Measuring Skill in the Mutual Fund Industry*

Jonathan B. Berk
Stanford University and NBER

Jules H. van Binsbergen
Stanford University and NBER

February 1, 2013

Abstract

Using the dollar-value that a mutual fund adds as the measure of skill, we find that the average mutual fund adds about $2 million per year and that this skill persists for as long as 10 years. We further document that investors recognize this skill and reward it by investing more capital with better funds. Better funds earn higher aggregate fees, and there is a strong positive correlation between current compensation and future performance.

*We could not have conducted this research without the help of the following research assistants to whom we are grateful: Ashraf El Gamal, Maxine Holland, Christine Kang, Fon Kulalert, Ian Linford, Binying Liu, Jin Ngai, Michael Nolop, William Vijverberg, and Christina Zhu. We thank George Chacko, Rick Green, Ralph Koijen, David Musto, Paul Pfleiderer, Anamaria Pieschacon, Robert Stambaugh and seminar participants at Robeco, Stockholm School of Economics, Stanford University, University of Chicago, University of Toronto, Vanderbilt, Wharton, the NBER summer institute, and the Stanford Berkeley joint seminar for helpful comments and suggestions.

An important principle of economics is that agents earn economic rents if, and only if, they have a skill in short supply. As central as this principle is to microeconomics, surprisingly little empirical work has addressed the question of whether or not talent is actually rewarded, or, perhaps more interestingly, whether people without talent can earn rents. One notable exception is the research on mutual fund managers. There, an extensive literature in financial economics has focused on the question of whether stock picking or market timing talent exists. Interestingly, the literature has not been able to provide a definitive answer to this question. Considering that mutual fund managers are among the highest paid members of society, this lack of consensus is surprising because it leaves open the possibility that mutual fund managers earn economic rents without possessing a skill in short supply.

Given the importance of the question, the objective of this paper is to re-examine whether or not mutual funds earn economic rents without possessing skill. We find that the average mutual fund adds value by extracting about $2 million a year from financial markets. More importantly, this value added is persistent. Funds that have added value in the past keep adding value in the future, for as long as 10 years. It is hard to reconcile our findings with anything other than the existence of money management skill. We find that the distribution of managerial talent is consistent with the predictions of Lucas (1978): higher skilled managers manage larger funds and reap higher rewards. One of our most surprising results is that investors appear to be able to identify talent and compensate it: current *compensation* predicts future performance.

Our methodology differs from prior work in a number of important respects. First, our dataset includes all actively managed U.S. mutual funds, thereby greatly increasing the power of our tests. Prior work has used shorter time periods and focused attention exclusively on funds that only hold U.S. stocks. Second, in addition to evaluating managers using a risk model, we also evaluate managers by comparing their performance to the investor's alternative investment opportunity set — all available Vanguard index funds (including funds that hold non-U.S. stocks). Prior work that has benchmarked managers in this way, has not ensured that these alternative opportunities were tradable and marketed at the time.

Finally, many prior studies have used the net alpha to investors, i.e., the average abnormal return *net* of fees and expenses, to assess whether or not managers have skill. However, as Berk and Green (2004) argue, if skill is in short supply, the net return is determined in equilibrium by competition between investors, and not by the skill of managers. One might hypothesize, based on this insight, that the *gross* alpha (the average abnormal return before fees) would be the correct measure of managerial skill. However, the gross alpha is a *return* measure, not a *value* measure. That is, a manager who adds a gross alpha of 1% on a \$10 billion fund adds more value than a manager who adds a gross alpha of 10% on a \$1 million fund. Thus, a better measure of skill is the expected value the fund adds, i.e., the product of the fund's abnormal return (the return before fees minus the benchmark return) and assets under management (AUM). This measure consists of two parts: the amount of money the fund charges in fees (the percentage fee multiplied by AUM), plus the amount it takes from or gives to investors (the overall dollar under- or over-performance relative to the benchmark). The amount of money collected in fees by the fund can only come from one of two places — investors in the fund or financial markets. By subtracting the amount of money taken from investors from the fees charged, we are left with the money extracted from financial markets. That is, the *value added* of the fund.

The strongest evidence we provide for the existence of investment skill is the predictability we document in value added. We find that past value added can predict future value added as far out as 10 years, which is substantially longer than what the existing literature has found using alpha measures. To understand why our results differ from the existing literature, consider a concrete example. In his first 5 years managing Fidelity's Magellan fund, Peter Lynch had a 2% monthly gross alpha on average assets of about \$40 million. In his last 5 years, his gross alpha was 20 basis points (b.p.) per month on assets that ultimately grew to over \$10 billion. Based on the lack of persistence in gross alpha, one could mistakenly conclude that most of Peter Lynch's early performance was due to luck rather than skill. In fact, his value added went from less than \$1 million/month to over \$20 million/month, justifying his reputation as one of the most successful mutual fund managers of all time.

The advantage of using our measure of value added is that it quantifies the amount of money the fund extracts from financial markets. What it does not measure is how the mutual fund company chooses to distribute this money. For example, some have argued that Peter Lynch's success resulted from Fidelity's superior marketing efforts. Our measure provides no insight into what resources Fidelity brought to bear to maximize Magellan's value added. It simply measures the end result. Of course, marketing efforts alone are not sufficient to generate a positive value added. If Peter Lynch had had no skill, he would have extracted nothing from financial markets and our value added measure would be zero. The costs of all other input factors, including marketing, would have been borne by investors. In fact, Fidelity's marketing skills might very well have complemented Peter Lynch's stock picking skills, and thus played a role in the twenty fold increase in Magellan's value added. Consequently, our measure should be interpreted broadly as the resulting product of *all* the skills used to extract money from financial markets.

We benchmark managers against the investment opportunity set faced by a passive investor, in this case the net return of Vanguard's index funds. Consequently, our measure of value added includes the value these funds provide in diversification services. By benchmarking funds against the *gross* return of Vanguard's index funds (that is, the return before the cost Vanguard charges for diversification services) we can also measure value added without diversification services. By undertaking this analysis, we find that about half of the total value the mean fund adds is attributable to diversification services and the other half to market timing and stock picking.

The primary objective of this paper is to measure the value added of mutual funds. Our perspective is therefore different from many papers in the mutual fund literature that are primarily concerned with the abnormal returns that investors earn in the fund. Nevertheless, we do provide new insight on that question as well. Once we evaluate managers against a tradable benchmark, we no longer find evidence of underperformance. Over the time period in our sample, the equally weighted net alpha is 3 b.p. per month and the value weighted net alpha is -1 b.p. per month. Neither estimate is significantly different from zero. Notice that in equilibrium investors should be indifferent between indexing their money or using an active manager. Therefore, when a benchmark that does not include transaction costs is used, we should expect to see a negative net alpha. This is what we find when we use the Fama-French-Carhart portfolios as the alternative opportunity set. Of course, the lack of transaction costs in the benchmark should not affect the relative performance of funds, and so, importantly, our persistence results do not depend on the benchmark or risk adjustment we use. Neither does our result that the average fund adds value.

The rest of the paper is organized as follows. In the next section we briefly review the literature. In Section 2 we derive our measure of skill and in Section 3 we explain how we estimate it. We describe the data in Section 4. Section 5 demonstrates that skill exists. We then analyze how this skill is rewarded in Section 6. Section 7 investigates what portion of managerial skill is attributable to diversification services rather than other sources, such as stock picking or market timing. Section 8 shows the importance of using the full sample of active funds rather than the subset most researchers have used in the past. Section 9 concludes the paper.

1 Background

The idea that active mutual fund managers lack skill has its roots in the very early days of modern financial economics (Jensen (1968)). Indeed, the original papers that introduced the Efficient Market Hypothesis (Fama (1965, 1970)) cite the evidence that, as a group, investors in active mutual funds underperform the market, and, more importantly, mutual fund performance is unpredictable. Although an extensive review of this literature is beyond the scope of this paper, the conclusion of the literature is that, as investment vehicles, active funds underperform passive ones, and, on average, mutual fund returns before fees show no evidence of outperformance. This evidence is taken to imply that active managers do not have the skills required to beat the market, and so in Burton Malkiel's words: the "study of mutual funds does not provide any reason to abandon a belief that securities markets are remarkably efficient" (Malkiel, 1995, p. 571).

In a recent paper on the subject, Fama and French (2010) re-examine the evidence and conclude that the average manager lacks skill. They do find some evidence of talent in the upper tail of the distribution of managers. However, based on their estimate of skill (gross alpha), they conclude that this skill is economically small. In this paper, we argue that the economic magnitude of skill can only be assessed by measuring the total dollar value added not the abnormal return generated. As we will see, when the economic value added is calculated by multiplying the abnormal return by assets under management, a completely different picture emerges: a top 10% manager is able to use her skill to add about $24 million a year, on average.

Researchers have also studied persistence in mutual fund performance. Using the return the fund makes for its investors, a number of papers (see Gruber (1996), Carhart (1997), Zheng (1999) and Bollen and Busse (2001)) have documented that performance is largely unpredictable, leading researchers to conclude that outperformance is driven by luck rather than talent.[1] In contrast, we show that the value added of a fund is persistent.

Despite the widespread belief that managers lack skill, there is in fact a literature in financial economics that does find evidence of skill. One of the earliest papers is Grinblatt and Titman (1989), which documents positive gross alphas for small funds and growth funds. In a follow-up paper (Grinblatt and Titman (1993)), these authors show that at least for a subset of mutual fund managers, stocks perform better when they are held by the managers than when they are not. Wermers (2000) finds that the stocks that mutual funds hold, outperform broad market indices, and Chen, Jegadeesh, and Wermers (2000) find that the stocks that managers buy outperform the stocks that they sell. Kosowski, Timmermann, Wermers, and White (2006) use a bootstrap analysis and find evidence, using gross and net alphas, that 10% of managers have skill. Kacperczyk, Sialm, and Zheng (2008) compare the actual performance of funds to the performance of the funds' beginning of quarter holdings. They find that, for the average fund, performance is indistinguishable, suggesting superior performance gross of fees, and thus implying that the average manager adds value during the quarter. Cremers and Petajisto (2009) show that the amount a fund deviates from its benchmark is associated with better performance, and that this superior performance is persistent. Cohen, Polk, and Silli (2010) and Jiang, Verbeek, and Wang (2011) show that this performance results from overweighting stocks that subsequently outperform the stocks that are underweighted. Finally, Del Guercio and Reuter (2013) find that directly sold funds, that is, funds not marketed by brokers, do not underperform index funds after fees, thus implying outperformance before fees.

There is also evidence suggesting where this skill comes from. Coval and Moskowitz (2001) find that geography is important; funds that invest a greater proportion of their assets locally do better. Kacperczyk, Sialm, and Zheng (2005) find that funds that con-

[1] Some evidence of persistence does exist in low liquidity sectors or at shorter horizons, see, for example, Bollen and Busse (2005), Mamaysky, Spiegel, and Zhang (2008) or Berk and Tonks (2007).

centrate in industries do better than funds that do not. Baker, Litov, Wachter, and Wurgler (2010) show that, around earnings announcement dates, stocks that active managers purchase outperform stocks that they sell. Shumway, Szefler, and Yuan (2009) produce evidence that superior performance is associated with beliefs that more closely predict future performance. Cohen, Frazzini, and Malloy (2007) find that portfolio managers place larger bets on firms they are connected to through their social network and perform significantly better on these holdings relative to their non-connected holdings. Using holdings data, Daniel, Grinblatt, Titman, and Wermers (1997) find some evidence of stock selection (particularly amongst aggressive growth funds) but fail to find evidence of market timing. Kacperczyk, Nieuwerburgh, and Veldkamp (2011) provide evidence that managers successfully market time in bad times and select stocks in good times. These studies suggest that the superior performance documented in other studies in this literature is likely due to specialized knowledge and information.

Despite evidence to the contrary, many researchers in financial economics remain unconvinced that mutual fund managers have skill. This reticence is at least partly attributable to the lack of any convincing evidence of the value added that results from this talent. Our objective is to provide this evidence.

2 Theory and Definitions

Let R^n_{it} denote the excess return (that is, the net-return in excess of the risk free rate) earned by investors in the i'th fund at time t. This return can be split up into the return of the investor's next best alternative investment opportunity R^B_{it}, which we will call the manager's *benchmark*, and a deviation from the benchmark ε_{it}:

$$R^n_{it} = R^B_{it} + \varepsilon_{it}. \tag{1}$$

The most commonly used measure of skill in the literature is the mean of ε_{it}, or the *net alpha*, denoted by α_i^n. Assuming that the benchmark return is observed (we relax this assumption later), the net alpha can be consistently estimated by:

$$\hat{\alpha}_i^n = \frac{1}{T_i}\sum_{t=1}^{T_i}\left(R_{it}^n - R_{it}^B\right) = \frac{1}{T_i}\sum_{t=1}^{T_i}\varepsilon_{it}. \tag{2}$$

where T_i is the number of periods that fund i appears in the database.

As we pointed out in the introduction, the net alpha is a measure of the abnormal return earned by investors, not the skill of the manager. To understand why, recall the intuition that Eugene Fama used to motivate the Efficient Market Hypothesis: just as the expected return of a firm does not reflect the quality of its management, neither does the expected return of a mutual fund. Instead, what the net alpha measures is the rationality and competitiveness of capital markets. If markets are competitive and investors rational, the net alpha will be zero. A positive net alpha implies that capital markets are not competitive and that the supply of capital is insufficient to compete away the abnormal return. A negative net alpha implies that investors are committing too much capital to active management. It is evidence of sub-optimality on the part of at least some investors.[2]

Some have argued that the gross alpha, α_i^g, the abnormal return earned by fund i before management expenses are deducted, should be used to measure managerial skill. Let R_{it}^g denote the *gross* excess return, or the excess return the fund makes before it takes out the fee f_{it}:

$$R_{i,t}^g \equiv R_{it}^n + f_{it} = R_{it}^B + \varepsilon_{it} + f_{it}. \tag{3}$$

[2]For a formal model that relates this underperformance to decreasing returns to scale at the industry level, see Pastor and Stambaugh (2010).

The gross alpha can then be consistently estimated as:

$$\hat{\alpha}_i^g = \frac{1}{T_i}\sum_{t=1}^{T_i}\left(R_{it}^g - R_{it}^B\right) = \frac{1}{T_i}\sum_{t=1}^{T_i}\left(f_{it} + \varepsilon_{it}\right). \tag{4}$$

Unfortunately, just as the internal rate of return cannot be used to measure the value of an investment opportunity (it is the net present value that does), the gross alpha cannot be used to measure the value of a fund. It measures the return the fund earns, not the value it adds.

To correctly measure the skills that are brought to bear to extract money from markets, one has to measure the dollar value of what the fund adds over the benchmark. To compute this measure, we multiply the benchmark adjusted realized gross return, $R_{it}^g - R_{it}^B$, by the real size of the fund (assets under management adjusted by inflation) at the end of the previous period, $q_{i,t-1}$, to obtain the realized value added between times $t-1$ and t:

$$V_{it} \equiv q_{i,t-1}\left(R_{it}^g - R_{it}^B\right) = q_{i,t-1}f_{it} + q_{i,t-1}\varepsilon_{it}, \tag{5}$$

where the second equality follows from (3). This estimate of value added consists of two parts — the part the fund takes as compensation (the dollar value of all fees charged), which is necessarily positive, plus any value the fund provides (or extracts from) investors, which can be either positive or negative. Our measure of skill is the (time series) expectation of (5):

$$S_i \equiv E[V_{it}]. \tag{6}$$

For a fund that exists for T_i periods, this estimated value added is given by:

$$\hat{S}_i = \sum_{t=1}^{T_i}\frac{V_{it}}{T_i}. \tag{7}$$

The average value added can be estimated in one of two ways. If we are interested in the mean of the distribution from which value added is drawn, what we term the *ex-ante* distribution, then a consistent estimate of its mean is given by:

$$\bar{S} = \frac{1}{N} \sum_{i=1}^{N} \hat{S}_i, \tag{8}$$

where N is the number of mutual funds in our database. Alternatively, we might be interested in the mean of surviving funds, what we term the *ex-post* distribution. In this case, the average value added is estimated by weighting each fund by the number of periods that it appears in the database:

$$\bar{S}_W = \frac{\sum_{i=1}^{N} T_i \hat{S}_i}{\sum_{i=1}^{N} T_i}. \tag{9}$$

Before we turn to how we actually compute V_{it} and therefore S_i, it is worth first considering what the main hypotheses in the literature imply about this measure of skill.

Unskilled managers, irrational investors

A widely accepted hypothesis, and the one considered in Fama and French (2010), is that no manager has skill. We call this the *strong form* no-skill hypothesis, originally put forward in Fama (1965, 1970). Because managers are unskilled and yet charge fees, these fees can only come out of irrational investors' pockets. These managers can either invest in the index, in which case they do not destroy value, or worse than that, they can follow the classic example of "monkey investing" by throwing darts and incurring unnecessary transaction costs. So under this hypothesis:

$$S_i \leq 0, \text{ for every } i, \tag{10}$$

$$\alpha_i^n \leq -\mathrm{E}\left(f_{it}\right), \text{ for every } i. \tag{11}$$

Because no individual manager has skill, the average manager does not have skill either. Thus, this hypothesis also implies that we should expect to find

$$\bar{S} = \frac{1}{N}\sum_{i=1}^{N} \hat{S}_i \leq 0. \tag{12}$$

The latter equation can also be tested in isolation. We term this the *weak form* no-skill hypothesis. This weak-form hypothesis states that even though some individual managers may have skill, the average manager does not, implying that at least as much value is destroyed by active mutual fund managers as is created. We will take this two part hypothesis as the Null Hypothesis in this paper.

If managers are unskilled, then by definition, skill must be unpredictable. That is, under the strong form of the Null Hypothesis, positive past value added cannot predict future value added. Therefore, persistence of positive value added in the data implies a rejection of this Null Hypothesis. It may be tempting to conclude that because AUM is persistent, it is possible to observe persistence in value added, even if return outperformance relative to the benchmark is not persistent. However, if past outperformance was due to luck and therefore does not persist into the future, then $E_t[R^g_{i,t+1} - R^B_{i,t+1}] = 0$ implying that $E_t[q_{it}(R^g_{i,t+1} - R^B_{i,t+1})] = 0$. That is, value added is not persistent.

Skilled managers, rational investors

The second hypothesis we consider is motivated by Berk and Green (2004) and states that managers have skill that is in short supply. Because of competition in capital markets, investors do not benefit from this skill. Instead, managers derive the full benefit of the economic rents they generate from their skill. If investors are fully rational, then these assumptions imply that the net return that investors expect to make is equal to the benchmark return. That is:

$$\alpha^n_i = 0, \text{ for every } i. \tag{13}$$

Because fees are positive, the expected value added is positive for every manager:

$$S_i > 0, \text{ for every } i. \tag{14}$$

When investors cannot observe skill perfectly, the extent to which an individual manager actually adds value depends on the ability of investors to differentiate talented managers from charlatans. If we recognize that managerial skill is difficult to measure, then one would expect unskilled managers to take advantage of this uncertainty. We would then expect to observe the presence of charlatans, i.e., managers who charge a fee but have no skill. Thus when skill cannot be perfectly observed, it is possible that for some managers $S_i \leq 0$. However, even when skill is not perfectly observable, because investors are rational, every manager must still add value in expectation. Under this hypothesis, the average manager must generate value, and hence we would expect to find:

$$\bar{S} > 0. \tag{15}$$

We will take this hypothesis as the Alternative Hypothesis.

Some have claimed, based on Sharpe (1991), that in a general equilibrium it is impossible for the average manager to add value. In fact, this argument has two flaws. To understand the flaws, it is worth quickly reviewing Sharpe's original argument. Sharpe divided all investors into two sets: people who hold the market portfolio, whom he called "passive" investors, and the rest, whom he called "active" investors. Because market clearing requires that the sum of active and passive investors' portfolios is the market portfolio, the sum of just active investors' portfolios must also be the market portfolio. This observation immediately implies that the abnormal return of the average active investor must be zero. As convincing as the argument appears to be, it cannot be used to conclude that the average active mutual fund manager cannot add value. In his definition of "active" investors, Sharpe includes *any* investor not holding the market, not just active mutual fund managers. If active individual investors exist, then as a group active mutual fund managers could provide a positive abnormal return by making trading profits from individual investors who make a negative abnormal return. Of course, as a group individual active investors are better off investing in the market, which leaves open the question of why these individuals are actively trading.

Perhaps more surprisingly to some, Sharpe's argument does not rule out the possibility that the average active manager can earn a higher return than the market return even if all investors, including individual investors, are assumed to be fully rational. What Sharpe's argument ignores is that even a passive investor must trade at least twice, once to get into the passive position and once to get out of the position. If we assume that active investors are better informed than passive, then whenever these liquidity trades are made with an active investor, in expectation, the passive investor must lose and the active must gain. Hence, the expected return to active investors must exceed the return to passive investors, that is, active investors earn a liquidity premium.

3 Choice of Benchmarks and Estimation

To measure the value that the fund either gives to or takes from investors, performance must be compared to the performance of the next best investment opportunity available to investors at the time, which we have termed the benchmark. Thus far, we have assumed that this benchmark return is known. In reality it is not known, so in this section we describe two methods we use to identify the benchmark.

The standard practice in financial economics is not to actually construct the alternative investment opportunity itself, but rather to simply adjust for risk using a factor model. In recent years, the extent to which factor models accurately correct for risk has been subject to extensive debate. In response to this, mutual fund researchers have opted to construct the alternative investment opportunity directly instead of using factor models to adjust for risk. That is, they have interpreted the factors in the factor models as investment opportunities available to investors, rather than risk factors.[3] The problem with this interpretation is that these factor portfolios were (and in some cases are) not actually available to investors.

There are two reasons investors cannot invest in the factor portfolios. The first is straightforward: these portfolios do not take transaction costs into account. For example, the momentum strategy requires high turnover, which not only incurs high transaction costs, but also requires time and effort to implement. Consequently, momentum index funds do not exist.[4] The second reason is more subtle. Many of these factor portfolios were discovered well after the typical starting date of mutual fund databases. For example, when the first active mutual funds started offering size and value-based strategies, the alternative investment opportunity set was limited to investments in individual stocks and well-diversified index funds. That is, these active managers were being rewarded for the skill of finding a high return strategy that was not widely known. It has taken a considerable amount of time for most investors to discover these strategies, and so using portfolios that can only be constructed with the benefit of hindsight ignores the skill required to uncover these strategies in real time.

For these reasons we take two approaches to measuring skill in this paper. First, we follow the recent literature by adopting a *benchmark approach* and taking a stand on the alternative investment opportunity set. Where we depart from the literature, however, is that we ensure that this alternative investment opportunity was marketed and tradable

[3] See, for example, Fama and French (2010). Note that interpreting the benchmarks as alternative investment opportunities is not the same argument as the one made by Pastor and Stambaugh (2002) for using benchmarks.

[4] AQR introduced a momentum "index" fund in 2009 but the fund charges 75 b.p. which is close to the mean fee in our sample of active funds. It also requires a $1 million minimum investment.

at the time. Because Vanguard mutual funds are widely regarded as the least costly method to hold a well-diversified portfolio, we take the set of passively managed index funds offered by Vanguard as the alternative investment opportunity set.[5] We then define the benchmark as the closest portfolio in that set to the mutual fund. If R_t^j is the excess return earned by investors in the j'th Vanguard index fund at time t, then the benchmark return for fund i is given by:

$$R_{it}^B = \sum_{j=1}^{n(t)} \beta_i^j R_t^j, \tag{16}$$

where $n(t)$ is the total number of index funds offered by Vanguard at time t and β_i^j is obtained from the appropriate linear projection of the i'th active mutual fund onto the set of Vanguard index funds. By using Vanguard index funds as benchmarks, we can be certain that investors had the opportunity to invest in the funds at the time and that the returns of these funds necessarily include transaction costs and reflect the dynamic evolution of active strategies. Notice, also, that if we use this benchmark to evaluate a Vanguard index fund itself, we would conclude that that fund adds value equal to the dollar value of the fees it charges. Vanguard funds add value because they provide investors with the lowest cost means to diversification. When we use net returns on Vanguard index funds as the benchmark, we are explicitly accounting for the value added of diversification services. Because active funds also provide diversification services, our measure credits them with this value added.

Of course, one might also be interested in whether active funds add value over and above the diversification services they provide. In Section 7, we investigate this question by using the *gross* returns on the Vanguard index funds as the benchmark thereby separating diversification services from stock picking and market timing. As we will see, even without

[5]The ownership structure of Vanguard — it is owned by the investors in its funds — also makes it attractive as a benchmark because there is no conflict of interest between the investors in the fund and the fund owners. Bogle (1997) provides a brief history of Vanguard's index fund business.

including diversification services, value added is highly persistent and positive.

Second, we use the traditional *risk-based approach.* The standard in the literature implicitly assumes the riskiness of the manager's portfolio can be measured using the factors identified by Fama and French (1995) and Carhart (1997), hereafter, the Fama-French-Carhart (FFC) factor specification. In this case the benchmark return is the return of a portfolio of equivalent riskiness constructed from the FFC factor portfolios:

$$R_{it}^{B} = \beta_i^{mkt}\mathrm{MKT}_t + \beta_i^{sml}\mathrm{SML}_t + \beta_i^{hml}\mathrm{HML}_t + \beta_i^{umd}\mathrm{UMD}_t,$$

where $\mathrm{MKT}_t, \mathrm{SML}_t, \mathrm{HML}_t$ and UMD_t are the realizations of the four factor portfolios and $\beta_i^{\cdot}$ are risk exposures of the i'th mutual fund, which can be estimated by regressing the fund's return onto the factors. Although standard practice, this approach has the drawback that no theoretical argument exists justifying why these factors measure systematic risk in the economy. Fama and French (2010) recognize this limitation but argue that one can interpret the factors as simply alternative (passive) investment opportunities. As we point out above, such an interpretation is only valid when the factors are tradable portfolios.

We picked eleven Vanguard index funds to use as benchmark funds (see Table 1). We arrived at this set by excluding all bond or real estate index funds and any fund that was already spanned by existing funds.[6] Because the eleven funds do not exist throughout our sample period, we first arrange the funds in order of how long they have been in existence. We then construct an orthogonal basis set out of these funds by projecting the n^{th} fund onto the orthogonal basis produced by the first $n-1$ funds over the time period when the n^{th} fund exists. The mean plus residual of this projection is the n^{th} fund in the orthogonal basis. In the time periods in which the n^{th} basis fund does not exist,

[6]The complete list of all Vanguard's Index funds can be found here:
https://personal.vanguard.com/us/funds/vanguard/all?reset=true&mgmt=i.

Fund Name	Ticker	Asset Class	Inception Date
S&P 500 Index	VFINX	Large-Cap Blend	08/31/1976
Extended Market Index	VEXMX	Mid-Cap Blend	12/21/1987
Small-Cap Index	NAESX	Small-Cap Blend	01/01/1990*
European Stock Index	VEURX	International	06/18/1990
Pacific Stock Index	VPACX	International	06/18/1990
Value Index	VVIAX	Large-Cap Value	11/02/1992
Balanced Index	VBINX	Balanced	11/02/1992
Emerging Markets Stock Index	VEIEX	International	05/04/1994
Mid-Cap Index	VIMSX	Mid-Cap Blend	05/21/1998
Small-Cap Growth Index	VISGX	Small-Cap Growth	05/21/1998
Small-Cap Value Index	VISVX	Small-Cap Value	05/21/1998

Table 1: **Benchmark Vanguard Index Funds:** This table lists the set of Vanguard Index Funds used to calculate the Vanguard benchmark. The listed ticker is for the Investor class shares which we use until Vanguard introduced an Admiral class for the fund, and thereafter we use the return on the Admiral class shares (Admiral class shares have lower fees but require a higher minimum investment.)

*NAESX was introduced earlier but was originally not an index fund. It was converted to an index fund in late 1989, so the date in the table reflects the first date we included the fund in the benchmark set.

we insert zero. We then construct an augmented basis by replacing the zero in the time periods when the basis fund does not exist with the mean return of the basis fund when it does exist. We show in the appendix that value added can be consistently estimated by first computing the projection coefficients (β_i^j in (16)) using the augmented basis and then calculating the benchmark return using (16) and the basis where missing returns are replaced with zeros.

To quantify the advantages of using Vanguard funds rather than the FFC factor mimicking portfolios as benchmark funds, Table 2 shows the results of regressing each FFC factor mimicking portfolio on the basis set of passively managed index funds offered by Vanguard. Only the market portfolio does not have a statistically significant positive alpha. Clearly, the FFC factor mimicking portfolios were better investment opportunities than what was actually available to investors at the time. In addition, the R^2 of the regressions are informative. The value/growth strategy became available as an index fund after size, so it is not surprising that the R^2 of the SMB portfolio is higher than the HML portfolio. Furthermore, the momentum strategy involves a large amount of active trading, so it is unlikely to be fully captured by passive portfolios, which accounts for the fact that the UMD portfolio has the lowest R^2 and the highest alpha.

	MKT	SMB	HML	UMD
Alpha (b.p./month)	2	22	35	70
t-Statistic	0.83	2.80	3.37	3.38
Adjusted R^2	99%	74%	52%	15%

Table 2: **Net Alpha of FFC Portfolios:** We regress each FFC factor portfolio on the Vanguard Benchmark portfolios. The table lists the estimate (in b.p./month) and t-statistic of the constant term (Alpha) of each regression, as well as the R^2 of each regression.

Given that the alpha of the FFC factor mimicking portfolios are positive, and that they do not represent actual investable alternatives, they cannot be interpreted as benchmark portfolios. Of course, the FFC factor specification might still be a valid risk model for a U.S. investor implying that it will correctly price all traded assets in the U.S., including U.S. mutual funds investing in international stocks. For completeness, we will report our results using both methods to calculate the fund's alpha, but we will always interpret the Vanguard funds as benchmark portfolios and the FFC factor specification as an adjustment for risk.

4 Data

Our main source of data is the CRSP survivorship bias free database of mutual fund data first compiled in Carhart (1997). The data set spans the period from January 1962 to March 2011. Although this data set has been used extensively, it still has a number of important shortcomings that we needed to address in order to complete our study. We undertook an extensive data project to address these shortcomings, the details of which are described in a 35-page online appendix to this paper. The main outcome of this project is reported below.

Even a casual perusal of the returns on CRSP is enough to reveal that some of the reported returns are suspect. Because part of our objective is to identify highly skilled managers, misreported returns, even if random, are of concern. Hence, we procured additional data from Morningstar. Each month, Morningstar sends a complete updated database to its clients. The monthly update is intended to completely replace the previous update. We purchased every update from January 1995 through March 2011 and constructed a single database by combining all the updates. One major advantage of this database is that it is guaranteed to be free of survivorship bias. Morningstar adds a new fund or removes an old fund in each new monthly update. By definition, it cannot change an old update because its clients already have that data. So, we are guaranteed that in each month whatever data we have was the actual data available to Morningstar's clients at that time.

We then compared the returns reported on CRSP to what was reported on Morningstar. Somewhat surprisingly, 3.3% of return observations differed. Even if we restrict attention to returns that differ by more than 10 b.p., 1.3% of the data is inconsistent. An example of this is when a 10% return is mistakenly reported as "10.0" instead of "0.10". To determine which database is correct we used dividend and net asset value (NAV) information reported on the two databases to compute the return. In cases in which in one database the reported return is inconsistent with the computed return, but in which the other database was consistent, we used the consistent database return. If both databases were internally consistent, but differed from each other, but within 6 months one database was internally inconsistent, we used the database that was internally consistent throughout. Finally, we manually checked all remaining unresolved discrepancies that differed by more than 20 b.p. by comparing the return to that reported on Bloomberg. All told, we were able to correct about two thirds of the inconsistent returns. In all remaining cases, we used the return reported on CRSP.

Unfortunately, there are even more discrepancies between what Morningstar and CRSP report for total assets under management (AUM). Even allowing for rounding errors, fully 16% of the data differs across the two databases. Casual observation reveals that much of this discrepancy appears to derive from Morningstar often lagging CRSP in updating AUM. Consequently, when both databases report numbers, we use the numbers reported on CRSP with one important exception. If the number reported on CRSP changed by more than 8× (we observed a number of cases where the CRSP number is off by a fixed number of decimal places) and within a few months the change was reversed by the same order of magnitude, and, in addition, this change was not observed on Morningstar, we used the value reported on Morningstar. Unfortunately, both databases contained significant numbers of missing AUM observations. Even after we used both databases as a source of information, 17.2% of the data was missing. In these cases, we filled in any missing observations by using the most recent observation in the past. Finally, we adjusted all AUM numbers by inflation by expressing all numbers in January 1, 2000 dollars.

The amount of missing expense ratio data posed a major problem.[7] To compute the gross return, expense ratios are needed and over 40% of expense ratios are missing on the CRSP database. Because expense ratios are actually reported annually by funds, we were able to fill in about 70% of these missing values by extending any reported observation during a year to the entire fiscal year of the fund and combining the information reported on Morningstar and CRSP. We then went to the SEC website and manually looked up the remaining missing values on EDGAR. At the end of this process, we were missing

[7]Because fees are an important part of our skill measure, we chose not to follow Fama and French (2010) by filling in the missing expense ratios with the average expense ratios of funds with similar AUM.

only 1.6% of the observations, which we elected to drop.

Both databases report data for active and passively managed funds. CRSP does not provide any way to discriminate between the funds. Morningstar provides this information, but their classification does not seem very accurate, and we only have this information after 1995. We therefore augmented the Morningstar classification by using the following algorithm to identify passively managed funds. We first generated a list of common phrases that appear in fund names identified by Morningstar as index funds. We then compiled a list of funds with these common phrases that were not labeled as index funds by Morningstar and compiled a second list of common phrases from these funds' names. We then manually checked the original prospectuses of any fund that contained a word from the first list but was not identified as an index fund at any point in its life by Morningstar or was identified as an index fund at some point in its life by Morningstar but nevertheless contained a phrase in the second list. Funds that were not tracked by Morningstar (e.g., only existed prior to 1995) that contained a word from the first list were also manually checked. Finally, we also manually checked cases in which fund names satisfied any of these criteria in some periods but not in others even when the Morningstar classification was consistent with our name classification to verify that indeed the fund had switched from active to passive or vice versa. We reclassified 14 funds using this algorithm.

It is important to identify subclasses of mutual funds because both databases report subclasses as separate funds. In most cases, the only difference among subclasses is the amount of expenses charged to investors, so simply including them as separate funds would artificially increase the statistical significance of any identified effect. For funds that appear in the CRSP database, identifying subclasses is a relatively easy process — CRSP provides a separator in the fund name (either a ":" or a "/"). Information after the separator denotes a subclass. Unfortunately, Morningstar does not provide this information, so for mutual funds that only appear on the Morningstar database, we used the last word in the fund name to identify the subclass (the details of how we did this are in the online appendix). Once identified we aggregated all subclasses into a single fund.

We dropped all index funds, bond funds and money market funds[8] and any fund observations before the fund's (inflation adjusted) AUM reached \$5 million. We also dropped funds with less than two years of data. In the end, we were left with 6054 equity funds. This sample is considerably larger than comparable samples used by other researchers. There are a number of reasons for this. Firstly, we do not restrict attention to funds that hold only U.S. equity. Clearly, managerial skill, if it exists, could potentially be used to pick non-U.S. stocks. More importantly, by eliminating any fund that at any point holds a *single* non-U.S. stock, researchers have been eliminating managers who might have had the skill to opportunistically move capital to and from the U.S.[9] Second, the Morningstar database contains funds not reported on CRSP. Third, we use the longest possible sample length available.

When we use the Vanguard benchmark to compute abnormal returns we chose to begin the sample in the period just after Vanguard introduced its S&P 500 index fund, that is January 1977. Because few funds dropped out of the database prior to that date, the loss in data is minimal, and we are still left with 5974 funds.

5 Results

We begin by measuring managerial skill and then show that the skill we measure is persistent. As is common in the mutual fund literature, our unit of observation is the

[8]We classified a fund as a bond fund if it held, on average, less than 50% of assets in stocks and identified a money market fund as a fund that on average held more than 20% of assets in cash.

[9]It is important to appreciate that most of the additional funds still hold mainly U.S. stocks, it is just that they also hold some non-U.S. stocks. As we will discuss in Section 7 expanding the sample to all equity funds is not innocuous; not only is the statistical power of our tests greatly increased but, more importantly, we will show that managerial skill is correlated to the fraction of capital in non-U.S. stocks.

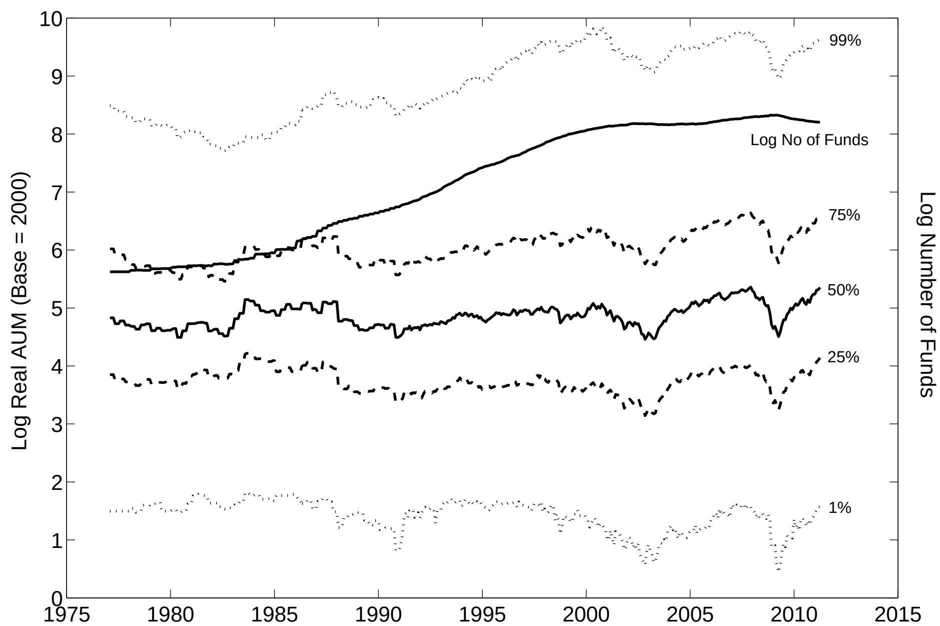


Figure 1: **Fund Size Distribution**
The graph displays the evolution of the distribution of the logarithm of real assets under management in $ millions (base year is 2000) by plotting the 1st, 25th, 50th, 75th and 99th percentiles of the distribution at each point in time. The smooth black line is the logarithm of the total number of funds.

fund not the individual manager. That is, we observe the dollar value the *fund* extracts from markets, or put another way, the fund's monopoly profits. We refer to these profits as "managerial skill" for expositional ease. Given that this industry is highly labor intensive, it is hard to conceive of other sources of these profits. However, it is important to keep in mind that this paper provides no direct evidence that these profits result from human capital alone.

5.1 Measuring Skill

We begin by first estimating S_i for every fund in our sample. Because S_i is the mean of the product of the abnormal return and fund size, one may have concerns about whether the product is stationary. Figure 1 allays such concerns because median inflation-adjusted fund size has remained roughly the same over our sample period. As the smooth solid line in the figure makes clear, growth in the industry's assets under management is driven by increases in the number of funds rather than increases in fund size.

Table 3 provides the cross-sectional distribution of S_i in our sample. The average fund adds an economically significant \$140,000 per month (in Y2000 dollars). The standard error of this average is just \$30,000, implying a t-statistic of 4.57. There is also large variation across funds. The fund at the 99th percentile cutoff generated \$7.82 million per month. Even the fund at the 90th percentile cutoff generated \$750,000 a month on average. The median fund lost an average of \$20,000/month, and only 43% of funds had positive estimated value added. In summary, most funds destroyed value but because most of the capital is controlled by skilled managers, as a group, active mutual funds added value.[10]

Thus far, we have ignored the fact that successful funds are more likely to survive than unsuccessful funds. Equivalently, one can think of the above statistics as estimates of the *ex-ante* distribution of talent. We can instead compute the time-weighted mean given by (9). In this case, we obtain an estimate of the *ex-post* distribution of talent, that is, the average skill of the set of funds actually managing money. Not surprisingly this estimate is higher. The average fund added \$270,000/month. When we use the FFC factor specification to correct for risk, we obtain very similar results.

It is tempting, based on the magnitude of our t-statistics to conclude that the Null Hypothesis (in both weak and strong form) can be rejected. However, caution is in order. There are two reasons to believe that our t-statistics are overstated. First, there is likely to be correlation in value added across funds. Second, the value added distribution features excess kurtosis. Even though our panel includes 6000 funds and 411 months, the sample might not be large enough to ensure that the t-statistic is t-distributed. However, under

[10]For the reasons pointed out in Linnainmaa (2012), our measures of value added underestimates the true skill of managers.

	Vanguard Benchmark	FFC Risk Measure
Cross-Sectional Mean	0.14	0.10
Standard Error of the Mean	0.03	0.03
t-Statistic	4.57	3.43
1st Percentile	-3.60	-3.93
5th Percentile	-1.15	-1.43
10th Percentile	-0.59	-0.77
50th Percentile	-0.02	-0.03
90th Percentile	0.75	0.70
95th Percentile	1.80	1.98
99th Percentile	7.82	6.76
Percent with less than zero	57.01%	59.70%
Cross-Sectional Weighted Mean	0.27	0.25
Standard Error of the Weighted Mean	0.05	0.06
t-Statistic	5.74	3.94
No. of Funds	5974	6054

Table 3: **Value Added ($\hat{S}_i$):** For every fund in our database, we estimate the monthly value added, $\hat{S}_i$. The *Cross-Sectional* mean, standard error, t-statistic and percentiles are the statistical properties of this distribution. *Percent with less than zero* is the fraction of the distribution that has value added estimates less than zero. The *Cross-Sectional Weighted* mean, standard error and t-statistic are computed by weighting by the number of periods the fund exists, that is, they are the statistical properties of $\bar{S}_W$ defined by (9). The numbers are reported in Y2000 $ millions per month.

the strong form of the Null Hypothesis, value added cannot be persistent. Consequently, if the value added identified in Table 3 results from managerial skill rather than just luck, we must also see evidence of persistence — managers that added value in the past should continue to add value in the future.

To test for persistence, we follow the existing literature and sort funds into deciles based on our inference of managerial skill. To infer skill at time τ, we construct what we term the *Skill Ratio* defined as:

$$SKR_i^\tau \equiv \frac{\hat{S}_i^\tau}{\sigma(\hat{S}_i^\tau)}, \tag{17}$$

where $\hat{S}_i^\tau = \sum_{t=1}^{\tau} \frac{V_{it}}{\tau}$ and $\sigma(\hat{S}_i^\tau) = \frac{\sqrt{\sum_{t=1}^{\tau}(V_{it}-\hat{S}_i^\tau)^2}}{\tau}$. The skill ratio at any point in time is essentially the t-static of the value added estimate measured over the entire history of the fund until that time.[11] We term the time period from the beginning of the fund to τ the *sorting period*. That is, the funds in the 10th (top) decile are the funds where we have the most confidence that the actual value added over the sorting period is positive. Similarly, funds in the 1st (bottom) decile are funds where we have the most confidence that the actual value added in the sorting period is negative. We then measure the average value added of funds in each decile over a specified future time horizon, hereafter the *measurement horizon*.[12]

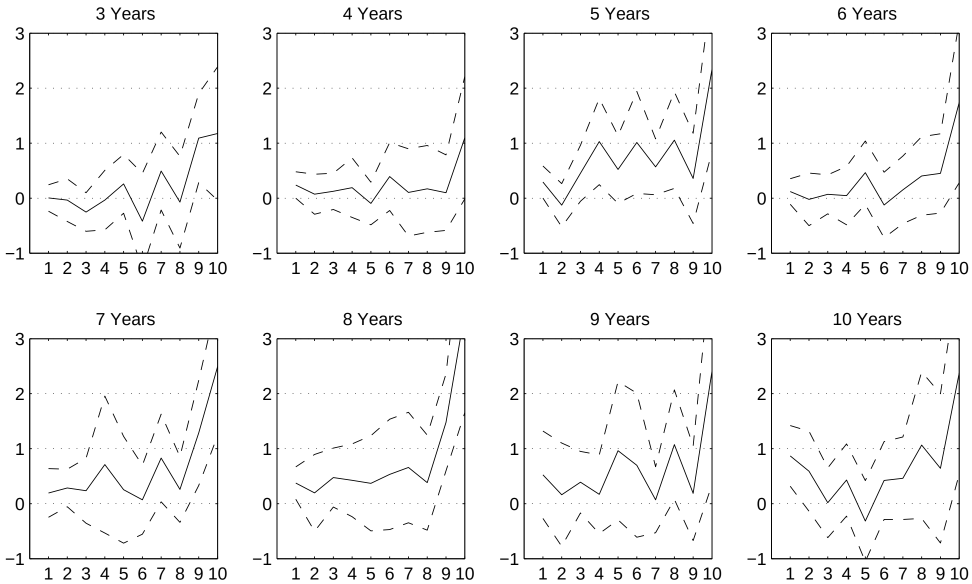
The main difficulty with implementing this strategy is uncertainty in the estimate of the fund's betas. When estimation error in the sorting period is positively correlated to the error in the measurement horizon, a researcher could falsely conclude that evidence of persistence exists when there is no persistence. To avoid this bias we do not use information from the sorting period to estimate the betas in the measurement horizon. This means that we require a measurement horizon of sufficient length to produce reliable beta estimates, so the shortest measurement horizon we consider is three years.

At each time τ, we use all the information until that point in time to sort firms into 10 deciles based on the skill ratio. We require a fund to have at least three years of historical data to be included in the sort. For each fund in each decile, we then calculate the value added, $\{V_{i,\tau}, \ldots, V_{i,\tau+h}\}$, over different measurement horizons, h, varying between 36 to 120 months using only the information in the measurement horizon. Because we need a minimum number of months, m, to estimate the fund's betas in the measurement horizon, we drop all funds with less than m observations in the measurement horizon. To remove the obvious selection bias, for the remaining funds we drop the first m value added observations as well, leaving the remaining observations exclusively in the horizon

[11]For ease of exposition, we have assumed that the fund starts at time 1. For a fund that starts later, the start date in the skill ratio is adjusted to reflect this.

[12]Similar results are obtained if we use the value added estimate itself to sort funds.

Panel A: Vanguard Benchmark



Panel B: FFC Risk Adjustment

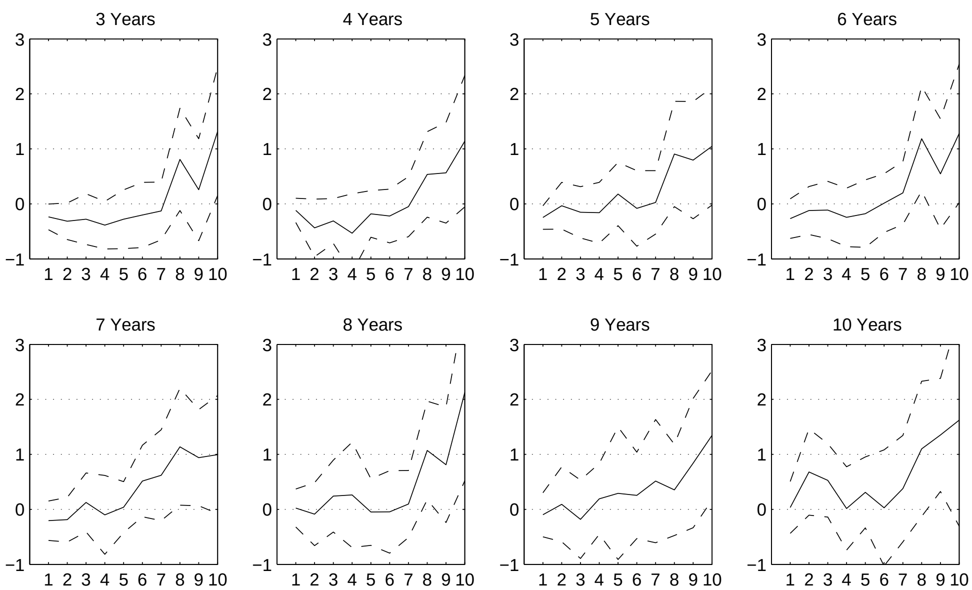


Figure 2: **Out-of-Sample Value Added**
Each graph displays average out-of-sample value added, $\hat{S}_i$ (in Y2000 \$ million/month), of funds sorted into deciles on the Skill Ratio, over the future horizon indicated. The solid line indicates the performance of each decile and the dashed lines indicated the two standard error bounds. Panel A shows the results when value added is computed using Vanguard index funds as benchmark portfolios and Panel B shows the results using the FFC risk adjustment.

Horizon	Value Added		Top Outperforms Bottom		Top in Top Half		Fraction of Total
Years	$ Mil	p-value (%)	Freq. (%)	p-value (%)	Freq. (%)	p-value (%)	**AUM** (%)
			Panel A: Vanguard Benchmark				
3	1.19	2.51	56.32	4.75	56.32	4.75	24.82
4	1.10	2.49	57.14	2.07	59.45	0.32	25.56
5	2.32	0.11	55.81	3.54	56.98	1.46	24.34
6	1.72	0.95	57.09	1.09	57.46	0.79	25.30
7	2.47	0.00	61.57	0.01	64.55	0.00	22.57
8	3.44	0.01	58.23	0.67	58.65	0.46	25.65
9	2.42	1.00	54.21	9.15	55.31	4.50	24.94
10	2.38	0.52	54.69	5.55	57.93	0.31	24.95
			Panel B: FCC Risk Adjustment				
3	1.30	1.33	56.13	0.47	57.63	0.06	17.93
4	1.13	3.01	58.14	0.02	57.72	0.05	19.50
5	1.03	2.68	59.60	0.00	58.79	0.01	17.88
6	1.27	2.22	58.85	0.01	56.50	0.28	19.38
7	0.98	3.37	59.71	0.00	56.12	0.44	17.91
8	2.13	0.42	59.12	0.01	57.14	0.13	19.01
9	1.35	1.12	56.51	0.18	55.15	1.09	16.10
10	1.62	4.67	58.91	0.01	56.74	0.22	21.83

Table 4: **Out-of-sample Performance of the Top Decile:** The two columns labeled "Value Added" report the average value added of the top decile at each horizon and the associated p-value. The next two columns report the fraction of the time and associate p-value the top decile has a higher value added realization than the bottom decile. The columns labeled "Top in Top Half" report the fraction of time the realized value added of the top decile is in the top half, and the final column reports the average fraction of total AUM in the top decile. All p-values are one tailed, that is, they represent the probability, under the Null Hypothesis, of the observed test-statistic value or greater.

$\{V_{i,\tau+m}, \ldots, V_{i,\tau+h}\}$. Because the Vanguard benchmark has at most 11 factors plus the constant, we use $m = 18$. We use $m = 6$ when we adjust for risk using the FFC factor specification. We then average over funds in each decile in each month, that is, we compute, for each decile, a monthly average value added. At the end of the horizon, funds are again sorted into deciles based on the skill ratio at that time, and the process is repeated as many times as the data allows.[13] At the end of the process, in each decile, we have a time series of monthly estimates for average value added. For each decile, we then compute the mean of this time series and its standard error. This procedure ensures that there is no selection bias because this strategy is tradable and could be implemented by an

[13]We choose the starting point to ensure that the last month is always included in the sample.

investor in real time. We therefore do not require funds to exist for the full measurement horizon. Finally note that this strategy uses non-overlapping data.[14]

Figure 2 plots the mean as well as the two standard error bounds for each decile and time horizon. From Figure 2 it appears that there is evidence of persistence as far out as 10 years. The point estimate of the average value added of 10th decile managers is positive at every horizon and is always the best preforming decile. The value added estimates are economically large. Although clearly noisy, the average tenth decile manager adds around \$2 million/month. Table 4 formally tests the Null Hypothesis that the value added of 10th decile is zero or less, under the usual asymptotic assumptions. The Null Hypothesis is rejected at every horizon at the 95% confidence interval, however, as we have noted above we have concerns about the validity of the t-test.[15]

If managers are skilled, and there are cross-sectional differences in the amount of skill, then relative performance will be persistent. Hence, we can use relative performance comparisons to construct a more powerful test of the strong form of the Null Hypothesis (that skill does not exist) by counting the number of times the 10th decile outperforms the 1st, and the number of times the 10th decile is in the top half.[16] As is evident from Table 4, the Null Hypothesis can be rejected at the 95% confidence level at almost all horizons. The FFC factor specification produces much more definitive results; with the sole exception of the nine-year horizon, the Null Hypothesis can be rejected at the 99% confidence level. Based on the results of this non-parametric test, we can definitively reject the strong form of the Null Hypothesis: skilled managers exist. Finally, note from the final column of Table 4 the disproportionate share of capital controlled by 10th decile

[14]Note that even after dropping the first m observations, the strategy is still tradable, because it is implementable at $\tau + m$.

[15]The earlier concerns are less important in this case because in each month we average over funds so the t-statistic is calculated using time series observations of the decile mean, thereby substantially reducing the effect of cross fund correlation and excess kurtosis.

[16]Because the volatility of the deciles varies, we restrict attention to tests where the probability under the Null Hypothesis is not a function of the volatility of the decile.

	Vanguard Benchmark	FFC Risk Measure
Equally Weighted	2.74	-3.88
t-statistic	0.73	-1.40
Value Weighted	-0.95	-5.88
t-statistic	-0.31	-2.35

Table 5: **Net Alpha (in b.p./month):** The table reports the net alpha of two investment strategies: Investing \$1 every month by equally weighting over all existing funds (*Equally Weighted*) and investing \$1 every month by value weighting (based on AUM) over all existing funds (*Value Weighted*).

managers. Investors reward skilled managers by providing them with more capital.

It might be tempting, based on our sorts, to conclude that all the skill is concentrated in 10th decile managers, that is, at most 10% of managers actually have skill. But caution is in order here. Our sorts are unlikely to separate skill perfectly. Although the estimates of value added in the other deciles are not significantly different from zero, they are almost all positive. Since we know that many managers destroyed value over the sample period, these positive point estimates imply that enough skilled managers are distributed throughout the other deciles to overcome the significant fraction of managers that destroy value.

Our persistence results stand in contrast to the existing literature that has found little evidence of persistence, mainly concentrated at horizons of less than a year and in poorly performing funds. The reason we find such strong evidence of persistence is due to our measure of skill — value added. If, instead, a return measure is used as the measure of skill, then the endogenous flow of funds obscures this evidence.

5.2 Returns to Investors

Given the evidence of skill, a natural question to ask is who benefits from this skill? That is, do mutual fund companies and managers capture all the rents, or are these rents shared with investors? Table 5 provides summary evidence. The average net alpha across all

	Deciles Sorted on Skill Ratio									
	Low	2	3	4	5	6	7	8	9	High
					Panel A: Vanguard Benchmark					
Value Added	-0.76	-0.68	-0.54	-0.39	-0.22	-0.05	0.24	0.54	1.19	2.03
Realized Net Alpha	-28	-9	-10	-7	-2	0	7	7	14	25
AUM	234	358	348	399	421	441	489	495	616	760
Age	141	220	198	217	209	238	247	242	240	234
Compensation	0.16	0.27	0.29	0.34	0.35	0.39	0.44	0.43	0.52	0.62
Fees	1.14	1.26	1.30	1.29	1.32	1.32	1.35	1.32	1.31	1.25
					Panel B: FCC Risk Adjustment					
Value Added	-0.91	-0.67	-0.53	-0.52	-0.33	-0.12	0.15	0.53	1.35	2.08
Realized Net Alpha	-35	-19	-14	-5	-2	1	7	10	17	24
AUM	193	263	261	402	413	472	446	505	700	727
Age	196	245	208	265	230	231	233	243	241	276
Compensation	0.19	0.21	0.22	0.33	0.34	0.39	0.38	0.43	0.57	0.60
Fees	1.38	1.28	1.28	1.25	1.23	1.24	1.30	1.32	1.27	1.24

Table 6: **Characteristics of Deciles Sorted on the Skill Ratio:** *Value Added* is the within decile average $\hat{S}_i$ in Y2000 \$ millions/month, *Compensation* is the within decile average of $\sum q_{it} f_{it}/T_i$, (in Y2000 \$ millions/month), *Fees* (in %/annum) is the within decile average $\sum f_{it}/T_i$, *Realized Net Alpha* (in b.p./month) is the realized (*ex-post*) abnormal return to investors, AUM is average assets under management (in Y2000 \$ millions) and *Age* is the average number of monthly observations in the decile.

funds is not significantly different from zero, so there is no evidence that investors share in the fruits of this skill.

Lower net alpha estimates are produced when the FFC factors are used as a measure of risk. In fact, on a value weighted basis, investors earned a significantly negative net alpha. But, as we have noted, relying on these estimates requires the additional assumption that this model correctly measures risk. If one instead interprets the FFC factor portfolios as the alternative investment opportunity set, then one would expect a negative alpha because these portfolios ignore transaction costs and were not necessarily available to investors at the time.

6 Skill and Labor Markets

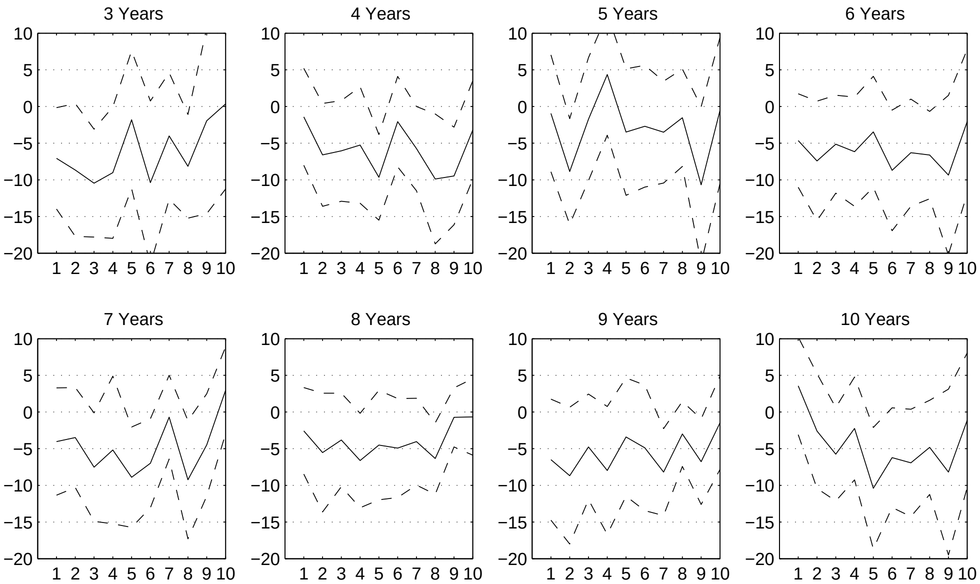
To see how realized value added is cross-sectionally distributed in the sample, we sorted funds into deciles based on the Skill Ratio calculated over the whole sample. The first row of Table 6 gives the (equal weighted) average value added in each decile. There is clearly large variation in the data. The worst decile destroyed almost $800,000 (Y2000) per month while the largest decile added about $2 million. Only the top 4 deciles added value, but funds in these deciles controlled 52% of the capital under management. Realized (ex post) net-alpha increases over the deciles. Interestingly, it is greater than zero for all the deciles that added value and is less than zero in all deciles that destroyed value. At least *ex-post*, managers that added value also gave up some of this value to their investors.

What about *ex-ante*? That is, would an investor who identified a skilled manager based on past data have expected a positive net alpha? To investigate this possibility, we calculate the average net alpha of investing in the out-of-sample sorts. That is, in each decile, for each fund and at each point in time, we calculate the net abnormal return,

$$\varepsilon_{it} = R_{it}^{n} - R_{it}^{B}$$

over measurement horizons of three to 10 years. As before, we drop the first m observations in the measurement horizon, where $m = 18$ or 6 months for the Vanguard Benchmark and the FFC factor specification, respectively. We then compute, at each point in time, the weighted average net abnormal return of each decile. At the end of the process, in each decile, we have a time series of monthly estimates for the weighted average net alpha of each decile. This time series represents the payoff of investing $1 in each month in a value weighted portfolio of funds in the decile. That is, the return of an extra marginal dollar invested in each decile. To get the average net alpha of this strategy, we compute the mean of this time series and its standard error. Figure 3 plots this mean as well as

Panel A: Vanguard Benchmark



Panel B: FFC Risk Adjustment

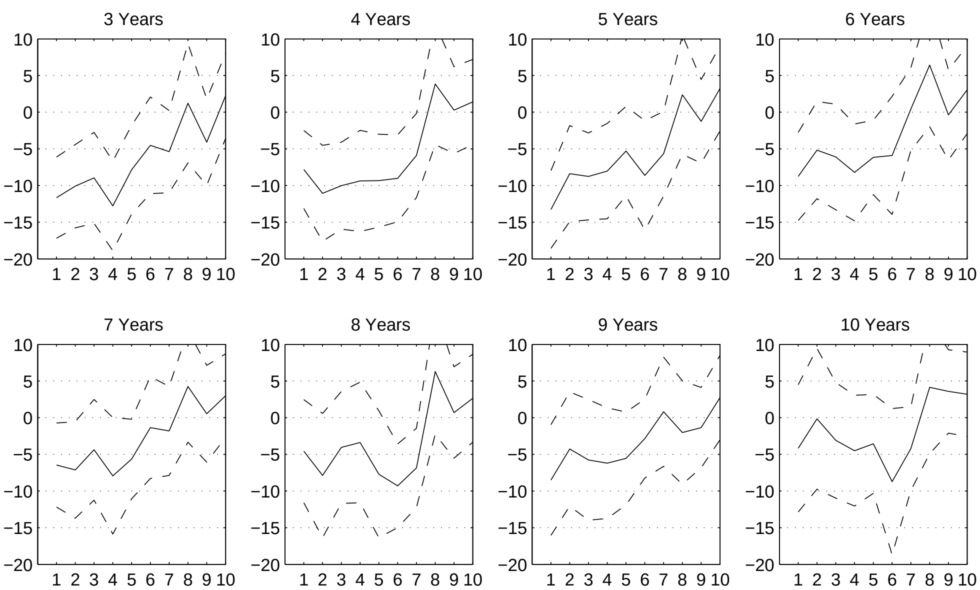


Figure 3: **Out-of-Sample Net Alpha**
Each graph displays the out-of-sample performance (in b.p./month) of funds sorted into deciles on the Skill Ratio, $\hat{S}_i$, over the horizon indicated. The solid line indicates the performance of each decile and the dashed lines indicated the 95% confidence bands (two standard errors from the estimate). Panel A shows the results when net alpha is computed using Vanguard index funds as benchmark portfolios, and Panel B shows the results using he FFC risk adjustment.

Horizon	**Net Alpha**		**Top Outperforms Bottom**		**Top in Top Half**	
Years	b.p.	t-stat.	Freq. (%)	p-value (%)	Freq. (%)	p-value (%)
			Panel A: Vanguard Benchmark			
3	0	0.06	54.21	13.82	58.42	1.21
4	-3	-0.95	55.76	5.15	57.14	2.07
5	-1	-0.11	52.33	24.68	52.33	24.68
6	-2	-0.43	54.91	5.84	55.27	4.56
7	3	0.99	58.58	0.29	60.08	0.06
8	-1	-0.29	53.17	18.16	56.54	2.56
9	-1	-0.46	53.11	16.64	56.41	1.97
10	-1	-0.24	47.90	78.71	52.10	24.74
			Panel B: FCC Risk Adjustment			
3	2	0.76	59.14	0.00	60.43	0.00
4	1	0.47	54.97	1.72	57.72	0.05
5	3	1.11	60.40	0.00	59.19	0.00
6	3	1.02	58.21	0.02	57.78	0.04
7	3	1.04	56.12	0.44	56.12	0.44
8	3	0.88	59.34	0.00	57.36	0.10
9	3	0.96	57.28	0.05	54.95	1.37
10	3	1.10	56.09	0.51	55.44	1.11

Table 7: **Out-of-sample Net Alpha of the Top Decile:** The columns labeled "Net-Alpha" report the weighted average net alpha (in b.p./month) of the top decile at each horizon and the associated p-value. The next two columns report the fraction of the time and associate p-value the top decile has a net alpha realization greater than the bottom decile. The columns labeled "Top in Top Half" report the fraction of time the realized net alpha of the top decile is in the top half. All p-values are one tailed, that is, they represent the probability, under the Null Hypothesis, of the observed test-statistic value or greater.

the two standard error bounds for each decile and time horizon.

Almost all net alpha estimates are not statistically significantly different from zero. As we show in Table 7, the point estimates of the tenth decile are very close to zero and mostly negative. The order statistics in Table 7 confirm the overall impression from Figure 3 that there is weak evidence of predictability in net alpha. Of the 16 order statistics, 7 of them have a one tailed p-value below 5% and only two are below 1%. So, at best, there is weak evidence that by picking the best managers, investors can get better returns than by picking the worst managers. The evidence appears more consistent with the idea that competition in capital markets drives net alphas close to zero.

In this case, there is a striking difference when we use the FFC factor specification as a risk adjustment. There is strong, statistically significant evidence of relative performance differences across the deciles (all but 3 of the order statistics are below 1% in this case). Both Figure 3 and Table 7 provide convincing out-of-sample evidence that investors could have done better by picking managers based on the Skill Ratio. These out-of-sample net alpha results are intriguing because they imply that either investors are leaving money on the table (not enough funds are flowing to the best managers resulting in positive net alphas), or investors do not care about the net alpha relative to the FFC factor specification, raising the possibility that the FFC factor specification does not measure risk that investors care about.

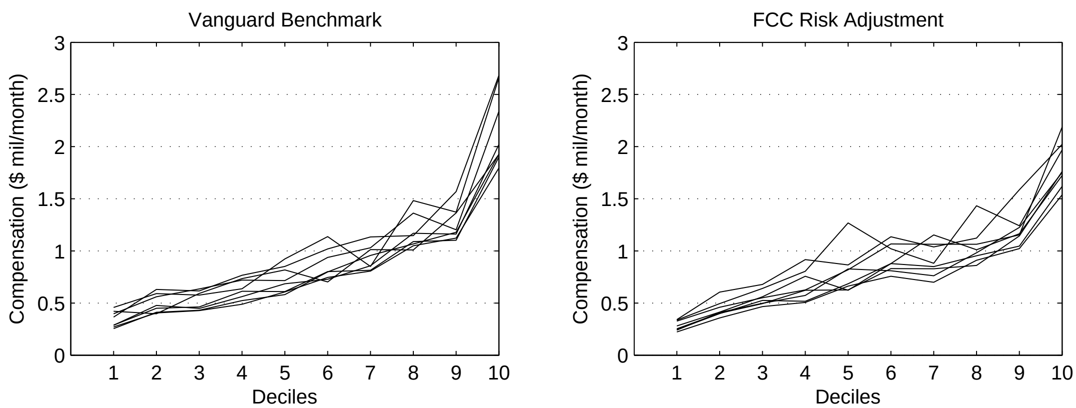


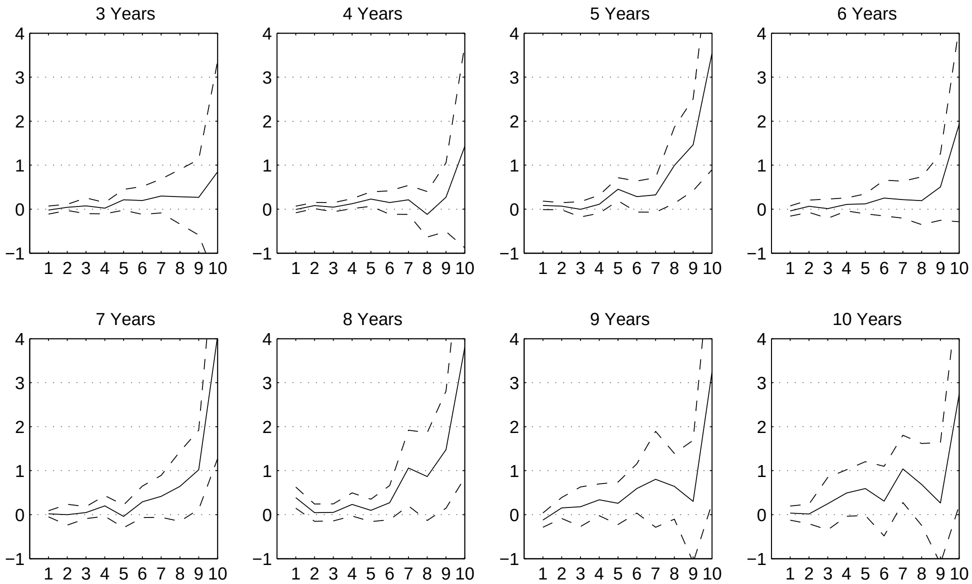
Figure 4: **Out-of-Sample Compensation**
The plots display the average out-of-sample monthly compensation of each decile sorted on the Skill Ratio using the Vanguard Benchmark and the FFC risk adjustment. Each line in the plots represents a different horizon, which varies between three and 10 years. For ease of comparison, the data sample (time period) is the same for both plots.

The evidence is consistent with efficient labor markets: funds managed by better managers earn higher aggregate fees. First, note in Table 6, that the percentage fee is relatively constant across the deciles. However, aggregate fees (compensation) across the deciles is close to monotonically increasing, especially in the extreme deciles where we have the most confidence of our estimates of value added. *Ex-post* there is a tight relationship between measured skill and compensation. And what about *ex-ante*? Once managers reveal their skill by adding value, do investors reward them with higher subsequent compensation? Figure 4 plots out-of-sample compensation and demonstrates that they do. Not only is compensation increasing in the deciles, but the average 10th decile fund earns considerably more than funds in the other deciles. Because the average fee does not differ by much across deciles, by choosing to allocate their capital to skilled managers, it is *investors* that determine these compensation differences, confirming a central insight in Berk and Green (2004). Figure 4 illustrates, again, that using the FFC risk adjustment makes a material difference. Compensation is still increasing in the deciles, but the differences are smaller than when the Vanguard benchmark is used.

If investors reward better funds with higher compensation, then they must be able to identify better managers *ex ante*. Thus, compensation should predict performance. To test this inference, we repeat the previous sorting procedure, except we use total compensation rather than the Skill Ratio to sort funds. That is, at the beginning of each time horizon, we use the product of the current AUM and fee to sort funds into the deciles and then follow the identical procedure we used before. Figure 5 summarizes the results and shows that current compensation does predict future performance. When managers are sorted into deciles by their current compensation, the relative difference in performance across the deciles is slightly larger than when the Skill Ratio is used (i.e., Figure 5 vs. Figure 2).

There is also increased monotonicity when the sorts are based on compensation rather than on the Skill Ratio. To formally document this difference, we count the number of times each decile outperforms the next lowest decile (in terms of value added). Table 8 reports the p-value of observing the reported numbers under the Null Hypothesis that there is no skill (so the probability is 1/2). The table confirms what the figures imply. While the Skill Ratio can identify extreme performers, it does not differentiate other funds very well. In contrast, investors appear to do a much better job correctly differentiating

Panel A: Vanguard Benchmark



Panel B: FFC Risk Adjustment

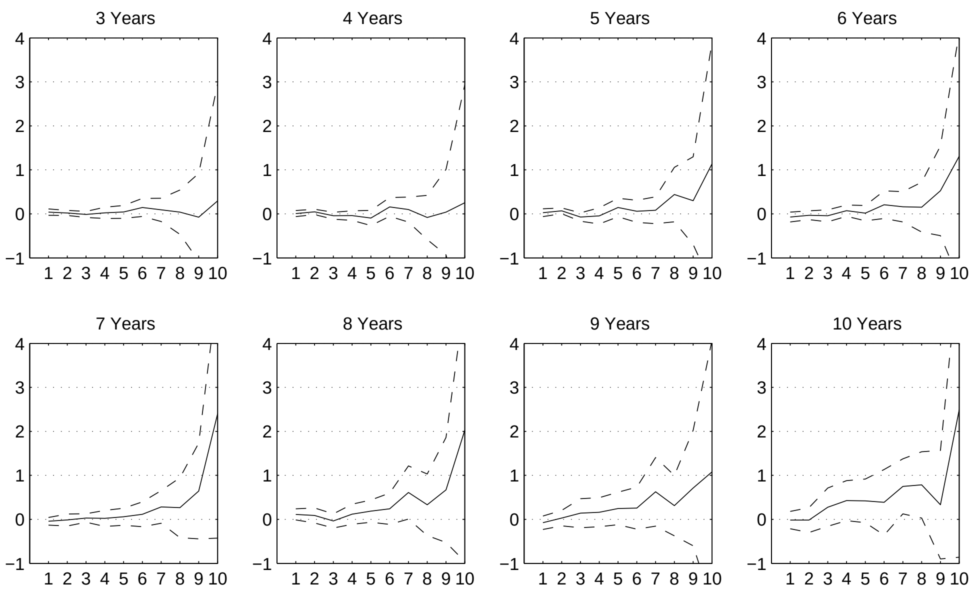


Figure 5: **Value Added Sorted on Compensation**

Each graph displays average out-of-sample value added, $\hat{S}_i$ (in Y2000 \$ million/month), of funds sorted into deciles based on total compensation (fees $\times$ AUM). The solid line indicates the performance of each decile and the dashed lines indicated the 95% confidence bands (two standard errors from the estimate). Panel A shows the results when value added is computed using Vanguard index funds as benchmark portfolios, and Panel B shows the results using the FFC risk adjustment.

	Vanguard Benchmark		FFC Risk Adjustment	
Horizon	Skill Ratio	Compensation	Skill Ratio	Compensation
3	18.55	9.17	2.76	72.15
4	3.51	1.87	12.20	4.89
5	18.61	0.02	8.40	48.81
6	4.57	9.23	0.58	3.24
7	15.92	5.85	0.53	17.52
8	4.16	3.79	4.88	25.58
9	28.61	14.71	5.97	3.90
10	53.02	0.25	16.38	4.82

Table 8: **Out-of-sample Monotonicity:** At each horizon, we calculate the number of times each decile outperforms the next lowest decile. The table shows the p-value (in percent) of the observed frequency under the Null Hypothesis that skill does not exist, i.e., that for a sample length of N months, the probability of the event is Binomial$(9N, 1/2)$.

all funds. We again see a difference when the FFC factor specification is used to adjust for risk. In this case, investors do not appear to differentiate as well, consistent with the prior evidence that compensation is not as highly correlated with subsequent performance.

For many years now, researchers have characterized the behavior of investors in the mutual fund sector as suboptimal, that is, dumb investors chasing past returns. Our evidence relating compensation to future performance suggests quite the opposite. Investors appear able to differentiate good managers from bad and compensate them accordingly. Notice from Figure 1 that real compensation for the top managers has increased over time, that is, fund size has increased while fees have remained constant. On the other hand, for median managers, real compensation has remained constant, suggesting that overall increases in compensation, in at least this sector of the financial services industry, are rewards for skill.

7 Separating Out Diversification Services

The Vanguard benchmarks are constructed from net returns while the funds' value added numbers are constructed using gross returns. Because Vanguard index funds provide diversification services, this means our value-added measure includes both the diversifica-

	Active Funds		Index Funds	
Benchmark	Vanguard Gross	Vanguard Net	Vanguard Gross	Vanguard Net
Number of funds	5974	5974	644	644
In Sample VA				
Mean ($mil/mon)	0.07	0.14	-0.05	0.03
t-statistic	2.49	4.57	-0.49	0.37
Weighted Mean ($mil/mon)	0.16	0.27	-0.02	0.11
t-statistic	3.46	5.74	-0.20	0.94
1st percentile	-3.83	-3.60	-6.18	-5.98
5th percentile	-1.27	-1.15	-1.20	-1.15
10th percentile	-0.64	-0.59	-0.53	-0.47
50th percentile	-0.02	-0.02	-0.01	0.00
90th percentile	0.61	0.75	0.50	0.76
95th percentile	1.55	1.80	1.56	1.82
99th percentile	7.56	7.82	4.83	4.83
In Sample Net Alpha				
Equally Weighted (b.p./mon)	-	2.7	-	-1.9
t-statistic	-	0.73	-	-0.41
Value Weighted (b.p./mon)	-	-1.0	-	1.3
t-statistic	-	-0.31	-	0.65
Persistence (p-value (%) of the top decile outperforming the bottom decile at each horizon)				
3 year horizon	4.31	4.75	-	-
4 year horizon	18.59	2.07	-	-
5 year horizon	0.90	3.54	-	-
6 year horizon	2.12	1.09	-	-
7 year horizon	0.00	0.01	-	-
8 year horizon	0.05	0.67	-	-
9 year horizon	4.88	9.15	-	-
10 year horizon	4.32	5.55	-	-

Table 9: **Performance of Active Funds and Index Funds:** The table computes the value added, net alphas and the p-value of the persistence order statistic that counts the number of times the top decile outperforms the bottom decile for the set of active mutual funds, and compares it the set of index funds (including the Vanguard index funds themselves). To separate the value added coming from diversification benefits vs. stock picking/market timing, we use two different benchmarks: (1) Vanguard index funds gross returns and (2) Vanguard index funds net returns, labeled "Vanguard Gross" and "Vanguard Net" in the table.

tion benefits as well as other skills and services that managers provide. Therefore, if an active manager chooses to do nothing other than exactly replicate a Vanguard benchmark fund, we would compute a positive value added for that fund equal to the diversification benefits it provides (i.e., the fees charged by Vanguard times the size of the fund). So a natural question to ask is what fraction of value added is compensation for providing diversification and what fraction can be attributed to other skills?

We answer this question by recomputing value added using the *gross* returns (including fees) of the Vanguard funds as the benchmark and comparing that to our earlier measures. The first two columns of Table 9 demonstrate that about half the value added is due to diversification benefits ($70,000 per month) and half ($70,000 per month) is due to other types of skill, such as stock picking and/or market timing. The non-diversification skills are also persistent. As the bottom panel in Table 9 demonstrates, when funds are sorted on the Skill Ratio computed using Vanguard gross returns as the benchmark, the top decile consistently outperforms the bottom decile.[17] Similar results are obtained when we use the *ex-post* distribution of skill (i.e., equation 9). Slightly less than half the value added can be attributed to diversification benefits.

Although Vanguard is widely regarded as the most efficient provider of diversification services, one might be concerned that Vanguard is not as efficient as the representative index fund. The evidence in Table 9 should allay such concerns. When value added of the average index fund is computed using Vanguard gross returns as the benchmark (third column of the table), the estimates are negative, implying that Vanguard is more efficient at providing diversification services than the average index fund. Vanguard's efficiency advantage implies that if we had used portfolios of representative index funds instead of just Vanguard's funds, our value added numbers would be larger.

[17]The other order statistics also support persistence and are available upon request.

8 Sample Selection

The existence of numerous selection biases that have influenced the inferences of prior mutual fund studies is well known. Consequently, we have been careful to make sure that this study is free of such biases. We therefore use the full set of actively managed mutual funds available to a U.S. investor at the time. This means we do not exclude funds that invest internationally, so the set of funds in our study is considerably larger than the set previously studied in the literature. One may therefore wonder to what extent the selection bias induced by restricting the dataset to funds that only invest in U.S. stocks would affect our inferences. Introducing this selection bias neither affects our persistence results (skilled managers remain skilled) nor does it affect our results regarding returns to investors: net alphas are not significantly different from zero when managers are evaluated against the tradable benchmark. However, it does affect our average value added results. In Table 10 we form subsamples of active and index funds based on their *ex post* average portfolio weight in international stocks. We find that active funds that invested more in international stocks added more value. Funds that restricted themselves to only investing in U.S. stocks (on average, less than 10% in non-U.S. stocks) added no value on average.

One potential concern is that our Vanguard benchmark funds do not appropriately define the alternative investment opportunity set for international stocks, even though we include all the international index funds that Vanguard offers. If this explanation is right, it implies that index funds that invest more internationally will add more value. Table 10 shows that this is not the case. We find that index funds that invested more in international stocks added less value. Therefore, the *ex post* selection of active funds that have invested more or less in international funds appears to be correlated with skill. It is unclear how investors could have known this fact *ex ante*, which is why we use the full cross-section of mutual funds traded in the U.S. in this study.

Active Funds

Frac. int.	No of funds	Vanguard BM Net Mean VA	Vanguard BM Net Mean VA TW	Vanguard BM Gross Mean VA	Vanguard BM Gross Mean VA TW
<10	3740	0	0.013	-0.045	-0.073
<30	4617	0.055	0.126	-0.002	0.025
<50	4817	0.074	0.159	0.016	0.056
<70	5002	0.106	0.219	0.045	0.113
<90	5236	0.133	0.266	0.073	0.158
$\leq$ 100	5974	0.135	0.269	0.072	0.162

Index Funds

Frac. int.	No of funds	Vanguard BM Net Mean VA	Vanguard BM Net Mean VA TW	Vanguard BM Gross Mean VA	Vanguard BM Gross Mean VA TW
<10	462	0.102	0.175	0.034	0.032
<30	485	0.094	0.167	0.028	0.021
<50	494	0.123	0.178	0.042	0.056
<70	509	0.121	0.177	0.041	0.051
<90	521	0.070	0.163	0.026	-0.003
$\leq$ 100	644	0.034	0.114	-0.025	-0.046

Table 10: **Fraction in International Funds and the Performance of Active Funds vs Index Funds:** The table computes value added numbers for funds with varying degrees of international stock exposure. We compute the numbers for active as well as passive funds. We use two different benchmarks: (1) Vanguard index funds net returns and (2) Vanguard index funds gross returns.

9 Conclusion

In this paper we reject the Null Hypothesis that mutual fund managers have no skill. We show that the average mutual fund generates value of about $2 million/year. This value added cannot easily be attributable to luck alone because it is persistent for as long as 10 years into the future. Furthermore, investors appear to be able to identify and correctly reward this skill. Not only do better funds collect higher aggregate fees, but current aggregate fees are a better predictor of future value added than past value added.

Our results are consistent with the main predictions of Berk and Green (2004). Investors appear to be able to identify skilled managers and determine their compensation through the flow–performance relation. That model also assumes that because rational investors compete in capital markets, the net alpha to investors is zero, that is, managers are able to capture all economic rents themselves. In this paper, we find that the average abnormal return to investors is close to zero. Further, we find little evidence that investors can generate a positive net alpha by investing with the best funds.

Appendix

A Benchmarks Funds with Unequal Lives

In this appendix, we explain how we construct our set of benchmarks. We show how to evaluate a fund relative to two benchmarks that exist over different periods of time. The general case with N benchmark funds is a straightforward generalization and is left to the reader.

Let R_{it}^g denote the gross excess return of active fund i at time t, which is stacked in the vector R_i^g :

$$R_i^g = \begin{bmatrix} R_{i1}^g \\ \vdots \\ R_{iT}^g \end{bmatrix}$$

and let R_{1t}^B denote the return on the first benchmark fund and R_{2t}^B the return on the second benchmark fund, which, over the time period in which they both exist, form the matrix R_t^B:

$$R_t^B = \begin{bmatrix} R_{1t}^B & R_{2t}^B \end{bmatrix}.$$

Assume that the first benchmark fund is available to investors over the whole sample period, while the second benchmark fund is only available over a subset of the sample, say the second half.

Let β denote the projection coefficient of R_{it}^g on the first benchmark fund's return, R_{1t}^B, and let

$$\gamma \equiv \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}.$$

denote the projection coefficients of R_{it}^g on both benchmark funds, R_{1t}^B and R_{2t}^B. Thus, during the time period when only the first benchmark exists, the value added of the fund at time t is:

$$V_{it} = q_{i,t-1}\left(R_{it}^g - \beta R_{1t}^B\right). \tag{18}$$

When both benchmark funds are offered, the value-added in period t is:

$$V_{it} = q_{i,t-1}\left(R_{it}^g - R_t^B\gamma\right). \tag{19}$$

Let there be T time periods and suppose that the second benchmark fund starts in period

$S+1$. The matrix of benchmark return observations is given by:

$$X = \begin{bmatrix} 1 & R_{11}^B & \cdot \\ \vdots & \vdots & \cdot \\ 1 & R_{1S}^B & \cdot \\ 1 & R_{1,S+1}^B & R_{2,S+1}^B \\ \vdots & \vdots & \vdots \\ 1 & R_{1T}^B & R_{2T}^B \end{bmatrix}$$

where $\cdot$ indicates a missing value. Let X^O denote the following orthogonal matrix:

$$X^O = \begin{bmatrix} 1 & R_{11}^B & \bar{R}_2^{BO} \\ \vdots & \vdots & \vdots \\ 1 & R_{1S}^B & \bar{R}_2^{BO} \\ 1 & R_{1,S+1}^B & R_{2,S+1}^{BO} \\ \vdots & \vdots & \vdots \\ 1 & R_{1T}^B & R_{2,T}^{BO} \end{bmatrix}$$

where:

$$\bar{R}_2^{BO} = \frac{\sum\limits_{t=S+1}^{T} R_{2t}^{BO}}{T-S}.$$

and where $R_{2,S+1}^{BO}, ..., R_{2,T}^{BO}$ are obtained by projecting R_{2t}^B onto R_{1t}^B:

$$R_{2t}^{BO} = R_{2t}^B - \theta R_{1t}^B \text{ for } t = S+1, ..., T$$

where,

$$\theta = \frac{cov\left(R_{2t}^B, R_{1t}^B\right)}{var\left(R_{1t}^B\right)}.$$

Finally, define:

$$\hat{X}^O = \begin{bmatrix} 1 & R_{11}^B & 0 \\ \vdots & \vdots & \vdots \\ 1 & R_{1S}^B & 0 \\ 1 & R_{1,S+1}^B & R_{2,S+1}^{BO} \\ \vdots & \vdots & \vdots \\ 1 & R_{1T}^B & R_{2,T}^{BO} \end{bmatrix}.$$

Proposition 1 *The value-added of the firm at any time t can be estimated as follows:*

$$V_{it} = q_{i,t-1}\left(R_{it}^g - \zeta_2 \hat{X}_{2t}^O - \zeta_3 \hat{X}_{3t}^O\right) \tag{20}$$

using a single OLS regression to estimate ζ:

$$\zeta = \left(X^{O\prime} X^{O}\right)^{-1} X^{O\prime} R_i^g.$$

Proof: The second and the third column of X^O are orthogonal to each other, both over the full sample as well as over the two subsamples. Because of this orthogonality and $X_{2t}^O = R_{1t}^B$, the regression coefficient ζ_2 is given by:

$$\zeta_2 = \frac{cov\left(R_{it}^g, R_{1t}^B\right)}{var\left(R_{1t}^B\right)} = \beta.$$

So for any $t \leq S$, (20) reduces to (18) and so this estimate of value added is consistent over the first subsample. Using the orthogonality of X^O,

$$\zeta_3 = \frac{cov\left(R_{it}^g, X_{3t}^O\right)}{var\left(X_{3t}^O\right)} = \frac{cov\left(R_{it}^g, R_{2t}^{BO}\right)}{var\left(R_{2t}^{BO}\right)},$$

rewriting

$$\gamma_1 R_{1t}^B + \gamma_2 R_{2t}^B = \gamma_1 R_{1t}^B + \gamma_2\left(\theta R_{1t}^B + R_{2t}^{BO}\right) = \left(\gamma_1 + \theta\gamma_2\right) R_{1t}^B + \gamma_2 R_{2t}^{BO}$$

and using the fact that linear projections are unique implies

$$\zeta_2 = \beta = \gamma_1 + \theta\gamma_2$$

and

$$\zeta_3 = \gamma_2.$$

So for $t > S$,

$$\begin{aligned} V_{it} &= q_{i,t-1}\left(R_{it}^g - \zeta_2 \hat{X}_{2t}^O - \zeta_3 \hat{X}_{3t}^O\right) \\ &= q_{i,t-1}\left(R_{it}^g - \left(\gamma_1 + \theta\gamma_2\right) R_{1t}^B - \gamma_2 R_{2t}^{BO}\right) \\ &= q_{i,t-1}\left(R_{it}^g - \gamma_1 R_{1t}^B - \gamma_2 R_{2t}^B\right) \end{aligned}$$

which is (19) and so the estimate is also consistent over the second subsample.

B Robustness

Table 11 reports the results of conducting our study within different subsamples of our data. We select the samples based on the time period and whether managers invest internationally. Even when we consider active funds that invest in U.S. equity only, we always use all 11 Vanguard index funds at the benchmark.

	All Equity			**U.S. Equity Only**		
	Beg-3/11	1/84-9/06	1/90-3/11	Beg-3/11	1/84-9/06	1/90-3/11
			Panel A: Vanguard Benchmark			
In Sample VA						
Mean ($mil/mon)	0.14**	0.28**	0.14**	-0.01	0.01	-0.02
Weighted Mean ($mil/mon)	0.27**	0.42**	0.29**	-0.00	-0.05	0.00
In Sample Net Alpha						
Equally Weighted (b.p./mon)	3	-7*	-5*	-1	-11**	-6**
Value Weighted (b.p./mon)	-1	-7*	-5*	-5	-12**	-8**
Total Number of Funds	5974	4599	5943	2731	2218	2700
			Panel B: FCC Risk Adjustment			
In Sample VA						
Mean ($mil/mon)	0.10**	0.15**	0.16**	-0.07	-0.03	-0.05*
Weighted Mean ($mil/mon)	0.25**	0.37**	0.29**	-0.05	-0.02	0.00
In Sample Net Alpha						
Equally Weighted (b.p./mon)	-4	-9*	-8*	-7**	-9**	-7*
Value Weighted (b.p./mon)	-6**	-7*	-5	-8**	-10**	-8*
Total Number of Funds	6054	4599	5943	2811	2218	2700

Table 11: **Subsample Analysis:** *Beg* is the beginning date of our sample, 1/77 for Vanguard Benchmark and 1/62 for FFC Risk Adjustment. * – t-statistic greater (in absolute value) than 1.96. **– t-statistic greater (in absolute value) than 2.54.

References

Baker, M., L. Litov, J. A. Wachter, and J. Wurgler (2010): "Can Mutual Fund Managers Pick Stocks? Evidence from Their Trades Prior to Earnings Announcements," *Journal of Financial and Quantitative Analysis*, 45(05), 1111–1131.

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 I. Tonks (2007): "Return Persistence and Fund Flows in the Worst Performing Mutual Funds," Working Paper 13042, National Bureau of Economic Research.

Bogle, J. C. (1997): "The First Index Mutual Fund: A History of Vanguard Index Trust and the Vanguard Index Strategy," Discussion paper, The Vanguard Group.

Bollen, N. P. B., and J. A. Busse (2001): "On the Timing Ability of Mutual Fund Managers," *The Journal of Finance*, 56(3), 1075–1094.

——— (2005): "Short-Term Persistence in Mutual Fund Performance," *Review of Financial Studies*, 18(2), 569–597.

Carhart, M. M. (1997): "On Persistence in Mutual Fund Performance," *Journal of Finance*, 52, 57–82.

Chen, H.-L., N. Jegadeesh, and R. Wermers (2000): "The Value of Active Mutual Fund Management: An Examination of the Stockholdings and Trades of Fund Managers," *Journal of Financial and Quantitative Analysis*, 35, 343–368.

Cohen, L., A. Frazzini, and C. Malloy (2007): "The Small World of Investing: Board Connections and Mutual Fund Returns," Working Paper 13121, National Bureau of Economic Research.

Cohen, R. B., C. K. Polk, and B. Silli (2010): "Best Ideas," *SSRN eLibrary*.

Coval, J. D., and T. J. Moskowitz (2001): "The Geography of Investment: Informed Trading and Asset Prices," *The Journal of Political Economy*, 109(4), pp. 811–841.

Cremers, K. J. M., and A. Petajisto (2009): "How Active Is Your Fund Manager? A New Measure That Predicts Performance," *Review of Financial Studies*, 22(9), 3329–3365.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers (1997): "Measuring Mutual Fund Performance with Characteristic-Based Benchmarks," *Journal of Finance*, 52, 1035–1058.

Del Guercio, D., and J. Reuter (2013): "Mutual Fund Performance and the Incentive to Generate Alpha," *Journal of Finance*, Forthcoming.

Fama, E. F. (1965): “The Behavior of Stock-Market Prices,” *Journal of Business*, 38(1), 34–105.

——— (1970): “Efficient Capital Markets: A Review of Theory and Empirical Work,” *Journal of Finance*, 25(2), 383–417.

Fama, E. F., and K. R. French (1995): “Size and Book-to-Market Factors in Earnings and Returns,” *Journal of Finance*, 50, 131–155.

——— (2010): “Luck versus Skill in the Cross-Section of Mutual Fund Returns,” *The Journal of Finance*, 65(5), 1915–1947.

Grinblatt, M., and S. Titman (1989): “Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings,” *The Journal of Business*, 62(3), pp. 393–416.

——— (1993): “Performance Measurement without Benchmarks: An Examination of Mutual Fund Returns,” *The Journal of Business*, 66(1), pp. 47–68.

Gruber, M. J. (1996): “Another Puzzle: The Growth in Actively Managed Mutual Funds,” *Journal of Finance*, 51, 783–810.

Jensen, M. C. (1968): “The Performance of Mutual Funds in the Period 1945–1964,” *Journal of Finance*, 23, 389–416.

Jiang, H., M. Verbeek, and Y. Wang (2011): “Information Content When Mutual Funds Deviate from Benchmarks,” *SSRN eLibrary*.

Kacperczyk, M., S. V. Nieuwerburgh, and L. Veldkamp (2011): “Time-Varying Fund Manager Skill,” Working Paper 17615, National Bureau of Economic Research.

Kacperczyk, M., C. Sialm, and L. Zheng (2005): “On the Industry Concentration of Actively Managed Equity Mutual Funds,” *The Journal of Finance*, 60(4), pp. 1983–2011.

——— (2008): “Unobserved Actions of Mutual Funds,” *Review of Financial Studies*, 21(6), 2379–2416.

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. (2012): “Reverse Survivorship Bias,” *Journal of Finance*, forthcoming.

Lucas, Robert E., J. (1978): “On the Size Distribution of Business Firms,” *The Bell Journal of Economics*, 9(2), pp. 508–523.

Malkiel, B. G. (1995): “Returns from Investing in Equity Mutual Funds 1971 to 1991,” *The Journal of Finance*, 50(2), pp. 549–572.

Mamaysky, H., M. Spiegel, and H. Zhang (2008): "Estimating the Dynamics of Mutual Fund Alphas and Betas," *Review of Financial Studies*, 21(1), 233–264.

Pastor, L., and R. F. Stambaugh (2002): "Mutual fund performance and seemingly unrelated assets," *Journal of Financial Economics*, 63(3), 315 – 349.

——— (2010): "On the Size of the Active Management Industry," Working Paper 15646, National Bureau of Economic Research.

Sharpe, W. F. (1991): "The Arithmetic of Active Management," *Financial Analysts Journal*, 47(1), pp. 7–9.

Shumway, T., M. B. Szefler, and K. Yuan (2009): "The Information Content of Revealed Beliefs in Portfolio Holdings," .

Wermers, R. (2000): "Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses," *The Journal of Finance*, 55(4), 1655–1703.

Zheng, J. X. (1999): "Testing Heteroskedasticity in Nonlinear and Nonparametric Regressions with an Application to Interest Rate Volatility," Working paper, University of Texas at Austin.