Delegated Monitoring of Fund Managers*

Simon Gervais†
University of Pennsylvania

Anthony W. Lynch‡
New York University and NBER

David K. Musto§
University of Pennsylvania

25 July 2002

*Financial support by the Rodney L. White Center for Financial Research is gratefully acknowledged. The authors would like to thank Zhiwu Chen, Judith Chevalier, Francesca Cornelli, Heber Farnsworth, Francisco Gomes, Gary Gorton, Jonathan Ingersoll, Richard Kihlstrom, Christine Parlour, Matt Spiegel, as well as seminar participants at Carnegie Mellon University, the London Business School, the London School of Economics, Queen's University, Washington University, the Wharton School, and Yale University for their comments and suggestions. All remaining errors are of course the authors' responsibility.

†Finance Department, Wharton School, University of Pennsylvania, Steinberg Hall - Dietrich Hall, Suite 2300, Philadelphia, PA 19104-6367, gervais@wharton.upenn.edu, (215) 898-2370.

‡Stern School of Business, New York University, 44 West Fourth Street, Suite 9-190, New York, NY 10012-1126, alynch@stern.nyu.edu, (212) 998-0350.

§Finance Department, Wharton School, University of Pennsylvania, Steinberg Hall - Dietrich Hall, Suite 2300, Philadelphia, PA 19104-6367, musto@wharton.upenn.edu, (215) 898-4239.

Delegated Monitoring of Fund Managers

Abstract

Because a money manager learns more about her skill from her management experience than outsiders can learn from her realized returns, she expects inefficiency in future contracts that condition exclusively on realized returns. A fund family that learns what the manager learns can reduce this inefficiency cost if the family is large enough. The family's incentive is to retain any given manager regardless of her skill but, when the family has enough managers, it adds value by boosting the credibility of its retentions through the firing of others. In this way, large fund families add value through cross-sectional reputation. As the number of managers grows the efficiency loss goes to zero.

1 Introduction

Open-end mutual funds intermediate much of consumers' investment in financial securities. The consumers' relationship to their funds' portfolio managers (just 'managers,' from here on) has been analyzed extensively as that of principals to their agents,[1] but as these analyses generally observe, there is a potentially important layer of agency between consumers and their funds' managers. A manager does not work directly for her investors, but rather for a mutual-fund organization such as Fidelity Management Corporation which contracts with many investors on the one hand and with several managers on the other. In this paper, we analyze how this contract structure adds value using a two-period model of the money-management problem.

The questions we address are similar to the ones about banks studied by Diamond (1984), but the economic considerations are quite different. Banks facilitate an entrepreneur's borrowing against his future output when the expense of observing his output is too much for individual investors. These investors must punish when underpaid, but a bank that monitors does not, and is itself punished less the more (statistically independent) loans it handles simultaneously. A manager does not need this service, since her output is her return, which is easily observed. Her challenge is rather to sell her expected output when the expense of observing her expectation is too much. In this context, she anticipates informational inefficiency in her future contracts which, as we show, can be reduced by a fund family, and more effectively so when the family handles more managers simultaneously.

Observing a manager's expectation regarding her own prospects is expensive because of how this expectation develops. A manager learns about her prospects from her money-management experience. That is, she begins with some sense of her ability to add value, but has a better sense after actually trying. So she does not know before a period of management all that she will know about her skill afterward. Also, she learns more about her ability than her portfolio return conveys, which is to say her performance to a future date will be an insufficient statistic for what she knows about her prospects on that date. It is cheap for outsiders to observe her portfolio returns but presumably expensive to observe what she learns about herself, so the manager expects inefficiency in future contracts if those contracts impute only the cheap information. The empirical evidence (e.g., Edelen, 1999) indicates that consumers get little or none of the surplus created by mutual-fund management, so the incidence of ex post inefficiency represents an ex ante expected cost to

[1] Important papers include Bhattacharya and Pfleiderer (1985), Starks (1987), Heinkel and Stoughton (1994), Admati and Pfleiderer (1997), Huddart (1999), and Das and Sundaram (2002).

the manager that she would like to reduce. Delegated monitoring reduces this cost by imputing the manager's future learning into future contracts.

A simple model of the contracting problem makes this point in two stages. In the first stage we show the benefit of perfect, costless monitoring: the manager's expected surplus is seen to improve, relative to the public-information case, when all her future private information informs her future contract. So perfect costless monitoring adds value, but this is an incomplete rationale for delegated monitoring of fund managers because, as Diamond (1984) observes, delegated monitoring brings agency problems of its own. So the second stage delegates monitoring to a self-interested third party. The self-interest derives from the cost of replacing a manager, which encourages monitors to retain even those managers they privately know not to be skilled. What restores the monitor's credibility, and therefore recovers the value of monitoring, is simultaneous monitoring of a number of funds. This result is similar to that for bank monitoring but it applies to money management for a different reason. In the banking context, monitoring works better when loans are numerous because the incidence of default by the bank is reduced. In the money-management context, monitoring works better with more than one manager because the firing of some of a fund family' managers can boost the expected ability of those remaining.

The empirical mutual-fund literature has explored some actions of managers' employers. Khorana (1996) documents the relation between return and continued employment, and Chevalier and Ellison (1999) show demotion and separation following poor return, and promotion following high return. The related literature on delegated portfolio management is primarily concerned with screening and portfolio choice. In Huberman and Kandel (1993) and Huddart (1999), risk-averse managers know their skill ex ante, and choose their portfolios knowing that investors will infer skill from their portfolios' returns. Both analyses find pooling and separating equilibria, where the separating equilibria involve skilled managers investing in the one risky asset more aggressively, so that, win or lose, they are identifiable ex post as the skilled ones. Heinkel and Stoughton (1994) also address the situation where a manager knows her ability and investors try to infer it; they conclude that managers must beat a benchmark return to be rehired. We depart from these papers by focusing on what managers learn from experience about their ability. This focus has an antecedent in Lynch and Musto (2002), which is not about monitoring but about fund flows. When investors infer that investment advisors replace underperforming strategies, the resulting theoretical relation between returns and net investment is seen to take the convex shape documented in the empirical literature (e.g., Ippolito, 1992), and fund flows are seen empirically to respond positively to evidence of replacement. Finally, Berk and Green (2002) show that the relationship between fund flows and past fund returns dissipates performance persistence in competitive rational markets.

The problem of credible delegated monitoring has been explored in contexts other than banking. For example, Baiman, Evans and Noel (1987) study the problem of credible auditing, in which a self-interested auditor may choose not to exert an effort while performing his duties. To our knowledge, delegated monitoring of money managers has not been treated theoretically. However, many empirical regularities are documented by Tufano and Sevick (1997). One notable finding is that the average mutual-fund board member, at one of the 50 largest sponsors, is on the board of 16 funds. This multiplicity of board seats, they report, is sometimes thought to give the fund sponsor too much influence over the directors. One contribution of our analysis is to show how this arrangement can actually be efficient, and how monitoring only one fund could create no value at all.

The rest of the paper is organized as follows. In section 2, we define and motivate the two-period model. In section 3, we show the efficiency loss incurred in the absence of monitoring. In section 4, we allow the manager to work in a fund family and discuss the agency problem that prevents effective monitoring by a one-fund family. We then show when and how a two-fund family can recover value, and generalize the result to N funds. Some alternate specifications and agency problems are discussed in section 5. Finally, section 6 summarizes and concludes. All proofs are in Appendix A.

2 The Model

There are two periods ($t = 1, 2$), in each of which a different large group of small risk-neutral investors shows up: a set of A investors each have a dollar to invest in the first period, and a new set of A investors each have a dollar to invest in the second period. In both periods, uninformed investments in financial markets yield an unconditional return which is either $x \in (0, 1]$ or $-x$ with equal probabilities. So uninformed investments have a zero expected return.[2]

Each period's investors can pool their money into a fund and hire a risk-neutral agent to manage it. The cost of operating the fund is $k > 0$, which is absorbed equally by each of the investors. This cost represents expenses incurred by the fund for transacting, hiring personnel, buying computer equipment, and so on. Once hired, the fund manager generates signals which are potentially

[2] Equivalently, we can think of these returns as *excess* returns after properly accounting for risk.

informative about the returns of traded financial assets.[3] Skilled managers generate excess expected returns. More precisely, a manager with skill $\alpha \in (0,1)$ improves the probability that returns are positive to $\frac{1+\alpha}{2} > \frac{1}{2}$, whereas unskilled managers cannot improve on the unconditional return distribution. Managers are skilled with probability ϕ, and the random variable

$$\tilde{a} = \begin{cases} \alpha, & \text{prob. } \phi \\ 0, & \text{prob. } 1-\phi \end{cases} \tag{1}$$

denotes a manager's skill, which is initially unknown. In each period t, we denote the event that the manager is hired by an indicator function $\tilde{h}_t$. The investors' return in period t is therefore[4]

$$\tilde{r}_t = \begin{cases} +x, & \text{prob. } \frac{1+\tilde{h}_t\tilde{a}}{2} \\ -x, & \text{prob. } \frac{1-\tilde{h}_t\tilde{a}}{2}. \end{cases} \tag{2}$$

Skill does not translate to better returns without spending k, which is assumed large relative to individuals' investible funds, so individuals are interested in investing their money with the manager. If the manager is skilled then their expected return, gross of fees, rises from zero to $\mathrm{E}\left[\tilde{r}_t \mid \tilde{h}_t = 1, \tilde{a} = \alpha\right] = x\alpha > 0$. Hiring an unskilled manager ($\tilde{a} = 0$), on the other hand, yields no excess return, as $\mathrm{E}\left[\tilde{r}_t | \tilde{h}_t = 1, \tilde{a} = 0\right] = 0$. Therefore, the expected surplus, net of the administrative cost k, created by a skilled manager with A under management is $Ax\alpha - k$ and the expected net surplus from a manager with probability p of being skilled is $Ax\alpha p - k$.

At the beginning of the first period neither the manager nor the investors knows whether the manager is skilled, so they all calculate the probability she is skilled to be ϕ, and therefore all calculate the expected net value the manager creates to be $Ax\alpha\phi - k$. At the beginning of the second period, investors update ϕ with public information, and the manager updates ϕ with what she has privately learned. The public information used by investors is the manager's first-period return. The manager's private information tells her exactly whether or not she is skilled, meaning that for her, ϕ goes to either 0 or 1. So the public learns something about whether or not the manager is skilled, but the manager learns everything. The assumption that managers learn their skill perfectly is for computational simplicity; it is not crucial to the analysis that the manager learn perfectly, only that she learns strictly more than the public. This captures the fact that she

[3] We assume the signals are free, i.e., there is no private effort cost, so as to avoid conflating the dynamics of interest here with moral hazards. A previous version of this paper (available upon request) had a private effort cost and results were similar.

[4] We could have incorporated the operating cost of the fund into the returns by subtracting k from both $+x$ and $-x$. This would not affect any of our results. In fact, this would be equivalent to assuming that k is proportional to A.

knows not only her return but also the portfolio that generated it, why she bought that portfolio, how the realized returns of different assets compare to her ex ante expectations, and so on.

At the beginning of each period the manager announces a compensation contract that she is willing to work for, and the period's investors accept or reject the offer. This setup mirrors the mutual fund industry in that funds publicly announce their fees and investors are free to choose whether they want to invest. Given that returns can take only two possible values in each period, this contract is fully characterized by the wage that will be paid by the investors to the manager in each of the two states. We make two assumptions about contracting. First, there is no negative compensation: instead of paying, the manager would just file bankruptcy. Second, we represent the empirical findings on mutual-fund performance by assuming that the contract leaves investors zero expected surplus.

Because investors and the manager agree at the time of first-period contracting that the probability of a high first-period return is $\frac{1+\alpha\phi}{2}$, they all calculate the same expected wage for a given contract, and therefore any contract with expectation $Ax\alpha\phi - k$ under these probabilities serves equally well. The manager cannot be hired in the first place if this quantity is negative so we assume that the parameters satisfy the restriction

$$\phi \geq \frac{k}{Ax\alpha}. \tag{3}$$

Because the manager has private information after the first period about the probability of a high second-period return (it is $\frac{1+\alpha}{2}$ if the manager learned she is skilled and $\frac{1}{2}$ otherwise), there is a signaling problem with second-period contracts. Among the contracts under which investors expect zero surplus, skilled managers prefer the contract with the lowest payment for low returns. The equilibrium contract pays the minimum for low returns, which under our assumptions is zero.[5] Thus, the second-period contract the manager offers takes the form

$$\tilde{w}_2 = \begin{cases} \omega, & \text{if } \tilde{r}_2 = +x \\ 0, & \text{if } \tilde{r}_2 = -x, \end{cases} \tag{4}$$

Since ω is the contract's only moving part, we sometimes refer to the second-period contract simply as ω. Under the assumption that investors get zero expected surplus, the ω offered by the manager is the one at which investors calculate an expected net return of zero.

[5] A formal proof that this contract is signaling-proof is provided in Appendix B. It is also worth mentioning that, in a model where skill gets impounded into performance only through costly effort, the contract that pays the manager only after positive returns would also provide the strongest incentives.

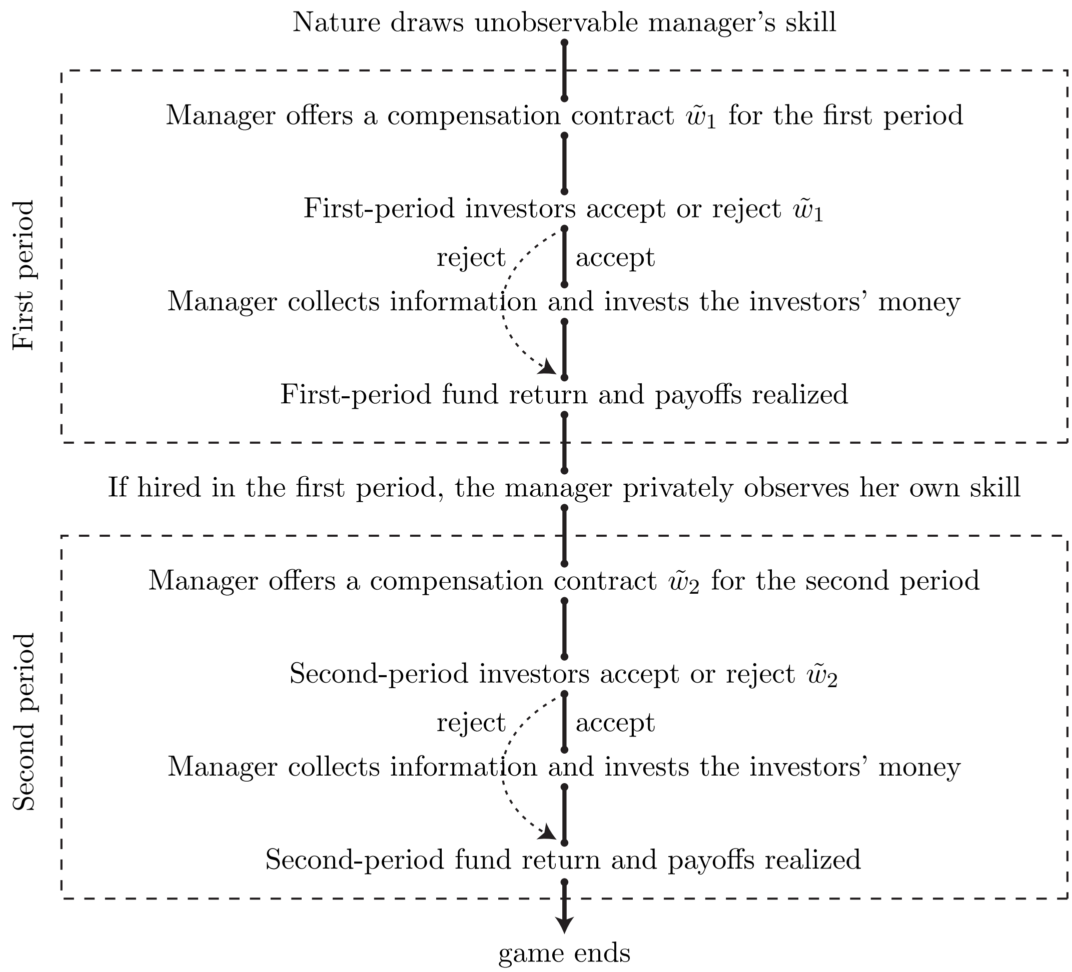

Figure 1: Sequence of events.

3 Equilibrium in the Absence of Monitoring

3.1 Strategies and Equilibrium

The chronology of this economy is in Figure 1. It begins with the manager offering investors to manage their money for one period in return for a compensation contract $\tilde{w}_1$. The investors are free to decide whether or not to hire the money manager. If hired, the manager collects information and invests the investors' money. The first period ends with the manager observing her ability and investors receiving their first-period payoffs. The second period repeats the process, except that second-period investors can use the first-period returns to update their beliefs about the manager's ability. The economy ends after the second-period payoffs.

The manager makes one decision per period, which is the compensation contract she offers, and the period's investors' one decision in each period is to accept or reject. Since the manager and investors are symmetrically informed at the outset, first-period contracting is straightforward: as mentioned above, the equilibrium contract has an expected value of $Ax\alpha\phi - k$, regardless of its form, and the investors accept it. From here on we focus exclusively on the second period, and accordingly drop the time subscript so as to minimize notation. The second-period contract has an important dependence on investors' beliefs about the manager.

Lemma 3.1 *Suppose that, at the beginning of the second period, investors assign a probability p that the manager is skilled. The manager is then hired by the investors if and only if*

$$p \geq \frac{k}{Ax\alpha}. \tag{5}$$

The manager's compensation in that period is then

$$\omega = \omega(p) \equiv \frac{2(Ax\alpha p - k)}{1 + \alpha p}. \tag{6}$$

The manager's expected net surplus, as calculated by investors, is $Ax\alpha p - k$, so condition (5) simply states that the manager will be hired when this expectation is non-negative. In that case, the manager extracts all surplus by charging $\omega(p)$ for a positive return. Note that $\omega(p)$ is increasing in p, so that both the probability of, and compensation for, a high return increase as the manager's apparent skill grows.

What is the manager's apparent skill? Investors start with prior belief ϕ and then update with the manager's realized return via Bayes' Rule.

Lemma 3.2 *After first-period fund returns of $+x$, investors assign a probability of*

$$\phi_{+x} \equiv \frac{(1+\alpha)\phi}{1+\alpha\phi} \tag{7}$$

to the manager being skilled. After first-period fund returns of $-x$, investors assign a probability of

$$\phi_{-x} \equiv \frac{(1-\alpha)\phi}{1-\alpha\phi} \tag{8}$$

to the manager being skilled.

The connection between skill and returns delivers $\phi_{+x} > \phi > \phi_{-x}$. Since ϕ was already high enough for the manager to be hired, this implies immediately that the manager is retained after good performance, and that her expected compensation conditional on good performance (but *not* on whether she is skilled) is her expected net surplus $Ax\alpha\phi_{+x} - k$. On the other hand, ϕ_{-x} could be either less or greater than $\frac{k}{Ax\alpha}$. The manager is fired if it is less, and retained if it is greater, with an expected compensation conditional on poor performance of $Ax\alpha\phi_{-x} - k$. The next section calculates the second-period compensation the manager expects at the outset, before anyone learns her ability, by combining these quantities.

3.2 Manager Profits

The manager's expected second-period compensation at the outset, which we denote π_0, is the probability of good performance times the expected compensation conditional on good performance, plus the probability of bad performance times the expected compensation, if any, conditional on bad performance. This is a simple calculation, but it is useful to arrange the result in a functional form that isolates the potential benefit of monitoring.

Proposition 3.1 *If $\phi_{-x} < \frac{k}{Ax\alpha}$, then*

$$\pi_0 = \phi(Ax\alpha - k) - \frac{(1-\alpha)\phi}{2}(Ax\alpha - k) - \frac{1-\phi}{2}k, \tag{9}$$

and if $\phi_{-x} \geq \frac{k}{Ax\alpha}$, then

$$\pi_0 = \phi(Ax\alpha - k) - (1-\phi)k. \tag{10}$$

There is one source of expected value creation and two potential sources of expected value destruction in second-period contracts. All value creation comes from the value $Ax\alpha - k$ created by skilled managers. The manager's initial probability of being skilled is ϕ, so the expected value creation from always hiring exactly the skilled managers is $\phi(Ax\alpha - k)$. The sources of expected value destruction are the firing of skilled managers and the retention of unskilled managers. Firing a skilled manager destroys $Ax\alpha - k$, i.e., the value the skilled manager would have created; retaining an unskilled manager destroys k, i.e., the administrative cost k paid for a manager who creates no value. If only good performers are retained (i.e., if $\phi_{-x} < \frac{k}{Ax\alpha}$), then the manager loses in expectation from being fired while skilled in the outcome where she does poorly but learns she is skilled, which has probability $\frac{(1-\alpha)\phi}{2}$; similarly, she loses in expectation from being retained while unskilled in the outcome where she does well but learns she is unskilled, which has probability $\frac{1-\phi}{2}$. If good and bad performers are all retained (i.e., if $\phi_{-x} \geq \frac{k}{Ax\alpha}$), then all loss comes from being retained while unskilled, which has probability $1 - \phi$.

This formulation shows how the manager loses from inefficient contracting, and what the manager would pay for a service that improves efficiency. For example, if $\phi_{-x} \geq \frac{k}{Ax\alpha}$ then the manager would pay up to $(1-\phi)k$ for a service that would, in a way that is credible to investors, simply fire her if she is unskilled. And in general she would pay for a service that decreases the incidence of being retained while unskilled by more than $\frac{Ax\alpha-k}{k}$ times the amount it increases the incidence of being fired while skilled. In the remainder of this paper we show how this service can be provided by delegated monitoring.

4 Delegated Monitoring

How can the manager incorporate her private learning into future contracts? The method we propose is to submit to delegated monitoring by working as one manager in a fund family. There are other forms delegated monitoring might take; we focus on fund families to show how this popular organizational form can deliver this important benefit. To the manager of the previous section we offer the opportunity to join a simplified fund family: for a cut of her wage, the fund family intermediates her initial contract with the public, learns her skill along with her, and then either retains or fires her. It is immediate that this arrangement recovers all value (gross of the family's cut) if the family retains exactly the skilled managers; what is more interesting and meaningful is whether it adds value in the context of realistic incentives for the family to retain or fire. We present the benchmark case of perfect monitoring, discuss the incentive problem, and then resolve the problem. The key to the resolution is seen to be the proliferation of funds within the family.

4.1 Fund Family Defined

The manager can join a fund family at the outset. To join, the manager negotiates a fraction μ of her second-period wages, if any, that will go to the family. The family offers contracts to the public. As before we assume that investors pay all expected surplus, so the first-period expected wage is again $Ax\alpha\phi - k$. The family has access to the same information as the manager and so learns her skill over the first period, and then retains or fires her. Second-period investors also pay all expected surplus so, if investors calculate a probability p that the manager is skilled, then her second period wage is $\omega(p)$.

The role of μ in the model is to represent the sensitivity of the family's wealth to the public's belief about the manager's skill, and to the manager's actual skill. Since $\omega(p)$ increases with p, the family's wealth increases with the public's belief, and since $\omega(p)$ is paid only for good returns, the family's wealth increases with the manager's skill, controlling for the public's belief. We do not model the negotiation over μ; we assume only that if monitoring creates expected value, then the negotiated μ splits this expected value between the family and the manager.

The level of μ does not represent the cost to the family of monitoring the manager. If monitoring is costly to the family then μ must recoup this cost in expectation to be economical (see section 5 for more discussion of this point). The cost of monitoring could in any case be confidentiality, not resources. In particular, it might be expensive for investors, but not the family, to learn what the manager learns because the manager must hide her beliefs and trading from the public to defend against front-running. It is intuitively safer to disclose this sensitive information to a short list of closely-watched board members, and the cost of disclosing to people with easy access to trading records and other operational data is presumably small.

The assumption that the family's income is exclusively from retained managers represents a replacement cost. Since the family would presumably recruit and earn revenue from a replacement for a fired manager, the implicit assumption is that this would be a wash, i.e., the replacement cost equals the subsequent revenue. This equality is for convenience and not important to the analysis; it is sufficient that the replacement cost be a significant deterrent to firing the initial manager. This same cost explains the zero net revenue from all managers in the first period. Income from the first period is unimportant to incentives in any case since first-period wages are the same regardless of the family's actions. Section 5 describes an alternate source of the agency problem, one fashioned after Ross (1977).

4.2 First-Best and the Agency Problem

The best monitoring can do is eliminate all contracting inefficiency by retaining a manager if and only if she is skilled. This gives the manager an expected second-period wage of $\phi(Ax\alpha - k)$, which can be compared to her expected wage in the absence of monitoring as derived in Proposition 3.1. If $\phi_{-x} < \frac{k}{Ax\alpha}$, then the expected wage improves on the unmonitored case by the amount $\frac{(1-\alpha)\phi}{2}(Ax\alpha - k) + \frac{1-\phi}{2}k$; if $\phi_{-x} \geq \frac{k}{Ax\alpha}$, then it improves by the amount $(1-\phi)k$. So first-best monitoring adds a positive amount of expected value which the family and manager can share through the μ they negotiate.

If a family's incentives were to retain only skilled managers then the analysis of delegated monitoring would end here. However, the incentives to fire are too weak because replacing the manager brings zero net revenue. So a manager gets no efficiency gains from being a family's only manager as the family would never fire its only source of net revenue.

One path to an incentive to fire unskilled managers is reputation. Reputation is typically viewed as a time series dynamic, and that dynamic could be important here. Firing is costly in the short run, but it may build reputation that pays off later through the credibility of future retentions. However, if the family can monitor more than one fund, it is not necessary to wait for reputation to pay off. Rather than time-series reputation, the important dynamic may be *cross-sectional* reputation where the firing of some managers pays off through the credibility of *simultaneous* retentions. The rest of this section shows how this can work, first in the simplest two-manager case, and then in the general N-manager case.

4.3 The Two-Fund Family

Instead of one manager in the family, now there are two managers, each with A dollars. The two managers have independently distributed skill and returns, and we allow them and the family to contract at the outset on a firing policy. The firing policies we allow are commitments by the family to fire a stated number of good performers, and a stated number of bad performers (both numbers could be zero), conditional on the performance realization. That is, with two managers there are three possible performance realizations — both good, both bad and one good and one bad — and the firing policy decides whether 0, 1 or 2 managers are fired if both have good first-period performance, how many are fired if both have bad performance, and whether the good performer, the bad performer or both are fired if one is good and the other bad. Commitment is assumed to be costless: it is costless for the family to contract on the firing policy, and the family honors the contract at no additional cost. If instead this is costly then the expected value created by monitoring must be at least this cost.

We do not allow all possible firing policies. In particular we do not allow policies where the family decides after the first period how many good performers or bad performers to fire. Such policies push the model into a complex signaling environment where the public must calculate what each possible combination of publicly-observable performance outcome and publicly-observable firing outcome implies about the privately-observable skill outcome. The number of calculations grows exponentially with the number of managers, quickly becoming unmanageable. With our restriction on firing policies the public sees exactly the firing outcome it expects, given the performance outcome, so the signaling complication is removed. Since complex signaling strategies could be useful to fund families, our results with this constrained set of policies can be viewed as lower bounds on the value families create as delegated monitors.

Firing policies add value to the extent they impute the family's private observation of skill.

Conditional on the performance realization the public knows how many good and how many bad performers will be fired but they don't know *which* good and bad performers will be fired. They also know that the family decides, the family knows who is skilled and, everything else equal, the family is better off retaining a skilled manager than an unskilled manager. So the firing policy adds value to the extent that the family has a choice whom to fire because the public learns about a manager from the fact that she was not chosen. We solve this problem by first narrowing in on the scenarios where the family has an informative choice to make, and then calculating the value added in the resulting equilibrium.

Several observations on the two-manager problem simplify the solution considerably. First, the family cannot add value by committing to firing both managers in a given outcome because the reputation created by firing both would not pay off for anybody. Second, the family does not commit to firing the poor performer (or to firing the good performer) when one performance is good and the other bad, because doing so does not tell the public anything more than it already knows from first-period returns. So the firing policy pays off only if both performances are good or both are bad, and the decisions to make at the outset are whether to fire 0 or 1 if both returns are good, and whether to fire 0 or 1 if both are bad.

If two managers make the same return $r \in \{-x, +x\}$ in the first period, then for each manager the public calculates the same probability ϕ_r that she is skilled. If the family must fire one, where it is common knowledge that the family retains a skilled manager if possible, then the probability that the retained manager is skilled is $1 - (1 - \phi_r)^2 = \phi_r(2 - \phi_r)$. Naturally, $\phi_r(2 - \phi_r) > \phi_r$ so this is a boost to the public's belief about the retained manager. Whether this boost is a good idea depends on the relative values of ϕ_r and $\frac{k}{Ax\alpha}$, as the next proposition summarizes.

Proposition 4.1 *If both managers have the same first-period return $r \in \{-x, +x\}$, then the second-period outcomes with and without delegated monitoring depend on ϕ_r and $\frac{k}{Ax\alpha}$ as follows.*

- $\phi_r(2 - \phi_r) < \frac{k}{Ax\alpha}$*: Neither manager works the second period, with or without delegated monitoring. Value created is zero.*

- $\phi_r < \frac{k}{Ax\alpha} \leq \phi_r(2 - \phi_r)$*: Neither works the second period without delegated monitoring, but in a fund family one is fired and the other works. Value created per manager is* $\frac{1}{2}\left[Ax\alpha\phi_r(2 - \phi_r) - k\right]$.

- $\phi_r^2 < \frac{k}{Ax\alpha} \leq \phi_r$*: Both work without delegated monitoring, but with a delegated monitor one is fired and the other works. Value created per manager is* $\frac{1}{2}\left(k - Ax\alpha\phi_r^2\right)$.

- $\frac{k}{Ax\alpha} \leq \phi_r^2$*: Both work with or without a delegated monitor. Value created is zero.*

If the managers are not monitored and therefore retained or fired simply on the strength of their returns, there are just two regions in which ϕ_r can fall. Either $\phi_r < \frac{k}{Ax\alpha}$ so they are both dismissed by investors, or $\phi_r \geq \frac{k}{Ax\alpha}$ so they are both retained. The proposition shows that monitoring subdivides these regions. If $\phi_r < \frac{k}{Ax\alpha}$ but $\phi_r(2-\phi_r) \geq \frac{k}{Ax\alpha}$, then one is fired and the other works. The boost is enough to allow some employment where none had been possible. If $\phi_r \geq \frac{k}{Ax\alpha}$ but $\phi_r^2 < \frac{k}{Ax\alpha}$ then again one is fired and the other works, but the reason is different. Monitoring creates value in this region because it reduces the expected loss from hiring unskilled managers more than it increases the expected loss from firing skilled managers.

All four scenarios are possible for bad performers, depending on the relation of ϕ_{-x} to $\frac{k}{Ax\alpha}$. After good performance, the only relevant question is whether $\phi_{+x}^2 < \frac{k}{Ax\alpha}$, in which case one manager goes. A useful perspective on the result is that monitoring adds value if $\frac{k}{Ax\alpha}$ is in the range $\left[\phi_r^2, \phi_r(2-\phi_r)\right]$. The width of this range is $2\phi_r(1-\phi_r)$, which is proportional to the variance of the manager's skill conditional on her return. So the measure of the space where monitoring adds expected value is linear in the uncertainty about managerial skill that is not resolved for investors by performance.

4.4 The N-Fund Family

We now generalize the fund family to N managers. In this general case, there are $N+1$ possible first-period outcomes, one for each possible number $M_{+x} \in \{0, 1, \ldots, N\}$ of good performers (the number of bad performers is then $M_{-x} = N - M_{+x}$). The firing policy dictates, for each of these outcomes, how many good performers to fire and how many poor performers to fire. As before, it is common knowledge that the family's incentive is to retain as many skilled managers as possible. We characterize the optimal firing policy for a given M_r, $r \in \{-x, +x\}$, and show that the efficiency outcome approaches first-best as N grows.

When choosing its firing policy, the fund family seeks to recover as much of the contracting inefficiency of section 3 as possible. The following proposition characterizes the optimal number $F_r(M_r)$ of M_r managers with the same first-period return r who will be fired. The proposition uses $B(I, J, p)$ to denote the binomial cumulative distribution function (cdf) for J draws, I successes and probability p of success.[6]

[6] That is, $B(I, J, p) = \sum_{i=0}^{I} \binom{J}{i} p^i (1-p)^{J-i}$.

Proposition 4.2 *If there is an integer F, $0 < F < M_r$, that satisfies*

$$B(F-1, M_r, 1-\phi_r) < \frac{k}{Ax\alpha} < B(F, M_r, 1-\phi_r), \tag{11}$$

then it is optimal for the family to commit to fire $F_r(M_r) = F$ of the M_r managers who make return $r \in \{-x, +x\}$ in the first period. If $B(0, M_r, 1-\phi_r) > \frac{k}{Ax\alpha}$, then the optimal number to fire is zero; if $B(M_r - 1, M_r, 1-\phi_r) < \frac{k}{Ax\alpha}$, then no managers work the second period regardless of monitoring.

The tradeoff used to derive this result is intuitive: as the family commits to fire a larger number of managers, the publicly perceived skilled of the remaining managers is higher (implying a larger per-manager surplus), but there are fewer managers working. The last part of Proposition 4.2 simply points out that, just as in the two-manager case, there are parameters for which firing no one is optimal, and parameters for which monitoring cannot create any value. As the following result demonstrates, these parameter ranges shrink as M_r increases.

Lemma 4.1 *For any $\phi_r > 0$ the optimal number F of managers to fire satisfies (11) for M_r large enough.*

A large enough fund family adds value because with positive probability it improves on the zero-firings case after any performance that might come from an unskilled manager, and it delivers the possibility of working again after any performance that might come from a skilled manager. In the previous section, we showed that monitoring by two-fund families adds value if $\frac{k}{Ax\alpha}$ lands in a range $2\phi_r(1-\phi_r)$ wide. Lemma 4.1 shows that M-fund families add value if $\frac{k}{Ax\alpha}$ is between $B(0, M_r, 1-\phi_r)$ and $B(M_r - 1, M_r, 1-\phi_r)$, i.e., it is between $\phi_r^{M_r}$ and $1-(1-\phi_r)^{M_r}$, which expands to the entire unit interval as M_r grows.

The important quantity in all the results is the ratio $\frac{k}{Ax\alpha}$ which, if we rewrite it as $\frac{k/A}{x\alpha}$, is seen to be the ratio of the per-dollar administrative expense to the per-dollar value added by skill. So as the cost of running a fund increases relative to the value added by skill, whether from k increasing or A or x decreasing, managers prefer policies with more ex-post firing. The effect of α by itself depends on whether performance is good or bad. If α is higher then there is less firing after good performance, for two reasons: the loss from firing a skilled manager is higher, and the likelihood that the performance came from a skilled manager is higher. The effect on firing after bad performance is ambiguous, because while the loss from firing a skilled manager is higher, the likelihood that the performance came from a skilled manager is now lower.

The firing policy described in Proposition 4.2 minimizes contracting inefficiencies for every possible first-period outcome. We now show that the remaining inefficiency goes to zero as the number of managers in the fund family grows. So, in the limit, the first-best outcome is achieved through monitoring. We start with the following observation.

Lemma 4.2 *As M_r grows, $\frac{F_r(M_r)}{M_r}$ converges to $1 - \phi_r$.*

As the number of managers with a given return increases, the fraction of managers fired converges to the fraction of these managers expected ex ante to be unskilled. This observation leads to our final result.

Proposition 4.3 *As the number of funds in the family grows, contracting efficiency converges to first-best.*

In the limit, the entire efficiency loss is recaptured and shared by the family. We reach this conclusion even though we consider only those firing policies that commit to the relation between two future observable and verifiable events: the number of managers with a given performance, and the number of those fired. If the contracting parties can do better by giving the family more ex post flexibility, then the convergence to first-best is faster but the limit is the same. Finally, it is interesting to note that the proofs to all the results in this section do not rely on the return distribution having two points; the same results would apply in any finite outcome space.

5 Additional Considerations

In this section we outline a few alternatives to the basic model: a per-manager monitoring cost the family must recover, the possibility of merging rather than replacing, and an alternate source of the agency problem that discourages firing.

5.1 Costly Monitoring

Suppose that monitoring a manager costs c. In this case the family loses money unless its expected per-manager income is at least c, so the negotiated μ must deliver at least this expectation to the family. This rules out families where the per-manager surplus from the firing policy is below c. If c exceeds all the efficiency losses in Proposition 3.1, i.e., $\frac{1-\alpha\phi}{2}(Ax\alpha - k) + \frac{1-\phi}{2}k$ if $\phi_{-x} < \frac{k}{Ax\alpha}$ and $(1-\phi)k$ otherwise, then the managers are better off managing alone and unmonitored than joining families, as even perfect efficiency does not recover c. If instead c is smaller than the relevant quantity, then there is a number N such that the per-manager surplus in families with at least N managers is above c. In this case, the operative question that a monitoring cost raises is whether managers can organize into large enough families to make delegated monitoring pay.

5.2 Merging Funds

In the analysis, the cost of recruiting a replacement manager discourages firing. In practice however, replacement is not strictly necessary because merging is an option. Fund families can, with the appropriate shareholder and regulatory approvals, merge fired managers' funds into retained managers' funds. If we add to our modelling assumptions that merging is costless then, as long as at least one manager is skilled, it is straightforward that the family fires all unskilled managers, retains all skilled managers and merges accordingly. This breaks down only when no manager is skilled, a contingency whose probability goes to zero as the number N of managers goes to infinity.

Merging is, however, not costless and could easily cost much more than replacement. In addition to the direct restructuring costs, there are costs from mismatching investors and diluting the value added by skill. Investors would resist mergers across investment objectives, and even within investment objectives, fund families lose price-discrimination power when they merge funds with different operating histories (Christoffersen and Musto, 2002). Also, the expected return $x\alpha$ from skill could shrink as assets grow, as in Berk and Green (2002), and if it falls below k/A then active management costs more than it produces. In short, this mechanism may restore some of the incentive to fire the unskilled but it replaces one set of costs with another.

5.3 Alternate Agency Problem

The agency problem associated with the firing of unskilled managers is more general than just replacement costs. Suppose there is no replacement cost and suppose instead that, in the style of Ross (1977), the family's expected income I depends on the manager's true and publicly-perceived probabilities of being skilled through the relation

$$I = \gamma_0(\text{perceived probability}) + \gamma_1(\text{true probability}).$$

As in Ross (1977), γ_0 captures short-term effects of the public's perception, which in the money-management business would include the reputation as well as new investment attracted to the fund and other funds in the complex, whereas γ_1 captures longer-term effects such as the effect of the manager's true skill on expected future fees and reputation. The pure-strategy equilibrium of a one-fund family depends on the relative size of γ_0 and γ_1. If $\gamma_0 < \gamma_1 \frac{\phi}{1-\phi}$ then, in equilibrium, the family replaces the manager if and only if she is unskilled, so efficiency obtains.[7] This equilibrium does not exist if $\gamma_0 > \gamma_1 \frac{\phi}{1-\phi}$, so the resulting outcome is inefficient.

In this context, committing the family ex ante to replace F of M managers who look the same to the public can be viewed as holding constant the number that γ_0 multiplies, and using the sensitivity γ_1 to incorporate the family's learning into second-period contracts. Thus, delegated monitoring recovers contracting efficiency even when the monitor's exposure to true skill is relatively small.

6 Conclusion

Decades of empirical studies have established that it is very hard to estimate portfolio-manager skill from realized portfolio returns. The portfolio manager herself has more than her portfolio's returns to learn from: she knows why she chose the initial portfolio, how and why she traded it, how the realizations of individual securities' returns compare with what she expected, and so on. She learns more from a period of experience than outsiders learn from a period of portfolio returns, so it stands to reason that she enters the next period knowing more about her skill than outsiders know. A manager with a good return could privately know she is *not* worth hiring, and a manager with a poor return could privately know she *is* worth hiring. We consider whether the manager would prefer to avoid getting into this situation in the first place, and whether a monitoring mechanism could serve this purpose. Our principal finding is that a monitoring mechanism like that of open-end mutual funds can improve the utility of managers, and that the multiple-fund monitoring noted by Tufano and Sevick (1997) is crucial to the effectiveness of monitoring.

The literature has analyzed the situation where managers know more about their skill than do outsiders. We step back from this situation to the beginning of a manager's career, when she intuitively knows little more than outsiders know about whether she adds value. If she expects the surplus in future contracts, then she expects a benefit from credible monitoring, but credibility is not automatic. In particular, a one-manager fund family has little credibility if it has discretion whether to fire its one manager and chooses not to, and it serves no purpose without discretion. We show how increasing the number of funds monitored by the monitor can work around this incentive problem. We make this point in the context of money management but the principle applies more generally. The key elements of the economic situation are that the agent learns about skill over time, that someone else could also learn at sufficiently low cost, and that a number of other agents

[7] If the public expects the family to retain only skilled managers, then the family gets γ_0 from retaining an unskilled manager and $(\gamma_1 + \gamma_0)\phi$ from replacing the manager with a rookie (with a rookie's probability ϕ of being skilled), so the belief is rational only for $\gamma_0 < \gamma_1 \frac{\phi}{1-\phi}$.

are in a similar situation. This could describe the situation of, for example, assistant professors.

We do not claim to have identified the only reason why fund managers may want to gather under the umbrella of a fund family. There are undoubtedly reasons other than the value that more credible monitoring creates (e.g., marketing reasons). However, our framework could also be useful to study other aspects of the money-management industry. For instance, it may be interesting to consider when a fund manager would be better off leaving a fund family to manage independently (e.g., start a hedge fund). Our analysis suggests that managers who believe their skill has been established beyond a doubt would see little benefit from monitoring, as would managers who perceive little difference between what they will privately learn about their skill and what their returns will convey.

The result that monitoring captures more value as the number of funds grows may relate to the recent development of "funds of funds."[8] Many funds have recently launched with the sole intention of investing in existing hedge funds. Given the close relation that investors can enjoy with the hedge funds they patronize, this trend may reflect a demand from hedge-fund managers for credible monitoring. How managers sort into those who manage independently, those who manage partly for funds-of-funds and those who manage within mutual-fund families is a promising area for future research.

We analyze the situation where a manager's performance is publicly observable from the moment she starts managing. In practice there is some discretion because the family could have the manager handle house money, or pretend money, for a while. The manager would intuitively have to cross a threshold of promise, as calculated from this practice experience, to start handling public money. Outsiders could infer this threshold and therefore know, as the model assumes, the manager's promise at that point. So the model still applies in the bigger picture where managers practice first, but it would be interesting to see what the optimal threshold is in this framework.

[8]We are grateful to Zhiwu Chen for pointing out this possibility to us.

Appendix A

Proof of Lemma 3.1

Let us denote the investors' information set at the beginning of period t by $\mathcal{I}_t$. Since investors assign a probability p that the manager is skilled, they expect her to generate an average net surplus of

$$\begin{aligned} & A\Big(\mathrm{E}\big[\tilde{r}_t \,|\, \mathcal{I}_t, \tilde{h}_t = 1, \tilde{a} = \alpha\big] \Pr\big\{\tilde{a} = \alpha \,|\, \mathcal{I}_t\big\} + \mathrm{E}\big[\tilde{r}_t \,|\, \mathcal{I}_t, \tilde{h}_t = 1, \tilde{a} = \alpha\big] \Pr\big\{\tilde{a} = 0 \,|\, \mathcal{I}_t\big\}\Big) - k \\ & \quad = \quad A\big[x\alpha p + 0(1-p)\big] - k \quad = \quad Ax\alpha p - k \end{aligned}$$

with their money if they hire her. The manager will be hired by the investors if and only if this last quantity is positive; this is equivalent to (5). In that case, since the investors assign a probability of

$$\begin{aligned} & \Pr\big\{\tilde{r}_t = +x \,|\, \mathcal{I}_t, \tilde{h}_t = 1, \tilde{a} = \alpha\big\} \Pr\big\{\tilde{a} = \alpha \,|\, \mathcal{I}_t\big\} + \Pr\big\{\tilde{r}_t = +x \,|\, \mathcal{I}_t, \tilde{h}_t = 1, \tilde{a} = 0\big\} \Pr\big\{\tilde{a} = 0 \,|\, \mathcal{I}_t\big\} \\ & \quad = \quad \frac{1+\alpha}{2}\phi + \frac{1}{2}(1-\phi) \quad = \quad \frac{1+\alpha p}{2} \end{aligned}$$

to the manager generating positive returns in the period, the manager extracts all of the surplus by announcing a compensation contract of

$$\omega_t = \omega(p) \equiv \frac{2(Ax\alpha p - k)}{1 + \alpha p}.$$

This completes the proof. ▮

Proof of Lemma 3.2

This result follows from a simple application of Bayes' rule:

$$\begin{aligned} \phi_{+x} \quad & \equiv \quad \Pr\big\{\tilde{a} = \alpha \,\big|\, \tilde{r}_1 = +x, \tilde{h}_1 = 1\big\} \\ & = \quad \frac{\Pr\big\{\tilde{r}_1 = +x \,\big|\, \tilde{a} = \alpha, \tilde{h}_1 = 1\big\} \Pr\big\{\tilde{a} = \alpha \,\big|\, \tilde{h}_1 = 1\big\}}{\sum\limits_{a=0,\alpha} \Pr\big\{\tilde{r}_1 = +x \,\big|\, \tilde{a} = a, \tilde{h}_1 = 1\big\} \Pr\big\{\tilde{a} = a \,\big|\, \tilde{h}_1 = 1\big\}} \\ & = \quad \frac{\frac{1+\alpha}{2}\phi}{\frac{1+\alpha}{2}\phi + \frac{1}{2}(1-\phi)} \quad = \quad \frac{(1+\alpha)\phi}{1+\alpha\phi}. \end{aligned}$$

Similarly,

$$\begin{aligned}
\phi_{-x} &\equiv \Pr\{\tilde{a} = \alpha \mid \tilde{r}_1 = -x, \tilde{h}_1 = 1\} \\
&= \frac{\Pr\{\tilde{r}_1 = -x \mid \tilde{a} = \alpha, \tilde{h}_1 = 1\} \Pr\{\tilde{a} = \alpha \mid \tilde{h}_1 = 1\}}{\sum_{a=0,\alpha} \Pr\{\tilde{r}_1 = -x \mid \tilde{a} = a, \tilde{h}_1 = 1\} \Pr\{\tilde{a} = a \mid \tilde{h}_1 = 1\}} \\
&= \frac{\frac{1-\alpha}{2}\phi}{\frac{1-\alpha}{2}\phi + \frac{1}{2}(1-\phi)} \quad = \quad \frac{(1-\alpha)\phi}{1-\alpha\phi}. \quad \blacksquare
\end{aligned}$$

Proof of Proposition 3.1

If $\phi_{-x} < \frac{k}{Ax\alpha}$, then the manager works only after a good first-period return so π_0 is the probability of a good first-period return times the expected surplus conditional on a good first-period return, i.e., $\frac{1+\alpha\phi}{2}\left[Ax\alpha\left(\frac{(1+\alpha)\phi}{1+\alpha\phi}\right) - k\right]$, which rearranges to (9).

If $\phi_{-x} \geq \frac{k}{Ax\alpha}$, then the manager works the second period no matter what, so π_0 is the same as the first-period expected surplus $Ax\alpha\phi - k$, which rearranges to (10). $\blacksquare$

Proof of Proposition 4.1

If $\phi_r(2-\phi_r) < \frac{k}{Ax\alpha}$, then Lemma 3.1 tells us that the probability that the retained manager is skilled, conditional on her return and the fact that she was retained, is too low for her to work the second period, so neither manager works the second period.

If $\phi_r < \frac{k}{Ax\alpha}$, we know from Lemma 3.1 that the managers cannot work the second period without monitoring, but if $\phi_r(2-\phi_r) \geq \frac{k}{Ax\alpha}$, then the retained manager's probability of being skilled is high enough to work. The total surplus for the retained manager is $Ax\alpha\phi_r(2-\phi_r) - k$, which is shared equally by the two managers at the outset.

If $\phi_r \geq \frac{k}{Ax\alpha}$ then, from Lemma 3.1 once again, both managers would work the second period after making r in the first if there were no firing policy, and would get the expected surplus $Ax\alpha\phi_r - k$. If the fund family commits to firing one of them then the expected surplus of the retained manager is $Ax\alpha\phi_r(2-\phi_r) - k$, and each manager's expected surplus at the outset, conditional on both managers making r in the first period, is $\frac{1}{2}\left[Ax\alpha\phi_r(2-\phi_r) - k\right]$. So the expected value added per manager by the firing policy is $\frac{1}{2}\left[Ax\alpha\phi_r(2-\phi_r) - k\right] - (Ax\alpha\phi_r - k) = \frac{1}{2}(k - Ax\alpha\phi_r^2)$. So this is the value added by the firing policy, and it is positive if and only if $\phi_r^2 < \frac{k}{Ax\alpha}$.

Finally, if $\phi_r^2 \geq \frac{k}{Ax\alpha}$, then the expected value created by firing one of the two managers, calculated in the previous paragraph, is negative. So neither is fired and both work (since ϕ_r is then greater than $\frac{k}{Ax\alpha}$). $\blacksquare$

Proof of Proposition 4.2

We prove the result for a given first-period return $r \in \{-x, +x\}$, and a given number M of managers who generate that return. Let U represent the number of the M managers who are unskilled. Suppose that the family commits to firing F of M managers who make return r, and let $T_1(F)$ and $T_2(F)$ be the probabilities that a manager calculates at the outset that she will be fired when skilled and retained when unskilled, respectively. The family maximizes expected value by minimizing total contracting inefficiency, i.e., by choosing F to minimize

$$T_1(F)(Ax\alpha - k) + T_2(F)k.$$

When the family's policy is to fire F managers, some skilled managers end up getting fired when fewer than F of the M managers are unskilled; that is, $F - U$ skilled managers get fired when $U < F$. Thus the expected number of skilled managers who get fired is

$$\sum_{U=0}^{F-1} \binom{M}{U} (1-\phi_r)^U \phi_r^{M-U} (F-U),$$

and so a manager's ex ante probability of being fired while skilled is

$$T_1(F) = \frac{1}{M} \sum_{U=0}^{F-1} \binom{M}{U} (1-\phi_r)^U \phi_r^{M-U} (F-U).$$

This means that the commitment to fire an additional manager increases the ex ante probability of being fired while skilled by

$$T_1(F+1) - T_1(F) = \frac{1}{M} \sum_{U=0}^{F} \binom{M}{U} (1-\phi_r)^U \phi_r^{M-U} = \frac{1}{M} B\big(F, M, 1-\phi_r\big).$$

Similarly, when the family's policy is to fire F managers, unskilled managers are retained when more than F managers are unskilled; that is $U - F$ unskilled managers are retained when $U > F$. The expected number of unskilled managers who are retained is

$$\sum_{U=F+1}^{M} \binom{M}{U} (1-\phi_r)^U \phi_r^{M-U} (U-F),$$

and so a manager's ex ante probability of being retained while unskilled is

$$T_2(F) = \frac{1}{M} \sum_{U=F+1}^{M} \binom{M}{U} (1-\phi_r)^U \phi_r^{M-U} (U-F).$$

Firing an additional manager increases the ex ante probability of being retained while unskilled by[9]

$$T_2(F+1) - T_2(F) = -\frac{1}{M}\sum_{U=F+1}^{M}\binom{M}{U}(1-\phi_r)^U\phi_r^{M-U} = -\frac{1}{M}\Big[1 - B\big(F, M, 1-\phi_r\big)\Big].$$

So if the family moves from a policy of firing F to a policy of firing $F+1$, total efficiency loss increases by

$$\big[T_1(F+1) - T_1(F)\big](Ax\alpha - k) + \big[T_2(F+1) - T_2(F)\big]\,k = \frac{1}{M}\Big[B\big(F, M, 1-\phi_r\big)Ax\alpha - k\Big].$$

This quantity is increasing in F and has the same sign as $B(F, M, 1-\phi_r) - \frac{k}{Ax\alpha}$. So if there is an interior solution for F, it is the first F for which $B(F, M, 1-\phi_r) > \frac{k}{Ax\alpha}$, that is the F that satisfies

$$B\big(F-1, M, 1-\phi_r\big) < \frac{k}{Ax\alpha} < B\big(F, M, 1-\phi_r\big). \tag{12}$$

For each of the $M - F$ retained managers, outside investors will calculate a probability of being skilled greater than $\frac{k}{Ax\alpha}$, since the probability that *all* retained managers are skilled is $B(F, M, 1-\phi_r)$ which, from (12), is greater than $\frac{k}{Ax\alpha}$. So F managers are fired and $M - F$ work the second period. If $B(0, M, 1-\phi_r) > \frac{k}{Ax\alpha}$ then firing zero is optimal. If $B(M-1, M, 1-\phi_r) < \frac{k}{Ax\alpha}$, then the probability that at least one manager is skilled is below $\frac{k}{Ax\alpha}$ so no manager can work no matter how many are fired. ∎

Proof of Lemma 4.1

The value of $B(0, M, 1-\phi_r)$ is ϕ_r^M, which asymptotes to 0 as M grows so, for M large enough, it is below $\frac{k}{Ax\alpha}$. Similarly, the value of $B(M-1, M, 1-\phi_r)$ is $1-(1-\phi_r)^M$, which asymptotes to 1 as M grows so, for M large enough, it is above $\frac{k}{Ax\alpha}$. ∎

Proof of Lemma 4.2

Let the random variable $\tilde{U}$ represent the number of unskilled managers when M managers generate the same first-period return $r \in \{-x, +x\}$. This random variable follows a binomial distribution $\text{Bin}\big(M, 1-\phi_r\big)$. Let $\epsilon_1 = \frac{1}{2}\frac{k}{Ax\alpha}$. From the weak law of large numbers we know that for any $\delta \in (0,1)$ there is an m_1 such that $M > m_1$ implies

$$\Pr\left\{\frac{\tilde{U}}{M} \le (1-\phi_r) - \delta\right\} < \epsilon_1 < \frac{k}{Ax\alpha}. \tag{13}$$

[9] Of course, this quantity is negative, reflecting the fact that firing an additional manager actually *decreases* the likelihood that an unskilled manager is retained.

From Lemma 4.1, we know that for M large enough ($M > m_1'$, say),

$$\frac{k}{Ax\alpha} < B\big(F, M, 1-\phi_r\big) = \Pr\left\{\frac{\tilde{U}}{M} \leq \frac{F}{M}\right\}. \tag{14}$$

Together, (13) and (14) imply that

$$\frac{F}{M} > (1-\phi_r) - \delta \tag{15}$$

for $M > \max(m_1, m_1')$.

Similarly, if we let $\epsilon_2 = \frac{1}{2}\left(1 - \frac{k}{Ax\alpha}\right)$, we know that for the same δ as above there is an m_2 such that $M > m_2$ implies

$$\Pr\left\{\frac{\tilde{U}}{M} \geq (1-\phi_r) + \delta\right\} < \epsilon_2 < 1 - \frac{k}{Ax\alpha}. \tag{16}$$

From Lemma 4.1, we also have $\Pr\{\tilde{U} \leq F-1\} = B\big(F-1, M, 1-\phi_r\big) < \frac{k}{Ax\alpha}$ for M large enough ($M > m_2'$, say), which is equivalent to

$$\Pr\left\{\frac{\tilde{U}}{M} \geq \frac{F}{M}\right\} > 1 - \frac{k}{Ax\alpha}. \tag{17}$$

Together, (16) and (17) imply that

$$\frac{F}{M} < (1-\phi_r) + \delta \tag{18}$$

for $M > \max(m_2, m_2')$. So, for $M > \max(m_1, m_1', m_2, m_2')$, (15) and (18) imply that $(1-\phi_r) - \delta < \frac{F}{M} < (1-\phi_r) + \delta$. Since we can make δ arbitrarily small, this means that $\frac{F}{M}$ converges to $1-\phi_r$, as desired. ∎

Proof of Proposition 4.3

There are two elements to the proof: (i) contracting efficiency for M managers who all make return $r \in \{-x, +x\}$ converges to first-best as M goes to infinity; (ii) the effect of the number N of managers on the distribution of M, combined with (i), implies convergence as N goes to infinity.

(i) Contracting efficiency for M managers who make r converges to first-best as M goes to infinity if the expected fraction of unskilled managers who are retained and the expected fraction of skilled managers who are fired converge to zero as M goes to infinity. As in the proof of Lemma 4.2, we let the random variable $\tilde{U}$ represent the number of unskilled managers when M managers generate the same first-period return $r \in \{-x, +x\}$.

Expected fraction of retained unskilled managers goes to zero. Let us denote the fraction of retained unskilled managers by $\tilde{f}_{\mathrm{U}} \equiv \frac{\max(0,\tilde{U}-F)}{M}$. For any $\delta \in (0,1)$, we have

$$\begin{aligned} \mathrm{E}\big[\tilde{f}_{\mathrm{U}}\big] &= \Pr\big\{\tilde{f}_{\mathrm{U}} \le \delta\big\}\mathrm{E}\big[\tilde{f}_{\mathrm{U}} \mid \tilde{f}_{\mathrm{U}} \le \delta\big] + \Pr\big\{\tilde{f}_{\mathrm{U}} > \delta\big\}\mathrm{E}\big[\tilde{f}_{\mathrm{U}} \mid \tilde{f}_{\mathrm{U}} > \delta\big] \\ &\le 1\cdot\delta + \Pr\big\{\tilde{f}_{\mathrm{U}} > \delta\big\}\cdot 1 \;=\; \delta + \Pr\big\{\tilde{f}_{\mathrm{U}} > \delta\big\} \\ &= \delta + \Pr\left\{\frac{\tilde{U}-F}{M} > \delta\right\} \;=\; \delta + \Pr\left\{\frac{\tilde{U}}{M} > \frac{F}{M} + \delta\right\}. \end{aligned} \tag{19}$$

From Lemma 4.2 (and its proof), we know we can choose an m such that $M > m$ implies both that $(1-\phi_r) - \frac{\delta}{2} < \frac{F}{M} < (1-\phi_r) + \frac{\delta}{2}$ and also that $\Pr\{\frac{\tilde{U}}{M} > (1-\phi_r) + \frac{\delta}{2}\} < \delta$. These inequalities imply

$$\delta > \Pr\left\{\frac{\tilde{U}}{M} > (1-\phi_r) + \frac{\delta}{2}\right\} = \Pr\left\{\frac{\tilde{U}}{M} > \left[(1-\phi_r) - \frac{\delta}{2}\right] + \delta\right\} \ge \Pr\left\{\frac{\tilde{U}}{M} > \frac{F}{M} + \delta\right\},$$

so that $\mathrm{E}\big[\tilde{f}_{\mathrm{U}}\big]$ in (19) is smaller than 2δ. Since we can make δ arbitrarily small, this implies that the expected fraction of retained unskilled managers goes to zero.

Expected fraction of fired skilled managers goes to zero. Let us denote the fraction of fired skilled managers by $\tilde{f}_{\mathrm{S}} \equiv \frac{\max(0,F-\tilde{U})}{M}$. For any $\delta \in (0,1)$, we have

$$\begin{aligned} \mathrm{E}\big[\tilde{f}_{\mathrm{S}}\big] &= \Pr\big\{\tilde{f}_{\mathrm{S}} \le \delta\big\}\mathrm{E}\big[\tilde{f}_{\mathrm{S}} \mid \tilde{f}_{\mathrm{S}} \le \delta\big] + \Pr\big\{\tilde{f}_{\mathrm{S}} > \delta\big\}\mathrm{E}\big[\tilde{f}_{\mathrm{S}} \mid \tilde{f}_{\mathrm{S}} > \delta\big] \\ &\le 1\cdot\delta + \Pr\big\{\tilde{f}_{\mathrm{S}} > \delta\big\}\cdot 1 \;=\; \delta + \Pr\big\{\tilde{f}_{\mathrm{S}} > \delta\big\} \\ &= \delta + \Pr\left\{\frac{F-\tilde{U}}{M} > \delta\right\} \;=\; \delta + \Pr\left\{\frac{\tilde{U}}{M} < \frac{F}{M} - \delta\right\}. \end{aligned} \tag{20}$$

Again, from Lemma 4.2, we know we can choose an m such that $M > m$ implies both that $(1-\phi_r) - \frac{\delta}{2} < \frac{F}{M} < (1-\phi_r) + \frac{\delta}{2}$ and also that $\Pr\{\frac{\tilde{U}}{M} < (1-\phi_r) - \frac{\delta}{2}\} < \delta$. These inequalities imply

$$\delta > \Pr\left\{\frac{\tilde{U}}{M} < (1-\phi_r) - \frac{\delta}{2}\right\} = \Pr\left\{\frac{\tilde{U}}{M} < \left[(1-\phi_r) + \frac{\delta}{2}\right] - \delta\right\} \ge \Pr\left\{\frac{\tilde{U}}{M} < \frac{F}{M} - \delta\right\},$$

so that $\mathrm{E}\big[\tilde{f}_{\mathrm{S}}\big]$ in (20) is smaller than 2δ. Since we can make δ arbitrarily small, this implies that the expected fraction of fired skilled managers goes to zero. This completes the first part of the proof.

(ii) To finish the proof, observe that the number $\tilde{M}$ of managers who generate a first-period return of r follows a binomial distribution with N draws and probability of success p equal to either $\frac{1-\phi\alpha}{2}$ for $r = -x$ or to $\frac{1+\phi\alpha}{2}$ for $r = +x$. Thus from the weak law of large numbers we know that $\frac{\tilde{M}}{N}$ converges in probability to p. From the first part of the proof, we know that for any $\delta \in (0,1)$ there is an m such that $\tilde{M} > m$ implies that the expected fraction of the $\tilde{M}$ managers who are skilled and fired or unskilled and retained is less than $\frac{\delta}{2}$. The convergence of $\frac{\tilde{M}}{N}$ to p means that there is an n such that $N > n$ implies that $\Pr\{\frac{\tilde{M}}{N} \leq \frac{m}{N}\} < \frac{\delta}{2}$ (since $\frac{m}{N}$ goes to zero and p stays the same as N grows). So the expected fraction of managers who are skilled and fired or unskilled and retained is less than $\frac{\delta}{2} + \frac{\delta}{2} = \delta$. Therefore, since we can make δ arbitrarily small, efficiency losses converge to zero as N goes to infinity. This completes the proof. ∎

Appendix B

Proof that $\tilde{w}_2$ in (4) is signaling-proof in the second period

At the beginning of the second period, the manager knows whether or not she is skilled. The investors don't, and so update the probability that the manager is skilled to, say, $\phi' \in (0,1)$. Any contract offered to the investors by a skilled manager can and has to be imitated by the unskilled manager (i.e., the managers pool in equilibrium): otherwise, the manager is immediately identified as unskilled, and she is not hired by the second-period investors.

In equilibrium, any contract offered to the second-period investors extracts all of the available surplus, denoted S for this proof. So, if the contract offered by the manager pays $\omega_{+x} \geq 0$ for positive returns, and $\omega_{-x} \geq 0$ for negative returns, it will be the case that, in this pooling equilibrium,

$$\frac{1+\alpha\phi'}{2}\omega_{+x} + \frac{1+\alpha\phi'}{2}\omega_{-x} = S,$$

which implies that ω_{+x} must satisfy

$$\omega_{+x} = \omega_{-x} + \frac{2(S-\omega_{-x})}{1+\alpha\phi'}. \tag{21}$$

The expected compensation of a skilled manager is then

$$\frac{1+\alpha}{2}\omega_{+x} + \frac{1-\alpha}{2}\omega_{-x}$$

which, using (21), is equal to

$$\frac{1+\alpha}{1+\alpha\phi'}S - \frac{\alpha(1-\phi')}{1+\alpha\phi'}\omega_{-x}.$$

This last expression is clearly decreasing in ω_{-x}, so the skilled manager will want to reduce it as much as possible. Given limited liability, it will be the case that the skilled manager prefers a contract with $\omega_{-x} = 0$ to any other contract allowing investors to break even. Thus the equilibrium second-period contract is of the form specified in (4). ∎

References

Admati, A., and P. Pfleiderer, 1997, “Does It All Add Up? Benchmarks and the Compensation of Active Portfolio Managers,” *Journal of Business*, 70, 323-350.

Baiman, S., J. H. Evans III, and J. Noel, 1987, “Optimal Contracts with a Utility-Maximizing Auditor,” *Journal of Accounting Research*, 25, 217-244.

Berk, J. B., and R. C. Green, 2002, “Mutual Fund Flows and Performance in Rational Markets,” Working Paper, University of California at Berkeley and Carnegie Mellon University.

Bhattacharya, S., and P. Pfleiderer, 1985, “Delegated Portfolio Management,” *Journal of Economic Theory*, 36, 1-25.

Chevalier, J., and G. Ellison, 1999, “Career Concerns of Mutual Fund Managers,” *Quarterly Journal of Economics*, 114, 389-432.

Christoffersen, S. E. K., and D. K. Musto, 2002, “Demand Curves and the Pricing of Money Management,” *Review of Financial Studies*, forthcoming.

Das, S. R., and R. K. Sundaram, 2002, “Fee Speech: Signaling, Risk-sharing and the Impact of Fee Structures on Investor Welfare,” *Review of Financial Studies*, forthcoming.

Diamond, D. W., 1984, “Financial Intermediation and Delegated Monitoring,” *Review of Economic Studies*, 51, 393-414.

Edelen, R. M., 1999, “Investor Flows and the Assessed Performance of Open-End Mutual Funds,” *Journal of Financial Economics*, 53, 439-466.

Heinkel, R., and N. Stoughton, 1994, “The Dynamics of Portfolio Management Contracts,” *Review of Financial Studies*, 7, 351-387.

Huberman, G., and S. Kandel, 1993, “On The Incentives for Money Managers: A Signalling Approach,” *European Economic Review*, 37, 1065-1081.

Huddart, S. 1999, “Reputation and Performance Fee Effects on Portfolio Choice by Investment Advisers,” *Journal of Financial Markets*, 2, 227-271.

Ippolito, R. A., 1992, "Consumer Reaction to Measures of Poor Quality: Evidence from the Mutual Fund Industry," *Journal of Law and Economics*, 35, 45-70.

Khorana, A., 1996, "Top Management Turnover: An Empirical Investigation of Mutual Fund Managers," *Journal of Financial Economics*, 40, 403-426.

Lynch, A. W., and D. K. Musto, 2002, "How Investors Interpret Past Fund Returns," Working Paper, New York University.

Ross, S. A., 1977, "The Determination of Financial Structure: The Incentive-Signalling Approach," *Bell Journal of Economics*, 8, 23-40.

Starks, L., 1987, "Performance Incentive Fees: an Agency Theoretic Approach," *Journal of Financial and Quantitative Analysis*, 22, 17-32.

Tufano, P., and M. Sevick, 1997, "Board Structure and Fee-Setting in the U.S. Mutual Fund Industry," *Journal of Financial Economics*, 46, 321-355.