000	0.6167	0.0039
Top in top half	Percent	p-value	Percent	p-value	Percent	p-value
ABR	0.5245	0.2518	0.4965	0.5000	0.4167	0.9588
LSE	0.5417	0.1393	0.6042	0.0048	0.6250	0.0022
EMN	0.5486	0.1056	0.5486	0.1056	0.5500	0.1176
ALL	0.6319	0.0005	0.6458	0.0002	0.5917	0.0177

It displays results for three distinct lengths for the sorting period: 3, 4, and 5 years. The top panel presents the average value added and p-value for those funds in the top decile of value added in the prior window. The middle panel shows the frequency that the average value added of top decile places in the top half of decile averages, and the p-value. The bottom window shows the frequency and that the top decile funds achieve superior average value added than relative to those funds in the bottom decile. For the middle and bottom panels, the reported p-values are one-tailed and correspond to the binomial distribution with $p = 0.50$

$$\text{SKR}_i^\tau \equiv \frac{\hat{S}_i^\tau}{\sigma(\hat{S}_i^\tau)}, \tag{5}$$

where $\hat{S}_i^\tau = \sum_{t=1}^{\tau} \frac{V_{it}}{\tau}$, $\sigma(\hat{S}_i^\tau) = \sqrt{\sum_{t=1}^{\tau} \frac{(V_{it} - \hat{S}_i^\tau)^2}{\tau}}$, and τ represents the number of months within each sorting window.

In essence, the skill ratio at the end of each sorting window is fundamentally the t-statistic of the value-added estimate calculated during the sorting window. Therefore, ranking funds according to the skill ratio rather than to the raw value-added estimate corresponds to ranking according to our statistical confidence in the fund's ability to produce value added. Once we sort the funds into deciles in the sorting window, we conduct three separate tests on the performance of the top decile funds in the horizon window. First, we calculate the average monthly value added for the top decile funds in the horizon window. Secondly, we calculate the proportion of months in which the average value added for the top decile exceeds the average value added of the bottom decile. Finally, we calculate the proportion of months in which the average value added for the top decile exceeds the median of average value added for all deciles.

Table 11 presents the results of these tests. We display results for three distinct lengths for the sorting period: 3, 4, and 5 years. In each case, we assign the same number of

years to the horizon window. Because there are relatively few funds in the early years of our sample, we limit total number of years to twelve in the cases of 3 and 4 year sorting periods (amounting to a total of four and three sorting periods, respectively) and to 10 years in the case of the 5-year sorting window (two sorting windows). In each case, we require a minimum of 18 observations in each of the sorting period and the horizon window, and we re-estimate betas in each window. We use the Vanguard funds as the benchmark. Furthermore, there are too few ELCC funds to form deciles, and so we ignore them as a distinct category but include them with the other funds in the 'ALL' category. The p-values for 'Top Decile' value added are based on the time series of the monthly averages. The p-values for 'Top Beats Bottom' and 'Top in Top Half' are based on the binomial distribution wherein the number of trials equals the number of months, the number of successes equals the number of observations in which the average value added of the top decile exceeds the median or the bottom decile, respectively, and the probability equals 0.50.

The numbers in Table 11 suggest that skill persists across time among the top decile of HMFs. Indeed, the results stand in sharp contrast to those of earlier tests. For example, the various figures of value added for the top decile funds far exceed those calculated in Table 7 for all funds: Looking at only the numbers for 'ALL' funds, the value-added estimates

Table 12 The table shows results of regressing monthly value added of all HMFs combined and four HMF sub-categories separately on various characteristics

	Value added				
	All funds	ABR	LSE	EMN	ELCC
AUM	− 0.00429***	− 0.00341*	− 0.00574***	− 0.000178	− 0.00836***
	(− 3.645)	(− 1.849)	(− 3.670)	(− 0.141)	(− 7.754)
Industry Size	− 0.000130***	− 0.000164*	0.000375	− 0.000416	− 0.00270
	(− 3.328)	(− 1.767)	(1.303)	(− 1.632)	(− 1.276)
Lagged Top Decile x AUM	0.00218***	0.00001	0.00369*	− 0.000681	0.00407***
	(3.355)	(0.0684)	(2.026)	(− 0.501)	(6.751)
Turnover	− 0.00161*	− 0.0240	0.0721	− 0.00120	0.377
	(− 1.887)	(− 0.775)	(1.152)	(− 1.456)	(0.915)
Expense Ratio	− 0.335	0.447	− 0.0814	− 0.961*	− 6.266
	(− 0.784)	(0.429)	(− 0.221)	(− 1.827)	(− 1.268)
Lagged Value Added	− 0.322	− 0.632	− 0.580**	0.561**	− 0.195
	(− 1.494)	(− 1.573)	(− 2.606)	(2.855)	(− 1.592)
Constant	9.743***	7.421	− 4.799	5.290**	33.13
	(3.555)	(1.455)	(− 1.080)	(2.183)	(1.556)
Year-month FE	Yes	Yes	Yes	Yes	Yes
Fund FE	Yes	Yes	Yes	Yes	Yes
Fund type FE	YES	–	–	–	–
Observations	16389	5576	7129	2663	1021

The dependent variable in all specifications, Value Added is calculated in the current period as gross alpha multiplied by lagged AUM, where the betas are calculated in the prior 24 months on a rolling basis. Lagged Top Decile indicates top decile performance in value added over the prior 24 months on a rolling basis and is calculated with gross-of-fee excess HMF returns in relation to the Vanguard index model. Industry Size variable is calculated as the sum of all AUM in each month for all funds and each of the four specified HMF sub-categories, scaled by the then-current level of the S&P500 Index. AUM, Turnover and Expense Ratio are lagged. AUM are given in millions of U.S. dollars inflation-adjusted to January 2000. Turnover and expense ratio are expressed as percentages. t-statistics are reported in parentheses and are calculated using standard errors that are double clustered by fund and year

*, **, and ***Indicate statistical significance at the 10%, 5%, and 1% level, respectively

here are $764,800, $1.563 Million, and $2.02 Million for windows of three years, four, and five, respectively, whereas the value added for 'ALL' in Table 7 is a measly $57,600. The p-values for the figures just cited all register at less than .001, suggesting strong statistical evidence that the top funds can persistently produce value added. The high figures for value added in the top decile, combined with the much lower numbers in the earlier table, suggests that a few funds create value whereas many others destroy it. The results in 'Top Beats Bottom' and 'Top in Top Half' also provide evidence for persistence in value added, though less consistently. The precise numbers vary from window and category specification, but in 9 of the 12 outcomes for the 'Top Beats Bottom' panel the p-values stand below 0.05, and the same is true in 5 of 12 outcomes in the 'Top in Top Half' panel.

Thus, this table provides the most encouraging piece of evidence in this paper for HMFs. Whereas the earlier results mainly portray HMFs as ineffective, the evidence for persistence suggests that a few funds possess skill consistently across periods. Moreover, in combination with the results from the bootstrap analysis, a picture emerges in which a few top funds produce gross alpha and generate value added, though the majority of funds fail to deliver either.

Determinants of value added

The prior section establishes that a small subset of HMF funds demonstrates persistent skill. An obvious follow-up question concerns the characteristics of these funds. In other words, how might an observer hope to identify such funds in advance? Therefore, we next conduct two investigations into this question. Table 12 analyzes the determinants of value added in a general sense, while Table 13 looks specifically at the traits of funds in the top decile of performance as measured by value added.

A contextual note is due. It should be apparent that the issue of how funds achieve value added is an important question. However, per the framework of Berk and Green (2004) this is *not* an issue that should concern investors, who should continue to pursue net alpha rather than skill per se. This is because greater skill may simply attract inflows until net

Table 13 The table shows results of logistic regression where the variable Top Decile takes the value of one if HMFs belong to top 10% in terms of value added in the current month and zero otherwise, where value added is based on betas calculated in the prior 24 months on a rolling basis

	Top Decile (Vanguard)		Top Decile (Carhart)	
Lagged Top Decile	0.915***	0.599***	0.752***	0.407***
	(6.826)	(4.994)	(4.571)	(2.925)
Expense Ratio		− 0.033		− 0.0239
		(− 0.358)		(− 0.252)
Turnover		0.0001		− 0.0005
		(0.0612)		(− 0.264)
AUM		0.0009***		0.0009***
		(11.94)		(11.22)
Constant	− 3.479***	− 3.633***	− 3.573***	− 3.752***
	(− 37.80)	(− 21.58)	(− 36.43)	(− 21.25)
Observations	16753	16426	16753	16426

Lagged Top Decile indicates top decile performance in value added over the prior 24 months on a rolling basis and is calculated with gross-of-fee excess HMF returns in relation to Vanguard index (Vanguard) or the Carhart Four-Factor model (Carhart). AUM is lagged fund size in millions of U.S. dollars inflation-adjusted to January 2000. Expense ratio is given in percentages. *t*-statistics are reported in parentheses and are computed using standard errors that are double clustered by fund and year

*, **, and ***Indicate statistical significance at the 10%, 5%, and 1% level, respectively

alpha approaches zero. Thus, it is incorrect to pose this problem from the point of view of investors.

Nevertheless, this issue will be of acute interest to the asset management industry, where greater value added eventually should generate larger fees. We are unaware of any existing research into this precise question besides that of Berk et al. (2017), who find that mutual fund families but not outside investors efficiently allocate capital across funds. Even so, there are two obvious strands of the active management literature that might offer relevant predictive variables for value added. Since value added equals gross alpha multiplied by size, the two relevant strands are (i) alpha predictability, and (ii) diseconomies of scale in active management. In the former literature, for obvious reasons attention focuses on net alpha rather than on gross alpha, the relevant metric in the value-added paradigm. For the latter literature, the question reduces to whether funds experience diseconomies of scale and, if so, whether any characteristics either dampen or else accelerate the impact on performance as AUM increases?

A few relevant papers address diseconomies of scale in the active management industry. For example, Chen et al. (2004) find that diseconomies of scale more quickly hurt the performance of funds that focus on small-cap stocks, presumably because of these stocks' limited liquidity, but

that at the same time membership in a fund family enables a fund to expand in AUM with a lesser impact on performance. Ferreira et al. (2013) likewise show diseconomies of scale for funds that target small US stocks, and Yan (2008) demonstrates stronger diseconomies among funds with greater turnover and a focus on growth stocks. Harvey and Liu (2021) find strong evidence of fund-specific effects in terms of diseconomies of scale.

Of course, the predictability of fund performance is an exceptionally popular topic. Jones and Mo (2021) offer a useful summary. According to Elton et al. (1993), expense ratio and turnover are negatively related to performance. Kacperczyk et al. (2008) argue that their measure of return gap predicts subsequent performance. Other papers claim that deviations from benchmark returns predict performance, such as the discussion of R^2 in Amihud and Goyenko (2013) and that of active share by Cremers and Petajisto (2009).

In Table 12, we study the determinants of value added in a panel regression where the standard errors are double clustered by fund and time. Of the relevant variables cited above, we include those in our data, namely, turnover and expense ratio. In addition we also incorporate lagged AUM and a measure of industry size that prior studies such as Bollen et al. (2021) find negatively influences fund performance. More specifically, we add as explanatory variables lagged AUM as well as a measure of the total assets under active management for the category in question, the latter scaled by a proxy for US equity market capitalization. These variables test whether diseconomies of scale hurt performance at the level of the individual fund and the overall active management space, respectively. In order that the analysis include measures of recent performance, which any asset allocator would naturally consider, we calculate lagged value added on a rolling 24-month basis (minimum 18 months for inclusion) for each month. Afterwards, we calculate value added in the following month with the betas calculated from the prior 24-month estimation window. For each month we then form top deciles for lagged value added (which just to reiterate, are based on the prior 24-month estimation window) as well as for current month value added. Table 12 also includes an interaction term for AUM and lagged top decile, which we discuss in more depth below.

Our Table 12 results for AUM and industry size are significant. For two of the four categories and also 'All' the coefficient for AUM is negative and significant at the 1%, and it is negative and significant at the 10% level for one more category. Only for the EMN funds, the best performing category, is this coefficient insignificant. Likewise, industry size is negative at the 1% level for 'All' and at 5% for ABR category. The results for industry size suggest that competition hurts performance. Expense ratio is negative and significant for only one category. The non-effect of the expense ratio here, in contrast to many studies that show

its negative influence on net alpha, is likely explained in that value added is a pre-expense measure. Turnover is negative and significant only for one category and at the 10% level. Finally, somewhat to our surprise the results for Lagged Value Added are mixed, with one coefficient negative and significant, one positive and significant, and three insignificant.

Table 12 results for AUM deserve special attention. The negative coefficients here, where value added is the dependent variable, carry a different interpretation from the more familiar result in prior studies in which AUM decreases net alpha. In the studies with net alpha, AUM might lower alpha while the funds can still credibly claim that the extra money goes to good use; that is, the 'marginal dollar' net alpha might be lower than before while the 'marginal dollar' gross alpha remains above zero, so that no value destruction occurs. However, this benign interpretation is impossible with a negative effect of AUM on value added. Per the arguments of Berk and Green (2004), once AUM exceeds the capacity of managers to generate alpha, the worst-case scenario is that they invest additional dollars passively and that the value added therefore plateaus at some upper threshold. In fact, a positive coefficient on AUM should hold according to the value-added paradigm as a minimal condition for value creation (assuming AUM varies more than do expense ratios). Thus, the negative coefficient here in fact suggests that a large proportion of funds pursue value destroying strategies that earn negative gross alpha, and that additional funds therefore simply increase the magnitude of these losses. This must be acknowledged as a highly critical assessment on the HMF space, in which the market for active management fails in its basic function.

On the other hand, the interaction term Lagged Top Decile × AUM carries the opposite sign of AUM alone. The coefficient is positive and significant at the 1% level for 'All' and for one other category and at the 10% level for one more category. This is as it should be for an efficient market for active management. This result suggests that some small proportion of HMF funds pursue value-additive strategies. This unusual combination, which mixes a negative coefficient on AUM alone and a positive one for Lagged Top Decile × AUM, implies an HMF space with coexisting spheres of influence for efficiency and inefficiency in asset management industry.

Next, Table 13 performs a logistic regression that examines the characteristics of only the top decile funds. As in the prior table, we include size, expense ratio, and turnover. We also include as a variable Lagged Top Decile, again calculated on a rolling 24-month basis. This specification mirrors the prior table insofar as in each case a lagged version of the dependent variable serves as one of the independent variables (Lagged Value Added and Lagged Top Decile, respectively).

The sole strong result in Table 13 is that lagged top decile performance predicts subsequent top decile performance in all specifications. This is consistent with our prior table that analyzes top decile persistence in value added over multi-year periods. In a sense, the Lagged Top Decile variable may serve as a stand-in for characteristics that we have yet to identify. What these characteristics are, or whether certain funds simply persist well for idiosyncratic reasons, is a question ripe for further study. Furthermore, an interesting contrast arises between this table and the last: While there here appears to be persistence among the top decile funds, there was no evident pattern in the prior table of lagged value added predicting subsequent value added. Among the other variables, turnover and expense ratio are insignificant. The coefficient for AUM is positive and significant, though, admittedly, this outcome may arise mechanistically insofar as a larger firm with randomly strong alpha simply by stint of its size is more likely than a smaller but similarly lucky fund to achieve top decile status in terms of value added.

Conclusions

This paper examines to what extent the value-added approach to mutual funds can reconcile the recent large inflows into hedged mutual funds with the poor appraisal of their performance in academic studies. While a neat and tidy outcome would have paralleled Berk and van Binsbergen (2015)'s result for the broader mutual fund industry and found positive value added for the HMF space, no such picture emerges. While we partially confirm the bleak assessment of HMFs as a whole, we augment this view with a silver lining by demonstrating that pockets of efficiency and value creation coexist alongside a larger body of inefficiency. Three separate tests in this paper reenforce this message: the bootstrapped gross alphas, the test for persistence in value added among the top decile funds, and then finally the test for determinants of value added in which assets under management in all funds and in top performing funds influence value added in polar opposite directions.

Thus, our picture differs from that of the Berk and Green (2004) market-efficient equilibrium in the mutual fund industry, in which aggregate value added should remain positive while net alphas equal zero. On the one hand, the overall value added for HMFs is negligibly different from zero, suggesting that investors pay high expenses for these funds while the managers deliver pre-fee returns no different from those available through benchmarks. Yet, the approximate 25% decline in the assets under management of HMFs and the concomitant death of many HMF funds indicates that the market may be in the process of moving towards this efficiency. While the Berk and Green (2004) equilibrium may hold in the long-run, cynics of the asset management

industry may have a strong point in the short-run. From the point of view of investors in mutual funds, our results cast doubt on any strategy that chases recent strong performance among some faction of funds, since it is likely that inflows and diseconomies of scale will drive net alphas to zero. Indeed, the complete lack of evidence of positive net alphas in our bootstrap is one of our most unambiguous results.

We further contribute to knowledge of HMFs by highlighting a highly concentrated and unstable state of the HMF space, in which the top five funds control almost one-third of assets, the bottom half control only 4%, approximately 40% of funds disappear by the end of the sample, and the average age of a fund is only 5 years. We show that the rapid expansion in the HMF area coincides with a greater deterioration in EW results relative to VW ones, and we conjecture that this outcome arises due to a greater temptation among small funds to mimic 'hot' strategies. We discuss the implications of these conditions for performance evaluation tests.

Thus, whereas equilibrium theories predict simply that net alphas equal zero and aggregate value added should be positive, our HMF study paints a rather chaotic picture of an embryonic market that gradually feels its way towards efficiency, with many funds attempting to enter and then exiting. Other studies that examine out-of-sample performance of mutual fund predictors, especially those that receive great fanfare among the investing public, should be mindful of this potential pattern.

In conclusion, while our paper achieves its main objectives, it also broaches several topics worthy of further development. For example, researchers should explore in more detail what types of funds and fund families manage to create value added, and, on the flip side, what tell-tale warning signs might indicate a fund attempting to capitalize on naive investors who chase investment fads.

References

Agarwal, V., N.M. Boyson, and N.Y. Naik. 2009. Hedge funds for retail investors? An examination of hedged mutual funds. *Journal of Financial and Quantitative Analysis* 44 (2): 273–305.

Amihud, Y., and R. Goyenko. 2013. Mutual fund's r^2 as predictor of performance. *The Review of Financial Studies* 26 (3): 667–694.

Badrinath, S., and S. Gubellini. 2011. On the characteristics and performance of long-short, market-neutral and bear mutual funds. *Journal of Banking & Finance* 35 (7): 1762–1776.

Barras, L., O. Scaillet, and R. Wermers. 2010. False discoveries in mutual fund performance: Measuring luck in estimated alphas. *Journal of Finance* 65 (1): 179–216.

Berk, J.B., and R.C. Green. 2004. Mutual fund flows and performance in rational markets. *Journal of Political Economy* 112 (6): 1269–1295.

Berk, J.B., and J.H. van Binsbergen. 2015. Measuring skill in the mutual fund industry. *Journal of Financial Economics* 118 (1): 1–20.

Berk, J.B., J.H. Van Binsbergen, and B. Liu. 2017. Matching capital and labor. *The Journal of Finance* 72 (6): 2467–2504.

Bollen, N.P., J. Joenväärä, and M. Kauppila. 2021. Hedge fund performance: End of an era? *Financial Analysts Journal* 77: 109–132.

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

Chen, J., H. Hong, M. Huang, and J.D. Kubik. 2004. Does fund size erode mutual fund performance? the role of liquidity and organization. *American Economic Review* 94 (5): 1276–1302.

Clifford, C., B. Jordan, and T.B. Riley. 2013. Do absolute-return mutual funds have absolute returns? *The Journal of Investing* 22 (4): 23–40.

Cooper, M.J., H. Gulen, and P.R. Rau. 2005. Changing names with style: Mutual fund name changes and their effects on fund flows. *Journal of Finance* 60 (6): 2825–2858.

Cremers, K.M., and A. Petajisto. 2009. How active is your fund manager? A new measure that predicts performance. *The Review of Financial Studies* 22 (9): 3329–3365.

Elton, E.J., M.J. Gruber, S. Das, and M. Hlavka. 1993. Efficiency with costly information: A reinterpretation of evidence from managed portfolios. *The Review of Financial Studies* 6 (1): 1–22.

Fama, E.F., and K.R. French. 1995. Size and book-to-market factors in earnings and returns. *Journal of Finance* 50 (1): 131–155.

Fama, E.F., and K.R. French. 2010. Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance* 65 (5): 1915–1947.

Ferreira, M.A., A. Keswani, A.F. Miguel, and S.B. Ramos. 2013. The determinants of mutual fund performance: A cross-country study. *Review of Finance* 17 (2): 483–525.

Fung, W., and D.A. Hsieh. 2005. Extracting portable alphas from equity long/short hedge funds. In *The world of hedge funds: Characteristics and analysis*, 161–180. Singapore: World Scientific.

Harvey, C.R., and Y. Liu. 2021. Decreasing returns to scale, fund flows, and performance. *Duke I&E Research Paper*, (2017-13).

Huang, J.-Z., and Y. Wang. 2013. Should investors invest in hedge fund-like mutual funds? Evidence from the 2007 financial crisis. *Journal of Financial Intermediation* 22 (3): 482–512.

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

Jiang, G.J., and H.Z. Yüksel. 2019. Sentimental mutual fund flows. *Financial Review* 54 (4): 709–738.

Jones, C.S., and H. Mo. 2021. Out-of-sample performance of mutual fund predictors. *The Review of Financial Studies* 34 (1): 149–193.

Kacperczyk, M., C. Sialm, and L. Zheng. 2008. Unobserved actions of mutual funds. *The Review of Financial Studies* 21 (6): 2379–2416.

Kanuri, S. 2016. Hedged ETFs: Do they add value? *Financial Services Review* 25 (2): 181–198.

Kanuri, S., and R.W. McLeod. 2014. Performance of alternative mutual funds: The average investor's hedge fund. *Financial Services Review*. https://doi.org/10.2139/ssrn.2445279.

Kooli, M., and I. Stetsyuk. 2020. Are hedge fund managers skilled? *Global Finance Journal. Advance online publication.* https://doi.org/10.1016/j.gfj.2020.100574.

Kosowski, R., A. Timmermann, R. Wermers, and H. White. 2006. Can mutual fund "stars'' really pick stocks? New evidence from a bootstrap analysis. *Journal of Finance* 61 (6): 2551–2595.

Linnainmaa, J.T. 2013. Reverse survivorship bias. *Journal of Finance* 68 (3): 789–813.

Lo, A.W., and P.N. Patel. 2008. 130/30: The new long-only. *The Journal of Portfolio Management* 34 (2): 12–38.

McCarthy, D. 2013. Hedge funds versus hedged mutual funds: An examination of equity long/short funds. *The Journal of Alternative Investments* 16 (3): 6–24.

Pástor, L., and R.F. Stambaugh. 2012. On the size of the active management industry. *Journal of Political Economy* 120 (4): 740–781.

Sherrill, D.E., S.E. Shirley, and J.R. Stark. 2017. Actively managed mutual funds holding passive investments: What do ETF positions tell us about mutual fund ability? *Journal of Banking & Finance* 76: 48–64.

Yan, X.S. 2008. Liquidity, investment style, and the relation between fund size and fund performance. *Journal of Financial and Quantitative Analysis* 43 (3): 741–767.

Dr. Nathaniel Light is an Assistant Professor at St. John Fisher College. He earned a BA in Mathematics at Columbia University, MA in Economics at Central European University, MA in Financial Engineering at Temple University, and a doctorate degree in Finance at Temple University. Dr. Light's research interests include empirical asset pricing and capital markets. His research has been published in the Review of Financial Studies and presented at numerous seminars and conferences.

Dr. Ivan Stetsyuk is an Associate Professor at the University of Quebec in Montreal. He has published papers on the topics related to corporate governance, family business, and hedge funds. He is a member of the Eastern Finance Association, the Financial Management Association, the Fulbright Association, and the University Club of Montreal. Dr. Stetsyuk has earned a M.Sc. degree in Finance at Johns Hopkins University and Ph.D. in Finance at Temple University. He is a recipient of the Fulbright Scholarship.