New Special Study of the Securities Markets: Financial Intermediaries

Jonathan B. Berk* Jules H. van Binsbergen†

September 14, 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 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 a 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 an 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, 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

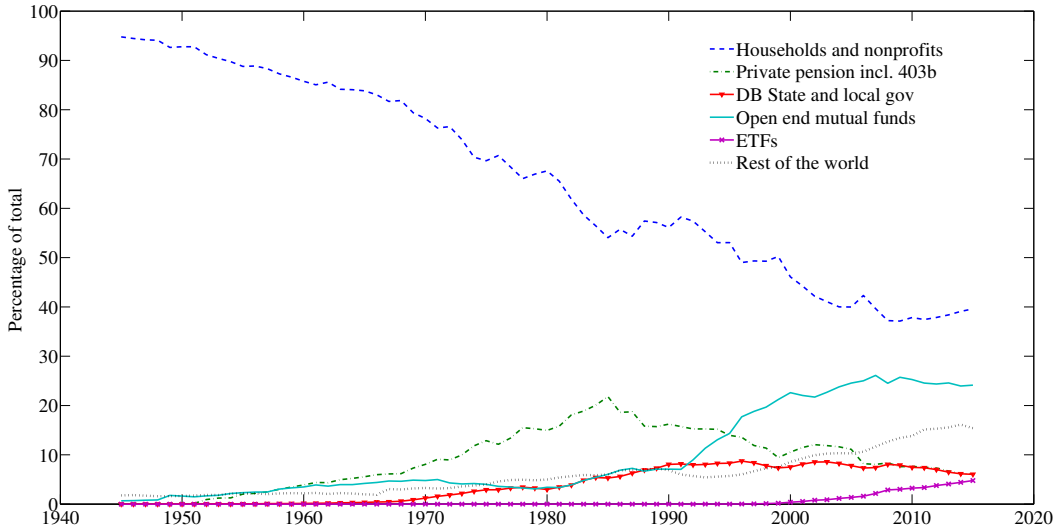

Figure 1: **Who Holds Corporate Equities?**
The graph shows the fraction of US corporate equity held by its largest investor groups.

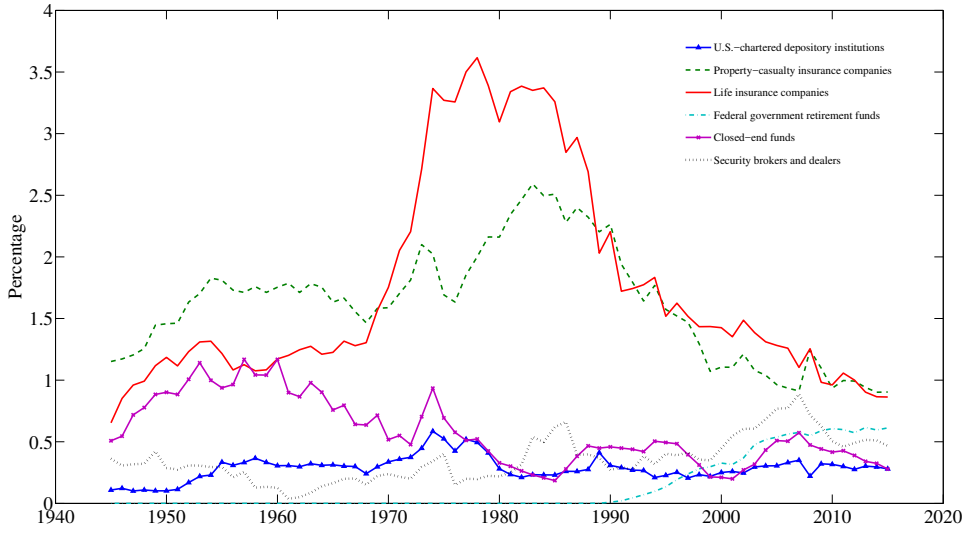

Figure 2: **Who Holds Corporate Equities? Continued**
The graph shows the fraction of US corporate equity held by various smaller investor groups.

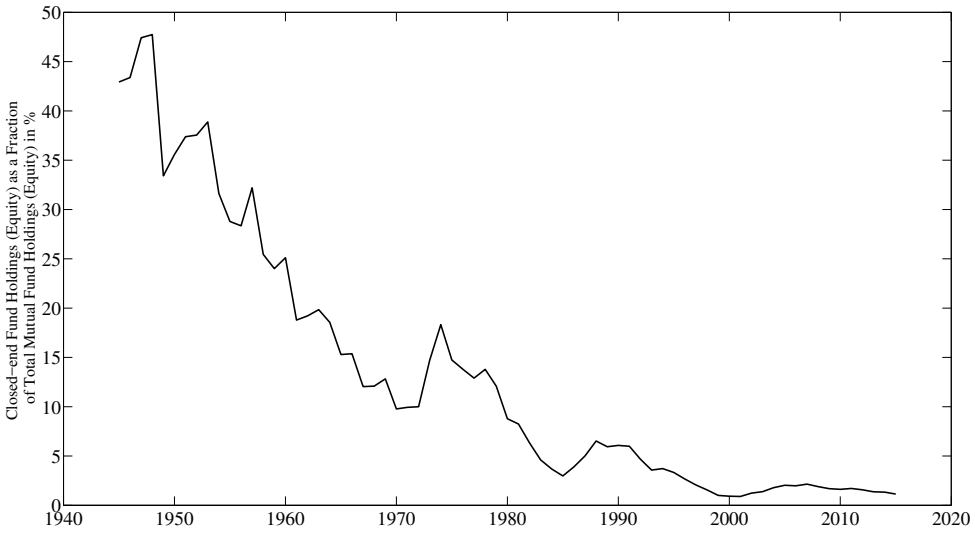

Figure 3: **Disappearing Closed End Funds**
The graph shows the value of equity holdings of closed-end funds as a fraction of the total equity holdings of open and closed-end mutual funds.

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 referred 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

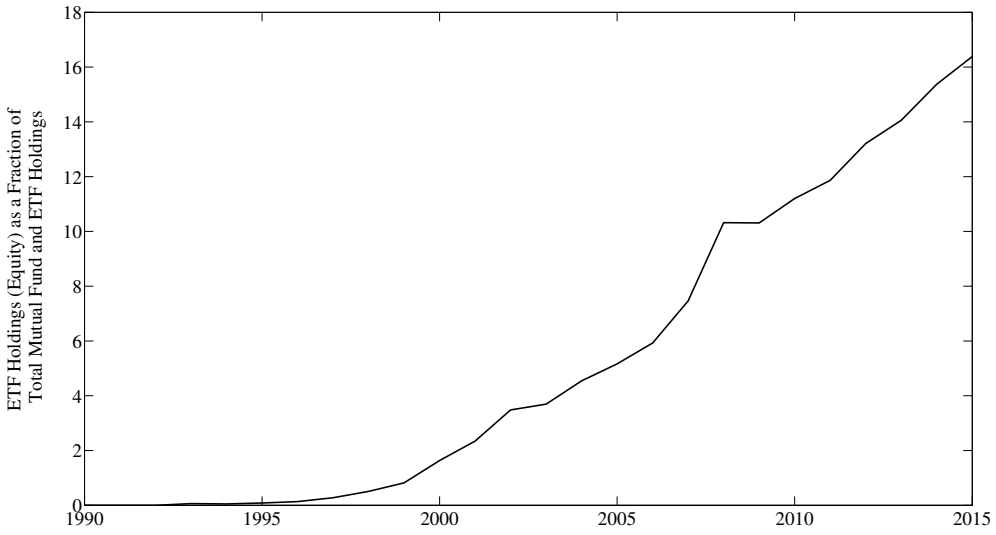

Figure 4: **The Rise of ETFs**
The graph shows the AUM invested with ETFs as a fraction of the total corporate equity holdings of open-end mutual funds, closed-end mutual funds and ETFs.

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. Competition and Rational Expectations

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. So far, financial economists have viewed investors who invest in stocks directly as fundamentally different from investors that invest in stocks through money managers. When investors invest in stocks directly, it is broadly accepted that the rational expectations equilibrium paradigm provides a very good description of how investors (and therefore prices) behave in practice. In fact, today, the rational expectations view is so common that when stock prices adjust in response to a piece of news, the change in the price in response to that news can be used as legal evidence of how valuable that piece of information is. On the other hand, when investors invest in stocks through money managers, a very common view is that they irrationally invest almost exclusively in investments that underperform their next best alternative ((Malkiel 1995, Carhart 1997, Fama and French

2010)). Further, because of this assumed investor irrationality returns are informative about the (lack of) investment skill of investment managers, as opposed to just risk. Also, because abnormal returns to investors have no persistence, financial economists labeled investors that acted upon outperformance as naive return chasers. This line of thinking took the lack of persistence as an exogenous fact unrelated to investor actions. The possibility that the flow performance relationship (the rational term for return chasing) is in fact what *causes* the lack of persistence, was never considered. What is arguably most surprising about this fundamentally different view of these two groups of investors, is the significant amount of overlap between the two groups. Unfortunately, this perspective of money management is still very pervasive today and continues to shape policymaking. As we illustrate in this paper, there is in fact little support for this view. We show that under the standard rational expectations assumption, many of the documented empirical findings can be explained. In fact, we show that the documented empirical patterns are exactly what we should expect to see in a competitive mutual fund market.

Applying the rational expectations equilibrium correctly goes a long way to better understanding the nature of competition in money management. Even though the paradigm was first presented in Muth (1961), it gained traction in finance in papers by Eugene Fama ((Fama 1965, Fama 1970, Fama 1976)) who labeled it the efficient market hypothesis. A key insight of the framework is that the expected return on a firm's stock is solely a reflection of the risk (appropriately measured) of that stock rather than of the quality of that firm's management. The high (low) quality of the firm's management is already reflected in the current high (low) stock (and bond) price of the firm, and thus leaves expected returns going forward unaffected. Put differently, firms with successful managers therefore already have a high market capitalization today, not a high expected return going forward.

Even though the literature has heavily debated whether prices incorporate *all* available information, there is little doubt that they reflect a large majority of it. Consequently, the cross-sectional distribution of firms' market capitalizations better measures the cross-sectional distribution of firm quality than the cross-sectional distribution of expected returns. As argued before, this idea is so widespread that when new information is released, it is common (including in our courts of a law) to measure the value of that new information by simply looking at the response of stock prices immediately upon the release of the information. The expected return subsequent to the release is never used for this purpose.

Surprisingly, in the money management literature the exact opposite way of thinking was widely adopted. Instead of focusing on the total value of the fund as being reflective of the skill of a mutual fund manager (the mutual fund counterpart to market capitalization)

the literature focused on return-based measures such as the abnormal return before or after management fees (the so-called gross and net alpha). By not appropriately applying the rational expectations framework, several important insights from that framework were missed. Changes in the size of mutual funds were seen as random and even irrational. However, once the rational expectations framework is applied correctly, it becomes clear that just as stock price changes happen as a rational response to new information regarding the quality of a firm, changes in total fund size happen in response to new information regarding the quality of a mutual fund manager. Perhaps even more importantly, the lack of an abnormal return to investors (a net alpha of zero) was erroneously seen as evidence of managers lacking skill, instead of what it really is: evidence of rational investors competing for managerial skill. Once the rational expectations framework is applied correctly, it becomes clear that the return to investors is unrelated to the skill of a manager in the same way that the expected return on a stock is unrelated to the quality of a company.

One may wonder why in two so closely related literatures two such different paradigms prevailed. One potential explanation is the way the original efficient markets papers were presented (Fama 1965, Malkiel 1995). The idea those papers put forward was that if stock prices reflect all available information, then no investor should be able to benefit from picking stocks. The fact that mutual fund managers deliver a zero net alpha to their investors was interpreted as evidence that not even people that are specialized in stock picking could pick stocks. This fact was seen as the ultimate evidence that stock markets are highly efficient. What this line of thinking misses however, is that it inconsistently applies the rational expectations framework. Once this discrepancy between the two literatures was put in place, it continued for several decades.

So how far does the analogy between stock markets and mutual funds go? The answer is very far. In both markets, investors compete with each other for positive net present value investment opportunities and by doing so eliminate them. For stocks, on seeing a mispricing investors compete to invest in the stock and through this competition drives stock prices to the right level. As a consequence, the expected return on the stock is solely driven by risk. For mutual funds, the mechanism is the same save for one difference: the price for a mutual fund is fixed, it is always equal to the value of the underlying securities. As a consequence, the market for mutual funds cannot equilibrate through prices. Instead, it equilibrates through quantities, that is, the AUM of the fund. The expected return to investors in both cases is only reflective of the risk of the investment and is unrelated to quality or skill.

Most readers will be familiar with how attractive investment opportunities in stock markets are rapidly competed away. If the price of the stock is too low relative to the expected

future cash flows, investors will all want to buy the stock, thereby increasing the price of the stock. This adjustment will stop when the price of the stock has risen so much that investors no longer view it as an attractive deal, and the price equals the present value of the cash flows. How does this work for mutual funds? We have argued that mutual funds also become less of an attractive investment opportunity as more and more investors compete for the skill of the manager. But why does a larger investor base decrease the attractiveness of the investment opportunity? The answer is decreasing returns to scale: as the fund size grows it becomes harder and harder for a mutual fund manager to find attractive investment opportunities for these new inflows. As the fund size (the assets-under-management) grows, the expected return on the fund will decrease. It will keep increasing until the abnormal expected return to investors (the net alpha) is 0. Similarly, if the net alpha is negative, funds will flow out, and the net alpha will increase until it is zero.

Applying the rational expectations framework correctly to mutual funds thus provides two important insights. The first insight is that the average abnormal return (or net alpha) that investors make by investing in a mutual fund does not teach us anything about the skill of the manager. Instead it teaches us something about the rationality of investors and/or the competition that they face. If net alphas are positive then the market for mutual funds is not very competitive as investors are leaving money on the table. If net alphas are negative, then investors are irrational as they are investing too much money with active managers and thus invest in negative net present value investment opportunities. The second insight is that just as the quality of a firm is reflected in the market capitalization of the firm, the skill of a mutual fund manager is reflected in the size of the fund that manager manages. If the fund size is large, the manager is highly skilled. If the fund size is small, the manager is much less skilled.

To further illustrate how rational expectations work in mutual funds, consider the following simple example based on Berk and Green (2004). Take a manager, let's call her manager 1, who can earn a 1.5% gross alpha (the abnormal return before fees are taken out) on a $5 billion fund. Because the manager does not have an infinite number of ideas and implements the ideas with the highest returns first, the fund's alpha deteriorates as the fund grows. Due to these decreasing returns to scale, the manager makes a 1% alpha when the fund's AUM is $10 billion and a 0.5% gross alpha when the fund size equals $15 billion. How much money does the manager extract from financial markets (what we term *value added*) for each of these three fund sizes? When the fund size is small and equal to $5 billion, the manager extracts $\$5 \times 0.015 = \$0.075$ billion, or $75 million. For the intermediate fund size, the value added equals $100 million (1% of $10 billion) and for the large fund size it equals $75 million (1.5% of $15 billion). These numbers are plotted in Figure 5, where the solid line represents

the value extracted from financial markets (or value added) and the dashed line represents the gross alpha.

Looking at the picture it is clear that the maximum value added occurs when the fund size equals $10 billion. For this size, the gross alpha is 1%. Before we turn to the problem

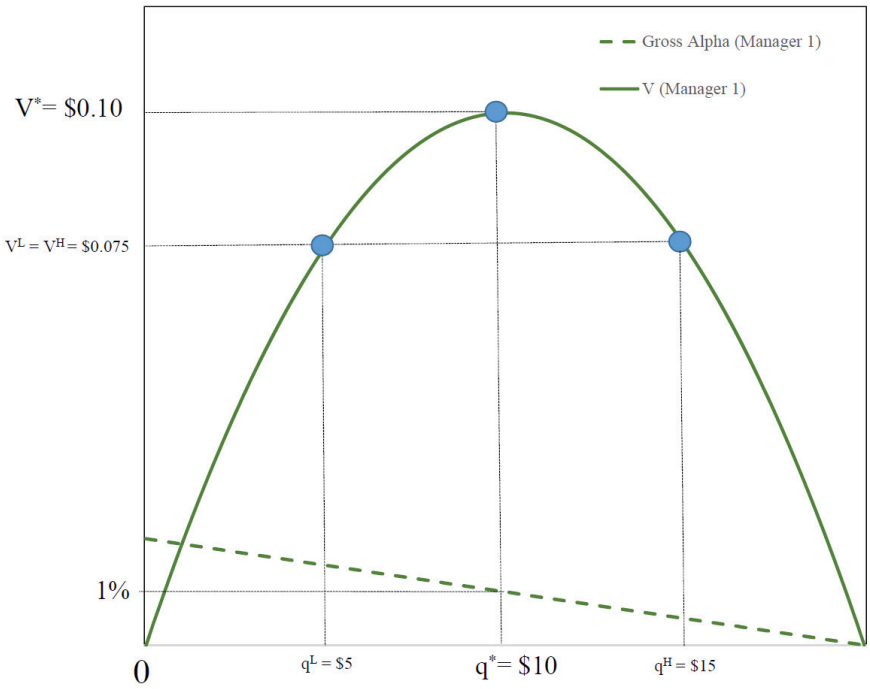

Figure 5: **Size, Value Added and Gross Alpha**
The graph shows the relationship between size and value added/gross alpha. This graph is an adjusted version of Figure 1 in Berk and van Binsbergen (2017).

that investors are solving, let us compare the manager above to another manager that is less talented. Manager 2 is also very good at delivering high gross alpha for small fund sizes, but is not nearly as talented in generating additional trading ideas as the fund size grows. That is, on a $5 billion fund manager 2 can only generate a 1% gross alpha, leading to a value added of $50 million. We plot the value added and the gross alpha of both managers in Figure 6. The graph shows that the gross alpha for both managers at the optimal amount of money is the same and equal to 1%. This implies that gross alpha is not a good measure of skill. After all, manager 2 is running out of ideas more quickly than manager 1, and is therefore less skilled. Where does this skill difference show up if not in gross alpha? It shows up in the amount of money the managers can handle, which is twice as large for manager 1 compared to manager 2. Because the gross alpha at the optimal amount of money is the same, but the optimal amount of money is twice as large for manager 1, this implies that manager 1's value added (the product of the size of the fund and the gross alpha) is twice

as large for manager 1. That is, manager 1 is able to extract twice as much money from financial markets compared to manager 2.

The conclusion that value added measures skill while gross alpha does not, follows only from the notion that both managers eventually run out of ideas, that is, they have decreasing returns to scale. Whether or not the rational expectations paradigm holds is irrelevant for this argument.

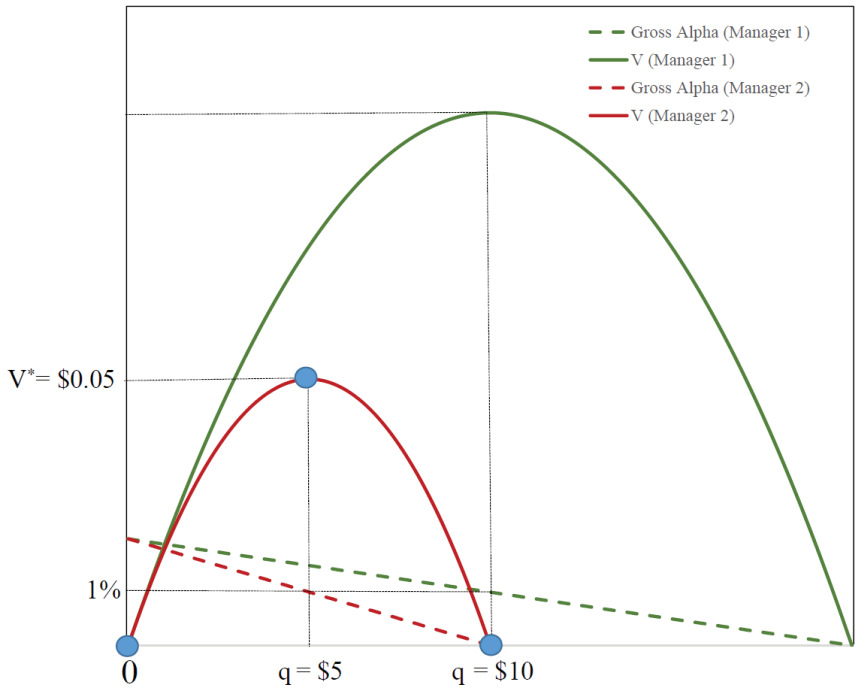

Figure 6: **Size, Value Added and Gross Alpha**
The graph shows the relationship between size and value added/gross alpha for two managers. Manager 1 is more skilled than Manager 2 while both make the same gross alpha on the first cent they invest. Manager 2 runs out of ideas more quickly than manager 1. That is, manager 2 is more affected by the decreasing returns to scale than manager 1 is. This graph is an adjusted version of Figure 2 in Berk and van Binsbergen (2017).

What about investors? Suppose that investors have rational expectations. This implies that if an attractive (positive NPV) investment opportunity presents itself, investors will pursue it until it is competed away. As a consequence, in equilibrium, the return that the investment delivers is solely a function of its risk. Put differently, investors in each fund must all earn a risk-adjusted expected return (or net alpha) of 0. If the risk-adjusted expected return is not 0, the equilibrium hasn't been reached yet. When the managers choose their percentage fee, f, to be equal to

$$f = 1\%,$$

investors invest $10 billion with manager 1 and $5 billion with manager 2. The net alpha, which is the difference between the gross alpha and the fee, will be zero for both managers and the mutual fund market equilibrates: investors have no incentive to either give money to or take money way from either fund. Although the net and gross alpha of the two managers is the same, the AUM of manager 1 is twice as large as that of manager 2, in equilibrium.

Before we proceed let us summarize the important insights that the rational equilibrium paradigm has delivered. First, because investors compete with each other for attractive investment opportunities, the net alphas are zero, always. That implies that net alphas cannot be informative on managerial ability as it does not differentiate across managers. Similarly, the gross alpha is also not informative. As argued above, the reason for why gross alpha, which is a return-based measure, does not measure managerial skill and value added does, is the same as the reason why present value measures dominate internal rate of return (IRR) measures when choosing between investment projects. Return-based measures simply fail to take into account the scale of the project. Only when there are constant returns to scale (an unreasonable assumption in investment management), can scale be ignored.

For ease of exposition, the framework we have presented is a static one where all players know the gross alpha that the manager delivers at each potential fund size. In reality investors (and managers) need to learn about this. This learning process implies that each time the manager does better than expected, investors will update their beliefs positively on the alpha the manager will deliver. As soon as investors update positively, there is a positive net alpha opportunity at the current fund size. As we have already seen, such an opportunity cannot survive in equilibrium. After all, as soon as the opportunity presents itself, investable money (that is in very large supply) will find its way there. The fund size will grow, and due to the decreasing returns to scale, the net alpha will be driven back down to zero. In summary, the fund size will change as the investors learn about the skill of the manager.

In addition to the fund size adjusting as investors learn, there is another potential equilibration mechanism: the fee the manager charges. If the manager increases (decreases) her percentage fee every time investors update positively (negatively) on the gross alpha she can deliver, the net alpha can be kept at zero while keeping the fund size constant over time. Even though this mechanism could also work, it is not the mechanism that we observe in the data: managers very rarely change their fees, implying that nearly all of the equilibration happens through the adjustment of the fund size.

So how can managers ensure that they extract the maximum amount of money from markets without changing their fee? From the point of view of the manager, the optimal

amount to invest in active management (i.e. where the value added graph peaks) remains the same regardless of the level of the fee. So what does the manager do if fees are too low and thus investors invest more than the optimal amount? The answer is he indexes the remaining money. This indexed money by definition earns a gross alpha of zero, and thus as the fraction of indexed money grows, the gross alpha of the fund as a whole decreases. Following the same logic as before, this process will continue until the gross alpha equals the fees and thus the net alpha is zero.

That brings us to the next important insight. The gross alpha and the size of the fund cannot be meaningfully interpreted in isolation and should be studied jointly. Let us illustrate this point with a simple example. Suppose there is an investment manager, call her manager A, whose optimal amount of money to manage is \$1 billion, at which point the gross alpha is 1%. So, at the peak of the value added graph, value added equals \$10 million. Suppose next that this manager is already managing the optimal amount of money and that the fee is 1%. The net alpha is 1%-1% = 0, and thus the fund is in equilibrium. Now consider another manager, manager B who is the twin sister of manager A. The only difference between her and her sister is that she chooses to charge a fee of 0.5%. At a fund size of \$1 billion, this would imply that the net alpha is 1%-0.5% = 0.5%. This is not an equilibrium. Investors want in on this opportunity and the fund size grows. Because the manager's optimal amount to actively invest is \$1 billion, the new inflows cannot be put to productive use. The manager therefore indexes the new inflows. When the fund size reaches \$2 billion, half of the money is indexed. Given that the indexed money does not earn a gross alpha (by definition), the gross alpha of the whole fund is 0.5 x 1% + 0.5 x 0% = 0.5%, the net alpha is zero, the fund is in equilibrium and like her twin sister her value added is \$10 million. Note that even though the managers are identical in their skill level, to a naive spectator the two managers may look very different: they have a different fund size and a different gross alpha. Only when we take the product of these two quantities does it become clear that we are dealing with two equally skilled managers.

This delivers yet another important insight: the percentage fee is irrelevant. Because the fund size adjusts to ensure that the net alpha is zero (i.e. the gross alpha is sufficiently high to cover the fees), it does not matter whether manager A and B choose a fee of 1% or 0.5%. With low fees, the fund is big and with high fees the fund is small but the money extracted from financial markets remains unchanged. It is often argued that the increase in AUM of mutual funds that charge a low fee is a result of the success of those funds in comparison with their high fee competitors. What the arguments above imply is that in equilibrium low-fee funds will automatically be larger, even if without any difference between the skill level of the managers and/or the net alpha that investors receive. Finally, because the fund

size adjusts in response to fee changes to ensure that the net alpha is always zero, regulating fees without also regulating the size of the fund is ineffective.

In the next section we will illustrate how well the framework presented above performs in the data. However, before we do so, we wish to address an often-heard argument for why the active mutual fund sector as a whole cannot add value. This argument is often referred to as Sharpe's arithmetic (Sharpe (1991)) and goes as follows. Suppose we split up the universe of investors in two groups, passive investors that invest in the market portfolio and all other investors which Sharpe labeled active investors. Because the first group earns the market return and because the two groups together must also hold the market, the active investors must by definition also earn the market return, and hence will not be able to beat it.

As it turns out, the reasoning above is flawed for two reasons. First, when defining "active" investors, Sharpe takes all investors that do not passively hold the market. In addition to active mutual funds, that also includes individual investors and investors in specialized index funds that do not exactly hold the market. So, if active mutual fund managers trade againt these other groups of investors, then active mutual funds can make trading profits at the expense of these investors. This does of course raise the question of why these investors are not indexing their money in a fund that holds the market.

What is perhaps more surprising is that the arguments in Sharpe (1991) leave open the possibility that active managers as a group can beat the market even if all investors are assumed to be fully rational. The reason 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 positive net alphas to investors going forward. The main roadblock to testing these two predictions is constructing an estimate of the funds' net alpha. Generally, financial economists have used two methods to convert the fund's returns into abnormal returns (net alphas) relative to the alternative investment opportunities investors have. The standard practice is not to construct the alternative investment opportunity itself, but rather to simply adjust for risk using a risk model. The problem with this approach is that the extent to which these risk models in fact appropriately correct for risk has been fiercely debated. As a consequence, researchers often choose to construct the alternative investment opportunity set after all. Even though, in principle, this addresses the problem of not knowing the appropriate risk model, the way researchers implement this 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 at least two arguments for why these often-used factor portfolios are not opportunities investors can actually invest in (Berk and van Binsbergen (2015)). First, the portfolios ignore transaction costs. The performance of a fund that incurs transaction costs cannot be compared to the performance of a theoretical alternative that does not. The second reason is more subtle and relates to the hindsight bias of the portfolios that are used. The typical factors researchers use were discovered in the late 1980's and 1990's and popularized by Fama and French (1996) and Carhart (1997). However, it is common to include data for these factors that start many years before their discovery date. If, in those earlier years, investors did not know about these portfolios, they do not represent a true alternative investment opportunity. By using these portfolios to benchmark managers, academics are essentially evaluating managers in 1970 using 1990's knowledge. Any manager who, in 1970, had discovered the trading strategies explored in these factors should be credited for this knowledge in the performance evaluation.

In summary, by evaluating fund performance against non-investable benchmarks, the academic literature has potentially biased the results against finding managerial skill. To assess the importance of this issue, we next evaluate the "performance" of the most commonly-used factor portfolios against the set of index funds offered by Vanguard. These index funds are by definition investable alternatives for investors. The reason why we choose the funds offered by the Vanguard company is that these index funds have the purpose of giving investors access to diversification at the lowest cost. Other often-used benchmarks provided by Morningstar, for example, do not share this objective. Moreover, not only is Vanguard the market leader, it is also the pioneer in the space of index investing. For example, the 11 funds listed in Table 1 span the set of all index funds offered by the firm between 1977 and 2011. In each case, the Vanguard fund was the first index fund to offer that particular strategy. As such, the introduction dates of these funds can be used to infer when these strategies became widely known to investors.

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). Source: Berk and van Binsbergen (2015), Table 1

*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 available index funds offered by Vanguard over the period 1977-2011.[1] The only portfolio with an insignificant positive alpha is the market. The alpha

[1]We start in 1977 because that was when Vanguard introduced its first index fund. Details of how the

	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. Source: Berk and van Binsbergen (2015), Table 2.

for all the other factors is positive and statistically significant. The numbers vary between 22 b.p. per month (for the size portfolio) and 70 b.p. (for momentum). As momentum is also the trading strategy with the highest transaction costs, it is perhaps not surprising that the "outperformance" measure is the largest. We can thus conclude that the factor mimicking portfolios represent a much better (theoretical) investment opportunity set than what was actually available to investors. We therefore argue that the correct way to benchmark mutual funds is to use the available Vanguard index funds.

In particular, to evaluate mutual fund performance, we construct each fund's benchmark as the closest portfolio spanned by the set of Vanguard index funds. Let R_t^j denote 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:

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

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.[2] As argued above, these benchmarks have two major advantages. First, they account for the industrial organization of the mutual fund sector through the dynamic discovery of various trading strategies. Second, the Vanguard returns are the actual returns that investors receive and thus are net of all transaction costs. We can thus be confident that these benchmarks represent actual investable alternative investment opportunities for investors.

Further, note that when we employ the benchmark above to measure the value added of one of the Vanguard index funds itself, it will be equal to the dollar fees that fund charges.

benchmarks are constructed can be found in Berk and van Binsbergen (2015).

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

The reason why Vanguard funds add value is that they give investors the lowest cost access to diversification services. Therefore, when we use as the benchmark the net returns of the Vanguard index funds, we explicitly account for the value added through such diversification services. Because active funds also provide diversification services, our measure credits them with this value added. We can also separate these diversification services from other skills by using gross returns on the Vanguard benchmarks.

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

$$\varepsilon_{it} \equiv R_{it}^n - R_{it}^B. \tag{2}$$

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

3.2.1. Net Alpha

Many researchers have argued that by investing in active mutual funds, investors underperform in the sense that their net alpha is negative (see Fama and French (2010)). Berk and van Binsbergen (2015) argue that this finding is largely driven by two very common empirical implementation choices. The first of these choices we have already discussed above. By using non-investable benchmarks, researchers have biased the performance measurement against finding skill. The second choice relates to sample selection. It has become common in mutual fund research to exclude from the sample mutual funds that hold foreign stocks. Further, most studies start the sample in the mid eighties while earlier data is readily available. Because of these two data sample restrictions researchers have dropped more than half of the observations. More importantly, the fraction of total mutual fund AUM that is in funds that exclusively hold U.S. stocks is strongly decreasing over our sample period and represents less than 25% of it by the end of our sample (in 2011). Put differently, academic researchers have focused on a fast-shrinking part of the industry. We can think of no reason for either of these two data selection choices.

In Table 3 we report both the equal-weighted and value-weighted alpha over our data sample. The table shows that they are not statistically different from zero.

Once we have corrected the literature's implementation choices by using the right data sample and by using the Vanguard benchmark, the numbers in Table 3 are consistent with the predictions of the rational expectations paradigm. Importantly, however, the rational expectations equilibrium has additional predictions. It implies that net alphas are not pre-

Equally Weighted	2.74
t-statistic	0.73
Value Weighted	-0.95
t-statistic	-0.31
Number of Funds	5974

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*). Source: Berk and van Binsbergen (2015), Table 6.

dictable. To test the validity of this prediction we sort firms into decile portfolios based on their historical net alpha and assess to what extent funds that have outperformed in the past will continue this outperformance in the future. (Berk and van Binsbergen 2017) show (see their Figure 5) that this indeed the case. There is no relation between past and future net alphas.

3.2.2. Skill

If active mutual funds have a net alpha of 0 and yet charge a fee, this must imply that they have skill. This skill is measured by the value added of the fund. So, the next question we need to address is how we estimate value added. To construct this measure, we first adjust the gross realized return of the fund by the realized return of the benchmark, $R^g_{it} - R^B_{it}$. 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:

$$V_{it} \equiv q_{i,t-1}\left(R^g_{it} - R^B_{it}\right). \tag{3}$$

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

As we have already argued, under the rational expectations paradigm, the only way to measure skill is value added. However, even when the rational expectations paradigm does not hold, this by no means implies that alpha measures can be used as a measure of skill. The observation that return measures do not appropriately adjust for scale holds whether or not the rational expectations paradigm is true. Further, value added always measures the amount of money extracted from markets. It is a consequence of the following simple adding-up constraint:

$$V_t = q_t \alpha^g_t(q_t) = q_t \alpha^n_t(q_t) + q_t f$$

where $\alpha_t^n(q_t)$ is the net alpha of the fund at time t as a function of the fund's 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 adding up constraint.

We begin by measuring the average value added of mutual fund managers over the period 1977-2011 in January 1, 2000 dollars.[3] The results in Table 4 show that mutual fund managers have skill. 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.

Well-performing funds have a higher likelihood of surviving compared to their less-performing counterparts. Therefore, first averaging by fund and then averaging across funds leads to estimates of what we call the *ex-ante* distribution of skill. If we compute an average without first averaging by fund, the estimate is different because surviving funds make up a larger part of the sample in this case (i.e. they are overrepresented). In that case, the resulting average is an estimate of the *ex-post* distribution of skill. It is the average skill level of the set of funds actually managing money. As expected, ex post mean is higher than the ex ante mean: the average fund added $270,000/month.

If managers are skilled, one would expect this skill to persist. Berk and van Binsbergen (2015) test for this persistence. In Figure 3 of that paper they demonstrate strong evidence of this persistence for horizons up to ten years. The paper shows (Table 4) that the Null Hypothesis that skill is not persistent can be rejected at the 95% confidence level at almost all horizons between 3 and 10 years. It also documents that managers in the top 10% control 25% of all invested capital implying that investors reward skilled managers by providing them with more capital.

Berk and van Binsbergen (2015) also demonstrate how competitive mutual fund markets

[3]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:** For every fund in our database, we estimate the monthly value added. 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 data set. The numbers are reported in Y2000 $ millions per month. This table is based on Table 3 in Berk and van Binsbergen (2015).

are. They show that if funds are ranked by the managerial compensation — the current size of the fund multiplied by the fee charged, performance is even more predictable. Because investors determine compensation (by determining the size of the fund) these results indicate that investors reward better managers with higher compensation. That means that investors are able to identify better managers *ex ante*. Investors appear to use more information to make this inference than what is contained in past returns.

In the past few decades financial economists have come to view mutual fund investors as naive, dumb and prone to the irrational "chasing" of past returns. The collective evidence we have presented here suggests quite the opposite. Investors use past returns to infer the skill of managers and rationally reallocate capital from bad managers to good managers. Because investors so fiercely compete with each other for skilled managers, they end up deriving no benefit from identifying this skill (i.e. the net alpha is zero), and the managers, because they have a skill in short supply, collect all the rents from their skill.

We can conclude that overall the data is consistent with the rational expectations framework. Markets are highly competitive and because investors do not bring anything to the table that is in short supply – after all, investable money is in very large supply – they do

not earn abnormal returns. Given that net alphas are not statistically significantly negative, there is also little evidence that too much capital is allocated to active managers. Our findings also suggest that there is very large cross-sectional variation in the level of skill. This cross-sectional variation can only be observed by using value added (fund size) to measure skill. Put differently, because the cross-sectional variation in fees is so low, the cross-sectional variation in gross alpha is low, and thus the large majority of the cross-sectional variation in skill is reflected in fund size, not gross alpha. Good managers manage large funds and bad managers manage (very) small funds. Because compensation is also primarily determined by fund size, good managers earn the highest compensation, which in turn is a good predictor of future dollar performance (value added).

3.3. Why Do Mutual Fund Firms Exist?

Next we address the question of whether the framework above is consistent with and/or can help shed light on the question of why mutual fund firms exist. If investors can allocate money to mutual funds (and thus managers) directly, is there a need for mutual fund firms to intermediate in this process? Berk, van Binsbergen, and Liu (2017) find that firms indeed play an important role in this intermediation process. The reason why there is room for intermediation is that executives in mutual fund firms seem better informed about managerial skill than investors are, and that they use this information to improve upon the capital allocation done by investors. Firms thus help investors allocate capital better.

More specifically, Berk, van Binsbergen, and Liu (2017) find that when a mutual fund firm decides to increase a manager's AUM by giving that manager an additional fund (a *promotion*), this increases that manager's value added. Similarly, the decision to take away a fund from a manager (a *demotion*) also leads to increases in subsequent value added. These capital reallocation decisions add at least $474,000 per manager per month compared to the counterfactual where managers would have kept their original capital allocation done by investors. This number represents about 30% of the total value that the industry adds.

Berk, van Binsbergen, and Liu (2017) further provide evidence that the improved capital allocation results from a unique informational advantage that insiders of the firm have relative to outsiders. If indeed only the insiders of the firm are privy to this informational advantage then capital reallocations that happen as a consequence of managers switching firms should not lead to increases in value added. This is indeed what Berk, van Binsbergen, and Liu (2017) find. Secondly, if promotions are based on private information of the firm, then they should not be predictable by variables easily observable to outsiders of the firm. After all, investors themselves can already observe those variables and thus can already adjust

the allocated capital accordingly. This is also confirmed in Berk, van Binsbergen, and Liu (2017). They find that past performance and past flows, which are both observable to outside investors, have very little predictive power for firms' promotion and demotion decisions. Finally, investors appear to be paying strong attention to these personnel decisions, as they are followed by inflows into the firm's funds. These inflows also allow the firm to capture the rents from these allocation decisions, as the total fee revenue goes up because of these additional invested funds.

One interesting finding in Berk, van Binsbergen, and Liu (2017) is that value added goes up after a demotion. As we have argued above, as long as the manager indexes all money above the optimal amount of capital, the value added should not change when the fund is larger than the optimal amount. One interpretation of this finding is therefore that mutual fund executives know the ability of fund managers better than they do themselves.

Finally, the idea that firm executives have unique information regarding the level of skill of managers can also shed light on the empirical finding that compensation better predicts future performance than past performance does. By successfully intermediating between mutual fund managers and the fund's investors, firms achieve a better capital allocation than investors would without such intermediation. As argued above, investors seem to recognize this advantage of the firm and invest more money in the firm's funds. Because of this improved capital allocation, fund size (and thereby fees in dollars) are a better predictor of future performance than the information in past performance.

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. 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. 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 suf-

ficient 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 separate 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.

3.7. Frictions

It is common in finance and economics to first study the properties of markets and equilibria in a frictionless setting, as this provides a useful benchmark. Once the frictionless benchmark is well understood, various frictions, if empirically relevant, can then be added to obtain a model that is closer to the data. A very illustrative example of such a framework is the seminal work by Modigliani and Miller, who evaluate capital structure and dividend decisions by firms. Even if such a frictionless framework ends up being far removed from the data, this by no means implies that the framework is not a useful benchmark to start with.

The important insight that our paper has delivered is that the neoclassical frictionless framework (with learning) actually provides a surprisingly accurate description of the behavior of the mutual fund market in the data. That is not to say that frictions could not be important, and could potentially make the fit of the model even better. One example of such a friction is taxes. Investors should care about after-tax returns, not pre-tax returns. As we do not observe the marginal tax rate of the marginal mutual fund investor, it is hard to adjust returns for taxes, but future work could focus on such adjustments. Other frictions that could be interesting to evaluate are settings where the adjustment in fund size is inhibited for one reason or another. If the size of the fund cannot freely adjust, this implies that the

equilibrating mechanism described in this paper cannot do its work. This could potentially change the predictions of the model and the excess rents that investors could earn.

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," .

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.

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.

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