



Measuring skill in the mutual fund industry[☆]



Jonathan B. Berk^{a,d,*}, Jules H. van Binsbergen^{b,c,d,**}

^a Graduate School of Business, Knight Management Center, Stanford University, Stanford, CA 94305, United States

^b The Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104, United States

^c Tilburg University, Netherlands

^d National Bureau of Economic Research, United States

ARTICLE INFO

Article history:

Received 15 April 2014

Received in revised form

22 October 2014

Accepted 19 November 2014

Available online 8 May 2015

JEL classification:

G23

G20

G11

J31

J24

Keywords:

Mutual funds

Managerial skill

Alpha

ABSTRACT

Using the value that a mutual fund extracts from capital markets as the measure of skill, we find that the average mutual fund has used this skill to generate about \$3.2 million per year. Large cross-sectional differences in skill persist for as long as ten years. Investors recognize this skill and reward it by investing more capital with better funds. Better funds earn higher aggregate fees, and a strong positive correlation exists between current compensation and future performance. The cross-sectional distribution of managerial skill is predominantly reflected in the cross-sectional distribution of fund size, not gross alpha.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

An important principle of economics is that agents earn economic rents if and only if they have a competitive advantage. As central as this principle is to microeconomics,

surprisingly little empirical work has addressed the question of whether or not competitive advantages are rewarded, or, perhaps more interestingly, whether people without a competitive advantage can earn rents. One notable exception is the research on mutual fund managers. An extensive literature

[☆] 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, Jiawei Li, Binying Liu, Jin Ngai, Michael Nolop, William Vijverberg, and Christina Zhu. We thank the referee, Peter DeMarzo, Darrell Duffie, George Chacko, Rick Green, Dirk Jenter, Ralph Koijen, David Musto, Lubos Pastor, Francisco Perez Gonzalez, Robert Novy-Marx, Paul Pfleiderer, Anamaria Pieschacon, Robert Stambaugh, Bill Schwert, and seminar participants at Robeco, University of Rochester, Stockholm School of Economics, Stanford University, University of Chicago, University of Toronto, Vanderbilt University, University of Pennsylvania, University of California San Diego, University of Wisconsin, Dartmouth College, Cheung Kong Graduate School of Business, Universitat Pompeu Fabra, Copenhagen Business School, Aalto School of Business, Columbia University, University of Texas Dallas, Michigan State University, London School of Economics, London Business School, Rotterdam School of Management, Tilburg University, the National Bureau of Economic Research summer institute, the Western Finance Association meetings, Jackson Hole Finance Conference, University of British Columbia Summer Conference and the Stanford Berkeley joint seminar for helpful comments and suggestions.

* Corresponding author at: Graduate School of Business, Knight Management Center, Stanford University, Stanford, CA 94305, United States. Tel.: +1 650 721 1280.

** Corresponding author at: The Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104, United States. Tel.: +1 215 573 1606.

E-mail addresses: jberk@stanford.edu (J.B. Berk), julesv@wharton.upenn.edu (J.H. van Binsbergen).

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 competitive advantage.

Given the importance of the question, the objective of this paper is to reexamine whether or not mutual fund managers are skilled. We find that the average mutual fund has added value by extracting about \$3.2 million a year from financial markets. Most important, cross sectional differences in value added are persistent for as long as ten years. We find it hard to reconcile our findings with anything other than the existence of money management skill. The cross sectional distribution of managerial talent documented in this paper 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, that is, current *compensation* predicts future performance.

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.¹ However, as Berk and Green (2004) argue, if skill is in short supply, the net alpha is determined in equilibrium by competition between investors, and not by the skill of managers. Some people have hypothesized, based on this insight, that the gross alpha (the average abnormal return before fees are subtracted) is the correct measure of managerial skill. However, the gross alpha is a *return* measure, not a *value* measure and so in this paper we argue that the gross alpha alone does not measure managerial skill either. 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. In fact, we prove that under the neoclassical assumptions that managers optimize, markets are competitive and investors rational, the only condition under which the gross alpha differentiates managers is if all funds are exactly the same size. This surprising fact follows directly from the observation that investor competition drives net alpha to zero. That immediately implies that the gross alpha is equal to the fee the fund charges its investors, that is, the fund manager chooses his gross alpha. As Berk and Green (2004, p. 1277) show, if managers are allowed to index part of their portfolio, then this choice is arbitrary. Therefore, no a priori reason exists for the gross alpha to even be correlated with managerial skill in the cross section.

We argue that the skill of a mutual fund manager equals the value his fund extracts from markets, the fund's gross excess return over its benchmark multiplied by assets under management (AUM), what we term the *value added* of the fund. Peter Lynch's career managing money provides an apt illustration of the differences between using alpha measures and value added to measure skill. In his first five years managing Fidelity's Magellan fund, Peter Lynch had a 2% monthly gross alpha on average assets of about \$40 million. In

his last five years, his gross alpha was 20 basis points (bp) per month on assets that ultimately grew to more than \$10 billion. Based on the lack of persistence in gross alpha, one could falsely conclude that most of Lynch's early performance was due to luck not skill. In fact, the value he added went from less than \$1 million/month to more than \$20 million/month.

Our measure of value added 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 Lynch's success resulted, at least in part, 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. However, marketing efforts alone are not sufficient to generate a positive value added. If Lynch had had no skill, he would have extracted nothing from financial markets and our value added measure would be zero because the costs of all other input factors, including marketing, would have to have been borne by investors. In fact, Fidelity's marketing skills could very well have complemented 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.

Our approach also differs from prior work in two other respects. First, our data set includes all actively managed US mutual funds, thereby greatly increasing the power of our tests. Prior work has focused attention exclusively on funds that hold only US stocks. Second, in addition to evaluating managers using a risk model, we also evaluate managers by comparing their performance with an alternative investment opportunity set – all available Vanguard index funds (including funds that hold non-US stocks). By using Vanguard funds as the benchmark we are guaranteed that this alternative investment opportunity set was tradable and marketed at the time. We fully acknowledge that there is a certain degree of arbitrariness in picking such a benchmark. We chose Vanguard because this company pioneered index fund investing and has a market-leading position in the index fund space.

The strongest evidence we provide for the existence of investment skill is the persistence of cross-sectional differences in value added. We provide evidence of this persistence as far out as ten years, which is substantially longer than what the existing literature has found using alpha measures. Perhaps the most surprising result in this paper is that investors appear to be able to detect this skill and use this information to invest their capital. Managerial compensation, computed as the product of AUM and the percentage fee, is primarily a function of fund size, so investors, by allocating their capital to funds, determine managerial compensation.² We find that a very strong positive cross-sectional correlation between managerial skill and managerial compensation exists. This observation

¹ See, for instance, Carhart (1997).

² For ease of exposition, we refer to the product of AUM and the percentage fee as managerial compensation. Because the product of AUM and the percentage fee also includes overhead expenses and firm profits, it is not the amount of money the investment manager takes home with him. Unfortunately, we do not have detailed data on the salaries paid to managers.

implies that investors are able to infer managerial quality. We confirm this inference by demonstrating that current compensation better predicts future value added than past value added does.

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 measure value added without diversification services. By undertaking this analysis, we find that about half of the total value the mean fund has added is attributable to diversification services and the other half to market timing and stock picking. Furthermore, we show that removing diversification services does not change our persistence results. We find that market timing and stock picking skills alone are also persistent.

The 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 return an investor can earn by investing in a mutual fund. Nevertheless, we do provide new insight into that question as well. Once we evaluate funds against a tradable benchmark, we no longer find evidence of the under-performance previously shown in the literature. Over the time period in our sample, the equally weighted net alpha is 3 bp per month and the value weighted net alpha is -1 bp per month. Neither estimate is significantly different from zero.

2. 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 under-perform the market, and, more important, mutual fund returns are 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 under-perform 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 the most influential paper on the subject of the existence of managerial talent, Carhart (1997) uses the net alpha earned by investors to measure managerial skill and concludes that there is no evidence of skilled or informed mutual fund managers. Two recent papers, Kosowski, Timmermann, Wermers, and White (2006) and Fama and French (2010), use alpha measures to obtain a cross-sectional distribution of managerial talent. The two studies reach differing conclusions on the extent to which skill exists and varies in the cross section. Researchers have also studied persistence in mutual fund performance. Using the return the

fund makes for its investors, a number of papers (Gruber, 1996; Carhart, 1997; Zheng, 1999; Bollen and Busse, 2001) find that performance is largely unpredictable.³

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 shows 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 these 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. Kacperczyk, Sialm, and Zheng (2008) compare the actual performance of funds with 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, Guercio and Reuter (2014) find that directly sold funds, that is, funds not marketed by brokers, do not under-perform index funds after fees, thus implying outperformance before fees.

Evidence also suggests 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 concentrate 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, Szeffler, 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 (2014) 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. The apparent reticence to accept the

³ 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).

idea that managerial skill exists is at least partly attributable to the lack of any convincing evidence of the existence of the value that should result from managers employing their talent. Our objective is to provide this evidence.

3. Definitions and hypotheses

Let R_{it}^n denote the 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 excess return of the investor's next best alternative investment opportunity, R_{it}^B , which we call the manager's benchmark, and a deviation from the benchmark ε_{it} :

$$R_{it}^n = R_{it}^B + \varepsilon_{it}. \quad (1)$$

The most commonly used measure of skill in the literature is the unconditional 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

$$\hat{\alpha}_i^n = \frac{1}{T_i} \sum_{t=1}^{T_i} (R_{it}^n - R_{it}^B) = \frac{1}{T_i} \sum_{t=1}^{T_i} \varepsilon_{it}, \quad (2)$$

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

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 is 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 suboptimality on the part of at least some investors.⁴

Some have argued that the gross alpha, α_i^g , the unconditional mean of the benchmark adjusted 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 percentage fee $f_{i,t-1}$ (charged from $t-1$ to t):

$$R_{it}^g = R_{it}^n + f_{i,t-1} = R_{it}^B + \varepsilon_{it} + f_{i,t-1}. \quad (3)$$

The gross alpha can then be consistently estimated as

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

Unfortunately, the gross alpha is a return measure, not a value measure. 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. In Section 4 we formally derive the surprising result that not only does the gross alpha not measure managerial skill, but it also need not even be positively correlated with skill.

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 :

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

where the second equality follows from Eq. (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 Eq. (5)

$$S_i \equiv E[V_{it}]. \quad (6)$$

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

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

The average value added across funds 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

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

where N is the number of mutual funds in our database. Alternatively, we could 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:

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

Before we turn to how we 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.

3.1. Unskilled managers

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. So, $q_{i,t-1} f_{i,t-1} = -E[q_{i,t-1} \varepsilon_{it}]$ implying

$$S_i = 0 \quad \text{for every } i. \quad (10)$$

⁴ For a formal model that relates this under-performance to decreasing returns to scale at the industry level, see Pastor and Stambaugh (2012).

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

$$\bar{S} = \frac{1}{N} \sum_{i=1}^N \hat{S}_i = 0. \quad (11)$$

Eq. (11) 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 can 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 take this two part hypothesis as the null hypothesis in this paper.

If no manager has skill, then by definition, there can be no cross-sectional variation in skill and managers cannot predictably outperform each other. 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 the null hypothesis.

3.2. Skilled managers

The alternative hypothesis we consider is that managerial skill exists and that some managers use this skill to extract money from markets. If managerial skill is difficult to measure, one would expect unskilled managers to take advantage of this uncertainty. In this case we would expect to observe the presence of charlatans, i.e., managers who charge a fee but have little or no skill. However, so long as charlatans do not choose to destroy value (they always have the option to index their money), we have

$$S_i \geq 0 \quad \text{for every } i. \quad (12)$$

To derive Eq. (12) we did not have to make an assumption on either the rationality of investors or the competitiveness of capital markets. Even if investors are irrational and are willing to accept a negative net alpha, value added itself cannot be negative. That is, as long as all managers optimize given their skill level, no manager, even an unskilled manager, will destroy value. That means that, on average, the total dollar fees collected cannot be less than the amount of money taken from investors due to under-performance, so the expected value added is never less than zero. Because, under this hypothesis, $S_i > 0$ for some managers, Eq. (12) implies that the average manager must generate value, and hence we would expect to find

$$\bar{S} > 0. \quad (13)$$

Some have claimed, based on Sharpe (1991), that in a fully rational 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 is used to imply that the abnormal return of the average active investor must be zero, what has become

known as Sharpe's Arithmetic. 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. 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 surprising 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 investors, 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.

4. Why alpha measures do not measure skill

We now formally show that under the standard neoclassical assumptions that underly modern asset pricing models, value added measures skill perfectly, while neither the gross alpha nor the net alpha are valid measures of skill. We start by defining what we mean by "measure." A variable is said to measure skill if it quantitatively measures the amount of money a manager extracts from markets. Of course, because skill is measured in dollars, under this definition a return measure will almost never measure skill. So to evaluate the usefulness of return measures, we define a *proxy* measure to be a variable that is proportional to skill. A proxy for skill is any positive linear transformation of a skill measure.

The standard neoclassical assumptions are that: (1) investors are rational, (2) financial markets are competitive, and (3) managers optimize. In addition, we assume that positive net present value investment opportunities are not in infinite supply, so that managers face decreasing returns to scale. That is, we assume that the alpha (before fees and expenses are deducted) that manager i generates by actively managing money is given by

$$\alpha_i^* = a_i - b_i q, \quad (14)$$

where $a_i > 0$ can be interpreted as the alpha on the first cent the manager actively invests, $b_i > 0$ is a parameter that captures the decreasing returns to scale the manager faces, both of which can vary by fund, and q is the amount of money the manager puts into active management. Under these assumptions Proposition 1 holds.

Proposition 1.

1. The net alpha never proxies for, or measures, managerial skill.

2. The only condition under which gross alpha measures managerial skill is if all managers set their fees to ensure that the AUM of all funds is exactly \$1. Gross alpha will proxy for managerial skill only under the condition that managers set fees to ensure that all funds have the same AUM.
3. Value added, the product of AUM and gross alpha, always measures skill.

Proof. The fact that the net alpha is not a valid proxy or measure of skill follows immediately from assumptions (1) and (2). As Berk and Green (2004) argue, if investors are rational and financial markets competitive (that is, investors compete with each other for positive present value opportunities), non-zero net alpha investment opportunities must be competed away. Thus the net alpha is zero for all managers, so it cannot measure skill, proving the first part of the proposition.

Under assumption (3), managerial skill, V_i , is the solution to the following optimization problem:

$$V_i \equiv \max_q q(a_i - b_i q),$$

where q is the amount of money the manager chooses to actively manage. The first order condition gives the optimal amount the manager chooses to actively manage, q_i^* :

$$q_i^* = \frac{a_i}{2b_i}, \quad (15)$$

so the skill of the manager is

$$V_i = q_i^*(a_i - b_i q_i^*) = \frac{a_i^2}{4b_i}, \quad (16)$$

which is a function of both parameters a_i and b_i . At the optimum, the alpha the manager makes on the actively managed part of his portfolio can be calculated by substituting Eq. (15) into Eq. (14), and is only a function of q_i :

$$\alpha_i^* = a_i - b_i \left(\frac{a_i}{2b_i} \right) = \frac{a_i}{2}. \quad (17)$$

The gross alpha of the fund as a whole is therefore

$$\alpha_i^g = \left(\frac{q_i^*}{q_i} \right) \alpha_i^*, \quad (18)$$

where q_i is the fund's total AUM. The ratio in parentheses reflects the fact that under assumption (3), a manager never chooses to actively manage anything other than q_i^* and thus indexes the amount $q_i - q_i^*$, which earns a zero alpha. Now, using Eqs. (15) and (17) to simplify Eq. (18) gives

$$\alpha_i^g = \frac{a_i^2}{4q_i b_i}, \quad (19)$$

which only equals V_i when $q_i = 1 \forall i$. So the only condition under which gross alpha measures skill is when all managers manage \$1. To proxy for skill, we need $V_i \propto \alpha_i^g$, which only occurs if $q_i = \bar{q} \forall i$, proving the second part of the proposition.

To prove the last part of the proposition, we use Eq. (19) to compute the product of gross alpha and AUM, what we call the value added of the manager:

$$q_i \alpha_i^g = q_i \frac{a_i^2}{4q_i b_i} = \frac{a_i^2}{4b_i}, \quad (20)$$

which is Eq. (16). \square

What the proposition highlights is the importance of scale in measuring managerial skill. The only time the gross alpha is informative about managerial skill is when all managers manage funds of the same size, making scale unimportant. A key insight is that because the net alpha is zero,

$$\alpha_i^g = \alpha_i^n + f_i = f_i, \quad (21)$$

implying the general condition that the gross alpha equals the fee the manager sets, and is therefore a choice variable unrelated to skill. The fee determines how much money investors choose to invest with the manager. Skill is a function of how much money the manager decides to *actively* manage. The two are unrelated because it is optimal for managers to index any excess funds.

For a relation to exist between the fee charged (i.e., the gross alpha) and managerial skill, managers must *choose* fees to cause this relation to happen. Of course, managers do not pick fees arbitrarily and so the question could arise as to whether, empirically, gross alpha is a good proxy for skill. Proposition 1 is informative on this question because it implies that gross alpha is only a good proxy for skill when managers choose their fees so that the cross sectional variation in AUM is low. However, in reality managers do not set fees this way, because AUM shows large cross-sectional variation in the data. That fact explains why prior research that has used gross alpha has failed to find cross sectional differences in skill.

Another implication of Eq. (21) is that, under the standard neoclassical assumptions underlying Proposition 1, managerial compensation (fee times AUM) also measures skill perfectly. Given the precision with which one can observe managerial compensation, an astute reader could question why we do not use this measure instead of value added to measure skill. In this paper we do not take a stand on whether investors are rational or whether markets are competitive. It is only under both of these conditions that the equivalence between the measures holds. When either one of these conditions is violated (as is the case under our null hypothesis), managerial compensation does not necessarily measure skill, but value added still does. In particular, managerial compensation is positive even under our null hypothesis. Thus, if we used managerial compensation to measure skill, we would not be able to differentiate between our null and our alternative hypotheses.

5. Choice of benchmarks and estimation

To measure the value that the fund either gives to or takes from investors, performance must be compared with 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 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. The problem with this interpretation is that these factor portfolios were (and in some cases are) not actually available or known 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.⁵ 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. We depart from the literature by ensuring that this alternative investment opportunity was marketed and tradable 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.⁶ 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

$$R_{it}^B = \sum_{j=1}^{n(t)} \beta_i^j R_t^j, \quad (22)$$

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, they reflect the dynamic evolution of active strategies so we can be certain that investors had the opportunity to invest in the funds at the time. In addition, the returns of these funds necessarily include transaction costs. Notice that if we use this benchmark to evaluate a Vanguard index fund itself, we would conclude that fund adds value equal to the dollar value of the fees it charges. Vanguard funds add value because they provide investors with low cost means to diversification. Consequently, 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 also credits them with this value added.

One could also be interested in whether active funds add value over and above the diversification services they provide. In Section 8, 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 including diversification services, value added is highly persistent and positive.

Our main motivation to use Vanguard index funds is to avoid hindsight bias in the selection of performance benchmarks. We do, however, acknowledge that choosing Vanguard as the benchmark necessarily involves a degree of arbitrariness. That is, other index funds and index fund providers could have been selected as well. To allay concerns that other providers offer diversification services at lower cost, we compare, in Section 8, the performance of all 644 index funds in our database with the Vanguard benchmark and find that, on average, other index funds have under-performed Vanguard. Despite this reassuring finding, one could argue that other low-cost international, small-cap, and value index funds could have existed before Vanguard offered such index funds. To address this concern, we searched our database and found the earliest self-proclaimed index fund of each type. In every case, Vanguard was first, that is, on the date Vanguard introduced each index fund in our benchmark, no other index fund of that type existed in our database. The advantage of using Vanguard funds is that because the company is viewed as both the market leader and a pioneer, we can be reasonably sure that when it introduces a new fund investors quickly become aware of the new fund's existence. As such, we believe it provides a reasonable description of what investors' actual alternative investment opportunity set was at the time.

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

$$R_{it}^B = \beta_i^{mkt} MKT_t + \beta_i^{sml} SML_t + \beta_i^{hml} HML_t + \beta_i^{umd} UMD_t, \quad (23)$$

where MKT_t , SML_t , HML_t , and UMD_t are the realizations of the four factor portfolios (excess return on the market, small minus big, high minus low, and up minus down) and β_i 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 a widely accepted theoretical argument has not been put forward to justify using these factors to 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. Such an interpretation is valid only when the factors are tradable portfolios.

We pick 11 Vanguard index funds to use as benchmark funds (see Table 1). We arrive at this set by excluding all bond

⁵ AQR introduced a momentum "index" fund in 2009, but the fund charges 75 bp which is close to the mean fee in our sample of active funds. It also requires a \$1 million minimum investment.

⁶ The ownership structure of Vanguard (it is owned by the investors in its funds) also makes it attractive as a benchmark because no conflict of interest exists between the investors in the fund and the fund owners. Bogle (2014) provides a brief history of Vanguard's index fund business.

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).

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 ^a
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

^a 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.

or real estate index funds and any fund that was already spanned by existing funds.⁷ Because the 11 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, 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 in (22)) using the augmented basis and then calculating the benchmark return using Eq. (22) and the basis in which missing returns are replaced with zeros.

To quantify the advantages of using Vanguard funds instead of 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 much better investment opportunities than what was available to investors at the time. The FFC specification includes both the size and value factor over the entire sample, whereas these strategies became available as index funds only in the nineties when Vanguard pioneered their introduction. They, therefore, contain a hindsight bias. That is, in the early part of our sample, they contain information about the performance of size and value that became generally known to investors only in the early nineties. 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

Table 2

Net alpha of Fama-French-Carhart (FFC) portfolios. We regress each FFC factor portfolio on the Vanguard benchmark portfolios. The table lists the estimate (in bp/month) and t -statistic of the constant term (alpha) of each regression, as well as the R^2 of each regression.

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

portfolios, which accounts for the fact that the UMD portfolio has the lowest R^2 and the highest alpha.

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

6. Data set

Our main data source is the Center for Research in Security Prices (CRSP) survivorship bias-free database of mutual funds 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 need to address to complete our study. We undertook an extensive data project to address these shortcomings, the details of which are described in an Internet Data 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 procure additional data from Morningstar. Each month,

⁷ The complete list of all Vanguard's index funds can be found at: <https://personal.vanguard.com/us/funds/vanguard/all?reset=true&mgmt=i>.

Morningstar sends a complete updated database to its clients. The monthly update is intended to replace the previous update. We purchase every update from January 1995 through March 2011 and construct 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, is the data available to Morningstar's clients at that time.

We then compare the returns reported on CRSP with what was reported on Morningstar. Somewhat surprisingly, 3.3% of return observations differ. Even if we restrict attention to returns that differ by more than 10 basis points (bps), 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 use dividend and net asset value 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 the other database was consistent, we use the consistent database return. If both databases are internally consistent, and differ from each other, but within six months one database is internally inconsistent, we use the database that is internally consistent throughout. Finally, we manually check all remaining unresolved discrepancies that differ by more than 20 bps by comparing the return with that reported on Bloomberg. All told, we are able to correct about two-thirds of the inconsistent returns. In all remaining cases, we use the return reported on CRSP.

Unfortunately, even more discrepancies exist between what Morningstar and CRSP report for total assets under management. 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 a factor eight (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 use the value reported on Morningstar. Unfortunately, both databases contain significant numbers of missing AUM observations. Even after we use both databases as a source of information, 17.2% of the data is missing. In these cases, we fill in any missing observations by using the most recent observation in the past. Finally, we adjust all AUM numbers by inflation by expressing all numbers in January 1, 2000 dollars.

The amount of missing expense ratio data poses a major problem.⁸ 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 reported annually

by funds, we are 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 go to the Securities and Exchange Commission (SEC) website and manually look up the remaining missing values on the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system. At the end of this process, we are missing only 1.6% of the observations, which we elect 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 its classification does not seem very accurate, and we have this information only after 1995. We therefore augment the Morningstar classification by using the following algorithm to identify passively managed funds. We first generate a list of common phrases that appear in fund names identified by Morningstar as index funds. We then compile a list of funds with these common phrases that are not labeled as index funds by Morningstar and compile a second list of common phrases from these funds' names. We then manually check the original prospectuses of any fund that contains a word from the first list but is not identified as an index fund at any point in its life by Morningstar or is identified as an index fund at some point in its life by Morningstar but nevertheless contains a phrase in the second list. Funds that are not tracked by Morningstar (e.g., exist only prior to 1995) that contain a word from the first list are also manually checked. Finally, we also manually check cases in which fund names satisfy any of these criteria in some periods but not in others even when the Morningstar classification is consistent with our name classification to verify that the fund has switched from active to passive or vice versa. We reclassify 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 in the form of either a colon (":") or a slash ("/"). Information after the separator denotes a subclass. Unfortunately, Morningstar does not provide this information, so, for mutual funds that appear only on the Morningstar database, we use the last word in the fund name to identify the subclass (the details of how we did this are in the Internet Data Appendix). Once identified, we aggregate all subclasses into a single fund.

We drop all index funds, bond funds, and money market funds and any fund observations before the fund's (inflation adjusted) AUM reached \$5 million.⁹ We also drop funds with less than two years of data. In the end, we are left with 6,054 funds.

Our sample is considerably larger than comparable samples used by other researchers. There are a number of reasons

⁸ Because fees are an important part of our skill measure, we choose not to follow Fama and French (2010) by filling in the missing expense ratios with the average expense ratios of funds with similar AUM.

⁹ We classify a fund as a bond fund if it holds, on average, less than 50% of assets in stocks and identify a money market fund as a fund that on average holds more than 20% of assets in cash.

for this. First, we do not restrict attention to funds that hold only U.S. equity. Clearly, managerial skill, if it exists, could be used to pick non-U.S. stocks. More important, by eliminating any fund that at any point holds a *single* non-U.S. stock, researchers have been eliminating managers who could have had the skill to opportunistically move capital to and from the U.S. As we show in the Internet Appendix, expanding the sample to all equity funds is not innocuous: The statistical power of our tests is greatly increased and, more important, managerial skill is positively correlated to the fraction of capital held in non-U.S. stocks. In addition, as Fig. 1 demonstrates, the fraction of AUM managed by funds that exclusively hold domestic stocks has dropped from 45% in 1977 to just 23% in 2011. So, by restricting attention to funds that exclusively hold domestic stocks, researchers have been focusing on a shrinking part of the mutual fund industry. This evidence is consistent with the idea that competition in the U.S. investment space has increased, thereby leading funds to find profitable investment opportunities abroad. Other evidence, also consistent with this idea, is provided in Pastor, Stambaugh, and Taylor (2015) for mutual funds and Dyck, Lins, and Pomorski (2013) for pension funds.

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 5,974 funds.

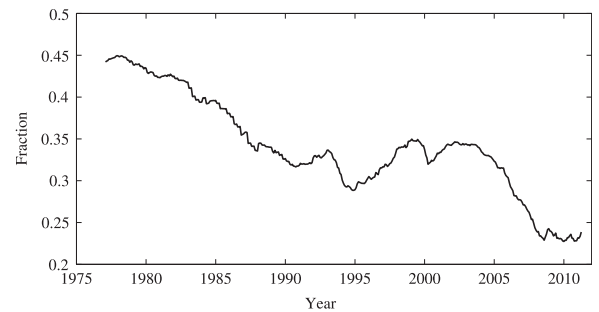


Fig. 1. Total assets under management (AUM) of domestic funds as a fraction of the AUM of all funds. The graph displays the AUM of mutual funds that exclusively hold domestic stocks as a fraction of the total AUM of all mutual funds, including those that invest in international stocks.

7.1. Measuring skill

We begin by first estimating average value added, S_i , for every fund in our sample. Because S_i is the mean of the product of the gross abnormal return and fund size, one can have concerns about whether the product is stationary. Fig. 2 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 mainly driven by increases in the number of funds, not increases in fund size.

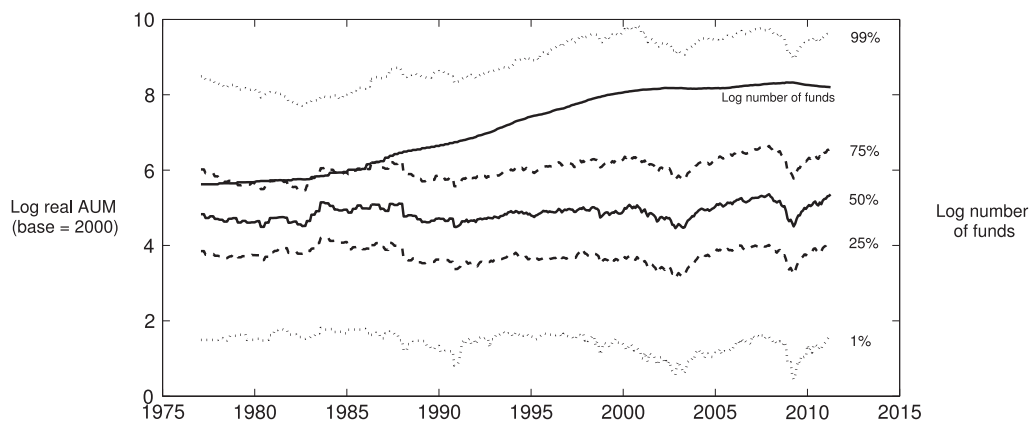


Fig. 2. Fund size distribution. The graph displays the evolution of the distribution of the logarithm of real assets under management (AUM) in millions of dollars (base year is 2000) by plotting the 1st, 25th, 50th, 75th, and 99th percentiles of the distribution at each point in time. The smooth black line is the logarithm of the total number of funds (not assets under management).

7. Results

As is common in the mutual fund literature, our unit of observation is the fund, not the individual manager. That is, we observe the dollar value the fund extracts from markets. We refer to this value as managerial skill for expositional ease. Given that this industry is highly labor-intensive, it is hard to conceive of other sources of this value added. However, it is important to keep in mind that this paper provides no direct evidence that this value results from human capital alone.

Table 3 provides the time-weighted mean of S_i (given by Eq. (9)) in our sample. The average fund added an economically significant \$270,000 per month, or \$3.2 million annually (in Y2000 dollars). The estimate of the mean of the ex ante distribution of talent, that is, Eq. (8), is \$140,000/month. Not surprisingly, this estimate is lower, reflecting the fact that unskilled managers go out of business sooner. When we use the FFC factor specification to correct for risk, we obtain very similar results.

Large variation also exists in the distribution across funds. The fund at the 99th percentile cutoff generated \$7.82 million

Table 3

Value added (\hat{S}_i). For every fund in our database, we estimate the average monthly value added, \hat{S}_i . The cross-sectional mean, standard error of mean, 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 of the weighted mean 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 Eq. (9). The numbers are reported in Y2000 \$ millions per month.

	Vanguard benchmark	FFC risk measure
Cross-sectional weighted mean	0.27	0.25
Standard error of the weighted mean	0.05	0.06
t -statistic	5.74	3.94
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%
Total number of funds	5974	6054

per month. 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

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, correlation in value added likely exists across funds. Second, the value added distribution features excess kurtosis. Even though our panel includes six thousand funds and 411 months, the sample might not be large enough to ensure that the t -statistic is t -distributed.

The most straightforward way to deal with the econometric shortcomings of our t -statistics is to use an alternative measure of statistical significance that does not have these issues. To derive this alternative measure, we exploit the fact that under the strong form of the null hypothesis, value added cannot be persistent. That is, under the null, managers that have added the most value in the past should not continue to add the most value in the future. 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 in the future when (1) top managers beat bottom managers, and (2) top managers are in the top half. The distribution, under the null, of both of these test statistics is Binomial($n, \frac{1}{2}$), where n is the number of future monthly observations.¹¹ Consequently,

we can calculate the p -value of each test statistic exactly, we do not need to rely on any large sample or asymptotic properties of the distribution. As a result, neither the excess kurtosis in returns nor the correlation across funds affects our calculations.

We operationalize the persistence tests as follows. 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 the Skill Ratio defined as

$$SKR_i^\tau \equiv \frac{\hat{S}_i^\tau}{\sigma(\hat{S}_i^\tau)}, \quad (24)$$

where $\hat{S}_i^\tau = \sum_{t=1}^{\tau} V_{it}/\tau$ and $\sigma(\hat{S}_i^\tau) = \sqrt{\sum_{t=1}^{\tau} (V_{it} - \hat{S}_i^\tau)^2/\tau}$. The skill ratio at time τ is essentially the t -statistic of the value added estimate measured over the entire history of the fund until time τ .¹² The time period from the beginning of the fund to τ is the *sorting period*. That is, the funds in the tenth (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 first (bottom) decile are funds for which we have the most confidence that the value added in the sorting period is negative. We then count the number of times the top decile beats the bottom decile and the number of times the top decile in the sorting period is one of the top five deciles over a specified future time horizon, called the *measurement horizon* hereafter.¹³

(footnote continued)

small samples as long as the distribution of value added is symmetric. The symmetry assumption turns out to be an accurate description of the data.

¹² For ease of exposition, we assume here 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.

¹³ Similar results are obtained if we use the value added estimate itself to sort funds.

¹⁰ For the reasons pointed out in Linnainmaa (2013), our measures of value added underestimates the true skill of managers.

¹¹ This result holds in general in large samples (because the sorting variable is unrelated to skewness and, therefore, is random), and holds in

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, as would occur if we would estimate the betas only once, 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 time to sort firms into ten 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 monthly value added, $\{V_{i,\tau}, \dots, V_{i,\tau+h}\}$, 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 $\{V_{i,\tau+m}, \dots, V_{i,\tau+h}\}$.¹⁴ 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.¹⁵ At the end of the process, in each decile, we have a time series of monthly estimates for average value added. We then compute, for each decile, the above order statistics as well as the mean and standard error of the time series.

We begin by reporting traditional measures of significance, the t -statistic of the mean. Fig. 3 plots the mean as well as the two standard error bounds for each decile and time horizon. From Fig. 3 it appears that evidence of persistence exists as far out as ten years. The point estimate of the average value added of tenth decile managers is positive at every horizon and is always the best performing 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 tenth decile managers 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.¹⁶

Next we report the results of our tests based on order statistics alone. As is evident from Table 4, the null hypothesis can be rejected at the 95% confidence level at almost all horizons.¹⁷ 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 tenth decile managers. Investors reward skilled managers by providing them with more capital.

It could be tempting, based on our sorts, to conclude that all the skill is concentrated in the tenth decile managers, that is, at most 10% of managers 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. Because 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.

We next exploit the fact that investors appear to be able to identify skilled managers to construct a sorting procedure that is better able to separate managers. From Fig. 4, we can conclude that once managers reveal their skill by adding value, investors reward them with higher subsequent compensation (total dollar fees). Because the average percentage fee does not differ by much across deciles, it is investors that determine these compensation differences (by choosing to allocate their capital to skilled managers), confirming a central insight in Berk and Green (2004).

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 instead of 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. Fig. 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 skill ratio is used (i.e., Fig. 5 versus Fig. 3).

There is also increased monotonicity when the sorts are based on compensation instead of skill ratio. To formally show

¹⁴ Even after dropping the first m observations, the strategy is still tradable, because it is implementable at $\tau+m$. So this procedure ensures that there is no selection bias. We therefore do not require funds to exist for the full measurement horizon. Finally, this strategy uses non-overlapping data.

¹⁵ We choose the starting point to ensure that the last month is always included in the sample.

¹⁶ 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 eliminating the effect of cross fund correlation and substantially reducing the excess kurtosis.

¹⁷ To ensure that these results are not driven by the small departures from symmetry, we preserve the skewness by demeaning the time series of decile monthly average value added and rerun the tests using the demeaned sample. The resulting p -values were distributed around 50%.

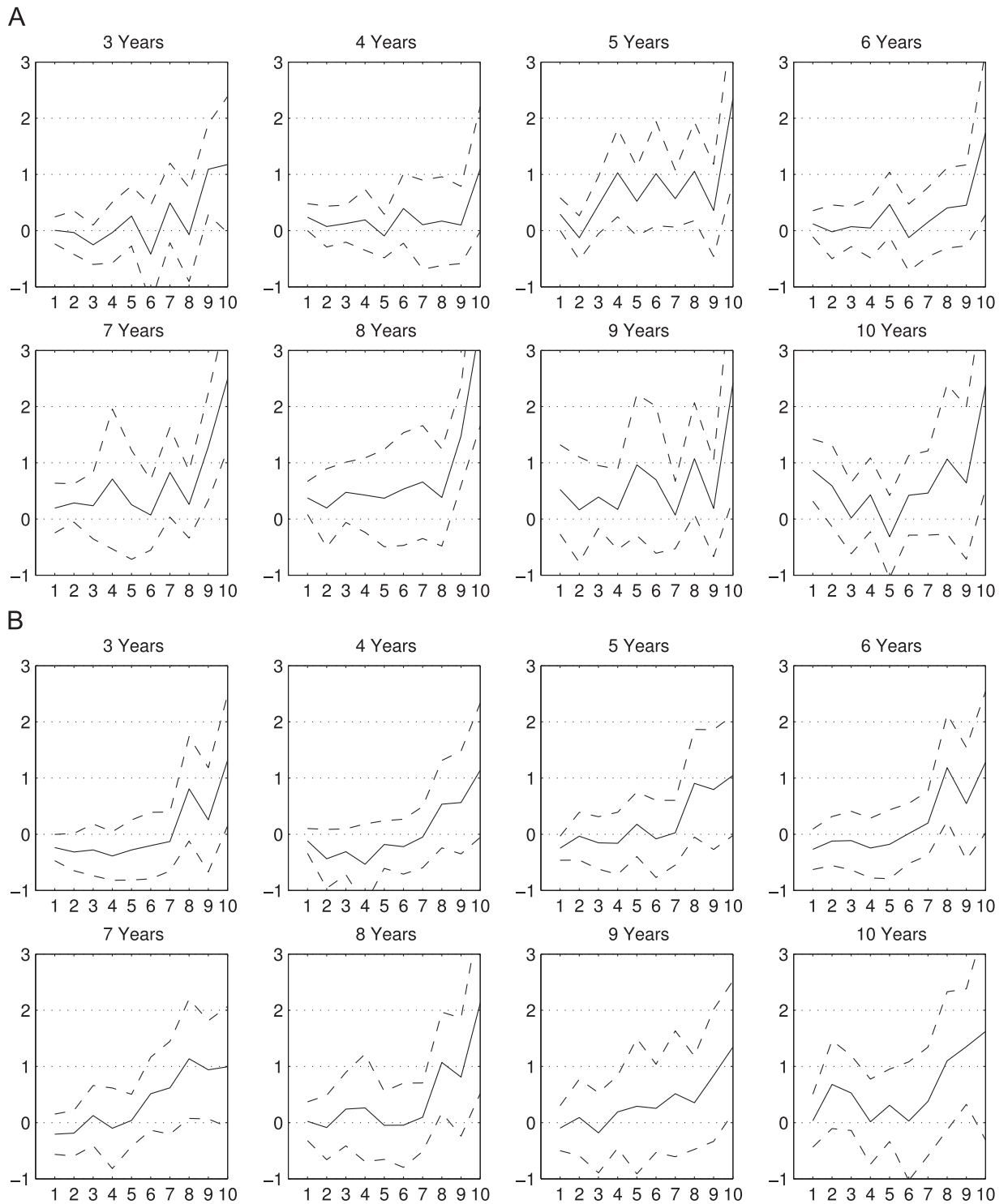


Fig. 3. Out-of-sample value added. Each graph displays average out-of-sample value added, \hat{S}_i (in millions of Y2000 dollars/month), of funds sorted into deciles on the skill ratio (horizontal axes), over the future horizon indicated by the graph title. The solid line indicates the performance of each decile and the dashed lines indicate 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 Fama-French-Carhart (FFC) risk adjustment.

Table 4

Out-of-sample performance of the top decile. The table reports the average value added of the top decile at each horizon and the associated p -value, the fraction of time and associated p -value that the top decile has a higher value added realization than the bottom decile, the fraction of time the realized value added of the top decile is in the top half, and 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.

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: FFC 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

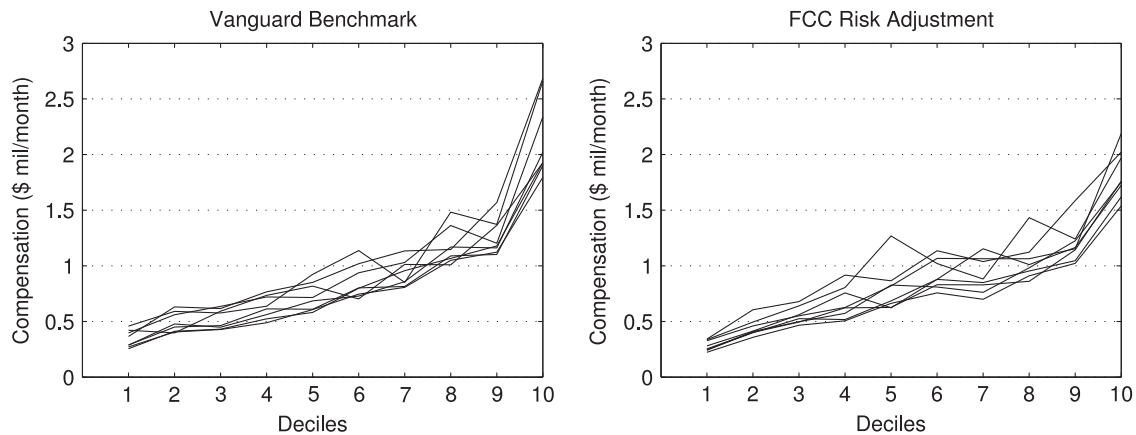


Fig. 4. Out-of-sample compensation. The plots display the average out-of-sample monthly compensation of each decile sorted on skill ratio using the Vanguard benchmark and the Fama-French-Carhart (FFC) risk adjustment. Each line in the plots represents a different horizon, which varies between three and ten years. For ease of comparison, the data sample (time period) is the same for both plots.

this difference, we count the number of times each decile outperforms the next lowest decile (in terms of value added). Table 5 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 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 all funds. Here we 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 evidence in Fig. 4 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 sub-optimal, that is, dumb investors chasing past returns. Our

evidence relating compensation to future performance suggests the opposite. Investors appear able to differentiate good managers from bad and compensate them accordingly. Notice from Fig. 2 that real compensation for the top managers has increased over time, that is, fund size has increased while fees have remained constant. In contrast, 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.

7.2. Returns to investors

Given the evidence of skill we have documented so far, a natural question to ask is: Who benefits from this skill? That is, do mutual fund companies and managers capture

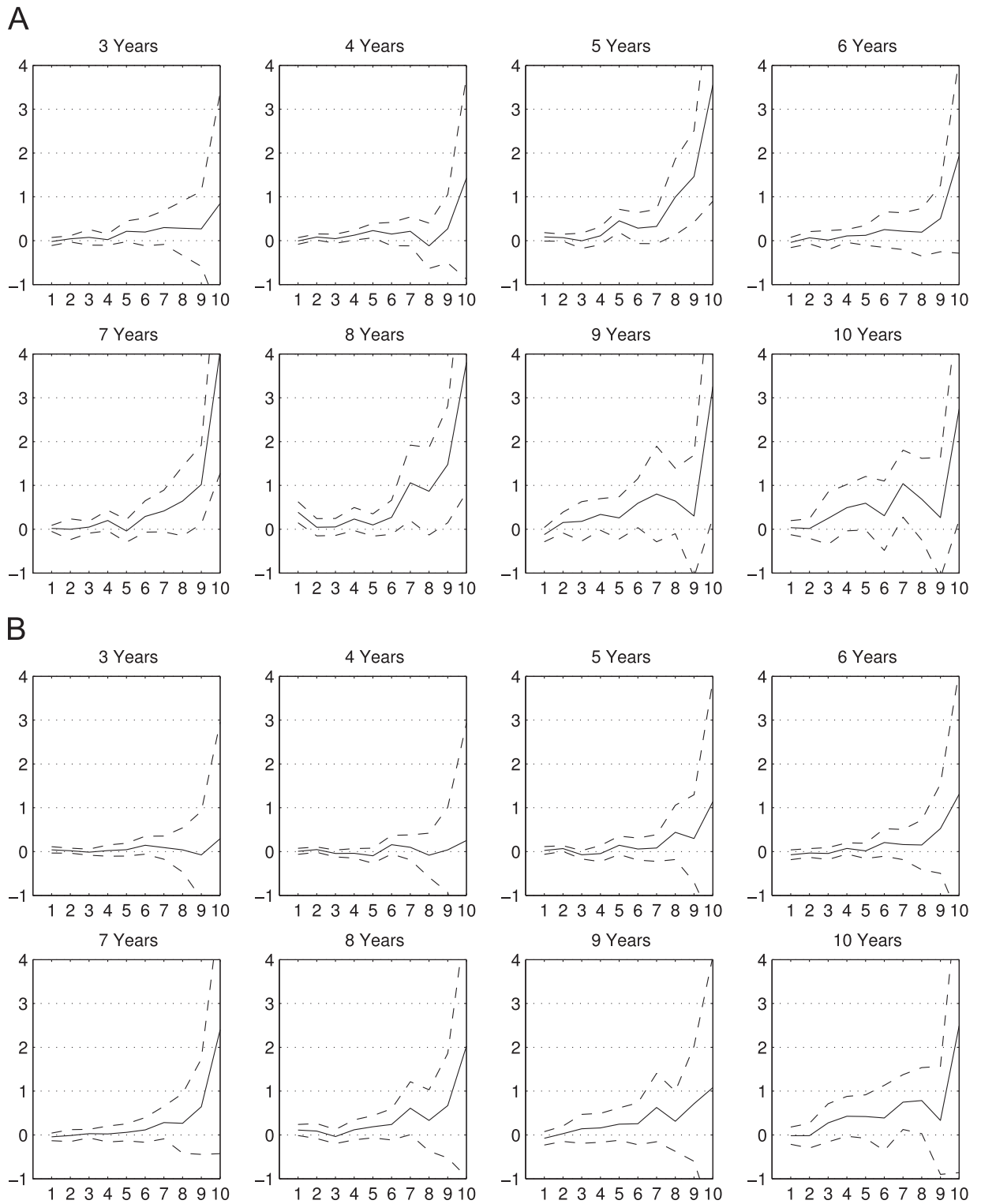


Fig. 5. Value added sorted on compensation Each graph displays average out-of-sample value added, \hat{S}_i (in millions of Y2000 dollars/month), of funds sorted into deciles based on total compensation (fees \times assets under management). The solid line indicates the performance of each decile, and 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 Fama-French-Carhart (FFC) risk adjustment.

Table 5

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., for a sample length of N months, the probability of the event is Binomial($9, 1/2$).

Horizon	Vanguard benchmark Skill ratio	Compensation	FFC risk adjustment 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 6

Net alpha (in bp/month): The table reports the net alpha of two investment strategies: Investing \$1 every month by equally weighting over all existing funds and investing \$1 every month by value weighting (based on AUM) over all existing funds.

	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

all the rents, or are these rents shared with investors? [Table 6](#) provides summary evidence. The average net alpha across all 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. On a value-weighted basis, relative to the FFC model, investors earned a significantly negative net alpha. But, relying on these estimates requires the additional assumption that this model correctly measures risk. If not, and instead one interprets the FFC factor portfolios as the alternative investment opportunity set, then one would expect a negative alpha because these portfolios were not available to investors as an alternative investment opportunity until the early nineties and they ignore transaction costs.

As we show in the Internet Appendix, very little evidence exists of persistence in net alpha when Vanguard is used as the benchmark. When funds are sorted into deciles based on the skill ratio, and the net alpha is measured over the same future horizons as in [Table 4](#), almost all net alpha estimates are not statistically significantly different from zero and they show no pattern across the deciles. The point estimates of the tenth decile are very close to zero and mostly negative. However, in this case, a striking difference emerges when we use the FFC factor specification as a risk adjustment instead of the Vanguard benchmark. As we show in the Internet Appendix, when the FFC factor specification is used as a risk adjustment, statistically significant differences exist across the deciles. At all horizons, the tenth decile always outperforms the first decile. These out-of-sample net alpha results imply that if investors truly measure risk using the FFC factor specification, they are leaving money on the table (not enough funds are flowing to the best managers resulting in positive net alphas). Alternatively, the results raise the

possibility that the FFC factor specification does not measure risk that investors care about. [Berk and van Binsbergen \(2013\)](#) develops this idea. In that paper, we use the information in fund flows to assess whether investors are actually using the asset pricing models that have been derived in the finance literature.

8. Separating out diversification services

The Vanguard benchmarks are constructed from net returns, and 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 diversification benefits and 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 services 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 with our earlier measures. The first two columns of [Table 7](#) demonstrate that about 40% of the value added is due to diversification benefits (\$110,000 per month) and 60% (\$170,000 per month) is due to other types of skill, such as stock picking or market timing. The non-diversification skills are also persistent. As the bottom panel in [Table 7](#) demonstrates, when funds are sorted on skill ratio computed using Vanguard gross returns as the benchmark, the top decile consistently outperforms the bottom

Table 7

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 with the set of index funds (including the Vanguard index funds themselves). To separate the value added coming from diversification benefits versus stock picking or market timing, we use two different benchmarks: Vanguard index funds' gross returns and Vanguard index funds' net returns.

Benchmark	Active funds Vanguard gross	Vanguard net	Index funds Vanguard gross	Vanguard net
Number of funds	5974	5974	644	644
In sample value added				
Weighted mean (\$mil/mon)	0.16	0.27	−0.02	0.11
t -statistic	3.46	5.74	−0.20	0.94
Mean (\$mil/mon)	0.07	0.14	−0.05	0.03
t -statistic	2.49	4.57	−0.49	0.37
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 (bp/mon)	−	2.7	−	−1.9
t -statistic	−	0.73	−	−0.41
Value Weighted (bp/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	3.47	4.75	−	−
4 year horizon	3.87	2.07	−	−
5 year horizon	11.84	3.54	−	−
6 year horizon	9.23	1.09	−	−
7 year horizon	0.00	0.01	−	−
8 year horizon	1.87	0.67	−	−
9 year horizon	23.39	9.15	−	−
10 year horizon	18.14	5.55	−	−

decile.¹⁸ The top decile is also distinctive because the fraction of value added attributable to diversification services is just 24% on average, lower than the average in the whole sample, indicating that stock selection and market timing skills are relatively more important for better managers. Similar results are obtained when we use the ex ante distribution of skill Eq. (8). Half the value added can be attributed to diversification benefits.

Although Vanguard is widely regarded as the most efficient provider of diversification services, one could be concerned that Vanguard is not as efficient as the representative index fund. The evidence in Table 7 supports the overall efficiency of Vanguard. 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.

9. Gross alpha

Proposition 1, combined with the fact that AUM shows large cross-sectional variation, implies that gross alpha should be a poor proxy for skill. The data supports this insight. Table 8 shows that the correlation between value added and gross alpha is small. The cross-sectional correlation between the fund by fund estimate of gross alpha (α_i^g) and skill (S_i) is only 23%, and the fraction of the cross-sectional variation in value added explained by gross alpha (the R^2) is just 5%.

One can think of two possible economies that are consistent with cross sectional variation in managerial skill. The first is one in which fees have high and AUMs have low cross-sectional variation and the second is one in which fees have low and AUMs have high cross-sectional variation. Because gross alpha equals the fee, it will proxy only for skill in the former world. What Table 8 shows is that we live in the latter world. The average value-weighted gross alpha in our sample is 6.5 bps/month, or about 80 bps/year, reflecting the fees active managers charge. The equally weighted number is higher, reflecting the fact that smaller funds charge higher fees. More important, there is very little persistence in gross alpha.

¹⁸ The other order statistics also support persistence and are reported in the Internet Appendix.

Table 8

Gross alpha $\rho(S_i, \alpha_i^g)$ is the cross-sectional correlation of estimated gross alpha and average value added. R^2 is the fraction of the cross-sectional variation in S_i explained by α_i^g . The rest of the table reports the average in-sample gross alpha as well as the results of the out-of-sample persistence test using the frequency with which the tenth decile beats the first decile. We follow exactly the same procedure we used in the original persistence tests except instead of using value added, we use the gross alpha to construct the skill ratio and we assess future performance using the value weighted gross alpha of the decile.

	Vanguard benchmark	FCC risk adjustment
<u>Explanatory power of gross alpha</u>		
$\rho(S_i, \alpha_i^g)$	0.23	0.24
R^2	0.05	0.06
<u>In sample gross alpha</u>		
Value weighted mean (bp/month)	6.5 ^a	0.66
Equally weighted mean (bp/month)	13 ^b	5
<u>Persistence</u> (p-value (%) of top decile outperforming bottom)		
3 year horizon	17.28	0.79
4 year horizon	34.19	1.36
5 year horizon	1.46	0.00
6 year horizon	2.67	0.06
7 year horizon	29.13	0.33
8 year horizon	18.16	0.32
9 year horizon	31.42	1.09
10 year horizon	59.00	6.19

^a t-statistic greater (in absolute value) than 1.96.

^b t-statistic greater (in absolute value) than 2.54.

When past gross alpha is used to predict future gross alpha, no evidence of persistence exists at most horizons.

When the FFC factor portfolios are used to adjust for risk, a different picture emerges. Although still positive, the gross alpha is not significantly different from zero and, not surprisingly, given the strong persistence in FFC net alpha estimates as shown in the Internet Appendix, there is evidence of persistence, although it is weaker than the equivalent evidence of persistence in value added (see Table 4). As already pointed out, these persistence results are evidence against the assumption that the FFC factors correctly price risk. In fact, in Berk and van Binsbergen (2013) we evaluate the FFC factor specification using flow data and fail to find evidence that the additional factors (beyond the market portfolio) in this specification capture risk that investors care about.

10. Conclusion

In this paper we use value added to measure the skill in mutual funds. We find that this method leads to a different conclusion about the existence of managerial skill. We show that the average manager is skilled, adding \$3.2 million per year. The evidence of skill that we uncover cannot easily be attributable to luck because cross-sectional differences in skill are persistent for as long as ten 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 also predict future value added.

We demonstrate that our measure of skill more accurately ranks managers than what is commonly used in the existing literature, the net and gross alpha. We prove that measuring skill by value added and gross alpha lead to

similar outcomes only when the cross-sectional variation in fund size is small relative to the cross-sectional variation in fees. Because, in reality, the cross-sectional variation in fund size is much larger than that of fees, only a small fraction of the cross-sectional distribution of skill is explained by gross alpha.

Our paper clarifies the role of the traditional measures that exist in the literature. We argue that the rationality of investors and the competitiveness of capital markets is measured by the net alpha. A positive net alpha implies that capital markets are not competitive. A negative net alpha implies that some investors are irrational in that they are committing too much money to active management. Because of the large cross-sectional variation that exists in AUM, we do not find a role for the gross alpha.

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. In that model, 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

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 :

$$R_i^g = \begin{bmatrix} R_{i1}^g \\ \vdots \\ R_{iT}^g \end{bmatrix}, \quad (25)$$

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 :

$$R_t^B = \begin{bmatrix} R_{1t}^B & R_{2t}^B \end{bmatrix}. \quad (26)$$

Assume that the first benchmark fund is available to investors over the whole sample period, and the second benchmark fund is available only 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

$$\gamma \equiv \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} \quad (27)$$

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

$$V_{it} = q_{i,t-1} (R_{it}^g - \beta R_{1t}^B). \quad (28)$$

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

$$V_{it} = q_{i,t-1} (R_{it}^g - R_t^B \gamma). \quad (29)$$

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

$$X = \begin{bmatrix} 1 & R_{11}^B & \cdot \\ \vdots & \vdots & \cdot \\ 1 & R_{1S}^B & \cdot \\ 1 & R_{1,S+1}^B & R_{2,S+1}^{BO} \\ \vdots & \vdots & \vdots \\ 1 & R_{1T}^B & R_{2,T}^{BO} \end{bmatrix}, \quad (30)$$

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

$$X^O = \begin{bmatrix} 1 & R_{11}^B & \bar{R}_2^{BO} \\ \vdots & \vdots & \vdots \\ 1 & R_{1S}^B & \bar{R}_2^{BO} \\ 1 & R_{1,S+1}^B & R_{2,S+1}^{BO} \\ \vdots & \vdots & \vdots \\ 1 & R_{1T}^B & R_{2,T}^{BO} \end{bmatrix}, \quad (31)$$

where

$$\bar{R}_2^{BO} = \frac{\sum_{t=S+1}^T R_{2t}^{BO}}{T-S} \quad (32)$$

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

$$R_{2t}^{BO} = R_{2t}^B - \theta R_{1t}^B \quad \text{for } t = S+1, \dots, T \quad (33)$$

where

$$\theta = \frac{\text{cov}(R_{2t}^B, R_{1t}^B)}{\text{var}(R_{1t}^B)}. \quad (34)$$

Finally, define

$$\hat{X}^O = \begin{bmatrix} 1 & R_{11}^B & 0 \\ \vdots & \vdots & \vdots \\ 1 & R_{1S}^B & 0 \\ 1 & R_{1,S+1}^B & R_{2,S+1}^{BO} \\ \vdots & \vdots & \vdots \\ 1 & R_{1T}^B & R_{2,T}^{BO} \end{bmatrix}. \quad (35)$$

Proposition 2. The value added of the firm at any time t can be estimated as

$$V_{it} = q_{i,t-1} (R_{it}^g - \zeta_2 \hat{X}_{2t}^O - \zeta_3 \hat{X}_{3t}^O) \quad (36)$$

using a single Ordinary Least Squares (OLS) regression to estimate ζ :

$$\zeta = (X^O X^O)^{-1} X^O R_i^g. \quad (37)$$

Proof. The second and the third column of X^O are orthogonal to each other, 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

$$\zeta_2 = \frac{\text{cov}(R_{it}^g, R_{1t}^B)}{\text{var}(R_{1t}^B)} = \beta. \quad (38)$$

So, for any $t \leq S$, Eq. (36) reduces to Eq. (28), and this estimate of value added is consistent over the first subsample. Using the orthogonality of X^O ,

$$\zeta_3 = \frac{\text{cov}(R_{it}^g, X_{3t}^O)}{\text{var}(X_{3t}^O)} = \frac{\text{cov}(R_{it}^g, R_{2t}^{BO})}{\text{var}(R_{2t}^{BO})}, \quad (39)$$

rewriting

$$\gamma_1 R_{1t}^B + \gamma_2 R_{2t}^B = \gamma_1 R_{1t}^B + \gamma_2 (\theta R_{1t}^B + R_{2t}^{BO}) = (\gamma_1 + \theta \gamma_2) R_{1t}^B + \gamma_2 R_{2t}^{BO} \quad (40)$$

and using the fact that linear projections are unique implies

$$\zeta_2 = \beta = \gamma_1 + \theta \gamma_2 \quad (41)$$

and

$$\zeta_3 = \gamma_2. \quad (42)$$

So, for $t > S$,

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

which is Eq. (29), and so the estimate is also consistent over the second subsample. \square

Appendix A. Data Description and Supplementary Results

A detailed description of the steps we took to clean the data and the additional Supplementary figures and tables associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.jfineco.2015.05.002>.

References

- Baker, M., Litov, L., Wachter, J.A., Wurgler, J., 2010. Can mutual fund managers pick stocks? Evidence from their trades prior to earnings announcements. *Journal of Financial and Quantitative Analysis* 45, 1111–1131.
- Berk, J.B., Green, R.C., 2004. Mutual fund flows and performance in rational markets. *Journal of Political Economy* 112, 1269–1295.
- Berk, J.B., Tonks, I., 2007. Return persistence and fund flows in the worst performing mutual funds. Unpublished working paper 13042. National Bureau of Economic Research, Cambridge, MA.
- Berk, J.B., van Binsbergen, J.H., 2013. Assessing asset pricing models using revealed preference. *Journal of Financial Economics*, forthcoming.
- Bogle, J.C., 2014. Lightning strikes: the creation of vanguard, the first index mutual fund, and the revolution it spawned. *Journal of Portfolio Management* 40, 42–49.
- Bollen, N.P.B., Busse, J.A., 2001. On the timing ability of mutual fund managers. *Journal of Finance* 56, 1075–1094.
- Bollen, N.P.B., Busse, J.A., 2005. Short-term persistence in mutual fund performance. *Review of Financial Studies* 18, 569–597.
- Carhart, M.M., 1997. On persistence in mutual fund performance. *Journal of Finance* 52, 57–82.
- Chen, H.-L., Jegadeesh, N., Wermers, R., 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., Frazzini, A., Malloy, C., 2007. The small world of investing: board connections and mutual fund returns. Unpublished working paper 13121. National Bureau of Economic Research, Cambridge, MA.
- Cohen, R.B., Polk, C.K., Silli, B., 2010. Best ideas (<http://ssrn.com/abstract=1364827>).
- Coval, J.D., Moskowitz, T.J., 2001. The geography of investment: informed trading and asset prices. *Journal of Political Economy* 109, 811–841.
- Cremers, K.J.M., Petajisto, A., 2009. How active is your fund manager? A new measure that predicts performance. *Review of Financial Studies* 22, 3329–3365.
- Daniel, K., Grinblatt, M., Titman, S., Wermers, R., 1997. Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance* 52, 1035–1058.
- Dyck, A., Lins, K.V., Pomorski, L., 2013. Does active management pay? New international evidence. *Review of Asset Pricing Studies* 3, 200–228.
- Fama, E.F., 1965. The behavior of stock market prices. *Journal of Business* 38, 34–105.
- Fama, E.F., 1970. Efficient capital markets: a review of theory and empirical work. *Journal of Finance* 25, 383–417.
- Fama, E.F., French, K.R., 1995. Size and book-to-market factors in earnings and returns. *Journal of Finance* 50, 131–155.
- Fama, E.F., French, K.R., 2010. Luck versus skill in the cross section of mutual fund returns. *Journal of Finance* 65, 1915–1947.
- Grinblatt, M., Titman, S., 1989. Mutual fund performance: an analysis of quarterly portfolio holdings. *Journal of Business* 62, 393–416.
- Grinblatt, M., Titman, S., 1993. Performance measurement without benchmarks: an examination of mutual fund returns. *Journal of Business* 66, 47–68.
- Gruber, M.J., 1996. Another puzzle: the growth in actively managed mutual funds. *Journal of Finance* 51, 783–810.
- Guercio, D.D., Reuter, J., 2014. Mutual fund performance and the incentive to generate alpha. *Journal of Finance* 69, 1673–1704.
- Jensen, M.C., 1968. The performance of mutual funds in the period 1945–1964. *Journal of Finance* 23, 389–416.
- Jiang, H., Verbeek, M., Wang, Y., 2011. Information content when mutual funds deviate from benchmarks. SSRN eLibrary (<http://ssrn.com/paper=1782692>).
- Kacperczyk, M., Nieuwerburgh, S.V., Veldkamp, L., 2014. Time-varying fund manager skill. *Journal of Finance* 69, 1455–1484.
- Kacperczyk, M., Sialm, C., Zheng, L., 2005. On the industry concentration of actively managed equity mutual funds. *Journal of Finance* 60, 1983–2011.
- Kacperczyk, M., Sialm, C., Zheng, L., 2008. Unobserved actions of mutual funds. *Review of Financial Studies* 21, 2379–2416.
- Kosowski, R., Timmermann, A., Wermers, R., White, H., 2006. Can mutual fund 'stars' really pick stocks? New evidence from a bootstrap analysis. *Journal of Finance* 61, 2551–2595.
- Linnainmaa, J.T., 2013. Reverse survivorship bias. *Journal of Finance* 68, 789–813.
- Lucas, R.E., 1978. On the size distribution of business firms. *Bell Journal of Economics* 9, 508–523.
- Malkiel, B.G., 1995. Returns from investing in equity mutual funds 1971 to 1991. *Journal of Finance* 50, 549–572.
- Mamaysky, H., Spiegel, M., Zhang, H., 2008. Estimating the dynamics of mutual fund alphas and betas. *Review of Financial Studies* 21, 233–264.
- Pastor, L., Stambaugh, R.F., 2012. On the size of the active management industry. *Journal of Political Economy* 120, 740–781.
- Pastor, L., Stambaugh, R.F., Taylor, L.A., 2015. Scale and skill in active management. *Journal of Financial Economics* 45–49, forthcoming.
- Sharpe, W.F., 1991. The arithmetic of active management. *Financial Analysts Journal* 47, 7–9.
- Shumway, T., Szeffler, M.B., Yuan, K., 2009. The information content of revealed beliefs in portfolio holdings. Unpublished working paper. University of Michigan, Ann Arbor, MI.
- Wermers, R., 2000. Mutual fund performance: an empirical decomposition into stock-picking talent, style, transactions costs, and expenses. *Journal of Finance* 55, 1655–1703.
- Zheng, L., 1999. Is money smart? A study of mutual fund investors' fund selection ability. *Journal of Finance* 54, 901–933.