Market Discipline and Internal Governance in the Mutual Fund Industry*

Thomas Dangl
ISK Vienna and
Vienna University of Technology
Coburgbastei 4/1, A-1010 Vienna
Phone: 0043-51818-970
thomas.dangl@iskwien.at

Youchang Wu
Department of Finance, University of Vienna
Bruennerstrasse 72, 1210 Vienna
Phone: 0043-1-4277-38211
youchang.wu@univie.ac.at

Josef Zechner
Department of Finance, University of Vienna
Bruennerstrasse 72, 1210 Vienna
Phone: 0043-1-4277-38072
josef.zechner@univie.ac.at

First version: January 2004
This version: November 2005

*A previous version of this paper was circulated under the title “Mutual Fund Flows and Optimal Manager Replacement”. We thank Michael Brennan, Engelbert Dockner, Robert Elliot, Marcin Kacperczyk, Kristian Miltersen, Neal Stoughton, Suresh Sundaresan, Russ Wermers, Stefan Zeisberger, participants of seminars at the University of Vienna, University of Tübingen, City University London, Lancaster University, University of Cologne, Norwegian School of Economics and Business Administration, UCLA, UC Berkeley, Stanford University, HEC Paris, the European Finance Association 2004 meeting in Maastricht, and German Finance Association 2004 meeting in Tübingen for valuable suggestions and comments. Youchang Wu and Josef Zechner gratefully acknowledge the financial support by the Austrian Science Fund (FWF) under grant SFB 010 and by the Gutmann Center for Portfolio Management at the University of Vienna.
Market Discipline and Internal Governance in the Mutual Fund Industry

Abstract

We develop a continuous-time model in which a portfolio manager is hired by a management company. Based on observed portfolio returns, all agents update their beliefs about the manager’s skills. In response, investors can move capital into or out of the mutual fund and the management company can fire the manager. Introducing firing rationalizes several empirically documented findings, such as the positive relation between manager tenure and fund size or the increase of portfolio risk before a manager replacement and the following risk decrease. The analysis predicts that the critical performance threshold which triggers firing increases significantly over a manager’s tenure and that management replacements are accompanied by capital inflows when a young manager is replaced, but may be accompanied by capital outflows when a manager with a long tenure is fired. Our model yields much lower valuation levels for management companies than simple applications of DCF methods and is thus more consistent with empirical observations.
The portfolio management industry has grown substantially over the last few decades, thereby generating increased interest among practitioners, regulators, and academics. The question of efficient governance of delegated portfolio management has attracted special attention. Most theoretical models have addressed this issue using the standard principal-agent paradigm to characterize the optimal contracts that alleviate the agency problem between fund investors and managers.\(^1\) Several empirical studies have investigated the effectiveness of incentive fees\(^2\) and, in a somewhat different vein, the role of funds’ boards of directors in controlling agency costs.\(^3\)

This paper takes a broader view of the governance mechanisms in the portfolio management industry by simultaneously taking into account internal governance as well as the disciplining effect of the product market. A key characteristic of the mutual fund industry is that fund investors are at the same time consumers. The open-end structure of mutual funds allows individual investors to “fire” the fund manager by withdrawing their money whenever they feel dissatisfied with the investment management services he provides. This not only disciplines the manager directly, as pointed out by Fama and Jensen (1983), but also gives the management company strong incentives to fire underperforming managers in order to avoid losing market share. We formalize both product market discipline and internal manager replacement in a fully dynamic framework. Our model shows that the interplay between these two alternative governance mechanisms is the key to understanding many phenomena observed in the mutual fund market.

In our model, the portfolio manager may have some stock picking ability that generates an abnormal expected rate of return by taking idiosyncratic risks. However, active portfolio management exhibits a diseconomy of scale. Furthermore, the manager’s ability to manage a specific fund is unknown to the management company, to the fund investors and to the manager himself. All agents in the model start with a common prior distribution of managerial ability and update their beliefs using the observed fund performance. Investors have perfect mobility, they can move money into or out of a fund without any cost. Within such

---


\(^2\)See, for example, Elton, Gruber, and Blake (2003).

\(^3\)See, for example, Tufano and Sevick (1997) for the case of open-end funds and Guercio, Dann, and Partch (2003) for the case of closed-end funds.
a framework, we address several main questions. First, how much capital is invested into or withdrawn from the fund for a given portfolio performance and how does that fund flow depend on managerial characteristics such as tenure or uncertainty of managerial ability? Second, given these product market forces, what is the optimal manager replacement policy for the fund management company and what are its main determinants? Third, what are typical valuation levels for portfolio management companies and how do they evolve over manager tenure? And fourth, what fund flows and portfolio risk changes are induced by a manager replacement?

We consider both the case in which the precision of the belief about managerial ability remains constant as well as the case in which the precision of the belief increases over time. For both cases learning implies a positive and convex relation between unexpected idiosyncratic portfolio returns and fund inflows.\(^4\) In the case in which the precision of the belief increases over time, the fund flow responses are stronger early in a manager’s career. By contrast, portfolio returns due to general market movements trigger fund outflows, as documented empirically by Warther (1995) and Fant (1999).

Several of this paper’s main results focus on the portfolio manager replacement decision. We derive an inverse relationship between the probability of manager replacement and past fund performance. This is in accordance with empirical findings in Khorana (1996) and Chevalier and Ellison (1999a). In our model manager turnover is more performance sensitive in the early years of a manager’s tenure and managers with longer tenure tend to manage larger funds and have a higher probability to retain their positions. These predictions are confirmed by Chevalier and Ellison (1999a), Chevalier and Ellison (1999b) and Fortin, Michelson, and Jordan-Wagner (1999).

The analysis also generates new predictions about manager replacement for which empirical evidence is not yet readily available. Whenever the precision of the belief about managerial ability increases over time, the critical ability level at which firing takes place increases substantially over the manager’s tenure. Thus, the management company may find it optimal to fire a manager who is believed to have above average ability. This is so since the high

\(^4\)The positive and convex relationship between performance and fund flows has been documented, for example, by Ippolito (1992), Gruber (1996), Chevalier and Ellison (1997), Sirri and Tufano (1998), Bergstresser and Poterba (2002), and Boudoukh, Richardson, Stanton, and Whitelaw (2003).
precision implies that it is highly unlikely that the manager will ever become a “star”, even if he is believed to be above average. It is better to hire a new manager with a lower expected ability but with more upside potential.

We find that the management company’s decision to fire a portfolio manager is accompanied by capital flows and by changes in the risk of the fund portfolio. For most parameter values a manager replacement is preceded by capital outflows and a portfolio risk increase, then followed by capital inflows and a portfolio risk decrease. However, if a manager with a sufficiently long tenure is fired and the volatility of managerial ability is sufficiently low, then the model predicts the opposite. In such cases we expect the manager replacement to be followed by capital outflows and an increase in the risk of the fund portfolio. The analysis also reveals how these responses vary cross-sectionally with parameters such as the uncertainty about managerial talent or the cost of a manager replacement.

In our model the management company represents a contingent claim on the talent of the portfolio manager. Thus, applying methods from contingent claims pricing, the paper provides a consistent framework to analyze the value of the management company. Very little is known about the market values and the value drivers of management companies. In a recent paper, Huberman (2004) provides some evidence from stock market listed management companies. The reported valuation levels are generally below five percent of the assets under management, which is much less than traditional DCF valuation methods would imply. By contrast, the real options approach derived in this paper produces significantly lower values than the DCF approach and is thus much more in line with empirically observed values. The numerical analysis also implies a particular time pattern of the value of the management company over the tenure of the portfolio manager.

Our theory is most closely related to Berk and Green (2004). As in their model, we also assume competitive provision of capital by investors, decreasing returns to scale in active portfolio management, and learning about managerial ability via past portfolio returns. We extend their model by explicitly distinguishing between the management company and the portfolio manager. While Berk and Green (2004) focus on a fund with an exogenous shut-down threshold, we analyze a fund management company as a contingent claim on the posterior belief about the portfolio manager’s ability in an infinite horizon, continuous-time
model. We emphasize the management company's governance role and allow the management company to control the belief process at any time by firing the portfolio manager. This allows us to derive the optimal replacement strategy and empirical fund flow and portfolio risk patterns associated with manager replacement. It also allows us to develop simple valuation expressions for the management company which can be calibrated to empirically observable parameter values.

Lynch and Musto (2003) develop a two-period model to explain the convexity of the flow-performance relation. Their key insight is that underperforming funds will change their strategies while those outperforming will not. Therefore bad past performance contains less information about future performance than good past performance does. Taking this effect into account, rational investors will be less sensitive to past performance when it is poor. Our model differs from the Lynch and Musto (2003) model in several respects. Lynch and Musto (2003) derive their results in a partial equilibrium setting without diseconomies of scale whereas we allow capital flows to adjust to equate risk adjusted expected future returns across funds. In this setup learning about managerial ability implies a convex relation between performance and fund flows even in the absence of manager replacement. In Lynch and Musto (2003) the convex relation between performance and fund flow only arises if underperforming funds adjust their strategies. Furthermore, we analyze the manager replacement decision in a fully dynamic framework, taking into account flow responses, replacement costs, and the option value of postponing the replacement.

Table 1 summarizes the main testable hypotheses generated by our model, in comparison with Berk and Green (2004) and Lynch and Musto (2003). It also indicates which hypotheses are supported by existing empirical evidence and which hypotheses remain to be tested.

The rest of the paper is structured as follows: Section 1 analyzes the determinants and dynamics of equilibrium fund size and fund flows. Section 2 derives the valuation model for the management company and examines the optimal manager replacement rule. Section 3 examines the distribution of manager tenure and the relations between manager tenure, manager turnover, fund size and the value of the management company. Section 4 analyzes fund flows and risk changes around the event of a manager replacement. Section 5 concludes.
Table 1: Testable hypotheses and empirical evidence

This table summarizes the main testable hypotheses generated by our model, in comparison with Berk and Green (2004) and Lynch and Musto (2003). It also indicates which hypotheses are supported by existing empirical studies and which of these hypotheses remain to be tested.

<table>
<thead>
<tr>
<th>Main predictions of this model</th>
<th>Berk-Green</th>
<th>Lynch-Musto</th>
<th>Empirical support?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive and convex flow - performance relation</td>
<td>yes</td>
<td>yes</td>
<td>yes$^a$</td>
</tr>
<tr>
<td>Negative tenure - flow relation</td>
<td>yes$^b$</td>
<td>yes$^b$</td>
<td>indirect$^c$</td>
</tr>
<tr>
<td>Negative fund size - flow relation</td>
<td>no</td>
<td>no</td>
<td>yes$^d$</td>
</tr>
<tr>
<td>Negative flow response to market return</td>
<td>yes$^d$</td>
<td>no</td>
<td>yes$^e$</td>
</tr>
<tr>
<td>Relation between fee, fund size and idiosyncratic risk</td>
<td>yes</td>
<td>no</td>
<td>untested$^f$</td>
</tr>
<tr>
<td>Negative performance - manager turnover relation</td>
<td>no</td>
<td>yes</td>
<td>yes$^g$</td>
</tr>
<tr>
<td>Increasing firing threshold over tenure</td>
<td>no</td>
<td>no</td>
<td>untested</td>
</tr>
<tr>
<td>Decreasing manager turnover over tenure</td>
<td>no</td>
<td>no</td>
<td>yes$^h$</td>
</tr>
<tr>
<td>Positive tenure - fund size relation</td>
<td>no</td>
<td>no</td>
<td>yes$^i$</td>
</tr>
<tr>
<td>Decreasing management company value to fund size ratio over tenure</td>
<td>no</td>
<td>no</td>
<td>untested</td>
</tr>
<tr>
<td>Flow and risk changes around manager turnover</td>
<td>no</td>
<td>no</td>
<td>preliminary$^j$</td>
</tr>
</tbody>
</table>

$^a$See, for example, Chevalier and Ellison (1997), Sirri and Tufano (1998), Bergstresser and Poterba (2002), and Boudoukh, Richardson, Stanton, and Whitelaw (2003).

$^b$Note, however, that Berk and Green (2004) and Lynch and Musto (2003) do not distinguish between fund age and manager tenure.

$^c$Chevalier and Ellison (1997) and Boudoukh, Richardson, Stanton, and Whitelaw (2003) find that younger funds attract more inflows and have higher flow sensitivity to past performance.

$^d$While this relation is inherent in Berk and Green (2004), they do not derive it explicitly.


$^f$Although supportive evidence has been found by Golec (1996), who documents a negative relation between fund size and fund risk and a positive relation between fund risk and management fee, our prediction that management fee is proportional to the product of fund size and squared idiosyncratic risk, i.e., Equation (14), remains to be tested.

$^g$See Khorana (1999a), and Chevalier and Ellison (1999b).


$^i$See Fortin, Michelson, and Jordan-Wagner (1999) and Chevalier and Ellison (1999b) for mutual funds and Boyson (2003) for hedge funds.

$^j$See Khorana (2001) for the change of fund risk around manager replacement and The Economist (2003) for anecdotal evidence on flow response to manager changes.

1 Learning and mutual fund flows

1.1 The dynamics of NAV

The net asset value per share, denoted by NAV, of an open-end fund is assumed to evolve as follows:

$$\frac{d\text{NAV}_t}{\text{NAV}_t} = [r + \lambda \sigma_m + \alpha_t - \delta_t - b_t]dt + \sigma_m dW_{mt} + \sigma_t dW_{it}$$  \hspace{1cm} (1)
where $r$ is the risk free rate, $\lambda$ is the market price of risk, $\alpha_t$ is the abnormal expected rate of return generated by the manager due to his stock-picking ability, $W_{mt}$ and $W_{it}$ are two uncorrelated standard Wiener processes driving the stochastic part of the market return and the idiosyncratic component of the fund’s return respectively. The subscript $t$ denotes the incumbent manager’s tenure, $\sigma_m$ is the fund’s constant exposure to market risk while $\sigma_{it}$ is the fund’s time-varying exposure to idiosyncratic risk, $\delta_t$ is the instantaneous dividend yield and $b_t$ is the instantaneous management fee ratio.

Our specification of the NAV dynamics explicitly accounts for active portfolio management. The term $r + \lambda \sigma_m$ is the fair return given the fund’s exposure to systematic risk. The abnormal expected rate of return, $\alpha_t$, can be interpreted as Jensen’s $\alpha$. The dividend yield, $\delta_t$, and the management fee ratio, $b_t$, are subtracted from the NAV return because they represent cash paid out to the investors and the management company respectively. The exposure to market risk, $\sigma_m$, is assumed to be constant since we do not intend to model the manager’s market-timing ability.\(^5\) In contrast to the constant exposure to market risk, the fund’s exposure to idiosyncratic risk, $\sigma_{it}$, can be changed by the manager at any time.

The abnormal expected rate of return is assumed to have the following functional form:

$$\alpha_t = \sigma_{it}(\theta_t - \gamma A_t \sigma_{it})$$

where $\theta_t$ is the incumbent manager’s stock-picking ability, $A_t$ is the market value of assets under management and $\gamma$ is a positive constant characterizing the decreasing return to scale of active portfolio management.

The manager’s ability, $\theta_t$, is assumed to be specific to a management company. Thus, we assume that the abnormal return is a joint product of the manager and the management company. A good manager in one company is not necessarily a good manager in another company because every company has its own organizational structures, research networks and business culture. This assumption implies that our model is essentially a matching model in the spirit of Jovanovic (1979). It allows us to abstract from the observable heterogeneity

---

\(^5\)There is little evidence showing that mutual fund managers have market-timing ability. However, recent studies using portfolio holdings data do provide support for the notion that some managers have superior stock-picking ability. See, for example, Grinblatt and Titman (1993), Daniel, Grinblatt, Titman, and Wermers (1997), Wermers (2000).
among the manager candidates available to replace the incumbent manager.

Our specification of the expected abnormal return has several desirable features. First, the expected abnormal return per unit of idiosyncratic risk, i.e., \( \theta_t - \gamma A_t \sigma_{it} \), is positively related to managerial ability \( \theta_t \), negatively related to fund size \( A_t \) and idiosyncratic risk \( \sigma_{it} \). This implies that active portfolio management exhibits diseconomies of scale and that the marginal return of taking idiosyncratic risk is decreasing. An important reason for the diseconomy of scale is the price impact of large portfolio transactions. Consider a manager who is able to identify a small number of undervalued stocks. If he is managing a small fund, he can invest the entire fund capital in these stocks and earn a high abnormal rate of return. However, if he is managing a large fund, doing so would move the prices of those stocks and erode his performance.\(^6\) Decreasing return to taking idiosyncratic risk is a necessary condition to rule out unlimited expected profit opportunities.

Second, there is an interaction between fund size and idiosyncratic risk. This has two implications: (1) The larger the fund size, the smaller the marginal return of taking idiosyncratic risk. This should be the case due to the larger price impact associated with larger fund size. (2) The more idiosyncratic risk a fund takes, the larger the diseconomy of scale it has to face. A fund with high idiosyncratic risk is likely to hold less liquid stocks and exhibit more concentrated holdings. Such funds would suffer the most from being large. However, a well-diversified fund, such as an index fund, will not be hurt as much by its large size.\(^7\)

\(^6\)Many authors have investigated this issue empirically. See for example, Perold and Salomon (1991), Indro, Jiang, Hu, and Lee (1999), Chen, Hong, Huang, and Kubik (2004).

\(^7\)Our formulation of the abnormal return can be interpreted in another way. Suppose the manager divides his assets under management into two parts: one inactive part with systematic risk \( \sigma_m \) and zero idiosyncratic risk, and one actively managed part with systematic risk \( \sigma_m \) and idiosyncratic risk \( \sigma_e \). The weights of these two components are \( 1 - w_t \) and \( w_t \) respectively. The inactive part delivers no abnormal return and does not suffer from any diseconomy of scale. The actively managed part produces an abnormal expected rate of return due to the managerial ability \( \theta_t \). But the abnormal return per unit of idiosyncratic risk decreases with the size of the actively managed part. It decreases faster when the idiosyncratic risk is higher, since the portfolio with higher idiosyncratic risk is less liquid. Therefore, the abnormal return of the fund can be written as

\[
\alpha_t = w_t \sigma_e (\theta_t - w_t \sigma_e \gamma A_t).
\]

This specification is equivalent to Equation (2) since the idiosyncratic risk of the whole portfolio is \( \sigma_{it} = w_t \sigma_e \).
1.2 Inference about managerial ability

We assume that neither the fund management company, nor the fund investors, nor the manager himself, can directly observe managerial ability, $\theta_t$. Therefore, the process $W_{it}$ is not observable either. In other words, when agents in our model observe a high (low) market-adjusted NAV return, they cannot be sure whether this is due to good (bad) luck or due to the manager’s ability. All the other terms in Equation (1) are assumed to be observable. The agents share a common prior belief: $\theta_0$ is normally distributed with a mean $a_0$ and variance $v_0$, and they all use Bayes’ rule to update their beliefs as they observe the realized NAV process.

We assume that the true ability, $\theta_t$, is changing randomly over time and can be described by a driftless Wiener process:

$$d\theta_t = \omega dW_{\theta t}$$

(3)

where $\omega$ is the instantaneous volatility of the true managerial ability, $W_{\theta t}$ denotes a standard Wiener process driving $\theta_t$. $W_{\theta t}$ is assumed to be uncorrelated with the Wiener processes driving the market return and idiosyncratic return, i.e., $W_{mt}$ and $W_{it}$, respectively. Equation (3) is motivated as follows. A manager may improve his investment skill by learning from his experience and this may lead to an upward drift of $\theta_t$. However, it is well-recognized that the business environment is changing rapidly over time, implying that old strategies and trading models can be outdated very quickly. Sometimes past experience might even be an obstacle to future success. When these two effects offset each other on average, we end up with a driftless process for managerial ability. Our specification also nests the special case of constant true managerial ability, which corresponds to $\omega = 0$.

To characterize the learning process, we substitute Equation (2) into Equation (1), and move all the directly observable terms to the left-hand side and denote the resulting expression by $d\pi_t$:

$$d\pi_t := \frac{dNAV_t}{NAV_t} - [(r + \lambda \sigma_m - \nu A_t \sigma_i^2 - \delta_t - b_t)dt + \sigma_m dW_{mt}].$$

(4)
Note that $W_{mt}$ is directly observable as long as the agents know both the true expected and realized market return, which we assume they do.

Using the new notation, we can rewrite Equation (1) as

$$d\pi_t = \theta_t \sigma_t dt + \sigma_t dW_t.$$  \hspace{1cm} (5)

Learning in our model consists of updating the belief about a manager’s time-varying $\theta_t$ from the observed history of $d\pi$. Given our model specifications, it follows directly from nonlinear filtering theory (see Lipster and Shiryayev (1978)) that the posterior distribution of $\theta_t$ is normal at every point of time. The posterior mean and variance of $\theta_t$, denoted by $a_t$ and $v_t$ respectively, evolve according to the following differential equation system:  

\begin{align*}
  da_t &= \frac{v_t}{\sigma_t} [d\pi_t - a_t \sigma_t dt], \hspace{1cm} (6) \\
v_t &= (\omega^2 - v_t^2) dt. \hspace{1cm} (7)
\end{align*}

We refer to $v_t$ as the uncertainty about managerial ability and to $\frac{1}{v_t}$ as the precision of the belief about managerial ability.

The posterior variance is a deterministic function of the manager’s tenure. It can be solved explicitly as follows,

$$v_t = \begin{cases} 
\omega \frac{v_0 - \omega + (v_0 + \omega)e^{2\omega t}}{\omega - v_0 + (v_0 + \omega)e^{2\omega t}} & \text{for } \omega > 0, \\
\frac{v_0}{\omega^2 t + 1} & \text{for } \omega = 0.
\end{cases} \hspace{1cm} (8)$$

One can easily see that $v_t$ converges monotonically to $\omega$ as $t$ goes to infinity. Intuitively, if we start with an a priori belief which has a high variance compared to the instantaneous volatility of true managerial ability, i.e., $v_0 > \omega$, then the precision of the belief will improve over time, i.e., $v_t$ will gradually decline. However, if $\omega > 0$, then $v_t$ can never go to zero since the true ability keeps changing over time. Therefore $v_t$ is bounded by $\omega$ and we have $\lim_{t \to \infty} v_t = \omega$. Similarly, if $v_0 < \omega$, $v_t$ will increase over time and converge to $\omega$.  

---


\(^9\)If $\omega = 0$, then $v_t$ converges to zero, i.e., the true ability will be perfectly known as the manager tenure $t$
Define $Z_t$ as the innovation process of unexpected returns to idiosyncratic risks such that its increment $dZ_t$ is a normalized measure of the deviation of $d\pi_t$ from its posterior mean, $a_t \sigma_{it} dt$:

$$dZ_t := \frac{d\pi_t - a_t \sigma_{it} dt}{\sigma_{it}}, \quad Z_0 = 0.$$  

Since $Z_t$ measures the unexpected idiosyncratic return, it represents the signal on which the updating of the belief is based. By construction $Z_t$ is a standard Wiener process conditional on the common information set of all agents in the model. Unlike the unobservable $W_t$ process, the $Z_t$ process is derived from an observable process and is thus observable. Rewriting the dynamics of $\text{NAV}_t$ and $a_t$ in terms of $dZ_t$, we have:

$$\frac{d\text{NAV}_t}{\text{NAV}_t} = \left[ r + \lambda \sigma_m + (a_t \sigma_{it} - \gamma A_t \sigma_{it}^2) - \delta_t - b_t \right] dt + \sigma_m dW_m + \sigma_{it} dZ_t,$$

$$da_t = v_t dZ_t,$$

where $v_t$ is given by Equation (8). Note that the instantaneous volatility of the posterior mean, $a_t$, is equal to the posterior variance, $v_t$.

### 1.3 Equilibrium fund size

We assume that after paying a one-time setup cost, the management company charges a proportional fee $b_t$ for its services. The operating cost of managing the fund, including the compensation to the fund manager, is assumed to be a fixed fraction $s$ of the total fee income. Therefore, the instantaneous net profit of the management company is $b_t (1 - s) A_t$. This specification is consistent with a linear sharing rule between the fund manager and the management company.

In the open-end fund market, investors will allocate more capital to funds whose abnormal expected rate of return is higher than the management fee and withdraw money from funds for which the opposite is true. For simplicity, we assume that such fund flows are free of charge, i.e., there is perfect capital mobility. Since information is symmetric, investors update their beliefs about managerial ability in the same way as the manager and the management goes to infinity.
company do, and they monitor the size and idiosyncratic risk of the fund to decide whether it is worthwhile to invest in it.

Due to the diseconomy of scale, the performance of funds tends to decrease after capital inflows whereas performance tends to improve after capital outflows. Assuming that the uncertainty about managerial ability contains only diversifiable risk, the size of every fund will adjust so that the abnormal expected return is equal to the management fee, as postulated by Berk and Green (2004):\(^{10}\)

\[
\sigma_{it}(a_t - \gamma A_t \sigma_{it}) = b_t. \tag{11}
\]

Given the instantaneous belief about managerial ability, the management company chooses a fee \(b_t\) and a level of idiosyncratic risk \(\sigma_{it}\) to maximize the value of the management company. Since we assume that these parameters can be adjusted costlessly and since the choice of these parameters does not affect learning, \(b_t\) and \(\sigma_{it}\) will be set at levels that maximize the instantaneous net fee income given the free capital movement constraint (Equation 11) and the constraints that \(b_t, \sigma_{it}\) and \(A_t\) must all be positive. Formally, the management company solves the following maximization problem,

\[
\max_{b_t, \sigma_{it}} \quad b_t(1 - s)A_t,
\]

s.t. \[\sigma_{it}(a_t - \gamma A_t \sigma_{it}) = b_t,\]

\[b_t > 0, \sigma_{it} > 0, A_t > 0.\]

Solving for \(A_t\) from Equation (11) and substituting into the objective function, i.e., the net fee income, we see immediately that the net fee income depends only on the ratio \(\frac{b_t}{\sigma_{it}}\), but not on \(b_t\) or \(\sigma_{it}\) per se. Since the net fee income is a quadratic function of \(\frac{b_t}{\sigma_{it}}\), it is maximized at \(\frac{b_t}{\sigma_{it}} = \frac{a_t}{2}\). When \(a_t \leq 0\), the problem has no solution, since the constraints cannot be met jointly. This means that managers whose estimated ability is non-positive get no assets to manage and are thus driven out of business. However, when \(a_t > 0\), the management company has the flexibility of choosing any combination of \(b_t\) and \(\sigma_{it}\) such that

\(^{10}\)Ippolito (1992), Edelen (1999) and Wermers (2000) provide supportive empirical evidence for this condition.
the ratio equals $\frac{a_t}{b_t}$. Therefore, our model does not determine a unique optimal value of $b_t$ or $\sigma_{it}$, but it does predict a one-to-one correspondence between $b_t$ and $\sigma_{it}$ for a given $a_t$.\footnote{Golec (1996) documents a positive relation between management fee ratio and fund idiosyncratic risk.}

To identify the influence of the belief about managerial ability on fund size, we assume that $b_t$ is constant over time and denote it by $b$.\footnote{In practice, management fees are usually very stable. Thus our assumption accords well with empirical evidence.} Clearly, when $b_t$ is fixed at $b$, $\sigma_{it}$ and $A_t$ adjust as the belief about managerial ability is updated over time. More specifically, we have the following proposition:

**Proposition 1.** The equilibrium size $A_t^*$ of an open-end fund and the optimal level of idiosyncratic risk $\sigma_{it}^*$ are given by

\[
A_t^* = \begin{cases} 
\frac{a_t^2}{4\gamma} & \text{if } a_t > 0 \\
\text{non-existent} & \text{if } a_t \leq 0
\end{cases}
\tag{12}
\]

\[
\sigma_{it}^* = \begin{cases} 
\frac{2b}{a_t} & \text{if } a_t > 0 \\
\text{non-existent} & \text{if } a_t \leq 0
\end{cases}
\tag{13}
\]

**Proof.** See the discussion preceding the proposition. \qed

Note that the equilibrium fund size $A_t^*$ is a convex function of the posterior mean of managerial ability $a_t$. This explains why the level of idiosyncratic risk $\sigma_{it}$ is negatively related to the posterior mean of managerial ability $a_t$. Since the marginal return of taking idiosyncratic risk is smaller and decreases faster for larger funds, managers with a higher estimated ability and thus managing larger funds, find it optimal to take less risk.

Solving for $a_t$ from Equation (13) and substituting into Equation (12), we obtain an equilibrium relation between management fee, fund risk and fund size:

\[
b = \gamma \sigma_{it}^2 A_t.
\tag{14}
\]

This equation predicts a linear relation between management fee and the product of the variance of a fund’s idiosyncratic returns and fund size, with the coefficient being a measure of the diseconomy of scale. This specification has the advantage that it does not include managerial
ability, which is unobservable. Since fee, size and idiosyncratic risk are all observable, one can directly test this equation. The above specification also suggests a novel way to identify diseconomies of scale. In our setup, the diseconomy of scale does not show up in diminished returns of large funds. However, it can be backed out from the above equation. Since the diseconomy of scale parameter is likely to differ across fund sectors, the coefficient in the regression should also differ cross-sectionally.

1.4 The dynamics of fund size and fund flows

By Ito’s Lemma, we can derive the dynamics of the equilibrium fund size from Equation (12) and (10):

$$\frac{dA_t}{A_t} = \frac{1}{4b\gamma A_t}(da_t)^2 + \frac{a_t}{2b\gamma A_t}da_t$$

$$= \frac{v_t^2}{4b\gamma A_t}dt + \frac{v_t}{\sqrt{b\gamma A_t}}dZ_t.$$ (15)

Equation (15) shows that the fund has a positive expected growth rate, which is positively related to the uncertainty about the managerial ability $v_t$, and negatively related to the fund size $A_t$. The positive expected growth rate is due to the convex relation between the equilibrium fund size and the estimated managerial ability, which implies that the fund size is more responsive to the upward adjustment than to the downward adjustment of the posterior mean of the managerial ability.

The net proportional fund flow is defined as the fund’s asset growth rate minus its NAV return (the sum of capital gain and dividend yield), which represents the asset growth rate in excess of the growth that would have occurred if the net fund flow had been zero and if dividends had been fully reinvested, i.e.,

$$FLOW := \frac{dA_t}{A_t} - \left(\frac{dNAV_t}{NAV_t} + \delta_t dt\right).$$ (16)

Substituting Equation (11) into (9), we get

$$\frac{dNAV_t}{NAV_t} = (r + \lambda \sigma_m - \delta_t)dt + \sigma_m dW_m + \sigma_{ii} dZ_t.$$ (17)
From Equations (15) (16) and (17), and noting that $\sigma_{it} = \frac{b}{\sqrt{b \gamma A_t}}$ according to Proposition 1, we obtain the following equation that characterizes the fund flow:

$$FLOW = \left( \frac{v_t^2}{4b \gamma A_t} - r - \lambda \sigma_m dW_{mt} + \frac{v_t - b}{\sqrt{b \gamma A_t}} dZ_t \right) dt - \sigma_m dW_{mt} + v_t \Gamma \left( \frac{v_t - b}{\sqrt{b \gamma A_t}} dZ_t \right).$$ (18)

Equation (18) implies a rich set of predictions on fund flow dynamics which are consistent with the stylized empirical facts. It shows that the net fund flows can be decomposed into three parts: the expected inflows, the response to unexpected market returns, and the response to unexpected idiosyncratic returns. Both the expected rate of inflows, $\frac{v_t^2}{4b \gamma A_t} - r - \lambda \sigma_m$, and the sensitivity of fund flows to unexpected idiosyncratic returns, $\frac{v_t - b}{\sqrt{b \gamma A_t}}$, are positively related to the uncertainty about managerial ability $v_t$ and negatively related to fund size $A_t$. For reasonable parameterizations, both these terms are positive, implying a positive and convex flow-performance relation.\(^{13}\) All these results are consistent with empirical findings documented, for example, by Chevalier and Ellison (1997) and Boudoukh, Richardson, Stanton, and Whitelaw (2003).

Equation (18) implies a negative flow response to the fund’s returns due to unexpected market movements. Such a relation is documented by Warther (1995) and Fant (1999) at the aggregate level for the mutual fund industry. The different responses of fund flows to market returns and the fund’s idiosyncratic returns can be understood in the following way: A fund attracts net inflows if and only if its internal growth rate, namely its NAV return, is less than its equilibrium growth rate determined by changing beliefs. While the positive unexpected market return $dW_{mt}$ has a positive impact on a fund’s NAV return, it has no influence on its equilibrium growth rate, because it does not affect the belief about managerial ability. Therefore, its influence on fund size must be offset by corresponding fund outflows.\(^{14}\) Things are different for the unexpected idiosyncratic return $dZ_t$. Higher idiosyncratic returns not only increase the asset value but also result in an upward revision of investors’ belief about managerial ability, with the latter effect generally dominating the former.

\(^{13}\) See Table 2 for our base-case parameterization.

\(^{14}\) In our model the investment opportunities for active portfolio management are independent of the size of the entire market. In practice, these opportunities are likely to increase with the size of the market. However, as long as the diseconomies of scale of active fund management are not completely offset by new investment opportunities arising when the market expands, the documented negative relation between market driven fund returns and fund flows will still obtain.
2 Valuation and manager replacement threshold

2.1 The valuation model

In the model developed in Section 1 the state of the world is determined by two variables, the posterior mean and variance of the manager’s ability, $a_t$, and $v_t$, respectively. Since $v_t$ is a deterministic function of manager tenure, the value of the management company, $F$, can be written as $F = F(a,t)$.

Given the relation between fund size and the belief about managerial ability it may become optimal for the management company to replace its portfolio manager. We assume that the ability of the new manager is again normally distributed with mean $a_0$ and variance $v_0$.

The manager may also quit voluntarily. Endogenizing the quitting decision would require modelling the manager’s outside options as well as ex-post bargaining between the manager and the management company in a stochastic differential game. We abstract from this complexity by assuming an exogenous quitting rate, described by a Poisson process, $q_t$, with a constant mean arrival rate $\mu$:

$$dq_t = \begin{cases} 0 & \text{with probability } 1 - \mu dt \\ 1 & \text{with probability } \mu dt. \end{cases}$$

The $q_t$ process is assumed to be uncorrelated with the $W_{mt}$, $W_{th}$, and $Z_t$ processes. Whenever the incumbent manager departs, either because he is fired or he leaves voluntarily, the management company incurs a cost for hiring a new manager. This cost can either be interpreted as a search cost or a cost for training the new manager and is assumed to be a fixed proportion $k$ of the initial value of the management company. The Bellman equation for the value of the management company can therefore be written as

$$F(a,t) = \max \{(1 - k)F(a_0,0), b(1 - s)A_t dt + e^{-rt}E[F(a,t) + dF(a,t)]\},$$

where the value function $F(a,t)$ represents the present value of all future net profits under the optimal replacement rule. On the right-hand side, the first term is the value of the management company when the manager is replaced. We call it the replacement value. The second term
is the continuation value, which consists of the immediate net profit in the period from $t$ to $t + dt$ and the expected company value at $t + dt$, discounted back at the risk free rate. The risk free rate is used because the change in the market value of the management company is driven by $dZ$ and $dq$, both of which are idiosyncratic.

The management company’s problem is to determine an optimal threshold for manager replacement. Since for $a_t \in (0, +\infty)$, $b(1 - s)A_t - r(1 - k)F(a_0, 0)$ is monotonically increasing in $a_t$ at any $t$, there exists a tenure-dependent threshold $a_t$, with continuation optimal when $a_t > a_t$ and replacement optimal when $a_t < a_t$.\(^{15}\)

In the continuation region the following no-arbitrage condition must hold

$$rF(a, t)dt = b(1 - s)A_t dt + E[dF(a, t)].$$ \tag{21}

Suppose that the function $F(a, t)$ is continuous and twice-differentiable. Using Ito’s Lemma and taking into account the possibility of job quitting, we have

$$dF(a, t) = \begin{cases} 
\frac{\partial F(a, t)}{\partial t} + \frac{1}{2} F_{aa}(a, t)v_0^2 \mu dt + F_a(a, t)v_0 dZ & \text{if } dq = 0 \\
(1 - k)F(a_0, 0) - F(a, t) & \text{if } dq = 1,
\end{cases}$$ \tag{22}

where $F_a(a, t)$ and $F_{aa}(a, t)$ denote the first and second order derivative of $F(a, t)$ with respect to $a$.

From Equation (22) and (19), we can see that the expected change of $F(a, t)$ is

$$E[dF(a, t)] = \frac{\partial F(a, t)}{\partial t} + \frac{1}{2} F_{aa}(a, t)v_0^2 \mu dt + (1 - k)F(a_0, 0) - F(a, t).$$ \tag{23}

Substituting Equation (23) into (21), we get the following partial differential equation for $F(a, t)$:

$$(r + \mu)F(a, t) = b(1 - s)A_t + \frac{\partial F(a, t)}{\partial t} + \frac{1}{2} F_{aa}(a, t)v_t^2 + (1 - k)F(a_0, 0),$$ \tag{24}

where $A_t$ and $v_t$ are given by Equation (12) and (8) respectively.

The partial differential equation (24) has to satisfy the following boundary conditions:

\(^{15}\)See Dixit and Pindyck (1994), pp. 128-130.
1. Value matching condition. At the replacement boundary, we must have
\[ F(a_t, t) = (1 - k)F(a_0, 0). \] (25)

2. Optimality condition. At the replacement boundary, we require
\[ F_a(a_t, t) \geq 0, \quad F_a(a_t, t)a_t = 0 \] (26)

3. Asymptotic condition:\(^{16}\)
\[ \lim_{a \to \infty} F_a(a, t) = \frac{(1 - s)a}{2\gamma(r + \mu)} \] (27)

The partial differential equation (24) generally has no closed-form solution. However, for the case in which \( \omega = v_0 \), it reduces to an ordinary differential equation and can be solved analytically. In this case the posterior variance \( v_t \) is constant over time so that there is only one state variable, \( a_t. \)\(^ {17}\) The Bellman equation is then given by
\[ (r + \mu)F(a) = \frac{(1 - s)a^2}{4\gamma} + \frac{1}{2}F_{aa}(a)v_0^2 + \mu(1 - k)F(a_0). \] (28)

The general solution of this ordinary differential equation is
\[ F(a) = F_0 + C(1)e^{-\sqrt{2(r+\mu)a/v_0}} + C(2)e^{\sqrt{2(r+\mu)a/v_0}}, \] (29)
where
\[ F_0 = \frac{\mu(1 - k)F(a_0)}{r + \mu} + \frac{(1 - s)[v_0^2 + a^2(r + \mu)]}{4\gamma(r + \mu)^2}. \]

and \( C(1) \) and \( C(2) \) are two constants that need to be determined by boundary conditions.

Using boundary conditions (25) to (27), we can derive the following proposition

\(^{16}\)See appendix A for an explanation of the optimality condition (26) and the derivation of the asymptotic condition (27).

\(^{17}\)As shown in the last section, even if \( v_0 \neq \omega \), \( v_t \) always converges to \( \omega \) if the manager’s tenure is sufficiently long. Therefore this special case can be regarded as a steady state of learning.
Proposition 2. When $v_0 = \omega$, the optimal manager replacement threshold $a$ and the initial value of the management company $F(a_0)$, are the solution to the following system of equations\(^{18}\):

$$a = \text{Max} \left[ 0, -\frac{v_0}{\sqrt{2(r+\mu)}} + \sqrt{\frac{4r\gamma(1-k)F(a_0)}{1-s} - \frac{v_0^2}{2(r+\mu)}} \right], \quad (30)$$

$$F(a_0) = \begin{cases} 
(1-s)[v_0^2 + a_0^2(r+\mu) + \sqrt{2(r+\mu)v_0ag - \sqrt{2(r+\mu)}(a_0-a)/v_0}] 
& \text{if } a > 0, \\
\frac{4\gamma(r+\mu)(r+\mu k)}{4\gamma(r+\mu)[(r+\mu k) - r(1-k)e^{-\sqrt{2(r+\mu)}a_0/v_0}]} 
& \text{if } a = 0.
\end{cases} \quad (31)$$

**Proof.** See Appendix B. \hfill \square

For the general case in which $\omega \neq v_0$, we cannot rely on the closed form solution derived in Proposition 2. Instead we use a binomial tree method to solve the partial differential equation (24) numerically. Since the state variable $a$, has a nonconstant volatility we use the approach developed by Nelson and Ramaswamy (1990) to construct a recombining tree and solve for the value of the management company and the replacement threshold recursively.

For the following numerical analysis we calibrate the model parameters to match empirically observed values wherever possible. We rely primarily on information from CRSP survivor-bias free U.S. mutual fund database (1961 to 2002). Parameters which are not directly observable are chosen to yield empirically reasonable values for fund size, portfolio risk, fund flow dynamics, and expected manager tenure. Appendix D describes the procedures used to calibrate the model for the numerical analysis.

Table 2 summarizes the base-case parameter values. The parameters are classified into three groups. The management fee ratio $b$, the systematic risk $\sigma_m$, and the market price of risk $\lambda$ affect fund flows, but have no influence on either the management company value or the replacement threshold. The variable cost $s$ and the measure of diseconomy of scale $\gamma$ are negatively related to the management company value but unrelated to the replacement threshold. Generally, there are two roots for this equation system, but only one of them makes economic sense, because the larger root of $a$ is bigger than $a_0$. This latter can be excluded because it implies that no manager will be employed at all.
Table 2: Base case parameter values

This table summarizes the parameter values for our base case scenario.

| Panel A: Parameters relevant for neither valuation nor replacement threshold |
|-------------------------|------------------|
| $b$                     | management fee ratio | 0.01 |
| $\lambda$              | market price of risk | 0.25 |
| $\sigma_m$             | systematic risk     | 0.12 |

| Panel B: Parameters relevant for valuation but not for replacement threshold |
|----------------------------|----------------------|
| $s$                       | variable cost as a percentage of fee income | 0.8 |
| $\gamma$                  | measure of diseconomy of scale                | $2 \times 10^{-8}$ |

| Panel C: Parameters relevant for both valuation and replacement threshold |
|------------------|----------------------------------|
| $r$              | risk free rate                   | 0.05 |
| $a_0$ $a priori$ | mean of managerial ability       | 0.2  |
| $v_0$ $a priori$ | variance of managerial ability    | 0.12 |
| $\omega$         | instantaneous volatility of true ability | 0.04 |
| $k$              | percentage replacement cost      | 0.05 |
| $\mu$            | mean arrival rate of job quitting | 0.05 |

threshold, due to the fact that the instantaneous profit is proportional to $s$ and $\frac{1}{\gamma}$ and that the replacement cost is proportional to the company value. The other parameters in Table 2 affect both the company value and the replacement threshold.

2.2 The optimal replacement threshold

In this subsection we analyze the management company’s optimal replacement policy. We start by focusing on the effect of manager tenure on the replacement threshold, $\omega_t$. Figure 1(A) displays the optimal replacement thresholds for different volatilities of managerial ability, $\omega = 0$, $\omega = 0.04$, $\omega = 0.08$, and $\omega = 0.12$, while the other parameters are fixed at the base case values. Consistent with Proposition 2, the firing threshold is constant over the manager’s tenure for the case in which $\omega = 0.12 = v_0$, i.e., when the precision of the belief about managerial ability remains constant over time.

In all the remaining cases in which $\omega < 0.12$, the precision of the belief about managerial ability increases over the manager’s tenure. Figure 1(A) reveals that for these cases the firm’s replacement threshold increases dramatically over the manager’s tenure. When the manager’s ability is constant over time (i.e., $\omega = 0$), the replacement threshold increases from approximately 0.08 for a newly hired manager by more than 300% to almost 0.30 for a manager
This figure plots the optimal manager replacement threshold as a function of manager tenure. In Panel (A), four different values of $\omega$ have been used while the other parameters are fixed at the base case levels as given in Table 2. In Panel (B), four different values of $v_0$ have been used while the other parameters are fixed at the base case levels as given in Table 2.

with a 25-year’s tenure. Thus, the management company adopts an increasingly tougher replacement rule as the precision of the belief improves over time. This result is driven by the dynamics of the value of the real option that the portfolio manager represents. As the precision of the belief about managerial ability increases over time, the value of the manager’s upside potential decreases. As a result, the management company finds it optimal to fire even a “good” manager since it becomes increasingly unlikely that this manager may become a “star”. The management company is therefore willing to replace such a manager, knowing that the expected ability of the new manager is lower than the ability of the current manager. The lower expected ability of the new manager is offset by the higher real option value associated with the new manager.

Figure 1(B) plots the optimal replacement thresholds for different variances of the prior belief, $v_0 = 0.16$, $v_0 = 0.12$, $v_0 = 0.08$, and $v_0 = 0.04$, while the other parameters are fixed at the base case values. When $v_0$ is high, the initial replacement threshold is low, but it increases fast over tenure. By contrast, when $v_0$ is low, the replacement threshold starts from a relatively high level but increases only slowly over tenure. Thus, when the initial uncertainty about managerial ability is high, the management company is more tolerant towards managers with a short tenure. However, its policy towards “older” managers is tougher than in the case of a low $v_0$, since a higher $v_0$ means a higher option value that can be gained by employing a new manager.
This figure illustrates how the optimal manager replacement threshold depends on the replacement cost \( k \) and the quitting density \( \mu \). The values of all parameters except the one under consideration are taken from Table 2.

Figure 2(A) illustrates how the optimal replacement threshold depends on the replacement cost \( k \). As expected, a higher replacement cost results in a parallel downward shift of the threshold. Figure 2(B) shows the effect of the manager quitting density \( \mu \) on the optimal replacement threshold. It indicates that the threshold for a “young” manager is relatively insensitive to the quitting density, whereas the firing threshold for seasoned managers decreases substantially when the quitting density goes up. To understand this, recall that the firing threshold of a seasoned manager is high because a new manager represents a significantly larger real option value. A high quitting density reduces this real option value and therefore makes a potential new manager relatively less attractive. This leads to a significant decrease in the firing threshold for managers with longer tenure.

3 Manager tenure, fund size, and the value of the management company

3.1 Manager tenure

Given the optimal manager replacement threshold, we can derive the expected value and distribution function of tenure for a new manager, as well as the probability density of manager
turnover and expected managerial ability conditional on tenure. While generally only numerical results are available, we have the following analytical results for the case of $v_0 = \omega$:

**Proposition 3.** When $v_0 = \omega$,

1. The expected tenure of the manager is given by
   \[
   T = \frac{1}{\mu} \left[ 1 - e^{-\frac{\sqrt{2\pi} (a_0 - a)}{v_0}} \right]; \tag{32}
   \]

2. The cumulative distribution function of the manager’s tenure is given by
   \[
   P(t) = 1 - e^{-\mu t} \left[ 2 \Phi \left( \frac{a_0 - a}{v_0 \sqrt{t}} \right) - 1 \right]; \tag{33}
   \]

3. The probability density of manager turnover conditional on tenure $t$ is given by
   \[
   f(t) = \mu + \frac{(a_0 - a) e^{-\frac{(a_0 - a)^2}{2v_0^2} \frac{t}{2}}} {\sqrt{2\pi v_0^2 t} \left[ 2 \Phi \left( \frac{a_0 - a}{v_0 \sqrt{t}} \right) - 1 \right]}; \tag{34}
   \]

4. The expected ability of a manager with tenure $t$ is given by
   \[
   E(a|t) = \frac{a_0 - 2a \left[ 1 - \Phi \left( \frac{a_0 - a}{v_0 \sqrt{t}} \right) \right]} {2 \Phi \left( \frac{a_0 - a}{v_0 \sqrt{t}} \right) - 1}, \tag{35}
   \]

where $\Phi(\cdot)$ denotes the cumulative standard normal probability, and $a$ is the constant optimal replacement threshold.

**Proof.** See Appendix C. \qed

Figure 3 plots the probability density of a manager departure conditional on tenure, $f(t)$, for four different values of $\omega$. It shows that the conditional density of a manager departure, $f(t)$, increases rapidly at the initial stage of the manager’s tenure, and then decreases over time. The non-monotonicity of the conditional departure density reflects the effect of learning and firing. Without learning and firing, the conditional departure density would be constant due to the constant quitting density. Since it takes some time for the management company
to learn about the manager’s ability, the initial density of firing is low, but it increases very quickly. As more and more incompetent managers are fired over time, managers who have survived longer will generally have a higher ability. Therefore, consistent with the empirical evidence discussed in the introduction, they have a lower probability of being fired, and their departures are more likely to be due to reasons unrelated to performance. The figure also shows that the probability of firing at the early stage of tenure is higher when the volatility of true managerial ability is low. This is due to the higher replacement threshold in these cases, as derived in the last section.

3.2 Fund size and the value of the management company

Despite the central role that portfolio management companies play in developed capital markets, very little is known about their market values and how it is influenced by competition and their historical performance. The model developed in the last section provides a theoretical framework to analyze these questions. In this subsection we derive the value of a management company relative to the assets under management and analyze its main determinants.
Figure 4: Expected managerial ability and fund size over tenure

The figure plots the expected managerial ability and fund size conditional on tenure, $E(a|t)$ and $E(A|t)$ respectively. Four different values of $\omega$ have been used. The other parameter values are as given in Table 2.

We start by analyzing the expected managerial ability and the expected fund size, conditional on the manager’s tenure. Figure 4(A) plots the expected managerial ability conditional on tenure, $E(a|t)$, for four different values of $\omega$. The expected ability increases over tenure for all values of $\omega$. Early in a manager’s tenure, it increases slightly faster when $\omega$ is lower, due to the higher replacement threshold associated with low $\omega$. However, after about 2.5 years, the slope becomes positively related to $\omega$. This is because the surviving managers are more likely to be highly talented when the volatility of the true ability is high. The increase of expected ability over tenure implies a positive relation between tenure and fund size, as plotted in Figure 4(B), consistent with the empirical results reviewed in the introduction.

We can now explore the evolution of the expected value of the management company, expressed as a fraction of the assets under management. Figure 5 plots this ratio as a function of the manager’s tenure, $E(FA|t)$. We first observe that this ratio is positively related to the volatility of the true managerial ability, $\omega$. This is because the option value reflected in the value of the management company is positively related to $\omega$, while the fund size is independent of $\omega$ for a given $a$. The value-to-size ratio also displays an interesting time-series pattern. It increases over the first short interval and then keeps decreasing over tenure. After 4 years, the ratio has decreased to about 5 percent if $\omega = 0$.

The decrease of the value-to-size ratio over tenure has to do with the fact that the fund’s growth options decrease over time. First, since the expected managerial ability is increasing
Figure 5: Expected value-to-size ratio over tenure

The figure plots the expected ratio of management company value to fund size conditional on tenure, $E[F_A|t]$. Four different values of $\omega$ have been used. The other parameter values are as given in Table 2.

Our analysis helps to shed light on a recently documented empirical puzzle. Standard valuation methods imply that the value of management companies should be between 20% and 30% of assets under management while empirical ratios are only around 2% to 4% (see Huberman (2004)). Our analysis demonstrates that extrapolating initial growth rates and assuming full dividend reinvestment in perpetuity is not appropriate for the valuation of fund management companies, since fund size is constrained by managerial ability and real options.

\footnote{The increase of the value-to-size ratio at the beginning of tenure is due to the fact that the replacement option initially has a low value since the probability of hitting the replacement threshold is close to zero when a manager just starts his tenure. The option becomes more valuable when it becomes more likely that it will be exercised.}
diminish over a fund’s life time. For mature funds, our model predicts ratios quite similar to values observed empirically, especially when $\omega$ is low.

## 4 Manager replacement, fund flows and portfolio risk

If fund flows are driven by learning about managerial ability and by the diseconomy of scale inherent in active portfolio management, then investors should react to manager replacements by either withdrawing or investing additional capital in the fund. Also, since fund size and portfolio risk are linked, one should expect portfolio restructurings to generate the necessary changes in portfolio risk. Since our model endogenizes optimal manager replacement, we can derive specific hypotheses about both fund flows and portfolio risk changes around manager replacements.

The response of fund flows to an observed manager change can be derived from Proposition 1. Since all new managers have the same expected ability $a_0$, the fund size must adjust from $\frac{a_t^2}{4\beta_I}$ to $\frac{a_0^2}{4\beta_I}$ when a manager with ability $a_t$ is replaced by a new manager at time $t$.\(^{20}\) This implies that the proportional fund inflow around a manager change is given by $\frac{a_0^2}{4\beta_I} - 1$. Depending on whether the posterior mean of the departing manager’s ability, $a_t$, is higher or lower than the a priori mean $a_0$, the flow response to a manager change can be either negative or positive.\(^{21}\)

From Proposition 1 we can also derive the percentage change of the fund’s idiosyncratic risk surrounding a manager change, which is given by $\frac{\Delta \sigma \approx a_t - 1}{\sigma_t} = \frac{a_t}{a_0} - 1$. Therefore, when a manager whose ability is inferred to be higher than the a priori mean is replaced, the fund’s idiosyncratic risk will go up, and vice versa.

The above results hold for manager changes due to either quitting or firing. In the following analysis, we consider only manager changes due to firing, which occur when the

\(^{20}\)Note that investors have no incentive to respond earlier in our simplified world without transaction costs, even if the manager change is fully anticipated.

\(^{21}\)The prediction that fund size will adjust instantaneously to reflect the new manager’s ability is admittedly quite strong, because in reality, managerial ability is not the only determinant of fund performance. However there is some anecdotal evidence that investors are quite sensitive to fund manager changes. For example, it was reported that when William von Mueffling, a star manager running the hedge-fund business of Lazard Asset Management company, resigned in January 2003, Lazard’s 4 billion hedge-fund business dwindled to less than 1 billion in just a few weeks (The Economist (2003)).
This figure plots the proportional fund inflows (Panel A) and the percentage change of a fund idiosyncratic risk (Panel B) when a fund manager is fired. Four different values of $\omega$ have been used. The other parameters values are as given in Table 2.

Proposition 4. If an open-end fund manager is fired by the management company, then

1. the fund will have a net proportional fund inflow of $\frac{a_t^2}{a_0^2} - 1$,
2. the fund’s idiosyncratic risk will change by a factor of $\frac{a_t}{a_0} - 1$.

Figure 6 plots the proportional fund flows (Panel A) and the proportional changes in a fund’s idiosyncratic risk (Panel B) when a manager is fired. One can see that manager firing induces a more dramatic fund inflow and a more significant risk decrease when the volatility of true managerial ability, $\omega$, is high. This is because high $\omega$ is associated with a lower replacement threshold. Furthermore, except for the case in which $\nu_0 = \omega$, the fund inflows and risk decreases after manager replacement become less dramatic as the fired manager’s tenure increases. If the volatility of true managerial ability is sufficiently low and the fired manager’s tenure is sufficiently long such that the optimal replacement threshold $a_t$ is above the a priori mean $a_0$, then a manager firing will be accompanied by a fund outflow and an increase of fund risk. However, even under those parameter values, most firings will occur before $a_t$ reaches $a_0$. Therefore manager replacement should in general be associated with fund inflows and a decrease of fund risk.
Proposition 4 also has implications on the flow-performance relation. It implies that bad performance is not necessarily associated with money outflows. It can result in money inflows if a manager replacement is triggered. This introduces additional convexity in the flow-performance relation. It is in contrast with the prediction of the Berk and Green (2004) model, which does not allow for manager replacement and assumes that the fund will be shut down whenever its performance reaches a lower bound.

Proposition 4 also provides an alternative explanation for an empirical finding documented by Khorana (2001), i.e., that the replacement of underperforming managers is preceded by an increase in the fund’s idiosyncratic risk and followed by a decline in the fund’s idiosyncratic risk. Although this is usually interpreted as evidence of gambling behavior of underperforming managers in the spirit of Chevalier and Ellison (1997) and Brown, Harlow, and Starks (1996), our result suggests that this may reflect the optimal risk-taking behavior of managers with different abilities.

Using our results on the determinants of the optimal replacement threshold presented in Subsection 2.2, we can derive further comparative static predictions on the flow responses and risk changes around manager firing. For example, the replacement threshold is negatively related to the replacement cost and the density of voluntary manager quitting. This implies that the net fund inflows and the decrease of fund risk after manager firing will be more significant when the replacement cost is high, for example due to manager entrenchment, and when the quitting density is high. The empirical test of these cross-sectional predictions is an interesting topic for future research.

5 Conclusion

This paper has developed a continuous-time model in which all parties involved learn about a portfolio manager’s ability from past returns. In response, investors can move capital into or out of a mutual fund, the portfolio manager can alter the risk of the portfolio and the management company can replace the portfolio manager. Thus, the model formalizes simultaneously the external governance of the product market and the internal governance mechanism in a fully dynamic framework, thereby generating a rich set of empirical predictions on mutual
fund flows, manager turnover and the value of management companies.

The results show that product market forces introduce strong incentives to replace poorly performing managers of open-end funds, even if it is costly to do so. Our analysis rationalizes several empirically documented findings, such as the positive relation between manager tenure and fund size, the decreasing probability and performance sensitivity of manager replacement over manager tenure, and the increase of portfolio risk prior to manager replacement followed by a subsequent risk decrease.

The management company is modelled as a contingent claim on the belief about the portfolio manager’s talents. We find that this implies much lower valuation levels for management companies than simple applications of traditional discounted cash flow methods suggest. Numerical results show that the value of the management company as a percentage of funds under management is relatively high early in the manager’s tenure, due to real option values, but decreases quickly to below five percent, which is largely consistent with empirical observations.

The analysis generates several new empirical predictions which have not yet been tested. First, our model predicts that the critical ability level, and thus the critical performance threshold at which the manager is fired, increases significantly over a manager’s tenure. This increase is particularly pronounced if the prior about a new manager’s ability is diffuse, if the probability that the manager leaves of his own will is small and if the volatility of managerial ability is low. Second, the analysis implies that management replacements should be accompanied by capital flows and by risk changes. Generally, a manager replacement should be preceded by capital outflows and a portfolio risk increase and followed by capital inflows and a portfolio risk decrease. However, these patterns become less pronounced the longer the fired manager’s tenure. For sufficiently long tenure the predictions may change signs. The replacement of a manager with a sufficiently long tenure therefore may be followed by a capital outflow and by a risk increase.

In addition to the specific results, this paper also contributes by developing a continuous-time valuation framework which allows for learning, competitive provision of capital and an optimal control chosen by the company. This basic framework can be used to explore a number of extensions and related issues. For example, we have assumed that the portfolio
performance is jointly generated by the management company and the manager, and that the contributions of these two parties cannot be separated. In practice the management company itself may possess specific expertise which contributes to portfolio performance and does not vanish with the departing manager. It would be interesting to allow for learning about the expertise of the management company as well as of the manager. Also, we have not considered the interaction between different funds of the same management company. If the funds use similar strategies, the diseconomies of scale due to capital flows into one fund would presumably spill over to depress the performance of other funds in the family. Furthermore, the number of funds in a fund family and the extent to which their strategies correlate are likely to influence the speed of learning. Another natural extension would be to endogenize the quitting decision of the portfolio manager and the contract between the manager and the management company.

Finally, we have assumed that capital can be moved freely into and out of the fund. In practice withdrawing capital or investing additional capital is associated with transactions cost, including costs such as front-end loads. In the limit, capital cannot be withdrawn from or injected into a fund at all, as it is essentially the case for closed-end funds. It would be interesting to see how such frictions affect our results and which cross-sectional predictions they generate for the mutual fund industry.

Appendix

A. Boundary conditions

The optimality condition (26) is a generalized version of the smooth-pasting condition normally used in option pricing literature, i.e.,

$$F_{a}(a_{t}, t) = 0.$$ 

We use the Kuhn-Tucker optimality condition instead of the standard smooth-pasting condition because $a$ is constrained to be nonnegative. When $a_{t} = 0$ the fund size becomes zero as well and learning can no longer take place. Thus, when $a_{t} = 0$ the fund is effectively “shut
down” by fund investors, even if the management company would prefer to keep the manager in place.

In practice, funds may face some additional constraints, for example, a minimum size or maximum risk constraint, which may potentially result in forced manager replacement at some $a_t > 0$. However, these types of constraints will not be binding in our model, since funds can always satisfy such constraints by lowering their fee ratio. One should also note that although fee adjustment can help to get round those constraints, it cannot be a substitute for the manager replacement in our model, since it has no influence on the net profit of the management company.

To study the asymptotic behavior of the value of the fund management company, we calculate the fundamental value of the underlying cash flow, i.e., the present value of discounted cash flows in the absence of the real option to replace the manager. Ignoring manager quitting and firing, the dynamics of the assets under management are given by Equation (15), so we can write $A_{t+\tau}$ as

$$A_{t+\tau} = A_t + \int_t^{t+\tau} \frac{v_l^2}{4b\gamma} dl + \int_t^{t+\tau} \frac{a_tv_l}{2b\gamma} dZ_l.$$  \hspace{1cm} (36)

Since $v_t$ converges to $\omega$ over time we write $v_l^2 = \omega^2 + \Delta_1(t)$. Consequently, the expected assets under management are

$$E(A_{t+\tau}) = A_t + \int_t^{t+\tau} \frac{\omega^2}{4b\gamma} dl + \int_t^{t+\tau} \frac{\Delta_1(l)}{4b\gamma} dl + E\left(\int_t^{t+\tau} \frac{a_tv_l}{2b\gamma} dZ_l\right),$$

$$= A_t + \frac{\omega^2\tau}{4b\gamma} + \Delta_2(t, \tau) + 0,$$ \hspace{1cm} (37)

where $\Delta_2(t, \tau) = \int_t^{t+\tau} \frac{\Delta_1(l)}{4b\gamma} dl$. The expected value of the stochastic integral vanishes.

Now we explicitly consider the possibility that managers quit their job and calculate the
fundamental value as the expected value of future cash flows discounted by the risk-free rate

\[
F_0(a,t) = \int_0^\infty (1-k)F(a_0)e^{-\mu \tau}e^{-r \tau}d\tau + b(1-s)\int_0^\infty e^{-\mu \tau}\left(A + \frac{\omega^2 \tau}{4b\gamma} + \Delta_2(t,\tau)\right)e^{-r \tau}d\tau
\]

(38)

\[
= \frac{\mu(1-k)F(a_0)}{r+\mu} + \frac{(1-s)[\omega^2 + a^2(r+\mu)]}{4\gamma(r+\mu)^2} + \Delta_3(t),
\]

where \(\Delta_3(t) = b(1-s)\int_0^\infty \Delta_2(t,\tau)e^{-(r+\mu)\tau}d\tau\) is independent of \(a\) and vanishes for \(t \to \infty\) or \(v_0 \to \omega\).

For large values of \(a\), the value of the management company \(F(a,t)\) converges to the fundamental value \(F_0(a,t)\) since the real option vanishes. Therefore the partial derivative \(\frac{\partial F(a,t)}{\partial a}\) converges to \(\frac{(1-s)a}{2\gamma(r+\mu)}\), as stated in Equation (27).

**B. Proof of Proposition 2**

**Proof.** First note that according to the asymptotic condition (27), \(C(2)\) must equal zero.

Given that \(C(2) = 0\), the value-matching condition \(F(a) = (1-k)F(a_0)\) implies that

\[
\frac{\mu(1-k)F(a_0)}{r+\mu} + \frac{(1-s)[\omega^2 + a^2(r+\mu)]}{4\gamma(r+\mu)^2} + C(1)e^{-\sqrt{2(r+\mu)a/v_0}} = (1-k)F(a_0). \tag{39}
\]

Furthermore, the initial value of the management company, i.e., \(F(a_0)\), also satisfies the general solution (29) (with \(C(2) = 0\)). Therefore we have

\[
F(a_0) = \frac{\mu(1-k)F(a_0)}{r+\mu} + \frac{(1-s)[\omega^2 + a^2(r+\mu)]}{4\gamma(r+\mu)^2} + C(1)e^{-\sqrt{2(r+\mu)a_0/v_0}}. \tag{40}
\]

Now we consider two alternative cases:

1. \(a > 0\).

   If \(a > 0\), then the smooth pasting condition \(F_a(a) = 0\) implies

   \[
   \frac{(1-s)a}{2\gamma(r+\mu)} + C(1)\frac{e^{-\sqrt{2(r+\mu)a/v_0}}}{v_0} = 0. \tag{41}
   \]
Therefore we have

\[ C(1) = \frac{(1-s)v_0 a}{(2+r+\mu)^2} e^{\sqrt{2(r+\mu)a/a_0}}. \]  

(42)

Substituting Equation (42) into (39) and collecting terms, we get

\[ a = -\frac{v_0}{\sqrt{2(r+\mu)}} + \sqrt{\frac{4r(1-k)F(a_0)}{1-s}} - \frac{v_0^2}{2(r+\mu)}. \]

Substituting Equation (42) into (40), we get

\[ F(a_0) = \frac{(1-s)[v_0^2 + a_0^2(r+\mu) + \sqrt{2(r+\mu)v_0 a e^{-\sqrt{2(r+\mu)(a_0-a)/v_0}}}]}{4\gamma(r+\mu)(r+\mu k)}. \]

2. \( a = 0. \)

If \( a = 0 \), then by Equation (39) we have

\[ C(1) = \frac{r(1-k)F(a_0)}{r+\mu} - \frac{(1-s)v_0^2}{4\gamma(r+\mu)^2}. \]

Substituting this equation into (40) and noting that \( a = 0 \), we have

\[ F(a_0) = \frac{(1-s)[v_0^2 + a_0^2(r+\mu) - v_0^2 e^{-\sqrt{2(r+\mu)(a_0-a)/v_0}}]}{4\gamma(r+\mu)[(r+\mu k) - r(1-k)e^{-\sqrt{2(r+\mu)(a_0-a)/v_0}}]}. \]

This completes our proof of Proposition 2. \( \square \)

C. Proof of proposition 3

Proof. To prove Proposition 3, first note that based on the reflection principle of the Wiener process, the unconditional density of \( a_t \) at any \( t \), given the optimal replacement threshold \( a \), the constant quitting density \( \mu \), and the zero correlation between quitting and the posterior
mean process $a_t$, is given by

$$
g(a_t) = \begin{cases} 
e^{-\mu t} & \frac{(a_t-a_0)^2}{2\sigma^2} - e^{-\frac{(a_t+a_0-2\sigma)^2}{2\sigma^2}} \\ 0 & \text{if } a_t \leq a.\end{cases}$$ (43)

Here $e^{-\mu t}$ represents the probability that the manager has not quit before $t$.

Therefore, the probability that the manager survives until $t$ is

$$P_{\text{survive}}(t) = \int_{-\infty}^{\infty} g(a)da = e^{-\mu t} [2\Phi(\frac{a_0-a}{\sqrt{\sigma}}) - 1],$$ (44)

and the cumulative distribution function of the manager’s tenure is

$$P(t) = 1 - P_{\text{survive}}(t) = 1 - e^{-\mu t} [2\Phi(\frac{a_0-a}{\sqrt{\sigma}}) - 1],$$

where $P(t)$ denotes the probability that the manager’s tenure is shorter than $t$.

From the cumulative distribution function $P(t)$, we can calculate the expected manager tenure $T$ as follows,

$$T = \int_{0}^{+\infty} t dP(t) = \frac{1}{\mu} [1 - e^{-\frac{\mu}{\sigma^2}}(a_0-a)].$$ (45)

The conditional probability that the manager who has survived until $t$ departs in the time interval $[t, t+dt]$ is given by $[P(t+dt) - P(t)]/[1 - P(t)]$. Dividing this conditional probability by $dt$ and letting $dt$ converge to zero, we get the conditional density of manager departure $f(t)$, which is given by Equation (34).

The expected managerial ability conditional on tenure $t$ can be calculated as follows,

$$E(a|t) = \int_{-\infty}^{+\infty} \frac{ag(a)}{P_{\text{survive}}} da$$

$$= \frac{a_0 - 2\sigma [1 - \Phi(\frac{a_0-a}{\sqrt{\sigma}})]}{2\Phi(\frac{a_0-a}{\sqrt{\sigma}})} 2\Phi(\frac{a_0-a}{\sqrt{\sigma}}) - 1.$$

This completes our proof of proposition 3.

34
D. Calibration

In our model, \( b \) represents the management fee as well as any other expenses charged by the management company. From the CRSP survivor-bias free US mutual fund database (1961 to 2002), we find that the average annual expense ratio equals 1.17%. Huberman (2004) reports for publicly traded money management firms a ratio of revenue to assets under management of 0.83%. We therefore calibrate the parameter \( b \) in our model to the average of these two values, that is to 1%. Huberman (2004) also documents an average net income to revenue ratio of 21% for his sample funds. Therefore we set our variable cost ratio, \( s \), to be 0.8.

The market price of risk, \( \lambda \), is defined as the excess rate of return of the market over the risk free rate divided by the market volatility. According to Dimson, Marsh, and Staunton (2002), the annual return on US government bonds over the period 1900 to 2001 was 4.8%, and the US equity return over the same period was 10.1%. We therefore assume a risk-free rate of 5% and an equity premium of 5%. Since the historical annual volatility of stock market return was approximately 20%, the market price of risk is set equal to 25%.

To obtain an empirically plausible estimate for the systematic risk of an average mutual fund, we first calculate the average mutual fund beta. We find that aggregate mutual fund returns exhibit a beta of 0.6. Since we also include bond and money market funds, this value is lower than those reported in previous studies (see for example Chen, Hong, Huang, and Kubik (2004)). Multiplying beta by the average market volatility of 20%, we get a systematic risk, \( \sigma_m \), of about 12%.

We set the diseconomy of scale, \( \gamma \), to be \( 2 \times 10^{-8} \) and the \( a \) priori mean of the managerial ability, \( a_0 \), to be 0.20. According to Proposition 1, these parameter values, together with a management fee of 1%, imply an initial fund size of 50 million and an initial idiosyncratic risk of 10%, which are very close to the empirical numbers calculated from the CRSP mutual fund database. From the CRSP database, we find that the average fund size during the starting year is 48.7 million dollars with an average expense ratio of 1.01%. The average idiosyncratic risk of funds less than two-years old is 9.4%.

\[22\] The idiosyncratic risk is estimated as follows: we first construct a portfolio of all funds less than two-years old and estimate its beta by averaging the betas obtained from year-by-year regressions; we then calculate the average return volatility of all funds less than two years old; finally, we estimate the average idiosyncratic risk by subtracting the systematic component from the average return volatility.
The remaining four parameters, the \textit{a priori} variance of managerial ability \(v_0\), the instantaneous volatility of true ability \(\omega\), the replacement cost \(k\), and the quitting density \(\mu\), are more difficult to calibrate. Therefore we adopt the following strategy. We first choose a base case set of parameter values which yields empirically reasonable fund size, portfolio risk, fund flow dynamics and expected manager tenure. We then provide comparative statics results by solving the model for different values around the base case.

For the base case scenario, we choose a \(v_0\) of 0.12, which corresponds to a standard deviation of 0.35. According to Proposition 1, a manager whose ability is believed to be one standard deviation above the \textit{a priori} mean will manage a fund of 373 million dollars with an idiosyncratic risk of 3.7%, while a manager with an ability two standard deviations above the \textit{a priori} mean will manage a fund of 1013 million dollars with an idiosyncratic risk of 2.2%.\textsuperscript{23} According to Equation (18), this parameterization also implies that a new fund has an expected inflow rate of 28% and a flow-performance sensitivity of 110%.\textsuperscript{24} We choose \(\omega\) to be 0.04 such that the \textit{expost} variance of ability declines to 0.08 after 5 years.

The quitting density is set equal to 0.05, which is about 30% of the annual manager turnover rate of 0.18 documented by Chevalier and Ellison (1999a). The replacement cost \(k\) is set to be 0.05. Given the other parameters summarized in Table 2, our simulation shows that the assumed values for the quitting density and the replacement cost lead to an expected tenure of 5.47 years, which accords well with the empirically documented annual turnover rate of 0.18.

\textsuperscript{23}At the end of year 2002, 93\% of funds in the CRSP database have a total net asset value of less than 1000 million dollars.

\textsuperscript{24}Boudoukh, Richardson, Stanton, and Whitelaw (2003) report an expected inflow rate of 36.6\% and a flow-performance sensitivity of 136.3\% for small and young equity funds.
References


