New Special Study of the Securities Markets: Financial Intermediaries

Jonathan B. Berk* Jules H. van Binsbergen†

March 16, 2017

Comments welcome

*Stanford University, Graduate School of Business and NBER; jbberk@stanford.edu.
†Wharton, University of Pennsylvania and NBER; julesv@wharton.upenn.edu.

1. Introduction

Equity markets have undergone several important changes in the past 70 years. Arguably one of the most important ones is how individuals hold equity. While in 1945 almost all corporate equities were held directly by households and nonprofits, today direct holdings by individuals make up less than 40% of holdings. In fact some studies argue that this number is closer to 20%. At the same time the market has witnessed enormous growth in the open-end mutual fund sector. From almost no presence at all in 1945, these funds now make up 25% of the market or more. In short, in the last 70 years there has been a large trend away from direct investing into delegated fund management. We view this trend as the single most important change in how investors use financial intermediaries. In 1945, when investors invested directly, the intermediary was a broker who was most likely paid as function of the number of trades he made. Today, investors give their money to fund managers or financial advisors, who then invest this money on investors' behalf in equity (and sometimes other) funds. These intermediaries are compensated based on the amount of assets under management (AUM). The move away from trade based compensation to AUM based compensation represents important progress. As we will argue, AUM based compensation contracts much better align the incentives of the money manager and her investors and is likely a primary factor in driving the trend from direct investing through brokers to indirect investing.

The invention of the mutual fund has made diversified investing accessible to essentially all investors. Previously, each individual investor had to construct diversified portfolios themselves, which involved an inefficient amount of trade given the amounts invested. Compared to that counterfactual, the mutual fund industry in all its diversity adds large amounts of value to investors.

Because of the rise in delegated money management, the bulk of this report will be devoted to that sector of the investment intermediary space. We believe the importance of delegated money managers is likely to keep rising as investors keep moving from direct investing into indirect investing. As we will argue in this report, the future regulation of equity markets relies on an in-depth understanding of the delegated money management equilibrium. Consequently, a large fraction of our report will focus on describing the equilibrium, and its implications for competition in the sector. We begin by first summarizing the important trends in the last 70 years.

2. The Last 70 Years

As we pointed out in the introduction, the single most important trend in the last 70 years is the secular decline of direct investing in equity markets and the concomitant rise of several other important players who hold equity on behalf of investors. Table L.223 of the Flow of Funds Accounts of the United States published by the Federal Reserve provides an overview of the amounts of corporate equity held by various types of investors. We compute how much each of these investors holds as a fraction of the total and plot these fractions for the six groups with the largest relative holdings in Figure 1. As we have already noted, the fraction held directly by households and non-profits has decreased from over 90% to about 40%. This downward trend is reflective of a equally important concomitant trend: rather than investing directly in markets, individual investors have increasingly chosen to allocate their money to investment managers. The fraction of equity held by open end mutual funds has increased to about 25%. The remainder can be explained by the rise of pension plan holdings (both defined contribution and defined benefit), as well as the rise of exchange traded funds (ETFs). Finally, holdings by foreigners have also increased.

French (2008) argues that the Fed uses the household and nonprofit sector as a residual. Its allocation is the aggregate value of corporate equity minus the combined values of the other sectors, implying that the household and nonprofit sector includes not only the publicly traded common equity held by households and nonprofits, but also preferred stock and closely held corporations. French (2008) uses various other data sources to separate these pieces, and argues that the fraction of public equity held by households is substantially lower than 40% and closer to 20% in 2007. The downward trend for these adjusted numbers up until 2007 is the same as the computations we present here. Based on these computations it is therefore not unreasonable to assume that since 2007, the fraction of equity held by households has not changed much.

In Figure 2 we show the holdings of the remaining 6 groups. All of these groups, which includes (among others) life insurance companies, properties/casualty insurance companies, and broker-dealers all have very small holdings of equity and by 2015 all these holdings are below 1%. Note also that over the entire sample, the closed-end funds only hold a very small fraction of the total. To gauge the trend in closed-end investing, we plot in Figure 3 the equity holdings of closed-end funds as a fraction of the total mutual fund holdings (including open-end and closed-end funds). The graph shows a marked decline in closed-end fund holdings, particularly in the seventies. In relative terms, closed-end funds have all but disappeared.

Finally, in Figure 4 we plot the corporate equity holdings of ETFs as a fraction of the total corporate equity holdings of open-end mutual funds, closed-end mututal funds and

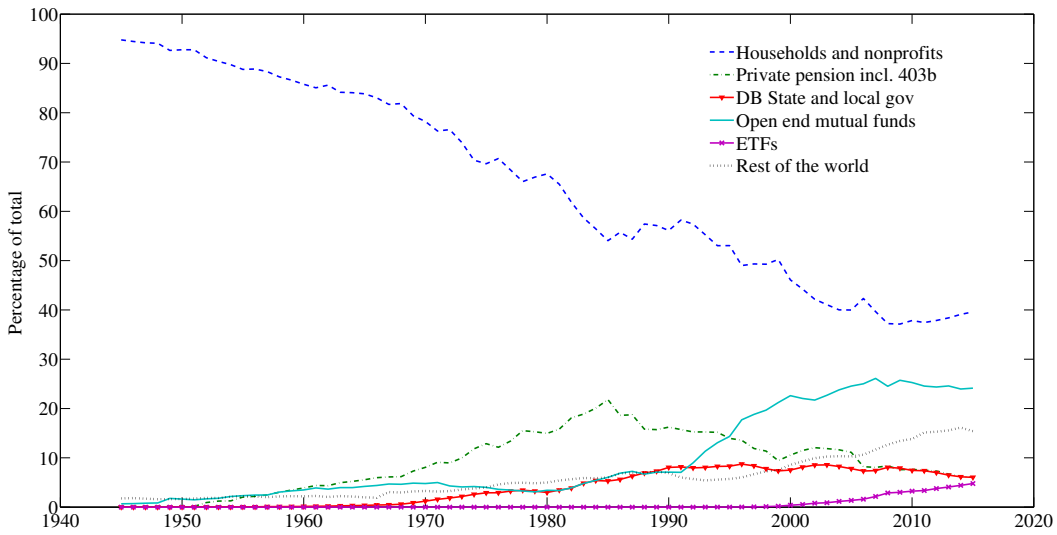

Figure 1: **Who Holds Corporate Equities?**
The graph shows the fraction of US corporate equity held by its largest investor groups.

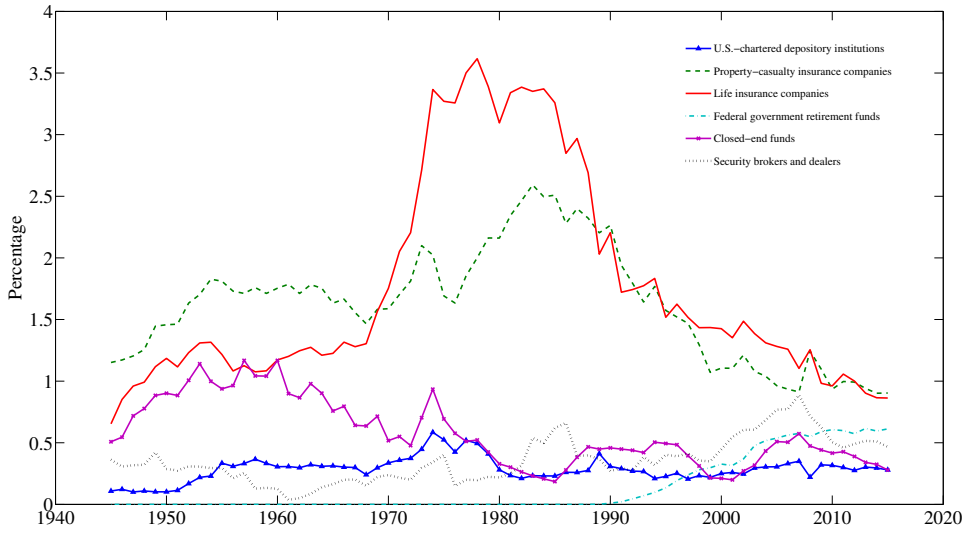

Figure 2: **Who Holds Corporate Equities? Continued**
The graph shows the fraction of US corporate equity held by various smaller investor groups.

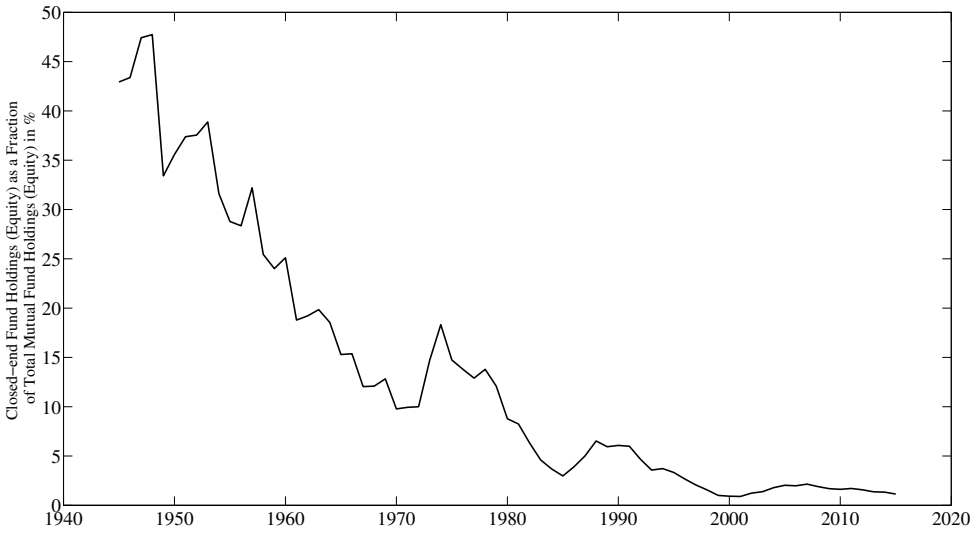

Figure 3: **Disappearing Closed End Funds**
The graph shows the value of equity holdings of closed-end funds as a fraction of the total equity holdings of open and closed-end mutual funds.

ETFs. The graph shows a clear upward trend. While ETFs were essentially non-existent in the early nineties, their fraction has increased to over 16% in 2015.

3. Money Management Firms

The explosive growth of the money management industry spurred a very large academic literature that studies this sector. The literature has largely been focused on answering two important questions: (1) whether investors are better off investing directly themselves or indirectly through a money manager, and (2) whether money managers add value by selecting stocks on behalf of investors. Until recently, the consensus view was that the answer to the first question is a qualified yes: investors are better off so long as they avoid active managers and invest in passive index funds. Further, the consensus is that the answer to the second question is no: money managers are no better at picking stocks than monkeys throwing darts at a dartboard.

In fact, both these conclusions are not correct. They are a result of inconsistently applying the rational expectations equilibrium concept (commonly refered to as "efficient markets") to delegated money management. In a series of research articles (Berk and Green 2004, Berk and van Binsbergen 2015, Berk and van Binsbergen 2017) we have demonstrated that when the rational expectations equilibrium is consistently applied to both direct and indirect investing, a different picture emerges. Specifically, the answer to the first question is that

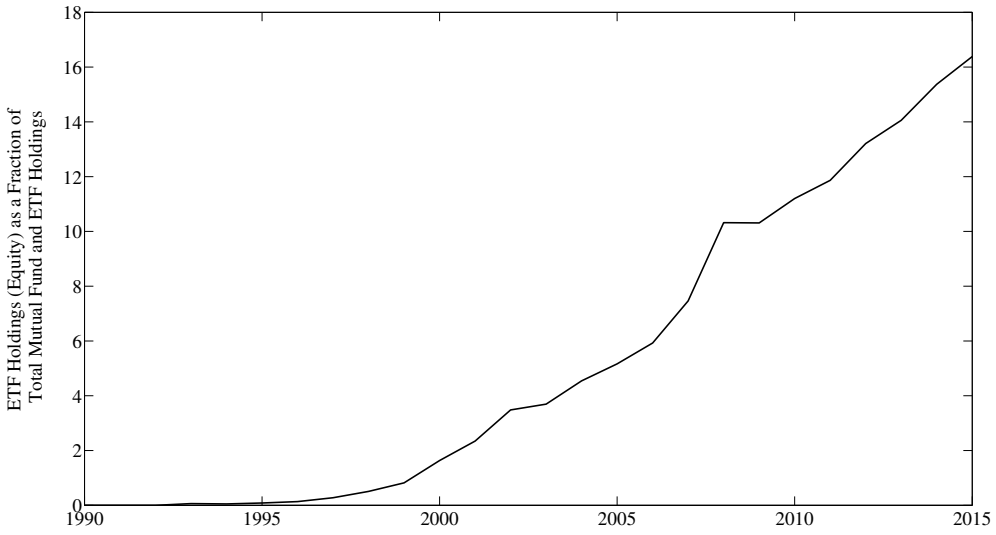

Figure 4: **The Rise of ETFs**
The graph shows the AUM invested with ETFs as a fraction of the total corporate equity holdings of open-end mutual funds, closed-end mutual funds and ETFs.

market competition implies that in equilibrium investors are indifferent between active and passive investing and the answer to the second question is yes. Active mutual fund managers add considerable value, but competition between investors ensures that this value accrues to the intermediaries rather than to investors. Because of the importance of these results, we will describe them in detail.

3.1. The Rational Expectations Equilibrium in Money Management

The primary question we have been asked to address is the competitiveness of the money management industry. To answer this question it is essential to first understand the nature of the competition in this industry. Until recently, many financial economists maintained a rather schizophrenic view of investors. When investors invest directly in stocks, the widely accepted view is that the rational expectations equilibrium so closely approximates the actual equilibrium, that changes in stock prices in reaction to news can be used as evidence in a court of law as a measure of the value of that news. On the other hand, when investors invest indirectly in stocks through a money manger, the generally accepted view was that in this market, investors are naive. Consequently, according to this view, they choose to invest almost exclusively in negative net present value investments (Malkiel 1995, Carhart 1997, Fama and French 2010). Furthermore, investor naivete is so dominant that the market does not equilibrate and so returns reflect things other than risk, in particular, returns

reflect managerial skill (or lack thereof). Moreover, because fund flows into mutual funds are known to be highly predictable based on past performance, but past performance has little or no predictability for future performance, researchers concluded that mutual fund investors acted on information that was worthless, what the literature has named "return chasing." What is particularly perplexing about this dichotomous view of investors is that there is substantial overlap in the investors in both markets. This historical view of the money management industry is unfortunately still so pervasive it continues to influence policy. As we argue below, there is little evidence to support it. By correctly characterizing the nature of competition in the industry, all of the above puzzles can be resolved. The data is consistent with an industry that is highly competitive. In fact, by one measure, more competitive even than stock markets.

The key to understanding the nature of competition in the money management industry is to apply the rational expectations paradigm consistently to the industry. Although the equilibrium concept was first proposed by Muth (1961), it was popularized in financial economics in a series of articles authored by Eugene Fama (Fama 1965, Fama 1970, Fama 1976), and became known as the "efficient market hypothesis." An important, and much emphasized, implication of the rational expectations equilibrium is that in such an equilibrium, the quality of a firm's actions and decisions cannot be measured by the expected return subsequent to the action or decision. Because investors compete with each other for attractive investment opportunities, the price of the attractive stocks (and bonds) is bid up to reflect the successfulness of the firm. That is, investors reward a successful firm with a high market capitalization, not a high expected return going forward.

While there is little consensus in the literature as to whether or not stock prices reflect *all* (publicly) available information, there is widespread agreement that at a minimum, prices reveal a substantial fraction of the information, making the cross-sectional distribution of firm size (as measured by market capitalization) a much better measure of firms' success than the cross sectional distribution of subsequent returns. The extent to which this idea has become widely accepted can be gauged by the fact that the change in value of a company upon the release of public information is admissible in a court of law as evidence of the value of the information itself. Put differently, it is widely accepted today that the main implication of the rational expectations paradigm is that the impact of information on a firm is measured by the change in value that results instantaneously when the information is released, and not by the expected returns going forward after the information is released.

Although the rational expectations paradigm came to dominate how we understand the the pricing of financial assets, it had little influence on how researchers characterized the

delegated money management equilibrium. In that literature, exactly the opposite paradigm has prevailed. In analyzing the behavior of mutual funds, researchers ignored the total size of the fund and instead used the future realized return as the measure of the quality of the fund. The two paradigms have radically different predictions, and so the conclusions of the literatures about the value of information were radically different. Stock price reactions were viewed as highly informative while changes in mutual fund sizes were deemed random and uninformative. Because mutual fund investors caused these "random" changes by responding to returns, they were deemed naive return chasers. In addition, researchers found that, on average, mutual funds did not deliver extra returns to their investors, i.e., there was no outperformance, which was interpreted as implying that there was no information in mutual fund returns, leading to the widely accepted perception that mutual fund managers lacked skill.

Before we examine the implications of consistently applying the rational expectations paradigm to both kinds of financial products, it is worthwhile considering why financial economists schizophrenically applied different concepts to two closely related investment products. We believe the answer lies in the original work that argued for using the rational expectations equilibrium concept to price stocks (Fama 1965, Malkiel 1995). In arguing for the rational expectations paradigm, researchers took the position that stock prices impound all information. Under this assumption, no agent should be able to predict future performance, and so no agent should be able to make money picking stocks. To demonstrate the empirical validity of this position, they used the performance of mutual funds. They argued that because mutual funds did not deliver positive net alpha to their investors, mutual fund managers lacked stock picking skill. The implication they drew from this evidence was that if even the professionals who claimed to have the ability to pick stocks could not, stock prices must indeed impound all information. However, this argument uses the rational expectations paradigm inconsistently. These researchers used the net alpha to measure stock picking ability and thus incorrectly concluded that there was no value to additional information other than what was already impounded in prices. Once this inconsistency was put in place, it perpetuated, resulting in financial economists inconsistently applying the paradigm in the two literatures.

To understand how to apply the rational expectations paradigm consistently to mutual funds, it is useful to use the stock market as an analogy. Just as in stock markets, in mutual funds, investors compete for attractive investment opportunities. In this case, the attractive investment opportunities are skilled fund managers (rather than skilled firm managers) and the information of relevance is the degree to which mutual fund managers can successfully pick stocks. There is, however, an important difference between the two markets. Unlike

stock markets, in the money management market the price is fixed. That is, regardless of how skilled the manager who manages the fund is, when an investor invests in a mutual fund, the price the investor pays for the fund is always the market value of the fund's underlying assets. What this implies is that the market does not equilibrate through prices, it equilibrates in quantities (i.e. fund size). Other than that difference, the rational expectations equilibrium in the two markets have identical implications. Just as with stocks, market capitalization of the fund measures the quality of the manager, the expected return of the fund measures its risk.

The key to understanding how the mutual fund market equilibrates is to realize that, like all industries, the mutual fund industry faces decreasing returns to scale. Consider a situation in which it is known to all investors that a manager was delivering a positive net alpha to investors. As with stocks, such a situation cannot be an equilibrium outcome. All investors will want to take advantage of this positive net present value investment opportunity. The manager will experience an inflow of funds, the size of the fund will thus increase and because the manager's investment ideas are finite, eventually the additional money cannot be put to productive use. This lowers the return the manager makes. The market will be in equilibrium when enough new funds are invested so that investors in the fund no longer receive an extra return (i.e. the net alpha falls to zero). A similar situation plays out if investors know that a manager was delivering a negative net alpha. In that case funds flow out, causing the size of the fund to drop and thereby increasing the net alpha. The market will return to equilibrium when enough funds are withdrawn to drive the net alpha to zero.

In the rational expectations equilibrium, competition between investors ensures that the expected risk-adjusted excess return to investors (the net alpha) is zero, implying that the net alpha no more measures the quality of a mutual fund manager than the expected return of a stock measures the quality of a firm manager. For stocks, the equilibrium is reached by bidding up the price of a successful firm thereby lowering the expected return to its equilibrium level, whereas for mutual funds, the equilibrium is reached by increasing the size of the fund. As is the case for stocks, the cross-sectional distribution of mutual fund success is predominantly reflected in the cross-sectional distribution of *fund size* as opposed to the distribution of risk adjusted excess returns to investors (net alpha). Just as with stocks, more skilled managers manage larger funds, less skilled managers manage (very) small funds, and they all make comparable net alphas close to zero.

In stock markets the rational expectations paradigm is often tested by assessing the extent to which future returns are predictable given a news announcement. This predictability provides evidence of the competitiveness of stock markets, and the rationality of investors.

It is not informative about the quality of firm management. The same logic applies in the mutual fund space. If net alphas in the mutual fund space are not zero, we learn something about the rational expectations of investors and the competition they face. We learn nothing about the quality of the mutual fund manager.

Consider the following simple model of mutual fund management based on the model in Berk and Green (2004). Let us start with a mutual fund manager who can generate a gross alpha (the alpha before fees have been taken out) that depends on the amount of invested capital equal to:

$$\alpha^g(q) = a - bq. \tag{1}$$

In words: the manager extracts from financial markets an extra amount a on the first cent she manages. Because the manager's investment ideas are in finite supply and because she invests her best ideas first, the extra amount decreases at a rate b for every additional dollar the manager invests.[1] The total dollar amount this manager extracts from financial markets, what we term the value added, is the product of the gross alpha and assets under management:

$$V(q) \equiv q\ \alpha^g(q) = q\ (a - bq). \tag{2}$$

The optimal amount to invest maximizes this quadratic function. Taking first order conditions with respect to q and setting this equal to zero gives

$$q^* = \frac{a}{2b}, \tag{3}$$

implying that gross alpha at the optimum is

$$\alpha^g(q^*) = \frac{a}{2}. \tag{4}$$

The value added at the maximum is given by:

$$V^* = \frac{a^2}{4b}. \tag{5}$$

Figure 5 plots the value added (V) and gross alpha (α^g) as a function of q. The figure also shows the value added and gross alpha at the optimal amount of money. Before we consider

[1] For ease of exposition we assume that the gross alpha is a linear function of fund size q, but the arguments presented do not rely on this linearity.

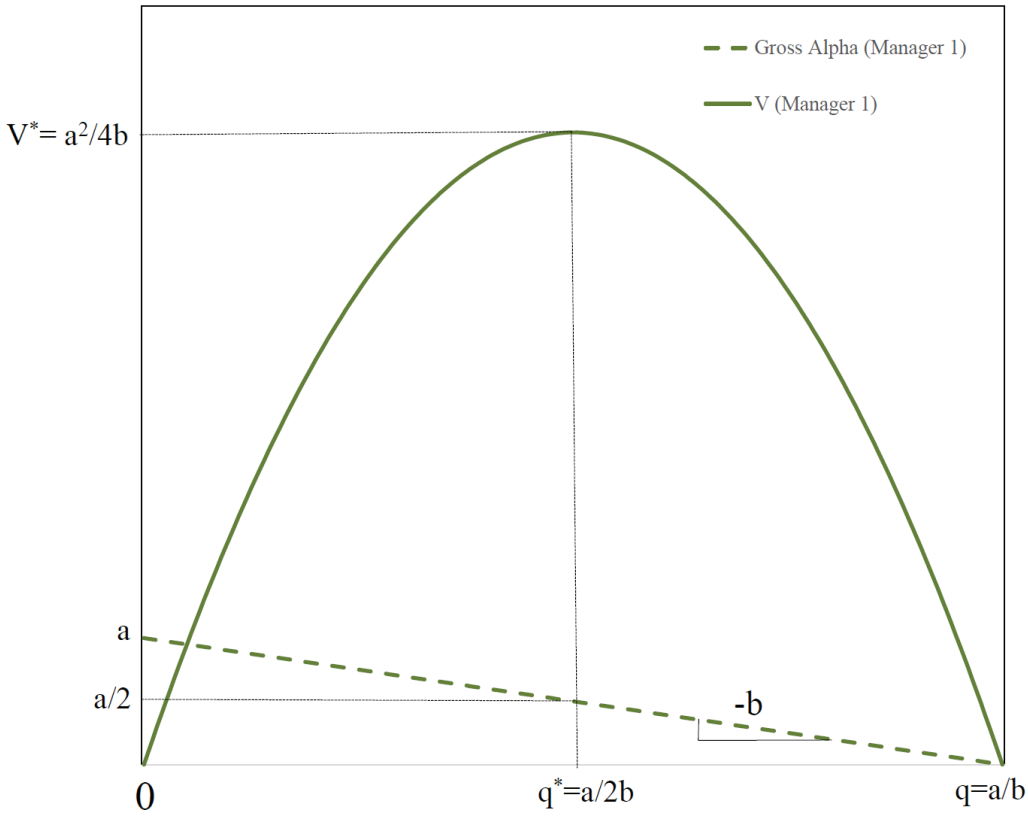

Figure 5: **Size, Value Added and Gross Alpha**
The graph shows the relationship between size and value added/gross alpha.

the investor's problem, it is worth comparing this manager to a manager of lower quality. Consider manager 2. Manager 2 runs out of good ideas more quickly than manager 1. That is, while manager 2 makes the same extra return on the first cent (equal to a), the rate at which the return deteriorates is twice as large and equal to $2b$ instead of b. In Figure 6 we plot the gross alpha of both managers as well as their value added. The optimal investment amount for manager 2 is twice as small as that for manager 1 and equal to $a/4b$. Because the gross alpha at the optimum is independent of the parameter b, both managers have the *same* gross alpha equal to $a/2$ at the optimum. Hence, gross alpha is not correlated with skill. If one were to use gross alpha as a measure of skill, we would come to the (wrong) conclusion that both managers are equally skilled. This result follows only from the fact that, if there are decreasing returns to scale, returns are not a good measure of (or proxy for) value and no assumptions related to rational expectations are required.

We now turn to investors and assume that they have rational expectations. A rational investor will chase any positive net present value investment opportunity. This implies that all assets earn an expected return commensurate with the risk of the asset. As a consequence,

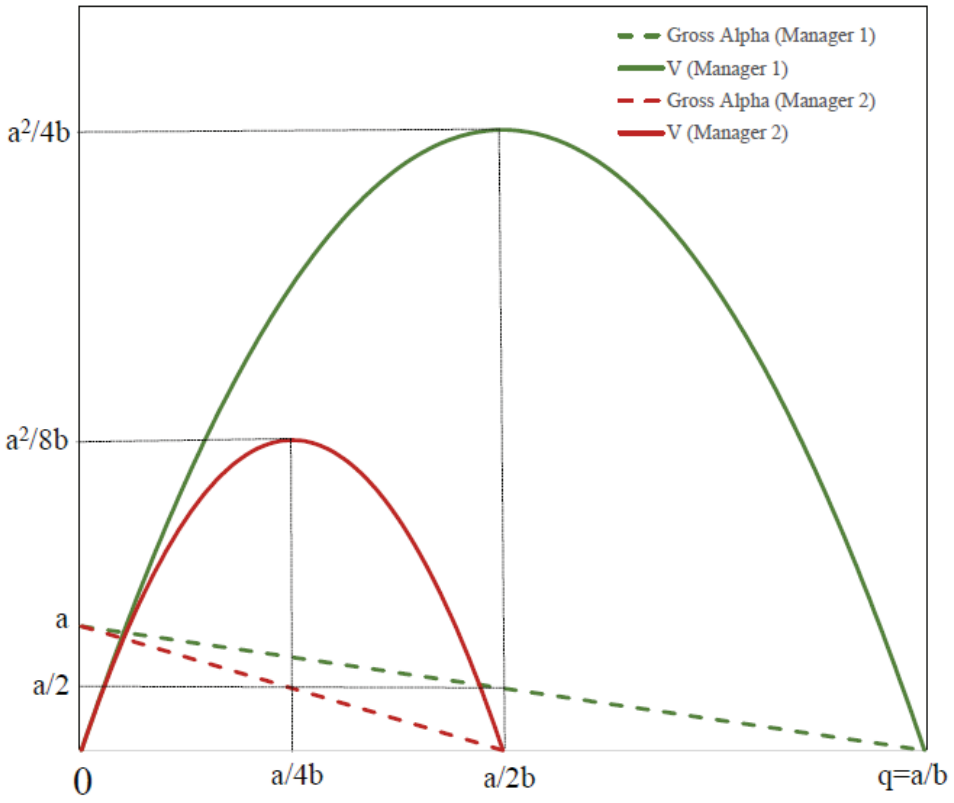

Figure 6: **Size, Value Added and Gross Alpha**
The graph shows the relationship between size and value added/gross alpha for two managers. Manager 1 is more skilled than Manager 2 while both make the same gross alpha on the first cent they invest. Manager 1 has an investment strategy that is more scalable than Manager 2. That is, manager 2's decreasing returns to scale parameter (b) is higher than manager 1's.

all funds must have net alphas of zero. If the manager picks her fee, f, equal to

$$f = \frac{a}{2},$$

investors will choose to invest $q^* = \frac{a}{2b}$ in the fund. As a consequence the fund's net alpha will be zero and the market will be in equilibrium. In the case of the second manager, the fee will be the same, but in this case the equilibrium investment is $\frac{a}{4b}$. That is, although both managers have the same gross (and net) alpha, manager 1 has twice as much money in equilibrium as manager 2. The total amount of money manager 1 extracts from markets is $V^* = \alpha^g(q^*)q^* = \frac{a}{2} \times \frac{a}{2b} = \frac{a^2}{4b}$ which is twice the amount that manager 2 extracts, $\frac{a}{2} \times \frac{a}{4b} = \frac{a^2}{8b}$, reflecting the fact that manager 1 is twice as skilled.

Before turning to the dynamics, it is worth emphasizing the essential characteristics of this equilibrium. Note that net alphas are always zero, so net alphas are not informative on managerial ability. Similarly, the gross alpha is also uninformative. The intuition for why

these return measures fail to measure skill follows from the same logic as to why present value measures should be used in place of internal rate of return (IRR) measures when making investment decisions. These days, nearly all textbooks in finance point out that IRR measures are flawed because they do not properly take into account the scale of the project. As we have already seen, the scale of a mutual fund is an endogenous quantity that is determined in equilibrium. Therefore, just as the IRR does not help us rank investment projects (we need the present value for that), the alpha does not help us rank managers. A necessary (but not sufficient) condition under which return measures can be used to make an investment decision is when the investment opportunity under consideration is infinitely scalable. While such an assumption might be a reasonable approximation when considering a very small investor in a large market, one would be hard pressed to argue that mutual fund managers fit this description. Although most financial economists would likely agree that it is unnecessary to test the hypothesis that positive NPV opportunities are in infinite supply in the economy as a whole, the mutual fund literature has nevertheless spent considerable effort testing this hypothesis in the mutual fund space. Not surprisingly, the literature has come to the conclusion that making the assumption that mutual funds face constant returns to scale is not very realistic. There is now mounting evidence that, all else equal, the return performance of a fund deteriorates with fund size (see Pastor, Stambaugh, and Taylor (2015) and Pastor, Stambaugh, and Taylor (2014)).

Thus far, the model we have derived is restrictive because we assumed that both managers and investors know the skill level (production function) of the manager. In reality, neither a nor b are likely to be known to either investors or managers. For simplicity, assume that managers and investors are symmetrically informed about the production function, and let $a_t = E_t[a]$ and $b_t = E_t[b]$ denote the conditional expectations of the parameters in the production function. Clearly, both a_t and b_t will change over time as the participants learn. Consequently, the optimal amount of capital changes, which, in the above equilibrium would require managers to continuously change their fees to ensure that investors would be willing to invest the optimal amount of capital. This equilibrium dynamic is counterfactual. Fees do not respond to information, fund size does.

The key to understanding how managers maximize the value they extract without continuously adjusting their fees is to consider the manager's problem. Notice that from the manager's perspective, it is always suboptimal to invest anything other than q^* in active management. Consequently, he will continue to invest this amount, regardless of the fee, by either borrowing money (if possible) when the fee is too high and so investors choose to invest less than q^*, or indexing the excess money when the fee is too low and investors provide more capital than q^*.

Let us explicitly consider the second case. Suppose that the fund size at time t, q_t is larger than $q_t^* = \frac{a_t}{2b_t}$ because $f < \frac{a_t}{2}$. The optimal strategy for the manager is to put q_t^* into active management and index the difference, $q_t - q_t^*$. In this case, the indexed money earns no alpha, so the equilibrium gross alpha on the whole fund is given by:

$$\left(\frac{q_t^*}{q_t}\right)\frac{a_t}{2} + \left(\frac{q_t - q^*}{q_t}\right)0 = \frac{a_t^2}{4q_t b_t}. \tag{6}$$

In equilibrium, the net alpha of the fund must be zero. Imposing this restriction gives

$$\frac{a_t^2}{4q_t b_t} - f = 0. \tag{7}$$

The equilibrium size of the fund is thus:

$$q_t = \frac{a_t^2}{4fb_t}. \tag{8}$$

Notice that the dynamic equilibrium where fees are fixed, shares an important characteristic with the static equilibrium: in both cases the gross alpha equals the fee charged, and therefore is not a reliable measure of managerial skill. Given that the size of the fund adjusts to ensure that the gross alpha and the fee are equal, it is not appropriate to think about gross alpha and fund size as two independent entities. Because they are related in equilibrium, the size of the fund and the gross alpha (i.e. the fee) are not separately identified by the parameters that determine managerial skill. Their product, on the other hand, is uniquely identified by those two parameters: regardless of the fee the manager chooses, the product of the size of the fund (q) and the equilibrium gross alpha (the fee) equals:

$$V_t^* = \frac{a_t^2}{4fb_t}f = \frac{a_t^2}{4q_t b_t}q_t = \frac{a_t^2}{4b_t}. \tag{9}$$

This product is what we call the *value added* of the fund. It is the correct way to measure managerial skill because it measures the manager's value added and is a function of only the skill of the manager.

In summary, with indexing and fixed fees, fund size adjusts to ensure that the gross alpha is sufficiently high to cover the manager's fees. There is a wide range of fees that all allow the manager to fully exploit her skill and extract the optimal amount of money from financial markets. What this implies is that the fee charged is irrelevant — managers can choose to charge a high fee and manage a small fund, or charge a low fee and manage a large fund. In both cases the amount the manager makes as well as the return the investors earn, is the

same. Because gross alpha must equal the fee in equilibrium, it too is irrelevant.

Finally, the equilibrium described above teaches another important lesson. Some have argued that investment managers should be more generous to their investors by lowering their fees thereby giving up a larger part of their performance to their investors. What the equilibrium shows is that it is not the manager's choice of fees that sets the return to investors equal to zero. It is competition between investors for good investment opportunities. The fee is *irrelevant* to this discussion. The only way a manager can be more generous to her investors is if the manager stops accepting money from new investors, thereby favoring old investors over new investors.

One implication of this insight is that regulating fees without also regulating the size of the fund is ineffective. To see this easily, consider the following simple example. Suppose a manager is currently managing the optimal amount of $q^* = \$100$ million. The manager charges a fee of 1% and generates a gross alpha of 1% on this amount. Her value added is thus \$1 million. Now suppose that regulations prohibit any fees higher than 0.5%. At this level of the fee, the net alpha to investors is positive and equal to 0.5%, which is not an equilibrium. More funds will flow in and the fund will grow. As the manager was already extracting the maximum amount of money from financial markets given her skill level (i.e. \$1 million), she cannot put this additional money to productive use and will thus index it. The fund grows to \$200 million of which \$100 million is invested as before and the second \$100 million is put in index funds. As a consequence, the gross alpha will drop to 0.5% (the equal-weighted average of the 1% gross alpha from before and the 0% gross alpha that results from indexing), and thus the net alpha will once again be 0% as both the gross alpha and the fee are 0.5%. At this point we are back to the equilibrium. As a consequence, the rents to investors have remained unchanged. The value-added of the manager has remained unchanged as a gross alpha of 0.5% on \$200 is also equal to \$1 million. In summary, other than the administrative costs associated with implementing the regulation, the regulation had no other effects.

Before we move on to documenting the empirical performance of this model, it is worth spending some time on an argument that is often cited for why mutual fund managers, as a group, must underperform (have a negative net alpha). Some researchers 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 immediately implies that the abnormal return of the average active investor must be zero. As convincing as the argument appears to be, it cannot be used to conclude that the average active mutual fund manager cannot add value. In his definition of "active" investors, Sharpe includes *any* investor not holding the market, not just active mutual fund managers. If active individual investors exist, then as a group active mutual fund managers could provide a positive abnormal return by making trading profits from individual investors who make a negative abnormal return. Of course, as a group individual active investors are better off investing in the market, which leaves open the question of why these individuals are actively trading.

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

3.2. Empirical Evidence

In this subsection we will demonstrate that the simple dynamic rational expectations equilibrium derived above is able to explain the important empirical regularities documented in the mutual fund literature, as well as resolve the most important puzzles. We will focus exclusively on the mutual fund sector because that is the only place in the money management space where the data is of very high quality. All mutual funds are required to report their results to the SEC, and these numbers must be verified by independent auditors. Other money managers are not subject to these strict reporting requirements, and so the resulting datasets are subject to self reporting biases.

We use the data set in Berk and van Binsbergen (2015). That data set, which covers the period from January 1962 to March 2011 is comprised of monthly observations compiled from combining two databases, the CRSP survivorship bias free mutual fund database and the Morningstar Principia database.

Our first objective is to test the implications of the rational expectations paradigm on investors. The equilibrium has two main implications. First, we should see net alphas of

zero and second, there should be no easy way to predict which funds will deliver superior risk adjusted returns. The main roadblock to testing these two predictions is constructing an estimate of the funds' net alpha. Generally two methods have been applied. The standard practice in financial economics is not to construct the alternative investment opportunity itself, but rather to simply adjust for risk using a risk model. In recent years, the extent to which risk 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. Although in principle this approach is a sensible way to address the issue of not knowing the correct model of risk, the way this approach is typically implemented in practice replaces one shortcoming with another. What researchers have typically done is assume that investors' next best investment opportunity is spanned by the factor mimicking portfolios in the Fama-French-Carhart factor specification (Fama and French 1996, Carhart 1997). That is, they have interpreted the factor mimicking portfolios in these factor specifications as investment opportunities available to investors, rather than risk factors.

There are two reasons why these factor portfolios are not investable opportunities. The first is straightforward. These portfolios do not include transaction costs. In essence, you cannot compare the performance of a fund that incurs transaction costs to a fund that does not. The second issue is more subtle. The factors that are typically used were identified in the the late 1980's and 1990's and popularized by Fama and French (1996) and Carhart (1997). However, most studies include data that begin at least 20 years before those factors were identified. In those earlier years, investors would not have known about these portfolios and obviously could not have invested in them. By using these portfolios to benchmark managers, researchers are effectively evaluating managers in 1970 using 1990's technology. Any manager who, in 1970, knew about the investment opportunities afforded by these portfolios should be given credit for this knowledge and the subsequent outperformance.

By benchmarking managers against non-investable benchmarks, researchers are effectively handicapping managers. To estimate the size of this handicap, we can evaluate the "performance" of the factor portfolios themselves against a set of passive, but investable, benchmarks. The most obvious set to use is the set of index funds offered by the Vanguard company. The advantage of using these funds is that they are constructed for the purpose of giving investors the least costly method to diversification. This explicit objective is not shared by alternative benchmarks constructed by companies such as Morningstar. Moreover, Vanguard is not only the market leader offering this service, it is also the pioneer in the space. For example, the 11 funds listed in Table 1 span the set of all index funds offered by the firm. In each case, the Vanguard fund was the first index fund to offer that particular strategy. That means that these funds are natural indicators to use to determine when a

strategy becomes widely known to all investors.

It is not uncommon in the mutual fund literature to use style benchmarks that are either identified by the fund itself or by external organizations such as Morningstar. The problem with using these benchmarks is that many funds regularly deviate from the style objectives they report and advertise. By simply projecting each fund on all available Vanguard index funds, such potential misclassifications are avoided.

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

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

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

Table 2 shows the results of evaluating the performance of each factor mimicking portfolio using the set of passively managed index funds offered by Vanguard over the period 1977-

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

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

2011.[2] Only the market portfolio does not have a statistically significant positive alpha. The other portfolios returned economically large excess returns, from 22 b.p. per month to as much as 70 b.p. for the momentum portfolio. Since this portfolio incurs the most transaction costs, it is not surprising that it exhibits the largest "outperformance." What the table shows is that the factor mimicking portfolios were much better investment opportunities than what was actually available to investors at the time. Given the flaws inherent in using the factor portfolios as benchmarks, we advocate using the Vanguard portfolios instead. We suggest that to evaluate mutual fund performance, one should construct a fund's benchmark by finding the closest portfolio in the set of Vanguard index funds. That is, if R_t^j is the excess return (over the risk free rate) earned by investors in the j'th Vanguard index fund at time t, then the benchmark return for fund i is given by:

$$R_{it}^B = \sum_{j=1}^{n(t)} \beta_i^j R_t^j, \tag{10}$$

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.[3] Using Vanguard index funds as the benchmark, recognizes the industrial organization of the mutual fund industry. The dynamic evolution of active strategies is therefore automatically accounted for. Thus one 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, also, that if we use this benchmark to evaluate a Vanguard index fund itself, we would conclude that that fund adds value equal to the dollar value of the fees it charges. Vanguard funds add value because they provide investors with the lowest cost means to diversification. 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 credits them with this value added.

Using this benchmark, we can now construct an empirical estimate of net alpha. If R_{it}^n is the return investors in the fund earn (i.e., the return after all fees are taken out), then define

$$\varepsilon_{it} \equiv R_{it}^n - R_{it}^B. \tag{11}$$

The time series average of ε_{it} is an estimate of the fund's net alpha.

[2]We start in 1977 because that was when Vanguard introduced its first index fund. Details of how the benchmarks are constructed can be found in Berk and van Binsbergen (2015).

[3]See Berk and van Binsbergen (2015) for a detailed description of the methodology used.

3.2.1. Net Alpha

Perhaps the most widely cited empirical fact in the mutual fund literature is the observation that investors in active mutual funds earn a negative alpha (see Fama and French (2010) and the references therein). The evidence in Berk and van Binsbergen (2015) demonstrates that this "fact" is an artifact of two implementation choices almost all research that studies this question has made. Neither choice can be justified on theoretical grounds. The most important is the one we identified above — researchers artificially handicapped funds by benchmarking them to non-investable strategies. The second is a data snooping bias. Almost all prior research documenting this underperformance dropped more than half of the observations by restricting attention to funds that only invest in U.S. stocks and starting the time series in the early eighties. We can think of no reason for this data selection procedure.

We begin by estimating the average net alpha by calculating the average ε_{it} across all funds by equally and value weighting the funds. As Table 3 shows, over this time period estimated net alpha is not statistically distinguishable from zero. For comparison purposes, the table recalculates these estimates under the two implementation choices made by the literature. Because the Fama-French-Carhart (FFC) factor portfolios can be constructed in 1962, in this case we begin the analysis in 1962. As one would expect, when funds are handicapped by requiring them to outperform a non-investable benchmark, the net alpha estimate drops. If one then also drops funds that do hold international stocks, the net alpha drops again. With both implementation choices the net alpha estimate is indeed negative and statistically significantly different from zero.

	Full Dataset		U.S. Equity Only	
	Vanguard Benchmark	FFC Factor Portfolios	Vanguard Benchmark	FFC Factor Portfolios
Equally Weighted	2.74	-3.88	-0.72	-6.70
t-statistic	0.73	-1.40	-0.20	-2.66
Value Weighted	-0.95	-5.88	-4.68	-8.34
t-statistic	-0.31	-2.35	-1.58	-3.70
Number of Funds	5974	6054	2731	2811

Table 3: **Net Alpha (in b.p./month):** The table reports the net alpha of two investment strategies: Investing \$1 every month by equally weighting over all existing funds (*Equally Weighted*) and investing \$1 every month by value weighting (based on AUM) over all existing funds (*Value Weighted*).

The evidence in Table 3 is consistent with the rational expectations equilibrium. But importantly, that equilibrium imposes further restrictions. It requires that any realized return in excess of the benchmark be unpredictable. To test this hypothesis we sort funds based

on their past estimates of net alpha and determine whether funds that have outperformed in the past are more likely to outperform in the future.

We operationalize the persistence tests as follows. To sort funds based on their estimated net-alpha at time τ, we compute the t-static of the net alpha estimate based on the entire history of the fund until that time:

$$\frac{\hat{S}_i^\tau}{\sigma(\hat{S}_i^\tau)}, \tag{12}$$

where $\hat{S}_i^\tau = \sum_{t=1}^{\tau} \frac{\varepsilon_{it}}{\tau}$ and $\sigma(\hat{S}_i^\tau) = \frac{\sqrt{\sum_{t=1}^{\tau}(\varepsilon_{it}-\hat{S}_i^\tau)^2}}{\tau}$. We term the time period from the beginning of the fund to τ the *sorting period*. That is, the funds in the 10th (top) decile are the funds where we have the most confidence that the estimated net alpha over the sorting period is positive. Similarly, funds in the 1st (bottom) decile are funds where we have the most confidence that the estimated net alpha in the sorting period is negative. We then report the subsequent net alpha estimate of an equally weighted portfolio consisting of all the funds in each decile over a specified future time horizon, hereafter the *measurement horizon*.

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

At each time τ, we use all the information until that point in time to sort firms into 10 deciles based on the historical t-statistic. 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 ε_{it} for each fund in each month, 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 $\{\varepsilon_{i,\tau+m}, \ldots, \varepsilon_{i,\tau+h}\}$. Because the Vanguard benchmark has at most 11 factors plus the constant, we use $m = 18$. We then average over funds in each decile in each month, that is, we compute, for each decile, an equally weighted average. At the end of the horizon, funds are again sorted into deciles based on the t-statistic at that time, and the process is repeated as many times as

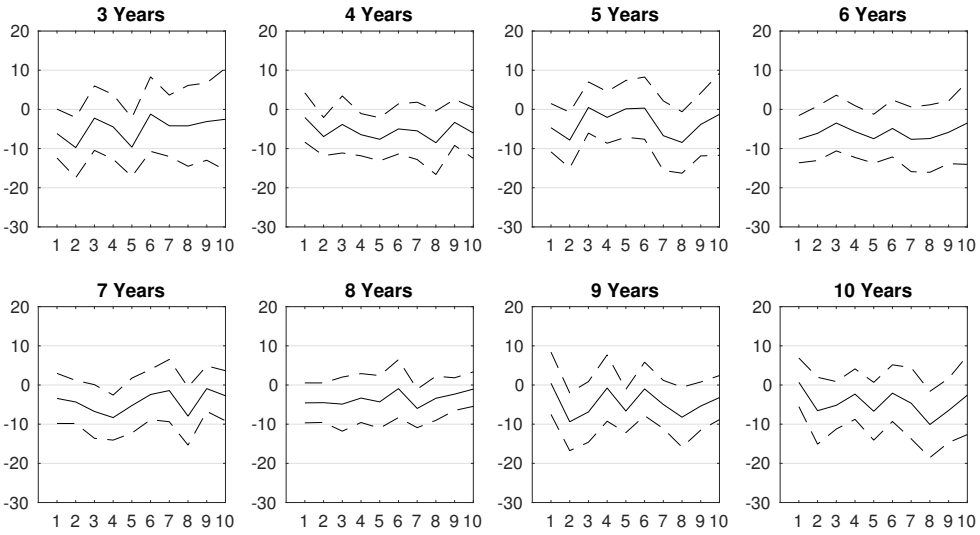

Figure 7: **Out-of-Sample Net Alpha**
Each graph displays the out-of-sample performance (in b.p./month) of funds sorted into deciles based on the historical t-statistic of the net alpha estimate over the horizon indicated. The solid line indicates the performance of each decile and the dashed lines indicated the 95% confidence bands (two standard errors from the estimate).

the data allows.

Figure 7 reports net alpha estimates as well as 95% confidence bounds. From the figure it is clear that net alpha is not predictable. Managers who historically outperformed their benchmarks are no more likely to deliver a positive net alpha to their investors than managers who underperformed historically.

3.2.2. Managerial Skill

A second implication of the rational expectations equilibrium is that, because managers earn fees and funds do not underperform, they must be skilled. To measure the skill of mutual fund managers, we need to construct an estimate of value added. To construct this measure, we first adjust the gross realized return by the realized net return of the benchmark, $R_{it}^g - R_{it}^B$. This quantity is then multiplied by the real size of the fund (assets under management adjusted by inflation) at the end of the previous period, $q_{i,t-1}$, to obtain the realized value added between times $t-1$ and t:

$$V_{it} \equiv q_{i,t-1}\left(R_{it}^g - R_{it}^B\right). \tag{13}$$

The time series average of V_{it} measures a fund's value added.

It is important to understand that while under the rational expectations paradigm the only measure of managerial skill is value added, value added always measures the amount of money extracted from markets regardless of whether the rational expectations paradigm holds. To understand why, notice that

$$V_t = q_t \alpha_t^g(q_t) = q_t \alpha_t^n(q_t) + q_t f$$

where $\alpha_t^n(q_t)$ is the net alpha of the fund as a function of fund size. The first term in the above equation is the amount of money the manager either gives to or takes from investors. The second term is the amount of money the manager takes for himself. Notice that there is *no other source of funds*. What this observation implies is that the money the manager takes in compensation (dollar fees) can only come from one of two places, either from skill (through stock picking) or from investors (by underperforming). So the sum of these two terms must equal the amount of money the manager makes from his stock picks. This observation relies on no assumption other than this budget constraint.

We begin by measuring the average value added of mutual fund managers over the period 1977-2011 in January 1, 2000 dollars.[4] The results are reported in Table 4 and indicate that mutual fund managers are highly skilled. The average fund adds an economically significant \$140,000 per month (in Y2000 dollars). There is also large variation across funds. The fund at the 99th percentile cutoff generated \$7.82 million per month and 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. The main insight is that most managers destroyed value but because most of the capital is controlled by skilled managers, as a group, active mutual funds added considerable value.

Successful funds are more likely to survive than unsuccessful funds. Consequently, one can think about the average value added of all mutual funds as estimates of the *ex-ante* distribution of talent. We can also compute the average V_{it} in the data set without first averaging by funds. Because surviving funds are overrepresented in this mean, we obtain an estimate of the *ex-post* distribution of talent, that is, the average skill of the set of funds actually managing money. Not surprisingly this estimate is higher. The average fund added \$270,000/month.

If managers are skilled, one would expect this skill to persist. To test for evidence of this persistence, we repeat the same procedure we used to test for persistence in net alphas. However, in this case we use past value added as the criterium for the sort. To infer skill at

[4]The data is available from 1962, but the analysis begins in 1977 because that is the year Vanguard offered its first index fund.

Cross-Sectional Mean	0.14
Standard Error of the Mean	0.03
t-Statistic	4.57
1st Percentile	-3.60
5th Percentile	-1.15
10th Percentile	-0.59
50th Percentile	-0.02
90th Percentile	0.75
95th Percentile	1.80
99th Percentile	7.82
Percent with less than zero	57.01%
Overall Mean	0.27
Standard Error of the Overall Mean	0.05
t-Statistic	5.74
No. of Funds	5974

Table 4: **Value Added ($\hat{S}_i$):** For every fund in our database, we estimate the monthly value added, $\hat{S}_i$. The *Cross-Sectional* mean, standard error, t-statistic and percentiles are the statistical properties of this distribution. *Percent with less than zero* is the fraction of the distribution that has value added estimates less than zero. The *Overall* mean, standard error and t-statistic are computed by computing the average value added in the dataset. The numbers are reported in Y2000 $ millions per month.

time τ, we construct the t-statistic of the estimate of value added, what we term the *Skill Ratio*, defined as:

$$SKR_i^\tau \equiv \frac{\hat{S}_i^\tau}{\sigma(\hat{S}_i^\tau)}, \tag{14}$$

where $\hat{S}_i^\tau = \sum_{t=1}^{\tau} \frac{V_{it}}{\tau}$ and $\sigma(\hat{S}_i^\tau) = \frac{\sqrt{\sum_{t=1}^{\tau}(V_{it}-\hat{S}_i^\tau)^2}}{\tau}$. As before, funds in the 10th (top) decile are the funds where we have the most confidence that the actual value added over the sorting period is positive. Similarly, funds in the 1st (bottom) decile are funds where we have the most confidence that the actual value added in the sorting period is negative. We then calculate the average value added for the funds in each decile over a specified future horizon. At the end of the horizon the procedure is repeated until the end of the dataset. The procedure is run for time horizons of 3-10 years using the same procedure to correct for the selection bias induced by dropping the first 18 observations.

Figure 8 plots the mean as well as the two standard error bounds for each decile for each time horizon. From Figure 8 it appears that there is evidence of persistence as far out as 10 years. The point estimate of the average value added of 10th decile managers is positive at every horizon and is always the best performing decile. The value added estimates are

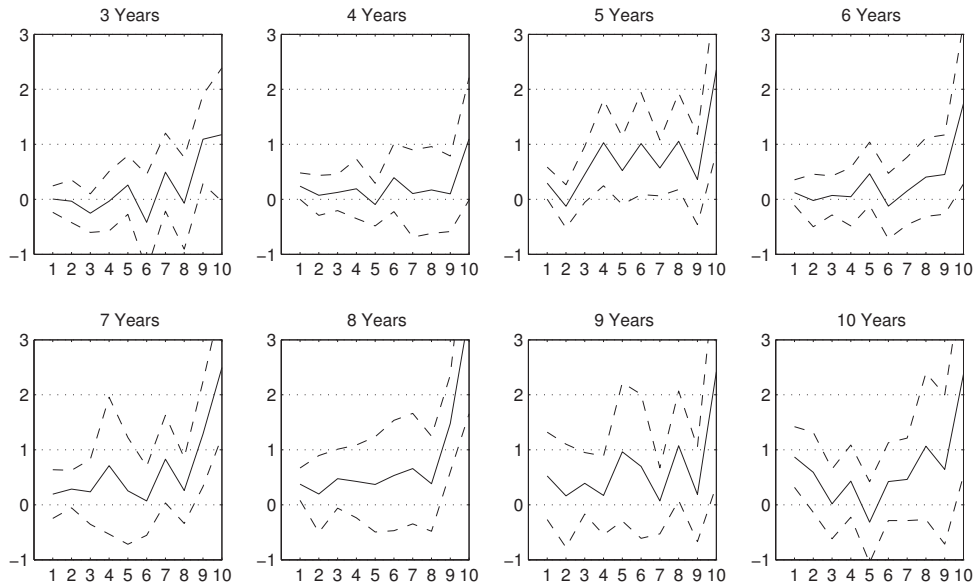

Figure 8: **Out-of-Sample Value Added**
Each graph displays average out-of-sample value added, $\hat{S}_i$ (in Y2000 $ million/month), of funds sorted into deciles on the Skill Ratio, over the future horizon indicated. The solid line indicates the performance of each decile and the dashed lines indicated the two standard error bounds.

economically large. Although clearly noisy, the average tenth decile manager adds around $2 million/month.

Next we report formal tests of the hypothesis that skill is not persistent. If skill is not persistent, managers that have added the most value in the past should not continue to add the most value in the future. Hence, we can use relative performance comparisons to construct a test of persistence by counting the number of times in the future (1) top managers beat bottom managers, and (2) top managers are in the top half.[5] The distribution, under the Null that there is no persistence, of both of these test statistics is Binomial(n,$\frac{1}{2}$), where n is the number of future monthly observations.[6] 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 is evident from Table 5, the Null Hypothesis that skill is not persistent can be rejected at the 95% confidence level at almost all horizons. Finally, note from the final column of Table 5 the disproportionate share of capital controlled by 10th decile managers. Investors

[5]Because the volatility of funds varies in the cross-section, we restrict attention to tests where the probability under the Null Hypothesis is not a function of fund volatility.

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

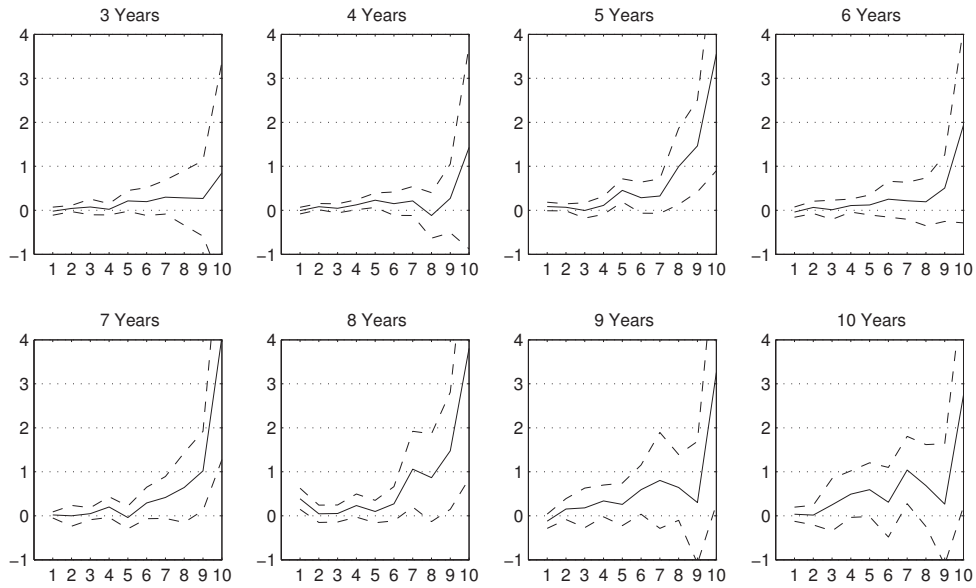

Figure 9: **Value Added Sorted on Compensation**
Each graph displays average out-of-sample value added, $\hat{S}_i$ (in Y2000 $ million/month), of funds sorted into deciles based on total compensation (fees × AUM). The solid line indicates the performance of each decile and the dashed lines indicated the 95% confidence bands (two standard errors from the estimate).

reward skilled managers by providing them with more capital. It might be tempting, based on our sorts, to conclude that all the skill is concentrated with a small fraction (10%) of managers. Two important cautions are in order. First, these 10% of managers control 25% of the total money invested. Second, our sorts are unlikely to separate skill perfectly. Although the estimates of value added in the other deciles are not significantly different from zero, they are almost all positive. Since we know that many managers destroyed value over the sample period, these positive point estimates imply that enough skilled managers are distributed throughout the other deciles to overcome the significant fraction of managers that destroy value.

The next figure demonstrates just how competitive mutual fund markets are. It repeats the above analysis except the sort is done using current compensation rather than the skill ratio. That is, rather than ranking funds by the skill ratio, funds are ranked by the managerial compensation at that time — the current size of the fund multiplied by the fee charged. The results of this procedure are reported in Figure 9. By comparing this figure to the prior one, one can see that sorting by compensation better predicts future performance. Although the skill ratio does a good job identifying the highest skilled managers, it does less well differentiating mediocre managers. Current compensation is better able to differentiate all

Horizon	Value Added		Top Outperforms Bottom		Top in Top Half		Fraction of Total
Years	$ Mil	p-value (%)	Freq. (%)	p-value (%)	Freq. (%)	p-value (%)	**AUM** (%)
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

Table 5: **Out-of-sample Performance of the Top Decile:** The two columns labeled "Value Added" report the average value added of the top decile at each horizon and the associated p-value. The next two columns report the fraction of the time and associate p-value the top decile has a higher value added realization than the bottom decile. The columns labeled "Top in Top Half" report the fraction of time the realized value added of the top decile is in the top half, and the final column reports the average fraction of total AUM in the top decile. All p-values are one tailed, that is, they represent the probability, under the Null Hypothesis, of the observed test-statistic value or greater.

managers. Because investors determine compensation (by determining the size of the fund) these results indicate that investors reward managers who are more skilled based on future performance with higher compensation today. That means that investors are able to identify better managers *ex ante*. Indeed, investors appear to use more information to make this decision than what is contained in the skill ratio.

There is also increased monotonicity when the sorts are based on compensation rather than on the Skill Ratio. To formally document this difference, we count the number of times each decile outperforms the next lowest decile (in terms of value added). Table 6 reports the p-value of observing the reported numbers under the Null Hypothesis that there is no skill (so the probability is 1/2). The table confirms what the figures imply. While the Skill Ratio can identify extreme performers, it does not differentiate other funds very well. In contrast, investors appear to do a much better job correctly differentiating all funds.

For many years now, researchers have characterized the behavior of investors in the mutual fund sector as suboptimal, that is, dumb investors chasing past returns. Our evidence relating compensation to future performance suggests quite the opposite. Investors appear able to differentiate good managers from bad and compensate them accordingly by directing capital to the better managers. Ultimately, investors themselves do not derive any benefit from recognizing this skill because the intense competition between them drives the net alpha of all funds to zero.

The overall conclusion is that the evidence is remarkably consistent with the rational ex-

Horizon	Skill Ratio	Compensation
3	18.55	9.17
4	3.51	1.87
5	18.61	0.02
6	4.57	9.23
7	15.92	5.85
8	4.16	3.79
9	28.61	14.71
10	53.02	0.25

Table 6: **Out-of-sample Monotonicity:** At each horizon, we calculate the number of times each decile outperforms the next lowest decile. The table shows the p-value (in percent) of the observed frequency under the Null Hypothesis that skill does not exist, i.e., that for a sample length of N months, the probability of the event is Binomial$(9N, 1/2)$.

pectation equilibrium paradigm. Markets appear to be highly competitive, so that investors are not able to earn abnormal returns. By the same token, funds do not underperform on average, so investors appear to rationally allocate the right amount of capital to active management. Skilled managers do exist, but there is large heterogeneity of skill. Investors recognize skill and thus direct capital towards the most skilled managers. Consequently, the most skilled managers manage the largest funds, and manager compensation, which is primarily determined by fund size, predicts future managerial dollar performance.

3.3. The Role of Mutual Fund Firms

An interesting question in money management is what role mutual fund firms play. Because investors invest directly in funds, why do mutual fund firms exist? Berk, van Binsbergen, and Liu (2017) argue that firms perform an important role intermediating between investors and managers. They argue, based on their empirical evidence, that executives in mutual fund firms are better informed about managerial skill than investors and they use this information to efficiently match managers to capital.

Berk, van Binsbergen, and Liu (2017) find that a decision to increase a portfolio manager's assets under management (AUM) leads to an increase in the manager's productivity as measured by value added. Similarly, they find that decisions to reduce managers' responsibilities by taking away assets also lead to increases in subsequent value added. They find that the decision to reallocate capital to a manager adds at least $474,000 per manager per month. By comparing this lower bound to the total value added in the industry, they find that the firm is responsible for at least 30% of the total value added of the average manager. Mutual fund firms appear to add substantial value by intermediating between investors and managers and thereby efficiently matching capital to skill.

Berk, van Binsbergen, and Liu (2017) provide supporting evidence in favor of their hypothesis that this value added derives from an informational advantage. First, external hires that involve a change in AUM do not lead to a detectable change in future value added, Second, past performance does not have much power to explain firm capital reallocation decisions suggesting that the firm's decision to reassign capital is based on factors not captured by past performance measures. Third, investors appear to be aware of this informational advantage because they respond to firms' capital reallocation decisions by investing additional capital in the firm's funds. These facts are consistent with the hypothesis, first theorized by Alchian and Demsetz (1972), that firm executives use factors that are not easily observable to people outside the firm to make personnel decisions. They are also consistent with the hypothesis in Gennaioli, Shleifer, and Vishny (2015) that investors trust mutual fund firms to make sound financial decisions on their behalf, and so when they see these firms making a managerial change they respond by investing additional capital.

The findings in Berk, van Binsbergen, and Liu (2017) suggest that mutual fund firms have private information about their managers that investors do not have. Firms use that private information to improve upon investors' capital allocations. Interestingly, the paper finds that the value added of managers goes up after a demotion, suggesting that mutual fund executives have a better knowledge of their manager's ability than the managers themselves. Finally, firms' private information might also help explain the finding that compensation better predicts future performance than the skill ratio. By intermediating between investors and managers, firms are able to use their private information to improve capital allocation. Investors, recognizing this skill, invest more money in the firm's funds (something that is also documented in the paper) thereby making fund size (and thus dollar fees) a better predictor of future performance than the information in past returns.

3.4. Compensation Contract

Viewed from a high level, the trend from direct investing to indirect investing is fundamentally a change in the compensation contract under which intermediaries work, more than a change in how investors invest. In 1945, investors hired a broker who executed trades and provided investment advice. Because most investors lacked any investment skill, they relied on such advice, and thus, one could view the broker as effectively managing the investors' portfolio. Viewed in that light, there is not much difference in the role of the intermediary today and in 1945. Instead of a broker managing an investor's equity investments, today a money manager or financial advisor performs the same role. What has fundamentally changed is how these intermediaries are compensated. In 1945 they were paid as a function

of the number of trades they executed. Today they are largely paid as a function of the amount of assets under management (and in some cases they also have a performance based component). We view this change as beneficial because, as we will argue, it more closely aligns incentives.

It is hard to understand how a compensation contract that is based on the number of trades could be optimal. Because trading incurs costs, from a manager whose objective is to maximize the amount of money she can extract from markets, the amount of trading should be minimized. But when such a manager is compensated in the number of trades, the compensation contract induces a conflict between maximizing the value she can add and her own compensation. In such an equilibrium an optimizing manager will trade too much, reducing the total value added. In a fully competitive market these costs will be borne by the manager, and so we would expect better managers to eschew such contacts. This observation is likely one reason for why the sector has moved away from such contracts.

In contrast, Berk and Green (2004) demonstrate the surprising result that a compensation contract that rewards managers as a function of AUM is optimal. At first glance, one would expect an optimal contract to depend on the manager's performance. But in the rational expectations equilibrium net alphas are zero. On average, regardless of their skill level, managers are expected to deliver the same abnormal return to investors. In contrast, compensating based on fund size does compensate for performance because investors react to returns by investing funds to ensure that net alpha is zero. Thus, compensating a manager based on fund size implies that that compensation will be a function of the market's perception of her skill level.

The crucial assumption in Berk and Green (2004) that delivers the above result is that managers are no better informed about their own ability as investors. There is evidence in Berk, van Binsbergen, and Liu (2017) that supports this assumption — that paper shows that when a manager is demoted (the firm lowers her AUM), her value added goes up. That means the manager must have been actively managing too much money. Since she could have chosen to index this capital, this result is consistent with the assumption that she does not know her own ability better than investors. When the assumption is not true it is unlikely that an contract that rewards managers in only AUM will be optimal. The reason is that a manager who is aware that she has more skill than the market is giving her credit for, will desire a performance-based contract. That is, she will prefer a contract that is at least partly a function of how she performs. Such a contract is ubiquitous in hedge funds and private equity, and we return to this issue below.

3.5. Index Funds

An important trend that is not visible in Figure 1 is the recent rise in index and/or passive investing. While index investing was close to non-existent 70 years ago, today the fraction of assets managed by open-end funds that designate themselves as index funds exceeds 20%. While this trend is notable in its magnitude, caution is in order when interpreting it. The lines between active and passive management have gradually blurred. For example, there are index funds that offer value or growth strategies at low fees. Take, for example, the Vanguard Value Index fund. Vanguard writes about this fund: "This fund invests in stocks of large U.S. companies in market sectors that tend to grow at a slower pace than the broad market; these stocks may be temporarily undervalued by investors." Given that such funds offer diversified strategies in specific sectors, investing in such a fund is not necessarily a passive strategy (especially if investors actively switch between such funds). Furthermore, as these strategies are based on sorting criteria such as the book-to-market ratios of the underlying stocks that change over time, substantial turnover is still required for such strategies. It is thus not obvious that these strategies should be classified as passive buy-and-hold strategies, what the profession usually associates with index investing. These days, Standard and Poor's (S&P) roudly advertises that they keep track of over 170,000 different indices, the vast majority of which are not proxies for or representative of the aggregate market portfolio. Because the costs to implement these strategies vary widely depending on the strategy, one should not expect the value different index funds add to be the same.

Even the funds that closely replicate a market index, such as the S&P 500, have important differences between them. For example, some index funds hold more cash than others to accommodate in and outflows, some S&P 500 index funds do not hold the full set of stocks in their portfolio to minimize on trading costs, and some funds engage in securities lending while others do not. In other words, even S&P 500 index funds are far from a homogeneous product.

An important issue to keep in mind for all mutual funds, but particularly for index funds, is that different investors in the same fund do not necessarily earn the same net return because all investors do not necessarily pay the same fee. Consider Vanguard's S&P 500 index fund, which features two classes, the so-called "Investor" class and the so-called "Admiral" class. There are substantial differences in fees between these two products. The Admiral class only charges 5bp per year, whereas the Investor class charges 16bp, a difference of 11bp. There are, however, good reasons for these differences to exist. The Admiral class is only available to investors who invest an amount bigger or equal than $10,000. Because there are fixed costs to running an account, it is not surprising that percentage fees are higher for

smaller accounts. It is hard to believe that Vanguard can actually cover the fixed expenses associated with things like customer service for an investor who merely invests $1000 and thus pays $1.60 in fees annually.

Berk and van Binsbergen (2015) show that about half the value added of active funds is attributable to diversification services. This implies that the value added of funds that just provide these services is significant. Fund size adjusts in equilibrium to the level of the fees. Because index funds charge low fees, we should expect these funds to be large, and thus we should expect to see these funds make up a large fraction of the market capitalization of mutual fund investments.

3.6. Other Sectors of the Money Management Industry

Thus far we have concentrated on the mutual fund sector because of the availability of high quality data. The other important sectors of the money management industry include hedge funds, private equity funds and venture capital funds. All investors in these sectors must satisfy the requirement to be classed as "qualified" investors. Because these requirements impose high net wealth and income constraints, investors in these sectors are either rich individuals or institutions. Given these facts, we think a legitimate question that any policy maker should answer before imposing any regulation on these sectors is why these investors should be protected. We do not take a stand on this issue, other than emphasizing the importance of answering the question before any policy is put in place.

Almost all the datasets researchers use to analyze these sectors suffer from one of two drawbacks. Either the data set is made up of data self reported by the management company, or it comes from investors investing in that management company.[7] In the former case there are serious reliability concerns. Bad performing funds might choose not to report at all, funds might time their reporting based on their performance and funds might choose to report some funds (the successful ones) and not others (the unsuccessful ones). A common problem in early mutual fund studies is that companies would seed funds, see how they do, and then report the results for only the successful funds (quietly shutting down the others). Today, this problem is largely solved in the mutual fund space because researchers have been careful to make sure that all funds that represented an investable strategy are included in the database. However, there is no equivalent process in other sectors of the money management industry.

Databases that have been put together based on information from investors do not suffer from these biases because the investor received the data by making investments in the funds.

[7] We need references here

However, because no one investor can invest in all funds in the sector, these databases represent a subset of the data. It is also likely that the subset contains selection biases. Clearly, ex-post successful investors are more likely to part with their data, which implies that the data will contain a bias in favor of ex-post successful funds.

Another important limitation is that even if the returns to investors are deemed of sufficient quality, no reliable data on gross returns are available. The reason for this is that both the fees and performance component are often negotiated per client. Even though most funds report that they charge a 2 and 20 fee schedule, many investors do not in fact pay this amount. This complicates the computation of returns and as a consequence our value added measure. It is therefore hard to assess the level and cross-sectional differences in skill across managers in this sector.

Finally, a last important difference is that hedge funds are not restricted by regulation to lever their investment strategies. This allows them to take aggressive bets even with little AUM. Mutual funds on the other hand usually do not take leverage, although a type of specialty mutual fund called a 30/130 fund is gaining popularity. Such funds go short 30% and extra long 130% in the strategy the fund is implementing.

Next we discuss how well the framework we have laid out so far can be applied to hedge funds. Overall, the literature has found that hedge fund performance to investors is similar or somewhat better than that of mutual funds (see Agarwal, Mullally, and Naik (2015) for an overview of the literature). Given the data selection issues discussed above, better performance should not be unexpected. Whether this actually translates to better risk adjusted returns to investors is not clear.

In other dimensions, the literature on hedge funds also finds results that are consistent with the framework we have discussed in this paper. Fung, Hsieh, Naik, and Ramadorai (2008) study funds-of-funds of hedge funds and find that alpha producing funds-of-funds experience far greater and steadier capital inflows than their less fortunate counterparts. In turn, these capital inflows adversely affect their ability to produce alpha in the future. These findings are in line with the rational expectations framework and decreasing returns to scale discussed above. Lim, Sensoy, and Weisbach (2015) find that younger and more scalable hedge funds have stronger flow-performance relations. This is also fully consistent with a world where rational investors learn about the ability of hedge funds over time. As the speed of learning slows with the age of the fund, so does the flow-performance relationship (Berk and Green 2004). Furthermore, if hedge funds employ highly scalable strategies, then a given outperformance warrants a larger adjustment to the size of the fund, relative to a strategy that is less scalable. Fung and Hsieh (1997) find that hedge funds follow strategies that

are dramatically different from mutual funds, and support the claim that these strategies are highly dynamic, suggesting that hedge funds are more active than mutual funds. This could imply that the value added of hedge funds is larger than that of mutual funds, though measurement of value added is complicated by a lack of fee data, as argued above.

The one question that the special study proposal raises that remains largely unanswered is the extent to which differences in the fee structure of mutual funds and hedge funds are justified by the differences in those vehicles. The existence of a different contract in the other sectors points towards examining the assumptions that underly the optimality of the mutual fund contract. In our opinion, the assumption in Berk and Green (2004) that is most likely violated in the other sectors is the assumption of symmetric information between investors and managers. That is, managers have as much information about their own ability as investors. When managers know more about their own ability than investors, they have an incentive, using the contract, to signal their ability. Because lesser ability managers can always mimic the contract of better managers, the resulting equilibrium will likely feature pooling, which limits the ability of better quality managers to seperate from lower quality managers. In such an equilibrium it therefore becomes optimal to offer a contract that rewards outperformance. That is, although the contract is the same across all managers, better managers are paid more because they do better. Viewed from this perspective, the fact that the contract in the other sectors is performance-based is evidence that managers in these sectors likely know their own abilities better. However, why the contract would feature an asymmetric payoff is unclear. Further research is required before any policy recommendations can be made.

4. Broker Dealers

As we have already discussed, the role of broker-dealers has changed dramatically in the last 50 years. The rise of discount brokers and delegated money management has meant that broker-dealers pay a much less important role as investment advisors. Today, their primary responsibility is to intermediate trading, rather than also provide investment advice. We view this development as positive, because it is hard to see, given the compensation contract, how broker-dealers could avoid the conflict of interest that incentivizes them to trade too much.

With the declining role of broker-dealers as investment advisors, the question of whether they should be subject to a fiduciary standard is less pressing. With that said, there is very little evidence on whether such a standard would be beneficial. At first glance, imposing such a standard would seem to be very low cost and since one would expect that customers benefit when broker-dealers act in their interests, the argument not to impose the standard

appears week. But, in reality, there is very little empirical or theoretical work that provides much insight beyond this observation. The work that does exist does not provide support for imposing this standard. Egan, Matvos, and Seru (2016) find no evidence of a difference in misconduct violations between broker-dealers and financial advisors, even though financial advisors are subject to a fiduciary standard already.

We can think of two reasons to proceed with caution. First, requiring somebody to act as a fiduciary does not mean they will in fact act that way. As we have already mentioned, the compensation contract that compensates dealers in the number of trades sets up a conflict of interest that we believe is more likely to influence behavior than a law imposing a fiduciary standard. Broker-dealers are also subject to other incentives that conflict with many of their clients. Second, imposing such a standard could be detrimental if it leads clients to believe that their brokers are representing their interests when in reality the conflict of interest means that they are not. One could argue that setting a "buyer beware" standard might better serve client interests. In summary, given the lack of evidence and uncertainty on whether a fiduciary standard would be beneficial, we think further investigation is needed before any policy determination can be made.

Unfortunately, we can provide no insight on the question of whether the relationship between a broker dealer and its customer is competitive or monopolistic.

5. Policy

This report suggests a number of topics that policy makers might consider. First we discuss issues related to regulations on fees and fund size. Second, we discuss the need for better quality data sources for the non-mutual fund sectors of money management.

5.1. Regulation on Fees and Fund Size

All of our conclusions rely on the assumption that the rational expectations equilibrium closely approximates the equilibrium in money management. We believe we have presented convincing evidence suggesting that this is indeed the case. In the case that a policy maker finds this assumption objectionable, we simply point out that such a position be consistent. That is, if one is to take the position that the rational expectations equilibrium does not describe the equilibrium in money management, then we believe one cannot also maintain the position that the rational expectation equilibrium does closely approximate the equilibrium in the stock market.

If indeed the rational expectations paradigm is an accurate description of the equilibrium in the money management industry, then the following statements are true:

1. Regulating the percentage fees charged by funds does not change the surplus (or absence thereof) that consumers extract from investing with those funds.

2. Regulating the fund size without regulating fees does not change the surplus (or absence thereof) that consumers extract from investing with those funds.

3. When fees and fund size are jointly regulated consumer surplus can be affected.

Perhaps the first question that should be answered is under what circumstances it is desirable to increase consumer surplus. However, even if we take as given that consumer surplus should be increased, our framework shows that regulating fees without regulating fund size is ineffective. As we showed in Section 3.1, if a regulation is imposed that puts and upper limit on the fee, fund size simply adjusts to the new level of the fee, once again driving the consumer surplus to zero.

What the framework shows is that it is not the level of fees that sets the return to investors equal to zero. It is competition between investors for good investment opportunities. This also means that managerial compensation (aggregate fees) is determined in equilibrium by this same competition, not by managers trying to fleece their investors. Importantly, even in the case when consumer surplus to investors is negative, this unfortunate state of the world does not derive from managers fleecing their investors. Instead it derives from decisions investors themselves make — investors are investing too much money with active managers. In this case regulating fees is unlikely to address the problem. A better approach might be to educate investors.

In summary, the only way to change managerial compensation through regulation would be to limit competition between investors in some way. One obvious method would be to limit the size of funds based on the fee they charge. Leaving aside that such regulation would favor existing investors over new investors, it is not obvious whether such interference in resource allocation is desirable from an economy wide perspective. Given the restrictions that regulation imposes on the leverage that mutual funds can take, the relation between size and value added plotted in Figure 5 shows that if size is restricted below the optimal size, the manager will not be able to extract the optimal amount of money from financial markets. That will likely negatively impact the informativeness of market prices. So it is not clear whether such policies are desirable. We leave this question for future research.

5.2. Data

As we have argued above, the current data sources for hedge funds, private equity and venture capital have several important limitations. Policy makers should consider implementing regulations to improve the quality of this data. Specifically, policy makers should consider requiring funds to report returns-to-investors, fees charged and fund size.

References

Agarwal, V., K. Mullally, and N. Y. Naik (2015): "Hedge Funds: A Survey of the Academic Literature," .

Alchian, A. A., and H. Demsetz (1972): "Production, Information Costs, and Economic Organization," *The American Economiic Review*, 62(5), 777–795.

Berk, J., and J. H. van Binsbergen (2017): "Mutual Funds in Equilibrium," *Annual Review of Financial Economics*, Forthcoming.

Berk, J. B., and R. C. Green (2004): "Mutual Fund Flows and Performance in Rational Markets," *Journal of Political Economy*, 112(6), 1269–1295.

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

Berk, J. B., J. H. van Binsbergen, and B. Liu (2017): "Matching Capital and Labor," *Journal of Finance*, (Forthcoming).

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

Egan, M., G. Matvos, and A. Seru (2016): "The Market for Financial Adviser Misconduct," Working Paper 22050, National Bureau of Economic Research.

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

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

——— (1976): "Efficient Capital Markets: Reply," *Journal of Finance*, 31(1), 143–145.

Fama, E. F., and K. R. French (1996): "Multifactor Explanations of Asset Pricing Anomalies," *Journal of Finance*, 51, 55–87.

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

French, K. R. (2008): "The Cost of Active Investing," *Journal of Finance*, 63(4), 1537–1573.

Fung, W., and D. A. Hsieh (1997): "Empirical characteristics of dynamic trading strategies: The case of hedge funds," *Review of financial studies*, 10(2), 275–302.

Fung, W., D. A. Hsieh, N. Y. Naik, and T. Ramadorai (2008): "Hedge funds: Performance, risk, and capital formation," *The Journal of Finance*, 63(4), 1777–1803.

Gennaioli, N., A. Shleifer, and R. Vishny (2015): "Money Doctors," *Journal of Finance*, 70(1), 91–114.

Lim, J., B. A. Sensoy, and M. S. Weisbach (2015): "Indirect incentives of hedge fund managers," *The Journal of Finance*.

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

Muth, J. F. (1961): "Rational Expectations and the Theory of Price Movements," *Econometrica*, 29(3), 315–335.

Pastor, L., R. F. Stambaugh, and L. A. Taylor (2014): "Do funds make more when they trade more?," Discussion paper, National Bureau of Economic Research.

Pastor, L., R. F. Stambaugh, and L. A. Taylor (2015): "Scale and Skill in Active Management," *Journal of Financial Economics*, 116, 23–45.

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