

Mutual Fund Flows and Performance in Rational Markets¹

Jonathan B. Berk
Haas School of Business
University of California, Berkeley
and
NBER
Email: berk@haas.berkeley.edu

and

Richard C. Green
Graduate School of Industrial Administration
Carnegie Mellon University
Email: rcgreen@cmu.edu

October 15, 2002

¹We would like to thank Hsiu-lang Chen, Gian Luca Clementi, Jeff Coles, Burton Hollifield, Mike Lemmon, David Musto, George Pennacchi, Uzi Yoeli and Lu Zhang for their helpful comments and suggestions.

Abstract

Mutual Fund Flows and Performance in Rational Markets

We develop a simple rational model of active portfolio management that provides a natural benchmark against which to evaluate observed relationship between returns and fund flows. We show that many effects widely regarded as anomalous are consistent with this simple explanation. In the model, investments with active managers do not outperform passive benchmarks because of the competitive market for capital provision, combined with decreasing returns to scale in active portfolio management. Consequently, past performance cannot be used to predict future returns, or to infer the average skill level of active managers. The lack of persistence in active manager returns does not imply that differential ability across managers is nonexistent or unrewarded, that gathering information about performance is socially wasteful, or that chasing performance is pointless. A strong relationship between past performance and the flow of funds exists in our model, indeed this is the market mechanism that ensures that no predictability in performance exists. Calibrating the model to the fund flows and survivorship rates, we find these features of the data are consistent with the vast majority (80%) of active managers having at least enough skill to make back their fees.

1 Introduction

One of the central mysteries facing financial economics is why financial intermediaries appear to be so highly rewarded in our economy, despite the apparent fierce competition between them and the uncertainty about whether they add value through their activities. Research into mutual fund performance has provided evidence that deepens this puzzle. Since Jensen (1968), studies have shown little evidence that mutual fund managers outperform passive benchmarks.

More recently, attention has focused on the relationship between fund flows and performance, and the lack of persistence in performance. This work has produced several findings that researchers have viewed as puzzling. The relative performance of mutual fund managers appears to be largely unpredictable using past relative performance. While some controversial evidence of persistence does exist (see Gruber (1996), Carhart (1997), Zheng (1999) and Bollen and Busse (2001)) it is concentrated in low liquidity sectors or at shorter horizons. Yet, mutual fund investors appear to chase performance. Flows into and out of mutual funds are related to lagged measures of excess returns (see Chevalier and Ellison (1997) or Sirri and Tufano (1998)).

Together, these regularities raise questions about the rationality of investors who place money with active managers despite their apparent inability to outperform passive strategies. For example, Daniel *et al* (1997) claim that “When a professional portfolio manager proposes an investment strategy based on fundamental analysis of equities, the presumption is that he or she expects the strategy to outperform simpler, purely mechanical, strategies based on stock characteristics like book-to-market, size and momentum.” Investors appear to devote resources to evaluating past performance of managers and to choose portfolio managers on that basis, even though future performance appears to be unrelated to past performance. Thus, Bollen and Busse (2001, p. 1), argue, “The existence of the mutual fund selection industry is predicated on the assumption that some mutual fund managers possess significant ability, and that this ability persists, allowing the astute investor to predict future performance based on past results.” Gruber (1996, p. 784), in his Presidential Address, states, “If managerial ability exists and is not included in the price of open-end funds, then performance should be predictable.” Sirri and Tufano (1998) argue that since persistence in performance is confined to certain sectors, the response of flows to performance should also be strongest there.

The evidence that performance does not persist is widely regarded as implying that superior performance is attributable to luck rather than differential ability across managers. Indeed, various researchers have interpreted this fact as evidence for “market efficiency.”

For example Malkiel (1995, p. 571) concludes that the “study of mutual funds does not provide any reason to abandon a belief that securities markets are remarkably efficient.” In their widely used text book, Ross, Westerfield and Jaffe (2002, p. 353) explain that if the market is “efficient in the semistrong form” average active mutual fund returns should be the same as the market as whole. They then continue: “the finding that they [active fund managers] do not outperform the market indices is consistent with semistrong-form and weak-form efficiency.” The implication is that stocks are so well priced that not even the best active managers can consistently make money by using their superior ability. From an economic point of view, this implication, were it true, would be troubling. Clearly, the most talented active managers have abilities that are in short supply that could be productively deployed in other activities. One would think these managers would choose to apply their talents in areas where they would reap the highest rewards. If all performance is due to luck, there should be no reason to reward these talents. Yet, in reality, managers appear to reap rich rewards from superior past performance.

In the face of this evidence many researchers have concluded that a consistent explanation of these regularities is impossible without appealing to behavioral arguments that depend on irrationality, or to elaborate theories based on asymmetric information or moral hazard. Yet the one thing that has been missing from this debate is a clear delineation of what a fully rational model with no moral hazard or asymmetric information implies about the characteristics of active portfolio managers. Before appealing to these additional effects we believe it makes sense to first establish which behaviors in the data are qualitatively and quantitatively consistent with more direct explanations.

Our simple model of active portfolio management and fund flows provides a natural benchmark against which to evaluate observed returns, flows and performance outcomes in this important sector of the financial services industry. Using this model we show that the effects discussed above are generally consistent with a fully rational and competitive market for capital investment and rational, self-interested choices by fund managers who have differential ability to generate abnormal returns.

The paper’s main insights are as follows. The fact that investments with active managers do not outperform passive benchmarks is a consequence of the competitiveness in the market for capital investment. If investors compete with each other for superior returns, they end up ensuring that none exist. A consequence of this insight is that the average skill level of all active managers cannot be inferred from their overall past performance: the fact that as a group they do not outperform passive managers need not imply that they lack skill. Furthermore, the lack of persistence does not imply that differential ability across managers is unrewarded, that gathering information about performance is socially wasteful, or that

chasing performance is pointless. It merely implies the provision of capital to the mutual fund industry is competitive.¹

Competition between investors implies a lack abnormal performance on average and a lack of persistence in performance. Yet this result is not inconsistent with a strong relationship between performance and flow of funds. On the contrary, the strong relationship between performance and flow of funds is the mechanism that ensures that there is no predictability in performance.

In our model performance is not persistent precisely because investors chase performance and make full, rational, use of information about funds' histories in doing so. In any given period, high performance is rationally interpreted by investors as evidence of the manager's superior ability, so new money flows to the fund to the point that expected excess returns going forward are competitive. This process necessarily implies that investors cannot expect to make positive excess returns going forward, which implies that superior performance cannot be predictable. Thus, the response of funds flow to performance is simply evidence that capital flows to investments where it is most productive.

As is standard in models of corporation finance, capital is supplied with perfect elasticity to managers whose differential ability allows them to identify positive net present value opportunities. Since these opportunities, or the ability to identify them, are the resource that is ultimately in scarce supply, the economic rents thus created flow through to the managers who create them, not to the investors who invest in them. In effect, the implication is precisely the opposite of Sirri and Tufano's intuition cited above. Any predictability in active manager's returns is evidence of a lack of competition in that sector—capital flows to that sector are not sufficiently immediate to compete away all the rents. *Ceteris paribus*, sectors with predictability should be associated with a smaller, rather than a larger, response of capital flows to performance. Similarly, the observed behaviors are quite consistent with the existence of skilled managers with differential ability. By calibrating our model to observed fund flows and survival rates reported in the literature, we show that the data on active mutual funds is consistent with the presence of significant managerial skill. Indeed, we find a surprisingly high level of skill—80% of managers have at least enough skill to make back their fees.

Our model is related to several recent papers, though none of these are directly aimed at reconciling the lack of return persistence with the responsiveness of flows to past performance. Bernhardt, Davies, and Westbrook (2002) study short-term return persistence in a model where privately informed managers trade to maximize funds under management. The

¹Ippolito (1989) finds evidence in support of the notion that the provision of capital in the mutual fund industry is competitive.

response of funds to performance in their model, however, is exogenous. In our model, the flow-performance relationship is endogenous, and consistent with no persistence in performance. Lynch and Musto (2002) endogenize the flow of funds in a two-period setting, where, as in our paper, investors and managers learn about manager's abilities and the profitability of their strategies from past returns. In their model differences in ability lead to persistent differences in performance. We show rational learning and strong response of flows to performance can be consistent with weak persistence in performance, or none at all. Nanda, Narayan and Warther (2000) develop a three-date model with heterogeneous managerial ability. They use the model to derive endogenously heterogeneous fee structures involving loads as managers compete for clients with different expected liquidity needs. In their model managerial ability is known, so there is no role for learning about managerial ability, and the response of the flow of funds to it. These are central concerns in our paper. Finally, in a corporate setting, Holmström (1999) considers a dynamic multi-period moral-hazard problem in which learning about managerial ability is very similar to that in our model. He characterizes the resulting optimal contract. Our goals are very different. There is no moral hazard or asymmetric information in our model, so the simple compensation scheme we consider is optimal, funds are allocated efficiently across managers, and all outcomes are first-best.

The paper is organized as follows. In the next section we motivate the paper by outlining a very simple thought experiment that communicates the intuition that underlies our model. In Section 3, we derive the model and introduce a simple parametric example that allows us, subsequently, to quantitatively characterize our results. We then use the model to characterize mutual fund survivorship in Section 4, along with the flow of funds and performance relations, life-cycle characteristics of funds, and the unobserved distribution of managerial ability in Section 5. Section 6 is the conclusion.

2 The Argument

We begin by considering a simple thought experiment. Imagine an economy with full information. Skilled investment managers exist who can generate positive, risk-adjusted, excess returns. Managers and investors alike know who these managers are. What would the returns these managers provide to investors look like?

In equilibrium, investors who choose to invest with active managers cannot expect to receive positive excess returns on a risk-adjusted basis. If they did, there would be an excess supply of capital to those managers. Every investor in the economy who held assets of

equivalent risk would want to sell those assets and invest with the active manager instead. Markets can only clear when the expected return to investors in these funds equals the expected return in alternative investment opportunities. That is, the risk-adjusted expected excess return to investing with a skilled active manager must be zero.

If skill or superior ability in active portfolio management could be deployed at an arbitrarily large scale without dissipating its effectiveness, then in a given risk class all funds in this hypothetical world would flow to the manager with the highest ability. That is, one manager would end up managing all the available investment capital leaving the manager with no investors to trade with, which contradicts the assumption that the manager can earn excess returns.

In light of the above argument, we assume that managerial ability to generate excess returns cannot be effectively employed at an arbitrarily large scale. Certainly, this is more consistent with the observed decentralization of the professional money management industry. If there are decreasing returns to scale in the use of managerial ability, funds will be invested with skilled managers to the point where the manager provides investors with expected returns equal to those available in passive alternatives.

This also suggests a mechanism the skilled manager can use to capture the rents in this idealized economy. He can manage a large fund and charge a fee that is proportional to the amount of assets under management. With this incentive scheme, investment will flow into the fund until the fund is so large that the risk-adjusted expected excess return is zero. Highly skilled managers will manage larger funds, earn more in fees, and extract more rents.

Note that in this economy investors earn the same returns in active and passive investments of the same riskiness and all returns are unpredictable. If there is cross-sectional variation in skill level, more funds flow to the higher skilled managers. Although the size of the fund is correlated with management ability, neither fund size nor management ability predicts investor returns. In this static environment there are no dynamics to the flow of funds.

Next consider a more realistic economy in which neither investors nor managers themselves can separate good managers from bad. In this case participants will infer ability from past returns. Investors initially will invest to the point that the expected excess risk-adjusted return from placing funds with a typical manager is zero. Subsequently, the lucky investors who happened to pick the skilled managers will earn superior returns. As time passes, however, and managers build a track record, differential skill levels among managers become apparent. Investors react by increasing their investment with the skilled managers, which allows these managers to extract more economic rents by collecting fees on assets under management. Learning from past returns, however, does not change the insight that in

equilibrium the flow funds ensures that investors do not expect to earn excess returns going forward. The expected return to investors is always the same. Conditional on perceived skill level, the risk-adjusted expected excess return to investing is zero.

In the second economy, since investors infer management skill from past returns, the flow of funds will be driven by managers' past returns. Nevertheless, all risk-adjusted excess returns are unpredictable. The response of the flow of funds to past returns, coupled with unpredictable future returns, is not a sign of investor irrationality. Returns are unpredictable *because* funds flow rationally in response to past returns.

Although this verbal discussion captures the basic intuition, there are implications of this theory that only a formal model can deliver. The rest of this paper, therefore, formalizes the argument and explores some of these implications.

3 The Model

In this section we introduce a simple model that allows us to solve recursively for the relationship between performance and the flow of funds. We assume that a population of active mutual fund managers constructs investment portfolios from the primitive assets in the economy. Investors reward managers by paying a fixed percent of assets under management every period. Mutual fund managers have heterogeneous ability to identify mispriced assets and generate superior performance.

All participants in the model are symmetrically informed. Managerial ability is unknown to both managers and investors, and they learn about this ability by observing the history of the managed portfolio's returns. The costs managers face are also common knowledge.

We will assume a manager's ability to achieve superior performance is limited to the idiosyncratic portion of the portfolio. That is, we rule out cases where superior ability derives from ability to predict the outcome of systematic uncertainty, and develop the model assuming that investors are all risk neutral. It is possible to embed the model in a more general one where investors are risk averse and active managers bear systematic risk as follows. Note that an investor's decision on the level of riskiness of his portfolio can be made separately from his decision to chase highly skilled managers, as long as "skill" does not involve market timing. A risk averse investor can first decide on what level of systematic risk is optimal. He can then look for skilled managers and invest with them, maintaining the same level of systematic risk by trading passive alternatives. So long as the investor knows the systematic risk of the manager's portfolio, he can adjust his own levels of systematic risk to ensure that his overall exposure remains the same once the investment with the

skilled manager is made. Since the systematic risk held by the manager is irrelevant to the investor's decision, we can assume that managers hold none—their portfolios consist of purely idiosyncratic risk. Investors evaluate this risk as if they were risk neutral.

Instead of placing funds with an active manager, investors can purchase passively managed portfolios, which we refer to as “the passive benchmarks.” These investments are available to all participants, including managers, with no costs or fees. They have an expected excess return equal zero (i.e., the expected return is the risk free rate). If investing in a managed fund, given available information, has positive expected excess returns, investors will supply capital to the fund with infinite elasticity. If the expected excess return of the fund is less than zero, investors will withdraw their money from the fund.

Managers act at every date to maximize fees, which in this setting will be equivalent to maximizing assets under management. Since, in equilibrium, assets under management will be increasing in expected returns, managers also attempt at every date to earn as high an expected return as possible and thus increase flows. The model is structured to preclude complicated dynamic strategies through which managers could manipulate investors' perceptions of their expected return. There is empirical evidence of such strategic behavior by managers in Brown, Harlow and Starks (1996) and Chevalier and Ellison (1997), who show that managers with good performance early in the year tend to reduce risk later, and those with bad performance do the reverse. While these agency issues are clearly of some importance, we believe it makes sense to ask first whether a simple, symmetric information model can the broader features of the performance-flow relationships.

To achieve high returns management must identify undervalued securities and trade to exploit this knowledge without moving the price adversely. To do this managers must expend resources and must pay bid-ask spreads that diminish the return available to pay out to investors. We assume these costs are independent of ability and are increasing in the size of the fund. While the manager faces increasing marginal costs of managing those funds he manages actively, he is free to invest any additional funds he can attract in the passive benchmark and avoid these costs. This will imply that there will be an optimal amount of money that he will actively manage. Funds in excess of this will be invested in the passive benchmark. Since he collects fees on these additional funds, they are a means through which he extracts rents associated with his superior ability.

Investors know the cost structure faced by the manager, and the amount of money under management. Accordingly, they can infer the manager's actions. They continue to supply funds to the manager because at the margin it allows them to participate in the excess returns attributable to the manager's ability. As the manager expands the scale of the fund, he will eventually earn the expected return that is available to investors in passive alternatives, and

thus the fund becomes unattractive net of fees.

Let q_{at} denote the funds being actively managed, and q_{It} the funds the manager indexes. The total funds under management are denoted $q_t = q_{at} + q_{It}$. We take the standard compensation contract in the mutual fund industry. The manager charges a fee of f per dollar managed. As we discuss later, there is no moral hazard in this model, and so this contract is optimal in the sense that it leads to first-best investment decisions. For simplicity, we restrict attention to funds without loads (exit and entry costs). The presence of loads has been attributed to manager's desire to separate investors with different liquidity needs (see Chordia (1996) and Nanda, Narayan and Warther (2000)), which is not a focus of this paper.

The manager incurs fixed costs of F each period he is in operation. These costs include overhead and any opportunity costs for management time. The fixed costs are not a function of the strategy, whether active or passive, adopted by the manager. Active management triggers additional variable costs, and these costs are increasing and convex in the funds under active management. Denote the variable costs incurred when actively managing q_{at} funds as $C(q_{at}) \geq 0$, where $C(0) = 0$ and $C(q) < q, \forall q$. Since we assume that there are decreasing returns to scale, $C'(q) > 0, \forall q$. To ensure a bounded solution, we also further assume that for every q , $C''(q) > 0$ and $\lim_{q \rightarrow \infty} C'(q) > 1$. These assumptions capture the notion that with a sufficiently large fund a manager will spread his information gathering activities too thin, or that larger transaction costs are imposed by market makers increasing their spreads.

We model differential ability by placing structure on the return the manager could earn absent costs and fees. Note that this is not the return actually earned by investors in the fund. That will be net of costs and fees. For a fund at date t ,

$$R_t = \alpha + e_t$$

denotes the excess return of the actively managed funds absent any costs or fees. One could think of this return as the return the manager makes on the first dollar he manages, before any costs or fees are incurred. The parameter α is unknown to both the managers and the market, and is the source of differential ability across managers. The errors, e_t , are normally distributed with mean zero and variance σ^2 . They are independently distributed through time. Denote the precision of the errors as $\omega = \frac{1}{\sigma^2}$. Investors learn about α by observing the realized excess returns the manager produces. This learning is the source of the relationship between performance and the flow of funds.

The excess total payout, over what would be earned on the passive benchmark, to in-

vestors in the fund will be:

$$TP_{t+1} = q_{at}R_{t+1} - C(q_{at}) - (q_{at} + q_{It})f.$$

The excess return to investors, per unit of funds invested in the portfolio, is then

$$\begin{aligned} r_{t+1} = \frac{TP_{t+1}}{q_t} &= \frac{q_{at}}{q_t}R_{t+1} - \frac{C(q_{at}) + q_{It}f}{q_t} \\ &= \frac{q_{at}}{q_t}R_{t+1} - c(q_t) \end{aligned} \tag{1}$$

where

$$c(q_t) \equiv \frac{C(q_{at}) + q_{It}f}{q_t} \tag{2}$$

denotes the unit cost associated with investing in the actively managed fund. The return r_t corresponds to the return empirically observed.

Investors supply funds with infinite elasticity to investment opportunities that have positive excess expected returns. At each point in time, then, funds flow to each fund so that the expected return going forward is zero:

$$E_t(r_{t+1}) = 0. \tag{3}$$

Clearly, this implies there is no predictability or persistence in fund returns. We next derive the manager's optimal decision rule.

3.1 The Manager's Decision Problem

For a manager to choose to go into the mutual fund business, it must be his best available employment opportunity. Under our assumptions, the value of entering the mutual fund business is an increasing function of the initial amount of money under management. Thus, a fund size, q_0 , will exist, such that for all fund sizes larger than this, money management is the manager's best available opportunity.

We do not model the source of the market's priors about managers. This is presumably based on observable characteristics, such as those studied by Chevalier and Ellison (1999).² We simply assume that the market's prior is that the manager's ability, α , is normally distributed with variance η^2 . If the information that shapes investors' expectations about managers evolves continuously, managers will go into business the first time investors' are

²These authors show that the average SAT scores of the manager's undergraduate institution does have modest ability to predict performance.

willing to provide q_0 in funds. We assume that this occurs when the prior expectation of the manager's ability is ϕ_0 . To summarize, the priors for new managers regarding their ability, α , is normally distributed with mean ϕ_0 and variance η^2 . The precision of the prior will be denoted $\gamma = \frac{1}{\eta^2}$.

We further assume that there are non-recurring fixed costs associated with setting up a mutual fund. Such costs include, but are not limited to, the cost to the manager of convincing investors of his ability despite having no track record. This implies that the opportunity cost of opening a fund is greater than the opportunity cost of continuing to operate. Accordingly, there will also be a minimal efficient scale of operation for the fund, \bar{q} , with $\bar{q} < q_0$. If $q_t < \bar{q}$, the manager will be unable to cover his fixed costs, F , and the fund will shut down.

Consider a fund in the cohort that began operating at date 0. The funds under management are observable, and the cost function is common knowledge. Investors update their posteriors based on the history of observed returns as Bayesians. Let

$$\phi_t \equiv E(R_{t+1} \mid R_1, \dots, R_t),$$

then taking expectations of both sides of (1), and using (3), gives:

$$\phi_t = \frac{q_t}{q_{at}} c(q_t). \quad (4)$$

The timing convention is as follows. The fund enters period t with q_{t-1} funds under management and estimate of managerial ability ϕ_{t-1} . Managers and investors observe r_t (from which they can infer R_t), and update their estimate of the manager's ability by calculating ϕ_t . Then funds flow into or out of the fund to determine q_t .

Managers act to maximize total fees, subject to the constraint that the expected return per dollar invested is no less than what investors could earn on their own by indexing. Finally, we rule out borrowing or short positions in the passive benchmark. This also excludes the possibility that the manager can borrow to cover operating costs, and thus experiment to build a track record. It is descriptive of open-ended mutual funds, which have been the focus of most of the empirical work on the flow of funds. Consequently, the manager will shut down the fund if his fees fail to cover his fixed costs, F .

Formally, the manager solves:

$$\begin{aligned} \max_{q_{at}, q_{It}} \quad & (q_{at} + q_{It})f \\ \text{s.t.} \quad & \end{aligned} \quad (5)$$

$$E_t\{TP_{t+1}\} = q_{at}\phi_t - C(q_{at}) - (q_{at} + q_{It})f \geq 0 \quad (6)$$

$$(q_{at} + q_{It})f \geq F \quad (7)$$

$$q_{at} \geq 0 \quad (8)$$

$$q_{It} \geq 0. \quad (9)$$

The first constraint in the manager's problem, (6), ensures the fund is viable from the investors' perspective: it offers a competitive return. The second constraint, (7), requires it to be viable from the manager's viewpoint: he is able to cover his fixed cost. The last two constraints rule out borrowing and short positions.

3.2 The Manager's Policies

We now analyze the solution to this problem, derive an expression for the unit cost function, $c(q_t)$, and characterize the flow of funds.

The left-hand side of the investor participation constraint, (6), is decreasing in the total funds under management, while both the objective function and the fixed cost recovery constraint (7) are increasing in this quantity. Thus, the investor participation constraint will bind if there is a feasible solution other than $q_{at} = q_{It} = 0$. Consequently, at the optimal allocation of funds:

$$q_{at}^* \phi_t - C(q_{at}^*) - (q_{at}^* + q_{It}^*)f = 0 \quad (10)$$

or

$$\phi_t = \frac{C(q_{at}^*) + q_t f}{q_{at}^*}, \quad (11)$$

where we have used $q_t = q_{at}^* + q_{It}^*$.

By inspection it is obvious that the investor participation constraint, (6), will never be met with $q_{at} = 0, q_{It} > 0$. At any non-zero solution, then, there are two possibilities for the optimal quantities, q_{at}^*, q_{It}^* . Either both are positive or $q_{at}^* > 0, q_{It}^* = 0$. We will consider each case separately.

When $q_{at}^* > 0, q_{It}^* > 0$, the first-order conditions for the optimal quantities are, respectively:

$$f - \lambda_t^p [C'(q_{at}^*) - \phi_t + f] - \lambda_t^f f = 0 \quad (12)$$

$$f - (\lambda_t^p + \lambda_t^f) f = 0 \quad (13)$$

where λ_t^p and λ_t^f are the Lagrange multipliers for the participation constraint (6) and fixed cost recovery constraint (7), respectively. Expression (13) immediately implies that $\lambda_t^p + \lambda_t^f = 1$. Then from (12),

$$C'(q_{at}^*) = \phi_t \quad (14)$$

Equating (14) and (11) gives the amount of funds the manager puts into active management,

$$\begin{aligned} \frac{C(q_{at}^*) + q_t f}{q_{at}^*} &= C'(q_{at}^*) \\ \Rightarrow q_{at}^* C'(q_{at}^*) - C(q_{at}^*) &= q_t f \\ \Rightarrow G(q_{at}^*) &= q_t \end{aligned} \tag{15}$$

$$\Rightarrow q_{at}^* = G^{-1}(q_t) \tag{16}$$

where $G(q) \equiv \frac{qC'(q) - C(q)}{f}$. This function is strictly increasing, and hence invertible, by the convexity of the costs since $G' = \frac{q}{f}C'' > 0$. The existence of q_{at}^* is guaranteed because $G(0) = 0$ (since $C(0) = 0$) and $\lim_{q \rightarrow \infty} G(q) = \infty$ (since $G(q) > qC'(q) - q = q(C'(q) - 1)$ and $\lim_{q \rightarrow \infty} C'(q) > 1$). Equation (16) defines q_{at}^* in terms of q_t . That is, knowing q_t alone allows observers to infer q_{at}^* , because the cost structure is also known.

The manager will invest in the actively managed strategy to the point where the marginal cost is equal to the excess expected return, (14). Substituting q_{at}^* gives us q_{It}^* , the amount the manager indexes:

$$\begin{aligned} q_{It}^* &= q_t - q_{at}^* \\ &= q_t - G^{-1}(q_t) \end{aligned} \tag{17}$$

When the manager is constrained because the non-negativity constraint (9) binds so that $q_{It}^* = 0$, the participation constraint will determine the maximum amount of funds the manager can attract and still give investors their reservation expected return. Setting $q_{at}^* = q_t$ in (11) gives

$$\phi_t = \frac{C(q_t) + q_t f}{q_t} = c(q_t), \tag{18}$$

where the second equality follows from (2). So in this case the per unit cost equals the excess return.

Whether or not the manager is constrained by (9) depends on the amount of money under management. There is a critical level q^c , such that whenever $q_t > q^c$, $q_{It}^* > 0$. To determine q^c , we simply solve for the level of q_t such that in the unconstrained problem, the manager chooses to set $q_{It}^* = 0$. Using (16) this implies that q^c solves

$$q^c = G^{-1}(q^c). \tag{19}$$

Finally we need to consider the case when it is uneconomical to run the fund, that is, $q_{at}^* = 0$ and $q_{It}^* = 0$. This case occurs whenever the total fees collected by the manager do

not exceed the fixed costs, and (7) cannot be satisfied. This will occur whenever

$$q_t f < F \Rightarrow q_t < \frac{F}{f}.$$

So there will be a minimal efficient scale of the fund, which we denote $\bar{q} = \frac{F}{f}$. Whenever q_t drops below \bar{q} the fund goes out of business. Note that if there are additional costs associated with going into the mutual fund business initially, $q_0 > \bar{q}$. The following proposition summarizes the manager's decision rule.

Proposition 1 *Let the maximum amount of capital a manager can raise at time t be denoted q_t and let q^c solve (19). Then if $q_t < \frac{F}{f}$, the fund shuts down. Otherwise, the manager puts*

$$q_{at}^* = \begin{cases} q_t & \text{if } q_t < q^c \\ G^{-1}(q_t) & \text{o.w.} \end{cases} \quad (20)$$

into active management and indexes the rest:

$$q_{It}^* = \begin{cases} 0 & \text{if } q_t < q^c \\ q_t - G^{-1}(q_t) & \text{o.w.} \end{cases} \quad (21)$$

The manager's choice can be described as follows. To stay in business he must have enough assets under management to simultaneously cover his fixed costs and ensure investors an expected excess return of at least zero. Assuming that this is possible, the manager would like to set the level of funds in active management such that marginal costs equals his expected return. If he does reach this level of active management, putting more money into active management is suboptimal. He is better off indexing the additional money and thus avoiding the trading or information gathering costs. In some cases he cannot raise enough funds to ensure that the marginal costs of active management equals his expected return because at this level of assets, expected excess returns to investors are negative. By reducing the amount of assets under management, and thus the costs incurred, the return to investors is raised.

In this form, the model also suggests that the (idiosyncratic) risk characteristics of funds should evolve through time. As managers establish that they have ability, they attract more funds, but their portfolios become more like the passive benchmarks. When active management involves taking higher idiosyncratic risks, then younger funds will tend to be the ones taking more of this kind of risk, and will tend to be the extreme performers.

3.3 Contract Optimality

The assumed managerial compensation contract is optimal in that it leads to first-best, efficient investment. This might seem surprising in that the contract does not depend directly on performance. There is no moral hazard or asymmetric information in this model, however. The manager's allocation of funds, q_t , between active and passive investments, q_{at}, q_{It} , is assumed to be observable to the market and enters directly in constraint (6). Formally, this means that there is no incentive compatibility constraint in the manager's problem. Maximizing the total return to the portfolio, through (6), increases the total fees the manager can earn. Thus, the incentives associated with the fee structure in our model are not in any way distortionary.

This is not a model of “closet indexing,” because the manager's choices are transparent and can be directly contracted upon. The allocation of capital across managers is efficient, and outcomes are first-best. Managers are providing a valuable service. They expand their actively managed portfolio to the point where marginal benefits equal marginal costs. The extra funds they attract beyond this level would otherwise be invested in passive alternatives in any case. The extra funds contribute to higher compensation for the manager, of course, but not inefficiently. Alternatively, managers could simply choose the optimal level for active management, q_{at}^* , close to the fund to additional flows, and charge differential fees based on past history.

Undoubtedly, investors would prefer an environment where they shared in the benefits derived from managerial ability. Managerial talent, however, is the resource in scarce supply. As in most models in corporation finance, the manager extracts the net present value of his services. Since those services are efficiently deployed, within the broader class of repeated single-period contracts the compensation scheme we assume is optimal.

Absent moral hazard or asymmetric information, the only reason multiperiod contracts might lead to superior outcomes would involve more efficient risk sharing. If there are differences in risk aversion between managers and investors then multiperiod contracts, if they are enforceable, can improve risk sharing. This would require multiperiod contracts to be enforceable—managers and investors would need to be forced to work or invest when it would be suboptimal to do so. If both parties are free to recontract at each point, then the multiperiod contracts would collapse into a sequence of one-period contract such as ours. Intertemporal commitment to such a contracts might be difficult to enforce on the manager's side. It may or may not be difficult to enforce on the investors' side, through loads or other devices. In any case, optimal intertemporal risk sharing is not the focus of our analysis.

3.4 Fund Dynamics

In Section 3.2 we showed that the fund will only continue to operate when $q_t \geq \bar{q}$. Note that the amount investors are willing to invest at time t is determined by ϕ_t , using (11). This implies that there exists a $\bar{\phi}$ such that investors will only be willing to invest more than \bar{q} when $\phi_t > \bar{\phi}$. Any time ϕ_t drops below this value, the fund shuts down and investors withdraw all funds.

Using Proposition 1, the portion of funds under active management is given by

$$h(q_t) \equiv \frac{q_{at}^*}{q_t} = \begin{cases} 1 & \text{if } q_t < q^c \\ \frac{G^{-1}(q_t)}{q_t} & \text{o.w.} \end{cases} \quad (22)$$

Similarly, from (11), the per unit cost of managing all funds, $c(q_t)$, is

$$c(q_t) = \frac{C(q_{at}^*) + q_t f}{q_t} = \begin{cases} \frac{C(q_t)}{q_t} + f & \text{if } q_t < q^c \\ \frac{C(G^{-1}(q_t))}{q_t} + f & \text{o.w.} \end{cases} \quad (23)$$

Substituting (22) and (23) into (4) implies that at each point in time,

$$\phi_t = \frac{c(q_t)}{h(q_t)} = \begin{cases} \frac{C(q_t)}{q_t} + f & \text{if } q_t < q^c \\ \frac{C(G^{-1}(q_t)) + q_t f}{G^{-1}(q_t)} & \text{o.w.} \end{cases} \quad (24)$$

We can use these expressions to obtain a simple recursive expression for the relationship between the flow of funds and performance.

Proposition 2 *The evolution of ϕ_t and the change in the amount of money under management at any date, as a function of the prior performance, is the solution to*

$$\phi_t = \phi_{t-1} + \frac{r_t}{h(q_{t-1})} \frac{\omega}{\gamma + t\omega} \quad (25)$$

$$\frac{c(q_t)}{h(q_t)} = \frac{c(q_{t-1})}{h(q_{t-1})} + \frac{r_t}{h(q_{t-1})} \frac{\omega}{\gamma + t\omega} \quad (26)$$

if

$$\phi_t - \bar{\phi} \geq 0. \quad (27)$$

Otherwise, $q_t = 0$.

Proof: It is straightforward to show from DeGroot (1970) (Theorem 1, p. 167), that the

mean of investor's posteriors will satisfy the following recursion:

$$\phi_t = \frac{\gamma + (t-1)\omega}{\gamma + t\omega} \phi_{t-1} + \frac{\omega}{\gamma + t\omega} R_t \quad (28)$$

where ϕ_0 is the mean of the initial prior. Next, use (24), evaluated at both t and $t-1$ to substitute for ϕ_t and ϕ_{t-1} in this expression. Finally, use the fact that the realized return is given by $r_t = h(q_{t-1})R_t - c(q_{t-1})$ (combining equations (1) and (22)) to substitute for R_t . These substitutions yield the desired result. QED

The difference equation (26) that gives the relationship between performance and funds flow is very simple, and yet it clearly can match several features of fund flows. First, a smaller t increases the weight on r_t , so flows to younger funds respond much more dramatically to performance than flows to more mature funds. Investors learn more with each observation. The relationship will also be nonlinear because under our assumptions the cost function is nonlinear. Finally, in the cross section for a given cohort, fund flow will be particularly responsive to extremely bad performance, simply because it causes investors to withdraw all their funds.

4 Survival Probabilities

Survival bias in mutual funds has been an important area of empirical investigation. Clearly, empirical studies that ignore funds that failed will have biased estimates of the expected returns for the surviving funds (see, for example, Carhart *et al* (2001)). More importantly, the survival rates themselves communicate information about the abilities of active managers. For example, one could examine the hypothesis that no skilled managers exist by comparing the fit of our model with $\alpha = 0$ to its performance with parameters of the prior distribution freely estimated. Survival rates are natural moments to focus on in such a test. With this in mind we derive explicit expressions for the survival rates in the model.

Note that Proposition 2 implies that a fund goes out of business whenever $\phi_{t-s} < \bar{\phi}$, $s = 1, \dots, t-1$. Thus, we can compute the unconditional probability that a fund will survive t years by calculating the probability that $\phi_t \geq \bar{\phi}$. We begin by conditioning on the true ability of the manager. Let the (unconditional) probability that a fund manager with ability α survives t periods be denoted $P_t(\alpha)$. The next proposition derives an expression for this probability:

Proposition 3 *Suppose a manager begins operating when the market's prior on her ability*

is ϕ_0 . The unconditional probability that a fund manager with true ability α survives t periods is

$$P_t(\alpha) = \int_{\bar{\phi}}^{\infty} f_t^\alpha(\phi) d\phi$$

where $f_t^\alpha(\phi)$, the density, conditional on management ability α of ϕ_t over the region $\phi_t \geq \bar{\phi}$, is defined recursively as follows:

$$f_t^\alpha(\phi) = \frac{\gamma + t\omega}{\omega} \int_{\bar{\phi}}^{\infty} f_{t-1}^\alpha(v) n^\alpha \left(\frac{\gamma + t\omega}{\omega} \phi - \frac{\gamma + (t-1)\omega}{\omega} v \right) dv \quad (29)$$

with the boundary condition that $f_1^\alpha(\phi) = \frac{\omega + \gamma}{\omega} n^\alpha \left(\frac{\gamma + \omega}{\omega} (\phi - \phi_0) + \phi_0 \right)$ and $n^\alpha(\cdot)$ is a Normal density function with mean α and precision ω , that is,

$$n^\alpha(u) = \sqrt{\frac{\omega}{2\pi}} e^{-\frac{1}{2}(u-\alpha)^2\omega}$$

Proof:

$$\begin{aligned} \Pr[\phi_t \geq \phi] &= \Pr[\phi_t \geq \phi \mid \phi_{t-1} \geq \bar{\phi}] \Pr[\phi_{t-1} \geq \bar{\phi}] \\ &= \Pr \left[\frac{\gamma + (t-1)\omega}{\gamma + t\omega} \phi_{t-1} + \frac{\omega}{\gamma + t\omega} R_t \geq \phi \mid \phi_{t-1} \geq \bar{\phi} \right] \Pr[\phi_{t-1} \geq \bar{\phi}] \\ &= \int_{\bar{\phi}}^{+\infty} \int_{\frac{\gamma+t\omega}{\omega}(\phi-v)+v}^{\infty} n^\alpha(u) du f_{t-1}^\alpha(v) dv \end{aligned}$$

Differentiating this expression with respect to ϕ , and thus eliminating the inner integral, gives the density in the proposition. The expression of $P_t(\alpha)$ then follows immediately. Finally, we need to derive f_1 . By (28) we have that

$$\phi_1 = \frac{\gamma}{\gamma + \omega} \phi_0 + \frac{\omega}{\gamma + \omega} R_1$$

where R_1 is distributed Normal $[\alpha, \sqrt{\frac{1}{\omega}}]$. This implies that $f_1(\phi)$ (the density of ϕ_1) is Normal $[\frac{\gamma}{\gamma + \omega} \phi_0 + \frac{\omega}{\gamma + \omega} \alpha, \frac{\sqrt{\omega}}{\gamma + \omega}]$. QED

The objects of study in the empirical literature are the overall survival rates in the data. It is now a simple matter to derive the unconditional probability that a fund will survive t periods:

Proposition 4 *The unconditional probability that a fund survives t periods is*

$$P_t = \int_{\bar{\phi}}^{\infty} F_t(\phi) d\phi$$

where $F_t(\phi)$, the unconditional density of ϕ_t over the region $\phi_t \geq \bar{\phi}$, is defined as:

$$F_t(\phi) = \int_{-\infty}^{\infty} f_t^u(\phi) n^{\phi_0}(u) du \quad (30)$$

where $n^{\phi_0}(u)$ is a Normal density function with mean ϕ_0 and precision γ , that is,

$$n^{\phi_0}(u) = \sqrt{\frac{\gamma}{2\pi}} e^{-\frac{1}{2}(u-\phi_0)^2\gamma}.$$

Proof: Recall that the prior distribution for α is Normal with mean ϕ_0 and precision γ , that is, $n^{\phi_0}(u)$. Then we have

$$\begin{aligned} P_t &= \int_{-\infty}^{\infty} P_t(u) n^{\phi_0}(u) du \\ &= \int_{\bar{\phi}}^{\infty} \int_{-\infty}^{\infty} f_t^u(\phi) n^{\phi_0}(u) du d\phi \\ &= \int_{\bar{\phi}}^{\infty} F_t(\phi) d\phi. \end{aligned}$$

QED

The density $F_t(\phi)$ will be important in subsequent analysis. It gives the distribution of the ϕ in period t for a manager with unknown ability who is still operating in period t .

A number of researchers have demonstrated that track record, beyond the most recent return, is an empirical determinant of survival.³ The role of learning in our model gives such a result.

For any given fund of age $t - 1$ with assets under management q_{t-1} , there exists a level of performance, denoted $\bar{r}(q_{t-1})$, such that if r_t is below this level, investors withdraw all funds. This critical realization can be determined by finding the r_t such that assets under management fall to \bar{q} . Using equation (26), this value is given by:

$$\bar{r}(q_t) = \left(\frac{c(\bar{q})h(q_{t-1})}{h(\bar{q})} - c(q_{t-1}) \right) \frac{\gamma + t\omega}{\omega} \quad (31)$$

³See Brown and Goetzmann (1995) and Carhart *et al* (2001).

Proposition 5 *The conditional probability that a fund of age $t - 1$ and size q_{t-1} survives through period t is given by*

$$1 - N^{\frac{\omega}{(h(q_{t-1}))^2}}(\bar{r}(q_{t-1}))$$

where

$$N^{\frac{\omega}{(h(q_{t-1}))^2}}(r) = \frac{1}{h(q_{t-1})} \sqrt{\frac{\omega}{2\pi}} \int_{-\infty}^r e^{-\frac{1}{2} \left(\frac{u}{h(q_{t-1})} \right)^2 \omega} du$$

is a Normal distribution with mean zero and precision $\frac{\omega}{(h(q_{t-1}))^2}$.

Proof: From (1), (3) and the distribution of R_t , r_t is distributed normally with mean zero and precision $\frac{\omega}{(h(q_{t-1}))^2}$. The result now follows directly from the fact that the probability of survival is the probability that $r_t > \bar{r}(q_{t-1})$. QED

Undertaking the recursive substitution in (26), will give the relation between the critical return at which the fund exits and its return history:

$$\frac{c(q_t)}{h(q_t)} = \frac{c(q_0)}{h(q_0)} + \sum_{i=1}^t \frac{r_i}{h(q_{i-1})} \frac{\omega}{\gamma + i\omega}.$$

Substituting this into (31) and simplifying yields

$$\bar{r}(q_t) = - \left[\frac{c(q_0)}{h(q_0)} - \frac{c(\bar{q})}{h(\bar{q})} \right] h(q_{t-1}) \frac{\gamma + t\omega}{\omega} - \sum_{i=1}^{t-1} \frac{h(q_{t-1})}{h(q_{i-1})} \left(\frac{\gamma + t\omega}{\gamma + i\omega} \right) r_i \quad (32)$$

so the critical level, $\bar{r}(q_{t-1})$, can be written as a function of the history of past returns. This implies that fund survival will depend on past performance and age, or alternatively on the current size of the fund and age.

5 Solutions for a Parametric Cost Function

The equilibrium zero-profit condition in our model gives a one-to-one mapping from ϕ_t to q_t . This means that $F_t(\cdot)$ from Proposition 4 also implies the distribution of q_t . These facts allow us to analytically derive unconditional relation between performance and flows, and compare these to the findings in the literature, such as the plots based on non-parametric estimates in Chevalier and Ellison (1997).

To derive an explicit expression linking performance to flows, we must pick a functional form for the cost structure. Assume $C(q) = aq^2$, so the variable cost function is quadratic.

We will further assume that the constraint $q_{It}^* \geq 0$ never binds. We derive necessary conditions for this below. The extension to the general case is straightforward, but tedious, and can be found in the appendix. Then under these assumptions we have

$$G(q) = \frac{q(2aq) - aq^2}{f} = \frac{aq^2}{f} \quad (33)$$

so

$$G^{-1}(q_t) = \sqrt{\frac{q_t f}{a}} \quad (34)$$

Using (22) and (23) gives

$$h(q_t) = \sqrt{\frac{f}{aq_t}} \quad (35)$$

$$\frac{c(q_t)}{h(q_t)} = 2\sqrt{aq_t f}. \quad (36)$$

Substituting (36) into (24) and solving provides the relation between market expectations of ability and assets under management:

$$q_t = \frac{\phi_t^2}{4af}. \quad (37)$$

To justify the assumption that $q_{It}^* > 0$ it must be the case that $\bar{q} > q^c$. Since the fixed cost constraint, (7), is binding at \bar{q} , this implies that the interior solution will always be implemented so long as

$$\bar{q} = \frac{F}{f} > q^c \quad (38)$$

Solving (19) gives

$$q^c = \frac{f}{a}. \quad (39)$$

Substituting and simplifying provides the following condition on primitives that guarantees that the interior solution will always be optimal.

$$F > \frac{f^2}{a}. \quad (40)$$

Using Proposition 1, the manager will allocate funds as follows:

$$q_{at}^* = \frac{\phi_t}{2a}. \quad (41)$$

and

$$q_{It}^* = \frac{\phi_t}{4a} \left(\frac{\phi_t}{f} - 2 \right). \quad (42)$$

Using Proposition 2, (35), and (36), (38) after some simplification, yields the following expression for the flow of funds as a function of the manager's performance so long as the manager remains in business:

$$\frac{q_t - q_{t-1}}{q_{t-1}} = \frac{r_t}{f} \left(\frac{\omega}{\gamma + t\omega} \right) + \frac{r_t^2}{4f^2} \left(\frac{\omega}{\gamma + t\omega} \right)^2 \quad (43)$$

Note, the flow of funds is zero at an excess return of zero. This suggests the empirically observed shape characteristics of the flow of funds to performance relationship. Because of the quadratic relationship between funds flow and excess return, funds respond more dramatically to extreme performance than to mediocre performance. While this quadratic behavior slows down the response to extreme negative performance, that will be offset in the cross section by funds closing when performance is extremely bad. That is, for r_t negative enough, the flow of funds is simply $-q_t$. Notice also that while age, t , and size, q_t , will tend to be correlated, they work in opposite directions on the absolute flow of funds. The absolute flow of funds is proportional to the size of the fund, so the flow of funds when expressed as a percentage, is independent of the fund size. Age, however, is still important, and since older funds will tend to be larger, if the age of the fund is not controlled for, then percentage change in the flow of funds will be more sensitive for smaller funds. The rest of this section is devoted to deriving explicit expressions for the cross-sectional relation between performance and the flow of funds.

Either the flow of funds responds continuously to r_t , or r_t is so bad that investors withdraw all the funds. As we saw in the previous section, the latter will occur for any r_t below a critical realization. However, in this section it is more convenient to write the critical realization as a function of ϕ_t rather than q_t . Denote this critical realization as $r^*(\phi_{t-1})$. It is determined by finding the r_t such that the market's posterior on the manager's ability falls to $\bar{\phi}$. Using equation (25), this value is given by solving:

$$r^*(\phi_{t-1}) = (\bar{\phi} - \phi_{t-1})h(q_{t-1})\frac{\gamma + t\omega}{\omega}$$

Substituting from (35) and (37), and simplifying, gives

$$r^*(\phi_{t-1}) = 2\left(\frac{\bar{\phi} - \phi_{t-1}}{\phi_{t-1}}\right) \left(\frac{\gamma + t\omega}{\omega}\right) f. \quad (44)$$

When r_t is below $r^*(\phi_{t-1})$ the percentage change in assets is -100%. When r_t is above $r^*(\phi_{t-1})$, (43) can be used to determine the flow of funds. In summary then the overall percentage change in the assets under management is given by

$$\frac{q_t - q_{t-1}}{q_{t-1}} = \begin{cases} -1 & \text{if } r_t < 2\left(\frac{\bar{\phi} - \phi_{t-1}}{\phi_{t-1}}\right) \left(\frac{\gamma + t\omega}{\omega}\right) f \\ \frac{r_t}{f} \left(\frac{\omega}{\gamma + t\omega}\right) + \frac{r_t^2}{4f^2} \left(\frac{\omega}{\gamma + t\omega}\right)^2 & o.w. \end{cases} \quad (45)$$

Empirical studies, however, commonly consider the flow of *new* funds, that is, the percentage change in new assets, which is typically defined to be

$$\frac{q_t - q_{t-1}(1 + r_t)}{q_{t-1}} = \begin{cases} -1 & \text{if } r_t < 2\left(\frac{\bar{\phi} - \phi_{t-1}}{\phi_{t-1}}\right) \left(\frac{\gamma + t\omega}{\omega}\right) f \\ \left[\frac{1}{f} \left(\frac{\omega}{\gamma + t\omega}\right) - 1\right] r_t + \frac{1}{4f^2} \left(\frac{\omega}{\gamma + t\omega}\right)^2 r_t^2 & o.w. \end{cases} \quad (46)$$

Note that this measure uses q_{t-1} in the denominator rather than $q_{t-1}(1 + r_t)$. Unfortunately, this definition distorts the implications of very large negative returns that cause liquidation of the fund. For any $r_t < r^*(\phi_t)$, $q_t = 0$ and $\frac{q_t - q_{t-1}(1 + r_t)}{q_{t-1}} = -(1 + r_t)$. That is, under this definition, for these returns, the measure becomes *less* responsive the worse the performance. In the limit when $r_t = -100\%$, the measure gives no response in the flow of funds. In an effort to address this issue while still maintaining consistency with the empirical estimates for other returns, we set our measure of the flow of funds equal to -1 whenever liquidation occurs.

For a given negative realization of r_t , we can derive the set of funds that will liquidate from (44). That is, all funds for which

$$r_t < r^*(\phi_{t-1}) = (\bar{\phi} - \phi_{t-1})h(q_{t-1})\frac{\gamma + t\omega}{\omega} = 2\left(\frac{\bar{\phi} - \phi_{t-1}}{\phi_{t-1}}\right) \left(\frac{\gamma + t\omega}{\omega}\right) f.$$

Note that if $r_t \leq -2\left(\frac{\gamma + t\omega}{\omega}\right) f$ this inequality is always satisfied (since $\bar{\phi} \leq \phi_{t-1}$), so the fund is liquidated regardless of previous history. When $-2\left(\frac{\gamma + t\omega}{\omega}\right) f < r_t < 0$, we can solve for ϕ_{t-1} to give

$$\phi_{t-1} < \frac{\bar{\phi}}{1 + \frac{r_t}{2f} \left(\frac{\omega}{\gamma + t\omega}\right)}$$

For convenience, define

$$\rho(r) \equiv \begin{cases} \frac{\bar{\phi}}{1 + \frac{r}{2f} \left(\frac{\omega}{\gamma + t\omega}\right)} & r > -2\left(\frac{\gamma + t\omega}{\omega}\right) f \\ \infty & r \leq -2\left(\frac{\gamma + t\omega}{\omega}\right) f \end{cases}$$

then for a given negative realization of r_t , all funds with $\phi_{t-1} < \rho(r_t)$ are liquidated.

Empiricists have studied the reaction of the flow of funds to returns by conditioning on funds' ages. To generate the analogous theoretical relation in our model, we aggregate over all funds with the same age. Using (46), the expected flow of funds for a fund of age t , conditional on the fund being alive in period $t - 1$ and having realized return r in period t is given by

$$\begin{aligned}
N_t(r) = & \left(\frac{1}{f} \left(\frac{\omega}{\gamma + t\omega} \right) - 1 \right) \left(\frac{\int_{\max(\bar{\phi}, \rho(r))}^{\infty} F_{t-1}(\phi) d\phi}{\int_{\bar{\phi}}^{\infty} F_{t-1}(\phi) d\phi} \right) r \\
& + \frac{1}{4f^2} \left(\frac{\omega}{\gamma + t\omega} \right)^2 \left(\frac{\int_{\max(\bar{\phi}, \rho(r))}^{\infty} F_{t-1}(\phi) d\phi}{\int_{\bar{\phi}}^{\infty} F_{t-1}(\phi) d\phi} \right) r^2 \\
& - \left(\frac{\int_{\bar{\phi}}^{\max(\bar{\phi}, \rho(r))} F_{t-1}(\phi) d\phi}{\int_{\bar{\phi}}^{\infty} F_{t-1}(\phi) d\phi} \right). \tag{47}
\end{aligned}$$

Since $\rho(r) < \bar{\phi}$ for any $r > 0$, the flow of funds is quadratic for any positive excess return. When excess return are negative, the flow of funds is no longer quadratic, for very large negative returns it converges to -100%.

From (47) it is clear that larger fees imply less sensitivity. Fund age also attenuates fund sensitivity. Figure 1 plots the flow of funds relationship for funds of different ages. The shapes of the relation between performance and flows are reminiscent of what researchers have found empirically (see, for example, Chevalier and Ellison (1997, Fig. 1)).

5.1 Evolution of Fund Volatility

Consider the volatility of funds' returns as a function of the assets under management. The return to investors is given by

$$r_{t+1} = \frac{q_{at}R_{t+1} - C(q_{at}) - (q_{at} + q_{It})f}{q_t}. \tag{48}$$

The conditional variance of this return is therefore

$$\text{var}_t(r_{t+1}) = \text{var}_t \left(\frac{q_{at}R_{t+1} - C(q_{at}) - (q_{at} + q_{It})f}{q_t} \right) = \text{var}_t \left(\frac{q_{at}}{q_t} R_{t+1} \right) = \left(\frac{q_{at}}{q_t} \right)^2 \frac{1}{\omega} \tag{49}$$

where the t subscript denotes that the variance is conditional on funds currently under management, q_t . Assuming that managers have quadratic variable costs, using (16) and

Figure 1: **Flow of New Funds as a Function of Return:** The curves plot the response in the flow of new funds to the previous period's return (i.e., equation (47)). The steepest curve shows the response for a 2 year old fund (i.e., the return is from year 2 to year 3). The remaining curves show the response for a 5, 10 and 20 year old fund respectively. The parameter values used are as follows: $\phi^* = 0.03$, $\bar{\phi} = 0.065$, $f = 0.015$, $\gamma = 277$ and $\omega = 25$.

(34), gives the following expression for the volatility of investor returns

$$\text{var}_t(r_{t+1}) = \left(\frac{q_{at}}{q_t}\right)^2 \frac{1}{\omega} = \frac{f}{aq_t\omega} = \frac{1}{\omega} \left(\frac{2f}{\phi_t}\right)^2. \quad (50)$$

The conditional volatility is a decreasing function of the size of the fund or the perceived skill of the manager. Since past positive returns increase the size of the fund, the volatility of the fund will decrease. The opposite occurs with negative returns. Thus, the model delivers the relation between risk and past performance that has been attributed in past research to attempts by managers to mislead investors. In our model, funds with superior past performance invest a larger portion of their new capital in passive strategies and thus lower their overall volatility. Similarly, funds with poor performance increase their volatility because as funds flow out, they preferentially liquidate capital that was allocated to passive strategies.

5.2 Comparison to Empirical Results

An important question in financial economics is whether active portfolio managers have skill. Our model allows us to address this question, in part, by asking what the distribution of skills across managers would have to look like to generate observed outcomes in a rational model such as ours. That is the objective of this section.

We proceed by first tying down the model parameters that can be inferred directly from existing evidence. The parameters that govern the distribution of skill level (ϕ_0 and γ) are then inferred by matching two moments we have derived closed form expressions for—the survival probabilities and the flow of funds.

The parameter f is reasonably straightforward to fix through past empirical studies of mutual funds. We use $f = 1.5\%$. This is a bit higher than the averages reported in the literature for the expense ratio. For example, Chen and Pennacchi (2002) report average expense ratios for the funds in their study of 1.14%. Our use of a slightly higher number is intended to account for the amortized loads that are not included in the expense ratio. We set σ at 20% based on a reasonable estimate of historical levels of portfolio volatility. We impose (40) as an equality to ensure that at any quantity where the firm operates, it also invests in the passive benchmark to some degree. From combining (37) and (38) we get that $\bar{\phi} = 2f = 3\%$. These parameter values are summarized in Table 1 below. It is straightforward to see (using (30) and (47)) that the only remaining parameters that affect the survival probabilities and flow of funds relationship are ϕ_0 and γ .

Our approach is to match the two parameters governing the distribution of skill level

Variable	Symbol	Value
Percentage fee	f	1.5%
Prior precision	γ	277
(Prior stan. dev.)	(η)	(6%)
Return precision	ω	25
(Return stand. dev.)	(σ)	(20%)
Mean of prior	ϕ_0	6.5%
Exit mean	$\bar{\phi}$	3%

Table 1: **Parameter Values**

using an “eyeball metric” that matches the empirical survival rates and relation between the flow of funds and performance. Table 1 contains the estimates of ϕ_0 and γ that resulted from this process. Before discussing these estimates, we report the details of this moment matching process.

The lighter bars in Figure 2 represent quantity-weighted averages of survival rates for all funds in the CRSP mutual fund data base from 1969 to 1999. That is, each year a certain number of new funds appeared in the data base, and a certain fraction of these survived for one, two,..., or nineteen years. The numbers in the figure weight these fractions by the number of funds starting in that year relative to the number of funds appearing all years, and then adds across years.⁴ The dark bars are the matched survival rates computed using the probabilities from Proposition 4. In our model survival rates drop off geometrically with age, while in the data they fall in a more linear fashion. Consequently, we had to choose which survival probabilities to match. For obvious reasons there is much more data for younger funds (see Figure 2) which means the early data is estimated much more precisely. We also found the longer term data puzzling, in that the rates seem to drop linearly. This may be due to managerial turnover within the mutual fund. That is, good managers might be promoted, leaving the fund with less talent. To minimize these concerns we chose to match to the first five years of data.

To match the relatively high survival for the first few years requires that funds begin operating at a scale that is considerably higher than the size at which they liquidate, with correspondingly high expectations for managerial ability. The priors are that managers are expected to earn an annual excess return of 6.5% when they begin operating.

High precision in priors relative to the variability in returns is required to reproduce the fund-flow and performance relationship. The slope of this relationship is determined by the

⁴We are very grateful to Hsiu-lang Chen and George Pennacchi for providing us with the raw survival rates.

Figure 2: **Percentage of Surviving Funds:** The bars show the fraction of funds as a function of the number of years in business. Light (red) bars are the actual survival rates computed from the CRSP mutual fund data base. The dark bars is (blue) what the model predicts the survival rates should be. The scale is marked on the left hand axis. The line marks the total number of funds that could have survived at each age. The scale for this line is marked on the right hand axis.

fees, f , and the ratio of the precision in priors and returns. Figure 3 provides the explicit comparison between the model's results and behaviors documented in empirical studies that we used. In it the flow of funds relationship for two-year old funds is superimposed over the plot of the non-parametric estimates and 90% confidence intervals in Chevalier and Ellison (1997). The curve from the model seems to pick up the general curvature in the relationship. A notable discrepancy in this case is that the empirical curve does not seem to go through the origin, as is required in our model. That is, we would be quite close to the observed performance-flow relation with a parallel shift of our curve up or to the left. A natural explanation for this is the general growth in the mutual fund sector during their sample period, something that is missing in our model.

Figure 3: **Flow of Funds:** This plot shows the flow of funds for 2 year old funds produced by the model (using the parameters reported in Table 1) superimposed over the actual flow of funds plot for these funds as reported in Chevalier and Ellison(1997, Fig.1). Chevalier and Ellison report the estimated curve (middle line) as well as the 90% confidence intervals, the outer lines.

Figure 4 plots the prior distribution over management ability using the inferred parameter values. It also shows the level of the management fee (1.5%). If level of skill in the economy is defined as the fraction of managers who can generate an α in excess of the fees they charge, then this fraction is the area of the curve to the right of the line. About 80% of managers

satisfy this criteria — they generate value in that they can beat their fees on at least the first dollar they manage. Furthermore, the average manager has an α of 6.5%, implying that he can generate a return of 5% over his fees.

These estimates might appear high, given the skepticism in the academic literature about whether active managers add value at all. It is a direct consequence, however, of the very high survival rates and empirically observed flow of funds relationship. Furthermore, the other parameters implied by these estimates seem reasonable. For example, the implied value of the ratio $\frac{q_0}{\bar{q}}$ (the size of a new fund over the minimum fund size) is 4.7. So, for example, if the minimum fund size is \$5 million, the model implies that new funds would start around \$25 million. At $\bar{q} = \$5$ million, (38) implies periodic fixed costs F of \$75,000. Far from implying that most managers are unskilled, the data is consistent with a surprisingly high level of skill amongst active managers.

Figure 4: **Distribution of Management Skill:** The vertical line marks the level of the management fee — 1.5%. Approximately 80% of the area below the curve lies to the right of this line. The parameter values (mean and precision) are $\bar{\phi} = 0.065$ and $\gamma = 277$.

6 Conclusion

In this paper we derive a number of empirical predictions of a rational model for active portfolio management when managerial talent is a scarce resource and is dissipated as the scale of operations increases. Many of these predictions reproduce empirical regularities that often have been taken as evidence of investor irrationality or agency costs between managers and investors. We argue that not only is the rational model consistent with much of the

empirical evidence, but it is also consistent with a surprisingly high level of skill amongst active managers.

Appendix

In this appendix we generalize the example in Section 5 to allow for the possibility that the amounts of funds that the manager can raise constrains him to not index. We begin by considering this case, that is, q_t such that $\bar{q} \leq q_t < q^c$. Using (22) and (23) gives

$$h(q_t) = 1 \quad (51)$$

$$\frac{c(q_t)}{h(q_t)} = aq_t + f. \quad (52)$$

Substituting (52) into (24) and solving provides the relation between priors and assets under management:

$$q_t = q_{at}^* = \frac{\phi_t - f}{a}. \quad (53)$$

Recall from (39) that $q^c = \frac{f}{a}$. Combining this solution with the solution in Section 5 gives the following general solution:

$$h(q_t) = \begin{cases} 1 & \text{if } q_t < \frac{f}{a} \\ \sqrt{\frac{f}{aq_t}} & \text{o.w.} \end{cases} \quad (54)$$

$$\frac{c(q_t)}{h(q_t)} = \begin{cases} aq_t + f & \text{if } q_t < \frac{f}{a} \\ 2\sqrt{aq_t f} & \text{o.w.} \end{cases} \quad (55)$$

Using (53), (38) and (39), define

$$\phi_c \equiv aq^c + f = 2f \quad (56)$$

$$\bar{\phi} \equiv a\bar{q} + f = \frac{aF}{f} + f. \quad (57)$$

When $\phi_t < \bar{\phi}$, the fund is shut down, if $\bar{\phi} \leq \phi_t < \phi_c$, the manager is constrained, otherwise he indexes. Summarizing,

$$q_t = \begin{cases} 0 & \phi_t < \frac{aF}{f} + f \\ \frac{\phi_t - f}{a} & \frac{aF}{f} + f \leq \phi_t < 2f \\ \frac{\phi_t^2}{4af} & \text{o.w.} \end{cases} \quad (58)$$

$$q_{at}^* = \begin{cases} 0 & \phi_t < \frac{aF}{f} + f \\ \frac{\phi_t - f}{a} & \frac{aF}{f} + f \leq \phi_t < 2f \\ \frac{\phi_t}{2a} & \text{o.w.} \end{cases} \quad (59)$$

$$q_{It}^* = \begin{cases} 0 & \phi_t < 2f \\ \frac{\phi_t}{4a} \left(\frac{\phi_t}{f} - 2 \right) & o.w. \end{cases} \quad (60)$$

Next we derive the flow of funds equation. There are three new cases not considered in Section 5 — $q_t, q_{t-1} \leq q^c$, $q_{t-1} < q^c < q_t$ and $q_t < q^c < q_{t-1}$. We consider each case individually.

When $q_t, q_{t-1} \leq q^c$, Proposition 2, (54), (55) and (58) after some simplification, yields the following expression for the flow of funds as a function of the manager's performance:

$$\frac{q_t - q_{t-1}}{q_{t-1}} = \frac{r_t}{\phi_{t-1} - f} \left(\frac{\omega}{\gamma + t\omega} \right) \quad (61)$$

The same logic can be used to derive a similar expression in the other cases. When $q_{t-1} < q^c < q_t$ we get

$$\frac{q_t - q_{t-1}}{q_{t-1}} = \frac{(\phi_{t-1} - 2f)^2 + 2\phi_{t-1}r_t \left(\frac{\omega}{\gamma + t\omega} \right) + r_t^2 \left(\frac{\omega}{\gamma + t\omega} \right)^2}{4f(\phi_{t-1} - f)} \quad (62)$$

and when $q_t < q^c < q_{t-1}$ we get

$$\frac{q_t - q_{t-1}}{q_{t-1}} = \frac{4f}{\phi_{t-1}} \left(1 - \frac{f}{\phi_{t-1}} \right) - 1 + 2 \frac{r_t}{\phi_{t-1}} \left(\frac{\omega}{\gamma + t\omega} \right) \quad (63)$$

Summarizing,

$$\frac{q_t - q_{t-1}}{q_{t-1}} = \begin{cases} -1 & q_t < \frac{F}{f} \\ \frac{r_t}{\phi_{t-1} - f} \left(\frac{\omega}{\gamma + t\omega} \right) & \frac{F}{f} \leq q_t, q_{t-1} \leq \frac{f}{a} \\ \frac{4f}{\phi_{t-1}} \left(1 - \frac{f}{\phi_{t-1}} \right) - 1 + 2 \frac{r_t}{\phi_{t-1}} \left(\frac{\omega}{\gamma + t\omega} \right) & \frac{F}{f} \leq q_t < \frac{f}{a} < q_{t-1} \\ \frac{(\phi_{t-1} - 2f)^2 + 2\phi_{t-1}r_t \left(\frac{\omega}{\gamma + t\omega} \right) + r_t^2 \left(\frac{\omega}{\gamma + t\omega} \right)^2}{4f(\phi_{t-1} - f)} & \frac{F}{f} \leq q_{t-1} < \frac{f}{a} < q_t \\ \frac{r_t}{f} \left(\frac{\omega}{\gamma + t\omega} \right) + \frac{r_t^2}{4f^2} \left(\frac{\omega}{\gamma + t\omega} \right)^2 & o.w. \end{cases} \quad (64)$$

Denote the realization of r_t that sets $q_t = q^c$ as $r^c(\phi_{t-1})$. This critical realization is determined by finding the r_t such that the market's posterior on the manager's ability falls to ϕ_c . Using equation (25), this value is given by solving:

$$r^c(\phi_{t-1}) = (\phi_c - \phi_{t-1})h(q_{t-1}) \frac{\gamma + t\omega}{\omega}$$

Substituting (54) and (53) and simplifying gives

$$r^c(\phi_{t-1}) = \begin{cases} (\phi_c - \phi_{t-1}) \frac{\gamma+t\omega}{\omega} & \phi_{t-1} < \phi_c \\ 2 \left(\frac{\phi_c - \phi_{t-1}}{\phi_{t-1}} \right) \left(\frac{\gamma+t\omega}{\omega} \right) f & \phi_{t-1} \geq \phi_c \end{cases} \quad (65)$$

Using (44), (56), (57), (64) and (65) the percentage change in new funds, when

$$r_t \geq \begin{cases} \left(\frac{2f}{\phi_{t-1}} \right) \left(\frac{aF}{f} + f - \phi_{t-1} \right) \left(\frac{\gamma+t\omega}{\omega} \right) & \phi_{t-1} \geq 2f \\ \left(\frac{aF}{f} + f - \phi_{t-1} \right) \left(\frac{\gamma+t\omega}{\omega} \right) & \phi_{t-1} < 2f \end{cases}$$

is

$$\frac{q_t - q_{t-1}(1+r_t)}{q_{t-1}} = \begin{cases} \frac{\frac{r_t}{\phi_{t-1}-f} \left(\frac{\omega}{\gamma+t\omega} \right) - r_t}{\frac{(\phi_{t-1}-2f)^2 + 2\phi_{t-1}r_t \left(\frac{\omega}{\gamma+t\omega} \right) + r_t^2 \left(\frac{\omega}{\gamma+t\omega} \right)^2}{4f(\phi_{t-1}-f)}} - r_t & \phi_{t-1} < 2f \text{ and } r_t < (2f - \phi_{t-1}) \frac{\gamma+t\omega}{\omega} \\ \frac{(\phi_{t-1}-2f)^2 + 2\phi_{t-1}r_t \left(\frac{\omega}{\gamma+t\omega} \right) + r_t^2 \left(\frac{\omega}{\gamma+t\omega} \right)^2}{4f(\phi_{t-1}-f)} - r_t & \phi_{t-1} < 2f \text{ and } r_t \geq (2f - \phi_{t-1}) \frac{\gamma+t\omega}{\omega} \\ \frac{4f}{\phi_{t-1}} \left(1 - \frac{f}{\phi_{t-1}} \right) - 1 + 2 \frac{r_t}{\phi_{t-1}} \left(\frac{\omega}{\gamma+t\omega} \right) - r_t & \phi_{t-1} \geq 2f \text{ and } r_t < 2 \left(\frac{2f - \phi_{t-1}}{\phi_{t-1}} \right) \left(\frac{\gamma+t\omega}{\omega} \right) f \\ \frac{r_t}{f} \left(\frac{\omega}{\gamma+t\omega} \right) + \frac{r_t^2}{4f^2} \left(\frac{\omega}{\gamma+t\omega} \right)^2 - r_t & o.w. \end{cases} \quad (66)$$

and, following the convention in the main body of the paper, otherwise $\frac{q_t - q_{t-1}(1+r_t)}{q_{t-1}} = -1$. To generate the empirical distribution we need to aggregate over all funds with the same age. Using (66), the flow of funds in period t , conditional on the fund being alive in period $t-1$ and having realized return $r > 0$ in period t is given by

$$\begin{aligned} N_t(r) = & \left[\frac{\omega}{\gamma + t\omega} \left(\frac{\int_{\bar{\phi}}^{\max(\bar{\phi}, \rho^*(r, 2f))} \frac{F_{t-1}(\phi)}{\phi-f} d\phi}{\int_{\bar{\phi}}^{\infty} F_{t-1}(\phi) d\phi} + \frac{\int_{\max(\bar{\phi}, \rho^*(r, 2f))}^{\max(\bar{\phi}, 2f)} \frac{\phi F_{t-1}(\phi)}{2f(\phi-f)} d\phi}{\int_{\bar{\phi}}^{\infty} F_{t-1}(\phi) d\phi} + \frac{\int_{\max(\bar{\phi}, 2f)}^{\infty} \frac{F_{t-1}(\phi)}{f} d\phi}{\int_{\bar{\phi}}^{\infty} F_{t-1}(\phi) d\phi} \right) - 1 \right] r \\ & + \left(\frac{\omega}{\gamma + t\omega} \right)^2 \left[\frac{\int_{\max(\bar{\phi}, \rho^*(r, 2f))}^{\max(\bar{\phi}, 2f)} \frac{F_{t-1}(\phi)}{4f(\phi-f)} d\phi}{\int_{\bar{\phi}}^{\infty} F_{t-1}(\phi) d\phi} + \frac{\int_{\max(\bar{\phi}, 2f)}^{\infty} \frac{F_{t-1}(\phi)}{4f^2} d\phi}{\int_{\bar{\phi}}^{\infty} F_{t-1}(\phi) d\phi} \right] r^2 \\ & + \frac{\int_{\max(\bar{\phi}, \rho^*(r, 2f))}^{\max(\bar{\phi}, 2f)} \frac{(\phi-2f)^2}{4f(\phi-f)} F_{t-1}(\phi) d\phi}{\int_{\bar{\phi}}^{\infty} F_{t-1}(\phi) d\phi} \end{aligned} \quad (67)$$

and when $r < 0$ it is

$$N_t(r) = \left[\frac{\omega}{\gamma + t\omega} \left(\frac{\int_{\max(\rho^c(r, \bar{\phi}), 2f)}^{\max(\rho^c(r, \bar{\phi}), \rho^c(r, \bar{\phi}))} \frac{F_{t-1}(\phi)}{\phi-f} d\phi}{\int_{\bar{\phi}}^{\infty} F_{t-1}(\phi) d\phi} + \frac{\int_{\max(\rho^c(r, \bar{\phi}), \rho^c(r, 2f))}^{\max(\rho^c(r, \bar{\phi}), 2f)} \frac{2F_{t-1}(\phi)}{\phi} d\phi}{\int_{\bar{\phi}}^{\infty} F_{t-1}(\phi) d\phi} \right) \right] \quad (68)$$

$$\begin{aligned}
& + \left. \frac{\int_{\bar{\phi}}^{\infty} \frac{F_{t-1}(\phi)}{f} d\phi}{\int_{\bar{\phi}}^{\infty} F_{t-1}(\phi) d\phi} \right) - \left. \frac{\int_{\bar{\phi}}^{\infty} F_{t-1}(\phi) d\phi}{\int_{\bar{\phi}}^{\infty} F_{t-1}(\phi) d\phi} \right] r \\
& + \left(\frac{\omega}{\gamma + t\omega} \right)^2 \left[\frac{\int_{\bar{\phi}}^{\infty} \frac{F_{t-1}(\phi)}{4f^2} d\phi}{\int_{\bar{\phi}}^{\infty} F_{t-1}(\phi) d\phi} \right] r^2 \\
& + \frac{\int_{\bar{\phi}}^{\max(\rho^c(r, \bar{\phi}), \rho^c(r, 2f))} \left(\frac{4f}{\phi} \left(1 - \frac{f}{\phi} \right) - 1 \right) F_{t-1}(\phi) d\phi}{\int_{\bar{\phi}}^{\infty} F_{t-1}(\phi) d\phi} - \frac{\int_{\bar{\phi}}^{\max(\rho^c(r, \bar{\phi}), \rho^*(r, \bar{\phi}))} F_{t-1}(\phi) d\phi}{\int_{\bar{\phi}}^{\infty} F_{t-1}(\phi) d\phi}
\end{aligned}$$

where

$$\begin{aligned}
\rho^*(r, a) & \equiv a - \frac{\omega}{\gamma + t\omega} r \\
\rho^c(r, a) & \equiv \begin{cases} \frac{a}{1 + \frac{r}{2f} \left(\frac{\omega}{\gamma + t\omega} \right)} & r > -2 \left(\frac{\gamma + t\omega}{\omega} \right) f \\ \infty & r \leq -2 \left(\frac{\gamma + t\omega}{\omega} \right) f \end{cases} .
\end{aligned}$$

References

- Bernhardt, D., R. Davies, and H. Westbrook (2002), "Smart Fund Managers? Stupid Money? A Model of Strategic Mutual Fund Investment Decisions," working paper, University of Illinois Department of Economics.
- Bollen, N.P.B., and J.A. Busse (2001), "Short-term Persistence in Mutual Fund Performance," Working Paper, University of Utah.
- Brown, K., W. Harlow and L. Starks (1996), "Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry," *Journal of Finance*, **51**: 85-100.
- Brown, S.J. and W.H. Goetzmann, (1995), "Performance Persistence," *Journal of Finance*, **50**: 679-698.
- Carhart, M., (1997), "On Persistence in Mutual Fund Performance," *Journal of Finance*, **52**: 57-82.
- Carhart, M.M., J.N. Carpenter, A.W. Lynch and D.K. Musto (2001), "Mutual Fund Survivorship," Working Paper.
- Chevalier, J. and G. Ellison (1997), "Risk Taking by Mutual Funds as a Response to Incentives," *Journal of Political Economy*, **105**: 1167-1200.
- Chevalier, J. and G. Ellison (1999), "Are Some Mutual Fund Managers Better than Others? Cross-Sectional Patterns in Behavior and Performance," *Journal of Finance*, **54**: 875-900.
- Chen, H. and G. G. Pennacchi (2002), "Does Prior Performance Affect a Mutual Fund's Choice of Risk? Theory and Further Empirical Evidence," Working Paper, University of Illinois at Urbana-Champaign.
- Chordia, T., (1996), "The structure of mutual fund charges," *Journal of Financial Economics*, **41**: 3-39.
- Daniel, K., M. Grinblatt, S. Titman and R. Wermers, (1997), "Measuring Mutual Fund Performance with Characteristic-Based Benchmarks," *Journal of Finance*, **52**: 1035-1058.
- DeGroot, M.H. (1970). *Optimal Statistical Decisions*, McGraw-Hill, New York.

- Gruber, M.J.(1996), “Another Puzzle: The Growth in Actively Managed Mutual Funds,” *Journal of Finance*, **51**: 783-810.
- Holmström, B. (1999), “Managerial Incentive Problems: A Dynamic Perspective,” *Review of Economic Studies*, **66**: 169-182.
- Hugonnier, J. and R. Kaniel (2002), “Mutual Fund Portfolio Choice in the Presence of Dynamic Flows,” working paper.
- Ippolito, R. (1989), “Efficiency with costly information: a study of mutual fund performance,” *Quarterly Journal of Economics*, **104**: 1-23.
- Ippolito, R. (1992), “Consumer reaction to measures of poor quality: evidence from the mutual fund industry,” *Journal of Law and Economics*, **35**: 45-70.
- Jensen, M. (1968), “The Performance of Mutual Funds in the Period 1945-1968,” *Journal of Finance*, **23**: 389-416.
- Lynch, A. and D. Musto, (2002), “How Investors Interpret Past Fund Returns,” working paper, NYU.
- Malkiel, B.G., (1995), “Returns from Investing in Equity Mutual Funds 1971 to 1991,” *Journal of Finance*, **50**: 549-572.
- Mamaysky, H. and M. Spiegel, (2002), “A Theory of Mutual Funds: Optimal Fund Objectives and Industry Organization,” working paper.
- Nanda, V., M.P. Narayanan and V.A. Warther, (2000), ‘Liquidity, investment ability, and mutual fund structure,’ *Journal of Financial Economics*, **57**: 417-443.
- Ross, S.A., R.W. Westerfield and J.F. Jaffe, *Corporate Finance* (6th edition), McGraw-Hill Irwin (2002).
- Sirri, E.R. and P. Tufano (1998), “Costly Search and Mutual Fund Flows,” *Journal of Finance*, **53**: 1589-1622.
- Wermers, R. (2000), “An Empirical Decomposition into Stock-Picking Talent, Style, Transaction Costs and Expenses,” *Journal of Finance*, **55**: 1655-1694.
- Zheng, L.(1999), “Is Money Smart? A Study of Mutual Fund Investors’ Fund Selection Ability,” *Journal of Finance*, **54**: 901-933.