MUTUAL FUND PORTFOLIO CHOICE IN THE PRESENCE OF DYNAMIC FLOWS

Julien Hugonnier

Swiss Finance Institute, École Polytechnique Fédérale de Lausanne

Ron Kaniel

Fuqua School of Business, Duke University
CEPR

We analyze the implications of dynamic flows on a mutual fund's portfolio decisions. In our model, myopic investors dynamically allocate capital between a riskless asset and an actively managed fund which charges fraction-of-fund fees. The presence of dynamic flows induces "flow hedging" portfolio distortions on the part of the fund, even though investors are myopic. Our model predicts a positive relationship between a fund's proportional fee rate and its volatility. This is a consequence of higher-fee funds holding more extreme equity positions. Although both the fund portfolio and investors' trading strategies depend on the proportional fee rate, the equilibrium value functions do not. Finally, we show that our results hold even if investors are allowed to directly trade some of the risky securities.

Key Words: mutual funds, delegated portfolio management, incomplete markets, continuous time stochastic game, BSDEs.

1. INTRODUCTION

The impact of fee structures on a fund's portfolio decisions has typically been studied in a setting in which a manager receives an exogenously specified amount of money to manage for an exogenously specified time period. In the few models where investors do make allocation decisions, they are typically restricted to a one time decision that determines the fraction of their initial wealth that they delegate to the fund. Once this initial decision is made, investors are unable to delegate additional funds or to withdraw any part of the money under the fund's management.

However, investors can typically move money in and out of a fund dynamically, because they can purchase or redeem holdings in the fund at the fund's net asset value. Our objective is to understand how this feature impacts a fund's trading strategy. To this end, we study a dynamic, continuous-time economy with small investors and a fund manager. Small investors implicitly face high costs that preclude them from trading directly in

the equity market. These implicit costs can be related, for example, to the fact that the opportunity costs of spending their time in stock trading related activities are high. Specifically, actively trading multiple risky securities requires considerably more attention than trading in only one mutual fund.[1] Although investors are precluded from holding equity directly, they are allowed to dynamically allocate money between a mutual fund and a locally riskless bond. We impose the natural restriction that investors cannot short the fund and assume that the fund charges fraction-of-fund fees, whereby the manager receives a fixed proportion of the assets under management.[2]

To focus on the impact of flows on a fund's portfolio decisions although maintaining a tractable setup, we make a few simplifying assumptions. First, agents have complete information and observe each other's actions. Second, from the perspective of the fund markets are complete. Third, small investors are assumed to have a logarithmic utility function. Fourth, the fund manager is strategic whereas investors are not. Specifically, when investors determine their portfolio, they take the fund strategy as given. On the other hand, when the fund manager selects the fund's portfolio he takes into account what will be the investors' reaction to such a portfolio.

Because investors are myopic, the fraction of their wealth that they allocate to the fund at a given time equals the fund's net-of-fees Sharpe ratio at that time. One might be tempted to conjecture that this will induce the fund to pick a portfolio that maximizes the net-of-fees Sharpe ratio. This conjecture holds when the fund's investment opportunity set is constant. However, when the opportunity set is stochastic, it does not. The reason it breaks down is that it implicitly takes investors' wealth as given, ignoring the fact that a fund's trading strategy actually impacts the distribution of the investors' future wealth.

For a stochastic opportunity set, we show that accounting for the impact of a fund's trading strategy on the distribution of the investors' future wealth leads to flow-induced trading distortions on the part of the fund. There are two channels by which these distortions are manifested. First, the fund's stock to bond mix may change. Second, and even more interesting, the composition of the equity portfolio changes. In selecting a fund's dynamic trading strategy, the manager knows that in future states of the world where the opportunity set is good, so that the Sharpe ratio that he or she can provide is high, investors will allocate a larger fraction of their wealth to the fund, and as such the fund will receive higher management fees. These fund flow considerations lead the manager to distort the fund's equity portfolio to increase the correlation between the future opportunity set (i.e., the net-of-fees Sharpe ratio) and the investors' wealth. We demonstrate that these distortions can be significant by studying an example with a stochastic investment opportunity set where the equilibrium can be computed in closed form.

We show that, in equilibrium, the overall equity component of the fund increases with the fee rate, implying a positive relationship between a fund's fee rate and its volatility. Naturally, the investor's holdings of the fund are inversely related to the fee rate. Although both the fund's and the investors' portfolios depend on the fee rate, the investors' effective exposure to equity, as well as the market value of the fees, do not. When the fund is restricted to holding only equity this fee independence no longer holds. However, the positive relationship between fee rates and fund volatilities still prevails.

---

[1] This assumption is relaxed in Section 4 where we allow investors to trade freely in a subset of the stocks that are available to the fund.

[2] Fraction-of-fund fee is the predominant fee structure in the mutual fund industry; as of 1997, approximately 98% of all mutual funds were using a fraction-of-fund fees without any performance-based incentives.

We also analyze the implication of the model regarding both the contemporaneous and the lagged performance–flow relationship. We show that, consistent with empirical evidence (see, e.g., Chevalier and Ellison 1997) our model can generate a positive relationship between lagged performance and flows when the opportunity set is stochastic. The sensitivity of this relationship is lower for higher fees.

Extending the model to allow the manager to exert costly effort to increase the fund's return retains most of the results. However, we show that exerting effort and tilting the portfolio toward equity are partial substitutes for the manager, so that the fund's equity portfolio need not be monotone in the fee rate. In particular, for low fee rates the fraction invested in equity may be decreasing in the fee rate.

Fraction-of-fund fees are by far the predominant compensation contract in the mutual fund industry. However, some funds have a performance component in their compensation contract. Grinblatt and Titman (1989), Chen and Pennacchi (1999), Carpenter (2000), and Basak, Pavlova, and Shapiro (2007) have studied the optimal portfolio strategy of a manager receiving convex performance fees in a setting in which the manager receives an exogenous amount of money to manage at the initial date. An analysis of the equilibrium asset pricing implications of both fulcrum fees and asymmetric performance fees is conducted in Cuoco and Kaniel (2007). In that paper, both fund managers and unconstrained investors trade directly in equity markets, but investors who use the fund services make allocation decisions only at the initial date. High-water-mark fees, used in the hedge fund industry, are discussed in Goetzmann, Ingersoll, and Ross (2003).

Because the focus of this paper is the impact of dynamic flows on a mutual fund's portfolio decisions, we take the fee structure as given. However, it is important to emphasize that we are not taking a stance on whether fraction-of-fund fees are the optimal compensation contract in the setting of our model, but we instead rely on its widespread use as the motivation for our analysis. Papers that analyze these optimal contracting issues include Starks (1987), Roll (1992), Heinkel and Stoughton (1994), Admati and Pfleiderer (1997), Lynch and Musto (1997), Das and Sundaram (2002a,b), Dybvig, Farnsworth, and Carpenter (1999), and Ou-Yang (2003), among others.

In solving our model we rely on tools that were developed for the study of utility maximization problems under incomplete markets, as well as on the theory of backward stochastic differential equations and its links to the study of optimal control problems. For details on the study of utility maximization problems in incomplete or constrained financial markets, see Karatzas et al. (1991) and Cvitanič and Karatzas (1992). For a comprehensive survey of backward stochastic differential equations and some of their applications in finance, the reader can refer to El Karoui, Quenez, and Peng (1997), Ma and Yong (1999), or Yong and Zhou (1999).

The remainder of the paper is organized as follows. The economic setup is described in Section 2. In Section 3, we solve the investor's optimization problem and derive the optimal fund portfolio. Section 4 describes the equilibrium trading strategies, discusses the model's implications, and studies various extensions of the model. Section 5 considers the case in which the manager can exert costly effort to increase the fund's returns. Section 6 concludes. Appendix A contains all the proofs.

2. THE MODEL

We consider a continuous time economy on the finite time span $[0, T]$. The uncertainty is represented by a probability space $(\Omega, \mathbb{F}, \mathcal{F}, P)$ on which is defined an n-dimensional

standard Brownian motion B. The filtration $\mathbb{F} := (\mathcal{F}_t)_{0 \le t \le T}$ is the usual augmentation of the filtration generated by the Brownian motion, and we let $\mathcal{F} := \mathcal{F}_T$ so that the true state of nature is solely determined by the path of the Brownian motion.

In the sequel, all processes are assumed to be adapted to $\mathbb{F}$, and all statements involving random quantities are understood to hold either almost surely or almost everywhere. Furthermore, we shall assume that all the random processes introduced are well defined, without explicitly stating the regularity conditions ensuring this. We shall also make use of the following vectorial notation: a star superscript stands for transposition, $\| \cdot \|$ denotes the usual Euclidean norm, and $\mathbf{1}$ is an n-dimensional vector of ones.

2.1. Securities and Mutual Fund Dynamics

2.1.1. Securities. There is a single perishable good in units of which all quantities are expressed. The financial market consists of $n + 1$ long-lived securities. The first security is a locally riskless bond, whose price process, S_0, satisfies

$$dS_t^0 = r_t S_t^0 dt, \quad S_0^0 = 1,$$

for some interest rate process r. The remaining securities are risky. They are referred to as the stocks and their price process, $S := (S^i)_{i=1}^n$, satisfies

$$dS_t = \mathrm{diag}[S_t](a_t dt + \sigma_t dB_t),$$

for some vector-valued drift process a and some matrix-valued volatility process σ. We assume that the processes r, a, and σ are uniformly bounded and that the volatility σ admits a uniformly bounded inverse.

The assumptions imposed on the coefficients of the model imply that the market price of risk, $\xi := \sigma^{-1}[a - r\mathbf{1}]$, is bounded. As a result, the formula

$$\left. \frac{dQ}{dP} \right|_{\mathcal{F}_t} = M_t := \exp\left[-\int_0^t \xi_\tau^* dB_\tau - \frac{1}{2} \int_0^t \|\xi_\tau\|^2 d\tau \right]$$

defines an equivalent risk-neutral probability measure. Because the volatility matrix of the stocks is invertible this risk-neutral probability measure is uniquely defined, and it follows that the financial market is dynamically complete in the absence of trading constraints.

2.1.2. Mutual Fund Dynamics. Trading takes place continuously, and there are no frictions. A mutual fund trading strategy is a vector process, $\pi = (\pi^k)_{k=1}^n$, specifying the portfolio weight in each of the stocks. Given a fund trading strategy, the return on investments in the fund evolves according to

$$(2.1) \qquad \frac{dF_t}{F_t} = [1 - \pi_t^* \mathbf{1}] \frac{dS_t^0}{S_t^0} + \sum_{k=1}^n \pi_t^k \frac{dS_t^k}{S_t^k} - \gamma \, dt$$

$$= (r_t - \gamma) dt + \pi_t^* \sigma_t [dB_t + \xi_t dt] = (r_t - \gamma) dt + \pi_t^* \sigma_t dZ_t,$$

where the process Z is an n-dimensional standard Brownian motion under the risk-neutral probability measure, and the constant $\gamma > 0$ represents the rate at which fees are withdrawn from the asset under management.[3,4]

In what follows we use the notation $F = F^\pi$ to emphasize the dependence on the portfolio weight process chosen by the fund manager, and denote by Π_f the set of fund portfolio processes with the property that the associated solution to equation (2.1) is nonnegative.

2.2. Agents

2.2.1. Investors.   Given that the investors have the same utility function we will treat them, without loss of generality, as one representative investor. This investor, which we denote by i, is allowed to trade freely in the riskless security but can only access the market for risky securities through the fund.[5]

We denote by $W_0^i > 0$ the investor's initial capital and by ϕ_t the amount he holds in the mutual fund at time t. Because investors cannot short actively managed funds, we impose the constraint

$$\phi_t \geq 0, \quad 0 \leq t \leq T.$$

Given a fund portfolio $\pi \in \Pi_f$, and under the usual self financing condition, the investor's wealth process evolves according to

$$(2.2) \qquad dW_t^i = \left[W_t^i - \phi_t\right]\frac{dS_t^0}{S_t^0} + \phi_t \frac{dF_t^\pi}{F_t^\pi}$$
$$= r_t W_t^i \, dt + \phi_t \pi_t^* \sigma_t\left[dB_t + \xi_t^\pi \, dt\right]$$

with initial condition W_0^i, where the process ξ^π is the risk premium associated with an investment in the stocks through the mutual fund; that is

$$\xi_t^\pi := \xi_t - \gamma \frac{\sigma_t^* \pi_t}{\|\sigma_t^* \pi_t\|^2}.$$

In what follows, we let Θ_i^π denote the set of nonnegative processes, ϕ, such that, given π, the solution to equation (2.2) is a positive process.

The investor is assumed to have logarithmic preferences over terminal wealth. Given a fund portfolio process, his objective is thus to select a trading strategy so as to maximize the expected utility

$$E\left[\log\left(W_T^i\right)\right].$$

---

[3] In practice, both management fees and distribution fees (12b–1) are indeed deducted continuously from the assets under management. Some funds also impose either front or back end load fees, but we ignore these in our analysis.

[4] In practice, fund managers earn fees not only for the amount invested in the fund, but also for the returns earned on that amount. In discrete time the distinction between fees accrued on an ex-ante or ex-post basis is important, but in continuous time this difference vanishes. See Appendix B for details.

[5] The inability of the investor to trade stocks directly should be viewed as a reduced form representing the fact that it is more costly for him to trade stocks efficiently than it is for the fund. In Section 4.3, we relax this assumption by allowing the investor to trade directly in a subset of the stocks.

The fact that the fund value process is driven by n independent Brownian motions and that the investor is not allowed to short the fund imply that the investor effectively faces an incomplete market. The assumption of log utility is therefore critical because this utility function is the only one allowing for a closed form solution in our possibly non-Markovian framework.[6]

2.2.2. The Fund Manager.   In exchange for his services, the manager receives the fees that are generated by the investor's trading of the fund. Specifically, if the investor follows some trading strategy, ϕ, then the manager receives $\gamma \phi_t$ per unit of time, and the corresponding cumulative fee process is given by

$$\Phi_t := \int_0^t \gamma \phi_\tau d\tau.$$

To keep the model as simple as possible, we assume that the manager chooses the fund portfolio to maximize the initial market value

$$(2.3) \qquad E_Q\left[\int_0^T \frac{d\Phi_t}{S_t^0} \right] = E_Q\left[\int_0^T \frac{\gamma \phi_t}{S_t^0} dt \right]$$

of the future fees.[7] To justify this assumption, we argue as follows. The mutual fund should be viewed as being part of a financial services firm. In this case, equation (2.3) represents the contribution of the given fund to the market value of the firm, and our specification implies that the manager acts to maximize the value of the firm. The simplifying assumption we make is that we ignore agency conflicts between the manager and the shareholders of the firm.

2.3. Equilibrium

The focus of our analysis is on how dynamic flows impact a fund's portfolio decisions. As such, we view the investor as a *small* agent who does not realize that his investment decisions may impact the fund's portfolio decisions. The fund manager, on the other hand, is strategic. Specifically, we assume that in making his investment decisions, the manager takes into account the impact of the fund portfolio on the investor's demand.

As a result, we can view our model as a stochastic differential game where the manager's strategic advantage is that of a Stackelberg leader. This implies (see Bagchi 1984) that the equilibrium can be computed through the following procedure:

1. Solve the investor's utility maximization problem given an arbitrary fund portfolio process to compute the strategy

$$\phi^\pi = \underset{\phi \in \Theta_i^\pi}{\arg\max}\ E\left[\log\left(W_T^i \right) \right],$$

which is the investor's best response to the posted fund portfolio.

---

[6] Under the assumption of constant market coefficients, the model can also be solved with a constant relative risk aversion utility function for the investor. However, the results in this case are qualitatively similar to those with a logarithmic utility function.

[7] Note that maximizing the market value of the fees is different from assuming that the manager is risk neutral. In particular, if the manager is risk averse, earns a fraction of the fees, and can trade in the market without constraint, then his value function depends positively on the market value of the future fees, and it follows that the optimal fund portfolio is that which maximizes equation (2.3).

2. Determine the fund portfolio $\hat{\pi}$ given the investor's best response by solving the optimization problem

$$\sup_{\pi \in \Pi_f} E_Q \left[\int_0^T \frac{d\Phi_t^\pi}{S_t^0} \right],$$

where Φ^π is the cumulative fee process generated by the investor's best response trading strategy, $\phi^\pi \in \Theta_i^\pi$.

Injecting this fund portfolio process in the first step, we obtain a pair $(\hat{\pi}, \phi^{\hat{\pi}})$ of optimal trading strategies. By construction this pair is a Nash equilibrium of the game, and because the definition of the optimal fund portfolio process implies that the corresponding allocations are Pareto optimal, it has to be that this non-cooperative equilibrium is dominating.

3. OPTIMAL TRADING STRATEGIES

In solving for the optimal trading strategies of the two agents, we follow the sequential procedure described in the previous section. The first step is carried in Section 3.1, and the second and third in Section 3.2.

3.1. The Investor's Best Response

In this section, we assume that the fund portfolio is fixed and derive the investor's best response trading strategy.

Given a fund portfolio, π, and the dynamics (2.1), the net-of-fees Sharpe ratio of the fund is given by

$$(3.1) \qquad \vartheta_t^\pi := \frac{[r_t - \gamma + \pi_t^* \sigma_t \xi_t] - r_t}{\|\sigma_t^* \pi_t\|^2} = \frac{\pi_t^* \sigma_t \xi_t - \gamma}{\|\sigma_t^* \pi_t\|^2}.$$

Combining the fact that the investor's utility function is myopic with the fact that he cannot short the fund, it is natural to conjecture that the investor will hold the fund only if it has a positive net-of-fees Sharpe ratio, and that in those cases the proportion of his wealth that will be invested in the fund should equal that Sharpe ratio. The following proposition confirms this intuition.

PROPOSITION 3.1. *Let $\pi \in \Pi_f$ denote an arbitrary fund portfolio process and define a nonnegative process by setting*

$$(3.2) \qquad \phi_t^\pi := \left[\vartheta_t^\pi \right]^+ W_t^{i\pi} = \max \left[0, \vartheta_t^\pi \right] W_t^{i\pi},$$

where $W^{i\pi}$ is the corresponding solution to equation (2.2). Then ϕ^π belongs to Θ_i^π and constitutes the investor's best response trading strategy.

3.2. The Optimal Fund Portfolio

Proposition 3.1 shows that for a given fund portfolio, π, the investor's best response trading strategy generates the cumulative fee process

$$(3.3) \qquad \Phi_t^\pi := \int_0^t \gamma \phi_\tau^\pi \, d\tau = \int_0^t \gamma \left[\vartheta_\tau^\pi \right]^+ W_\tau^{i\pi} \, d\tau.$$

As a result, we have that the optimal fund portfolio solves the optimization problem with value function

$$(3.4) \qquad \hat{v}_0 := \sup_{\pi \in \Pi_f} v_0^\pi = \sup_{\pi \in \Pi_f} E_Q \left[\int_0^T \gamma \left[\vartheta_\tau^\pi \right]^+ \frac{W_\tau^{i\pi}}{S_\tau^0} \, d\tau \right]$$

where v_0^π denotes the market value of the cumulative fee process, Φ^π. Our first result establishes a useful technical property.

LEMMA 3.2. *Fix an arbitrary fund portfolio, $\pi \in \Pi_f$, and let $W^{i\pi}$ denote the corresponding best response wealth process. Then the process*

$$(3.5) \qquad M_t^\pi := \exp \left(\int_0^t \gamma \left[\vartheta_\tau^\pi \right]^+ d\tau \right) \frac{W_t^{i\pi}}{S_t^0 W_0^i}$$

with ϑ^π as in equation (3.1), is a strictly positive martingale under the risk-neutral probability measure.

Let $\pi \in \Pi_f$ denote an arbitrary fund portfolio. The integration of the stochastic differential equation (2.2) yields

$$\int_0^t \frac{d\Phi_\tau^\pi}{S_\tau^0} = W_0^i - \frac{W_t^{i\pi}}{S_t^0} + \int_0^t \left(\frac{\phi_\tau^\pi \pi_\tau^* \sigma_\tau}{S_\tau^0} \right) dZ_\tau.$$

Taking expectations under the risk-neutral probability measure and using the fact that, thanks to Lemma 3.2, the stochastic integral on the far right is a true martingale, we obtain that equation (3.4) can be written as

$$(3.6) \qquad \hat{v}_0 = \sup_{\pi \in \Pi_f} \left\{ W_0^i - E_Q \left[\frac{W_T^{i\pi}}{S_T^0} \right] \right\}.$$

Given the previous results, we can interpret the objective function in equation (3.6) as the difference between the investor's initial wealth and the minimal amount required by an agent who has complete access to financial markets to generate the investor's optimal terminal wealth, $W_T^{i\pi}$. The discrepancy between these two amounts generates the initial market value of the fees to be received by the manager. Thus, we conclude that the optimal fund portfolio is that which "maximizes the incompleteness" of the financial market to which the investor has access.

Combining the result of Lemma 3.2 with Girsanov theorem, we have that, for each π, the formula $Q^\pi(A) := E_Q[1_A M_T^\pi]$ defines an equivalent probability measure under which

$$Z_t^\pi := Z_t + \int_0^t \sigma_\tau^* \pi_\tau \left[\vartheta_\tau^\pi \right]^+ d\tau$$

is an n-dimensional standard Brownian motion. In conjunction with the nonnegativity of the investor's initial wealth, the definition of this new probability measure shows that equation (3.6) can be written as

$$(3.7) \qquad \hat{v}_0 = W_0^i \left\{ 1 - \inf_{\pi \in \Pi_f} E_{Q^\pi} \left[e^{-\int_0^T \gamma [\vartheta_\tau^\pi]^+ d\tau} \right] \right\}.$$

The approach we take to solve this stochastic control problem relies on the interpretation of its objective function as the solution to a backward stochastic differential equation and on the comparison theorems for such equations. Fix an arbitrary fund portfolio process, let

$$f(t, \pi_t) := [\vartheta_t^\pi]^+ = \frac{[\pi_t^* \sigma_t \xi_t - \gamma]^+}{\|\sigma_t^* \pi_t\|^2}$$

denote the positive part of the corresponding Sharpe ratio and consider the backward stochastic differential equation (BSDE in short)

$$(3.8) \qquad -dX_t^\pi = f(t, \pi_t)(\pi_t^* \sigma_t Y_t - \gamma X_t^\pi) dt - Y_t^* dZ_t, \quad X_T^\pi = 1$$

for the pair of processes (X^π, Y). The map f being bounded, it follows from standard results that the unique solution to this equation is given by

$$X_t^\pi = E_{Q^\pi} \left[e^{-\int_t^T \gamma [\vartheta_\tau^\pi]^+ d\tau} \,\middle|\, \mathcal{F}_t \right]$$

and

$$Y_t = e^{\int_0^t \gamma [\vartheta_\tau^\pi]^+ d\tau} \varphi_t$$

where φ is the integrand in the representation of the Q^π-martingale

$$e^{-\int_0^t \gamma [\vartheta_\tau^\pi]^+ d\tau} X_t^\pi = E_{Q^\pi} \left[e^{-\int_0^T \gamma [\vartheta_\tau^\pi]^+ d\tau} \,\middle|\, \mathcal{F}_t \right] = X_0^\pi + \int \varphi_s^* dZ_s^\pi$$

as a stochastic integral with the Brownian motion. Comparing the above expression with equation (3.7), it is now easily seen that $\hat{v}_0 = W_0^i(1 - X_0)$, where

$$(3.9) \qquad X_t := \operatorname{ess} \inf_{\pi \in \Pi_f} X_t^\pi = \operatorname{ess} \inf_{\pi \in \Pi_f} E_{Q^\pi} \left[e^{-\int_t^T \gamma [\vartheta_\tau^\pi]^+ d\tau} \,\middle|\, \mathcal{F}_t \right].$$

Because the terminal condition in equation (3.8) is constant, the comparison theorem (see, e.g., El Karoui et al. 1997, Theorem 2.2) suggests that, to solve for the fund portfolio, one should go through the following steps:

1. Perform the pointwise minimization of the driver

$$(3.10) \qquad g(t, \pi, x, y) := f(t, \pi)(\pi^* \sigma_t y - \gamma x)$$

of the BSDE (3.8) over the set of admissible fund portfolios to obtain a candidate optimal control as a function of (t, ω, x, y).[8]

---

[8] The volatility of the BSDE is not a choice variable because it is entirely determined by the driver and the terminal condition. This is the reason why it is sufficient to optimize the driver of the equation to obtain a candidate optimal control.

2. Plug this candidate optimal control back into the driver, and show that the resulting BSDE admits a solution whose trajectory coincides with the process of equation (3.9).

This approach to the study of stochastic control problems goes back to El Karoui (1981) and was introduced to finance by El Karoui et al. (1997). Since then, this method has been applied to various financial optimization problems by El Karoui and Rouge (2000), Mania and Schweitzer (2005), and Hu and Imkeller (2005), among others.

A straightforward, albeit tedious, computation shows that the minimizer of the mapping defined by equation (3.10) is uniquely given by

$$\hat{\pi}(t, x, y) := 1_{\{\|y/x\| \le \|\xi_t\|\}}[\sigma_t^*]^{-1} \frac{2\gamma \left(\xi_t - \frac{y}{x}\right)}{\|\xi_t\|^2 - \left\|\frac{y}{x}\right\|^2}.$$

Given this candidate optimal control, our task now consists in establishing that the risk-neutral BSDE associated with the driver

$$\hat{g}(t, x, y) := \min_{\pi \in \mathbb{R}^n} g(t, \pi, x, y) = g\big(t, \hat{\pi}(t, x, y), x, y\big)$$

$$(3.11) \qquad\qquad = -1_{\{\|y/x\| \le \|\xi_t\|\}} \frac{x}{4} \left\|\xi(t) - \frac{y}{x}\right\|^2$$

and the terminal condition 1 admits a solution which coincides with the process of essential infima defined in equation (3.9). This existence issue is the main difficulty of our approach. In particular, because the driver of the BSDE

$$(3.12) \qquad\qquad -dX_t = \hat{g}(t, X_t, Y_t)dt - Y_t^* dZ_t, \quad X_T = 1,$$

has quadratic growth in Y, the standard existence results do not apply. Nevertheless, relying on the results of Kobylanski (2000), we are able to establish the existence of a minimal solution, and this will prove sufficient for the purpose of this paper.

PROPOSITION 3.3. *Let $\beta := X^{-1}Y$, where (X, Y) denotes the minimal solution to equation (3.12), and define a vector-valued process by setting*

$$(3.13) \qquad\qquad \hat{\pi}_t := \hat{\pi}(t, X_t, Y_t) = 1_{\{\|\beta_t\| \le \|\xi_t\|\}}[\sigma_t^*]^{-1} \frac{2\gamma (\xi_t - \beta_t)}{\|\xi_t\|^2 - \|\beta_t\|^2}.$$

Then $\hat{\pi}$ belongs to the set Π_f of admissible fund portfolios and achieves the infimum in the stochastic control problem (3.7).

Proposition 3.1 suggests that, because the investor's demand for the fund is proportional to the fund's net-of-fees Sharpe ratio, the optimal fund portfolio should be that which maximizes the fund's net-of-fees Sharpe ratio. A straightforward computation shows that the unique maximizer of the fund's net-of-fees Sharpe ratio is

$$(3.14) \qquad\qquad \tilde{\pi}_t := \arg \max_{\pi \in \mathbb{R}^n} f(t, \pi) = [\sigma_t^*]^{-1} \frac{2\gamma \xi_t}{\|\xi_t\|^2}.$$

Comparing this with the optimal fund portfolio process defined by equation (3.13), we see that the intuition according to which $\tilde{\pi}$ should be optimal is false in general. As we now demonstrate, this intuition is valid if the manager's investment opportunity is nonstochastic.

COROLLARY 3.4. *Assume that the market price of risk process, ξ, is a constant vector. Then the unique solution to equation (3.12) is given by*

$$(3.15) \qquad X_t := \exp\left(-\frac{1}{4}\|\xi\|^2 (T-t)\right)$$

and $Y := (0, \ldots, 0)^* \in \mathbb{R}^n$. In this case, the optimal fund portfolio process coincides with the bounded process, $\tilde{\pi}$, defined in equation (3.14).

Combining equations (3.13) and (3.14) with Proposition 3.1, one might be tempted to conclude that, given constant market coefficients, the optimal fund portfolio is that which maximizes the investor's welfare. This is, of course, false, as it is straightforward to verify that, if the investor were to determine the fund's portfolio strategy while still paying the applicable management fees, he would choose it to maximize the product $\|\sigma_t^* \pi\|[\vartheta_t^\pi]^+$ and not the net-of-fees Sharpe ratio of the fund.[9]

4. EQUILIBRIUM

We now gather the previous results to obtain a description of the equilibrium. We start with the general case in which the coefficients are arbitrary and then proceed with the special case in which they are constant.

THEOREM 4.1. *Let $\beta := X^{-1}Y$, where (X, Y) denotes the minimal solution to equation (3.12). Then the equilibrium fund portfolio and the investor's equilibrium trading strategy are given by*

$$(4.1) \qquad \hat{\pi}_t = 1_{\{\mathcal{B}\}} [\sigma_t^*]^{-1} \frac{2\gamma(\xi_t - \beta_t)}{\|\xi_t\|^2 - \|\beta_t\|^2},$$

$$(4.2) \qquad \phi_t^{\hat{\pi}} = 1_{\{\mathcal{B}\}} (\|\xi_t\|^2 - \|\beta_t\|^2) \left[\frac{W_t^{i\hat{\pi}}}{4\gamma}\right],$$

where $\mathcal{B}$ denotes the set on which $\|\beta\| \le \|\xi\|$. The initial market value of the fees and the investor's expected utility are given by

$$\hat{v}_0 = W_0^i (1 - X_0),$$
$$(4.3) \qquad V_i = E\left[\log\left(W_0^i S_T^0\right) + \int_0^T \frac{1_{\{\mathcal{B}\}} \|\xi_t - \beta_t\|^2}{8} dt\right],$$

where W_0^i is the investor's strictly positive initial capital.

Equation (4.1) shows that in equilibrium the equity component of the fund is proportional to the fee rate. Thus, our model predicts that, ceteris paribus, funds with higher management fees will invest more in equities and hence will have a higher volatility. This is consistent with the recent findings of Cremers and Petajisto (2007, Table 4.2), who

---

[9] Using standard algebra we have that the value function of this problem is given by $\sup_{\pi \in \mathbb{R}^n} \|\sigma_t^* \pi\|[\vartheta_t^\pi]^+ = \|\xi_t\|$, but an application of the Cauchy–Schwartz inequality shows that there is no fund portfolio that attains this value. This is intuitive because otherwise the investor would be able to attain the same utility as in a complete financial market while still paying a nontrivial amount of fees.

document a positive relation between expense ratios and tracking error volatility, and of Gil-Bazo and Ruiz-Verdù (2008) who find a positive relation between expense ratios and fund volatility. Similarly, Carhart (1997, Table 4.1) shows that there is a positive relation between how aggressive a fund category is and the average expense ratio in that category. Elton et al. (1993), find that funds with high expense ratios invest less in bonds than funds with low expense ratios.[10]

As the fee rate increases, a larger part of the investor's portfolio is invested directly in the bond. This happens through two distinct channels. First, and as expected, the higher the fee rate the lower the fraction of the investor's wealth that is invested in the fund (equation [4.2]).[11] Second, because the fee is charged on assets under management, it applies to both the equity and the bond components of the fund. Obviously, holding the bond through the fund is more expensive for the investor than holding it directly. The increased fund allocation to equity decreases the investor's indirect bond holdings and thereby mitigates his reaction to a higher fee rate.[12]

The fact that the solution to equation (3.12) is independent of the fee rate implies that the investor's effective portfolio weights are independent of the fee rate. Furthermore, multiplying equation (4.2) by the fee rate shows that the fraction of the investor's wealth paid as management fees is fixed. Combining the earlier yields that the investor's wealth process, the cumulative fee process, and the value functions are independent of the fee rate. In our model, the manager's ability to select the fund portfolio and the investor's ability to adjust his holdings in the fund compensate each other so that the agents are equally well off whether one or the other is setting the fee rate.[13]

The second term on the right-hand side of equation (4.3) is the utility gain attributed to being able to trade the mutual fund, in addition to the bond. Even though the investor is a price taker and the manager is strategic, the two agents split the surplus. This result is also found in Das and Sundaram (2002b), who also model the fund manager as a Stackelberg leader, and is due to the following. First, in the setting of our model making the investor better off relative to the case where he could invest only in the bond is not necessarily inconsistent with making the manager better off. Second, the investor's ability to move funds in and out of the fund limits the manager's ability to take advantage of him.

Before analyzing further properties of the equilibrium, let us first specialize the results of Theorem 4.1 to the case of constant coefficients.

COROLLARY 4.2. *Assume that the market coefficients are constant. Then the equilibrium fund portfolio and trading strategy are given by*

$$\hat{\pi} = [\sigma^*]^{-1} \frac{2\gamma\xi}{\|\xi\|^2}, \qquad \phi_t^{\hat{\pi}} = \|\xi\|^2 \left[\frac{W_t^{i\hat{\pi}}}{4\gamma} \right].$$

---

[10] Studies analyzing the relationship between returns and expense ratios have typically found a negative relation between net-of-fees risk-adjusted returns and expense ratios (see, for example, Carhart [1997]). However, Ippolito (1989) finds an insignifcant relation between net-of-fees return in excess of the risk free rate and expense ratios. Considering net-of-fees unadjusted returns, Droms and Walker (1996) find a positive relation.

[11] This implication of the model is consistent with Sirri and Tufano (1998), who document that changes in expense ratios are inversely related to fund flows.

[12] This argument assumes that the fund does not invest more than 100% of its assets in the bond. Although we are not able to rule out this possibility, it is highly unlikely.

[13] The model is discontinuous at $\gamma = 0$, as can be seen from the portfolio holdings in Theorem 4.1 and the different fee-irrelevance results.

The initial market value of the fees and the investor's expected utility are given by

(4.4)
$$\hat{v}_0 = W_0^i \left[1 - \exp\left(-\frac{1}{4} \|\xi\|^2 (T - t) \right) \right],$$

$$V_i = \log\left(W_0^i S_T^0 \right) + \frac{1}{8} \|\xi\|^2 T,$$

where W_0^i denotes the investor's strictly positive initial capital.

Assuming a constant investment opportunity set and multiplying the fund portfolio by the investor's trading strategy, we get that the investor's effective equity portfolio weights are

$$\hat{\pi}_t \phi_t^{\hat{\pi}} = [\sigma^*]^{-1} W_t^{i\hat{\pi}} \frac{\xi}{2}.$$

Thus, the relative composition of the investor's effective equity portfolio coincides with the one he would have chosen if he had been able to trade freely in the stocks. However, his overall exposure to equity risk is smaller by a factor of two.

A stochastic opportunity set, on the other hand, results in a distortion to the equity component of the fund's portfolio, as Theorem 4.1 implies that the investor's effective equity portfolio weights are

$$\hat{\pi}_t \phi_t^{\hat{\pi}} = 1_{\{\mathcal{B}\}} [\sigma_t^*]^{-1} W_t^{i\hat{\pi}} \left(\frac{\xi_t - \beta_t}{2} \right).$$

The intuition behind this distortion is as follows: the manager knows that in states of the world where the instantaneous opportunity set is good (i.e., states in which the fund's Sharpe ratio is high) the investor invests a large fraction of his wealth in the fund, thereby generating large fees. This "flow hedging" consideration induces the manager to distort the fund's equity portfolio in such a way as to increase the correlation between the investor's wealth and the fund's net-of-fees Sharpe ratio. To quantify these distortions, we consider in the next section a model which has a stochastic investment opportunity set and for which the equilibrium can be computed in closed form.

An important aspect of the mutual fund industry that we do not consider in our analysis is competition among funds. However, following Breton, Hugonnier, and Masmoudi (2010), it is possible to show that, at least under the assumption of constant coefficients, this feature does not change the nature of the equilibrium.

Breton, Hugonnier, and Masmoudi (2010) studied a model similar to ours, with market constant coefficients and $m \geq 2$ mutual funds. They assumed that fund k charges fraction-of-fund fees at rate $\gamma_k > 0$ and that all the funds have access to the same pool of assets. In modeling the competition between mutual funds, they assumed that all the managers are strategic compared to the investor and that, given the investor's best response, the managers play a Nash game. Under these assumptions, Breton, Hugonnier, and Masmoudi (2010) show that there exists a unique Pareto optimal equilibrium in which the kth manager chooses

$$\pi_k = [\sigma^*]^{-1} \frac{2\gamma_k \xi}{\|\xi\|^2}.$$

This implies that all mutual funds offer the same risk/return tradeoff, and it follows that the investor is indifferent between having access to one or more mutual funds. The

intuition for this result is as follows. With multiple funds, the relevant variable for the investor is not the individual net-of-fees Sharpe ratios of the funds but their relative net-of-fees Sharpe ratios. This implies that the amount invested in fund k depends not only on the portfolio of that fund but also on the portfolio of the other funds. Now fix the portfolio of all but one fund and consider the optimization problem of the remaining fund. The fact that the investor cannot short the funds implies that, given the strategy of the other funds, it is always possible for the manager of the remaining fund to choose his portfolio in such a way that the investor does not invest with any of the other funds. Iterating this argument shows that the only possible equilibrium is one in which the funds offer the same risk return tradeoff.

This striking result implies that the equilibrium we identify in Corollary 4.2 is also an equilibrium for the case in which there are multiple mutual funds. A similar conclusion was reached by Cetin (2006) in a model with mean variance preferences and a single risky asset. For the case of stochastic market coefficients, the situation is much more complicated, because one has to study a stochastic game between the managers where the pay-offs are given in terms of a solution to a system of BSDEs.

Our model also has some implications for the contemporaneous relation between performance and fund flows. The empirical literature has mostly focused on the relationship between past returns and current flows measured over long horizons such as a quarter or a year. However, in some cases, such as Sirri and Tufano (1998), the potential relation between flows and contemporaneous returns is not controlled for. Our model allows us to show the importance of controlling for this relation.

Denote by $R_t(\Delta)$ the total return, net-of-fees, that the investor would obtain from investing in the fund between dates $t - \Delta$ and t, that is,

$$R_t(\Delta) := \frac{F_t}{F_{t-\Delta}}.$$

The measure of fund flows that we consider is the one typically used in the literature (see, e.g., Chevalier and Ellison 1997) and is given by

$$(4.5) \qquad \rho_t(\Delta) := \frac{\phi_t - R_t(\Delta)\phi_{t-\Delta}}{\phi_{t-\Delta}} = \frac{\phi_t}{\phi_{t-\Delta}} - R_t(\Delta).$$

The following result characterizes the contemporaneous relationship between performance and fund flows in a market with constant coefficients.

PROPOSITION 4.3. *Assume that the coefficients are constant. Then*

1. *As a function of the total return on the fund, the measure $\rho_t(\Delta)$ is inverse U-shaped function if $\|\xi\|^2 < 4\gamma$ and U-shaped otherwise.*
2. *For a given level of total return, the measure $\rho_t(\Delta)$ is increasing in the fee rate if $2\gamma^2 \geq (\Delta^{-1}\log(R_t(\Delta)) - r)\|\xi\|^2$ and decreasing otherwise.*

Given constant coefficients, both the investor and the mutual fund hold fixed portfolios. If the investor is non-levered, then he holds the bond both directly and indirectly through his holdings of the fund. Therefore, the bond constitutes a larger fraction of his portfolio relative to that of the fund. When the investor is levered, his portfolio is tilted more toward equities relative to the fund portfolio. As can be seen from Corollary 4.2, the investor holds a levered position in the fund if and only if $4\gamma < \|\xi\|^2$. The first part of the proposition thus shows that the direction of the wedge between the investor's portfolio

and the fund portfolio determines whether the relationship between flows and returns will be U-shaped or inverse U-shaped. This implies that the empirical specification should be flexible enough to allow the derivative of the flow measure with respect to returns to switch sign.

The second part of the proposition is a consequence of the fact that a higher fee rate implies that a larger part of the investor's portfolio is held directly in the bond. For example, following a negative fund return the investor needs to inject money into the fund to rebalance his portfolio back to the optimal composition. Holding this return level fixed, a higher fee rate implies a higher return on the investor's portfolio, which in turn implies less rebalancing on the part of the investor. Although the condition in the second part of the proposition depends both on the fund return and the fee rate, it is evident that there are three regions: for sufficiently low returns the flow is an increasing function of the fee rate; for sufficiently high returns, the flow is a decreasing function of the fee rate; and for intermediate values, the relationship is inverse U-shaped.

The results of the proposition provide guidance on the way to control for the contemporaneous relationship between fund flows and returns. First, one should include a contemporaneous return variable as one of the independent variables. Second, although the expense ratio is typically included as one of the control variables in these studies, the second part of the proposition suggests that one should also include an interaction term between the fee rate and contemporaneous returns.

The empirical study of the contemporaneous relationship between flows and returns has been mostly restricted to analyzing the relationship between aggregate security returns and aggregate flows to the mutual fund industry, thereby ignoring cross-sectional differences in expense ratios. Evidence documenting a positive contemporaneous relationship appears in Patel, Hendriks, and Zeckhauser (1991), Warther (1995), and Edelen and Warner (2001) at the quarterly, monthly, and daily horizons, respectively.

As mentioned earlier, a large part of the empirical literature has focused on analyzing the relationship between flows and lagged returns. With constant coefficients, it is straightforward to show that fund flows, as measured by $\rho_t(\Delta)$, and lagged returns are uncorrelated, as would be expected with a constant investment opportunity set. To better understand the implications of our model, we next consider in detail an example with a stochastic opportunity set. Specifically, we demonstrate that with a stochastic opportunity set our model can generate an increasing relation between fund flow and lagged returns.

4.1. A Closed Form Example

Consider a financial market that consists of three securities: a riskless bond with constant rate of return r and two stocks. Let S_i denote the price process of stock i and assume that

$$(4.6) \qquad dS_{1t} = r\,S_{1t}dt + \sigma_{11}\,S_{1t}[d\,B_{1t} + \xi_1 dt],$$

$$(4.7) \qquad dS_{2t} = r\,S_{2t}dt + \sigma_{21}\,S_{2t}[d\,B_{1t} + \xi_1 dt] + \sigma_{22}\,S_{2t}[d\,B_{2t} + \xi_{2t}dt],$$

where $(B_i)_{i=1}^2$ are two independent Brownian motions, $(\xi_1, \sigma_{11}, \sigma_{21}, \sigma_{22})$ are constants and the market price of risk, ξ_{2t}, follows the mean reverting dynamics

$$(4.8) \qquad d\xi_{2t} = \lambda\left(\overline{\xi_2} - \xi_{2t}\right)dt + \psi\,d\,B_{2t},$$

for some constants $(\lambda, \overline{\xi_2}) \in \mathbb{R}_+^2$ and $\psi \in \mathbb{R}$. To focus on the impact of a time-varying opportunity set, we assume that $\sigma_{22} \neq 0$.

For this two-dimensional model, the backward equation (3.12) can be solved in closed form as a function of time and ξ_{2t}. In the proof of the following proposition, we compute this closed form solution and then plug it into Theorem 4.1 to obtain an explicit description of the equilibrium.

PROPOSITION 4.4. *Assume that condition (A.9) holds. Then the equilibrium fund portfolio and the investor's trading strategy are given by*

$$\hat{\pi}_t = [\sigma^*]^{-1} \frac{2\gamma(\xi_t - \eta_t \mathbf{e}_2)^+}{\|\xi_t\|^2 - |\eta_t|^2}, \quad \phi_t^{\hat{\pi}} = \left(\|\xi_t\|^2 - |\eta_t|^2 \right) \frac{W_t^i}{4\gamma},$$

where

(4.9) $$\eta_t := 2\psi(2 A_1(t)\xi_{2t} + A_2(t))$$

for some deterministic functions $(A_i)_{i=1}^2$, which are defined in equations (A.7) and (A.8) of the appendix and we have set $\mathbf{e}_2 = (0, 1)^*$.

The presence of a time-varying opportunity set leads the fund to pick a portfolio that does not maximize the fund's net-of-fees Sharpe ratio even though the investor has a logarithmic utility function. As shown in the proposition, this induces the investor to invest a smaller fraction of his wealth in the fund relative to the case in which both risk premia are constant.

To facilitate the comparative statics, we focus on the case where the two stocks are uncorrelated, so that $\sigma_{21} = 0$. Flow hedging occurs as soon as η_t is nonzero and hence arises even in this case, because the functions A_i do not depend on the stock volatility. Assuming that $\sigma_{21} = 0$ and using equation (3.14), we have that the myopic fund portfolio, which results from the pointwise maximization of the fund's net-of-fees Sharpe ratio, is

$$\tilde{\pi}_{1t} = \frac{1}{\sigma_{11}} \frac{2\gamma\xi_1}{|\xi_1|^2 + |\xi_{2t}|^2}, \quad \tilde{\pi}_{2t} = \frac{1}{\sigma_{22}} \frac{2\gamma\xi_{2t}}{|\xi_1|^2 + |\xi_{2t}|^2}.$$

On the other hand, specializing the results of Proposition 4.4 to the case where the two stocks are uncorrelated we obtain the following:

COROLLARY 4.5. *Assume that $\sigma_{21} = 0$ and that condition (A.9) holds. Then the equilibrium fund portfolio is given by*

$$\hat{\pi}_{1t} = \frac{1}{\sigma_{11}} \frac{2\gamma\xi_1}{|\xi_1|^2 + |\xi_{2t}|^2 - |\eta_t|^2}, \quad \hat{\pi}_{2t} = \frac{1}{\sigma_{22}} \frac{2\gamma(\xi_{2t} - \eta_t)}{|\xi_1|^2 + |\xi_{2t}|^2 - |\eta_t|^2},$$

where the process η is defined as in equation (4.9).

Corollary 4.5 implies that, as long as σ_{11} and ξ_1 are both positive, the weight of stock 1 in the equilibrium fund portfolio will be higher than the one associated with the myopic portfolio. However, the fact that

$$\hat{\pi}_t \phi_t^{\hat{\pi}} = W_t^i \left(\frac{\xi_1}{2\sigma_{11}}, \frac{\xi_{2t} - \eta_t}{2\sigma_{22}} \right)^*$$

TABLE 4.1
Parameter Values for the Base Case (in Monthly Units)

	Symbol	Value
Investor's horizon	T	12 months
Mean reversion parameter	λ	0.0226
Long-term mean of the market price of risk	$\overline{\xi_2}$	0.0788
Volatility of the market price of risk	ψ	-0.0189
Risk premium of the first stock	ξ_1	0.0788
Stock volatilities	$\sigma_{11} = \sigma_{22}$	0.0436
Instantaneous fee rate	γ	0.00125

shows that the fraction of the investor's wealth that is effectively invested in stock 1 is the same as in the myopic strategy.

To further analyze the dependence of the fund's portfolio on the different components of the model, we focus on a calibrated example. The parameters we use for the stock and market price of risk dynamics are taken from Wachter (2002) and are summarized in Table 4.1.

Figure 4.1 displays the fund allocation to the stocks under both the optimal and the myopic strategies as a function of ξ_2. The top panel shows that for the calibrated parameters the magnitude of the difference of the stock 1 portfolio weight is small. In the figure it is hard to see, but the two allocations are slightly different because η is nonzero. Computing the ratio of the two strategies shows that over the plotted range the percentage difference between the optimal and the myopic strategies is at most eight basis points. As we will show later, the impact on the stock 1 portfolio weight becomes larger as the volatility of the market price of risk of stock 2 increases.

The bottom panel shows the corresponding allocations to the second stock. Here the differences are considerably more pronounced. For example, at the long-term level of the market price of risk ($\xi_2 = \overline{\xi_2}$), the optimal portfolio weight is 15% lower than the myopic portfolio weight. Throughout most of the plotted range the allocation to stock 2 is lower under the optimal strategy than under the myopic one.[14] Flow hedging considerations are the driving force behind these distortions. When $\psi < 0$ (implying $\eta > 0$ if $\xi_2 > 0$), the returns to stock 2 are negatively correlated with innovations to its Sharpe ratio. By reducing the allocation to stock 2, the manager is increasing the correlation between the investor's future wealth, which is directly impacted by the return of the fund, and future Sharpe ratios. This increased correlation in turn increases the market value of future fees, which depend on the product of the investor's wealth and the fund's net-of-fees Sharpe ratio. On the other hand, when $\psi > 0$ (implying $\eta < 0$ if $\xi_2 > 0$) the optimal strategy invests more in stock 2 than the myopic strategy, as can be inferred from Corollary 4.5. This flow hedging activity again increases the correlation between investor's wealth and future Sharpe ratios, leading to a higher market value of future fees.

[14] The *x*-axis of this figure is six standard deviations around the long-term mean of the market price of risk. To the left of -0.01 the equilibrium allocation to the second stock is higher under the optimal strategy. In that region, stock 2 is shorted under both strategies. Thus, a higher portfolio allocation again implies a less extreme position.
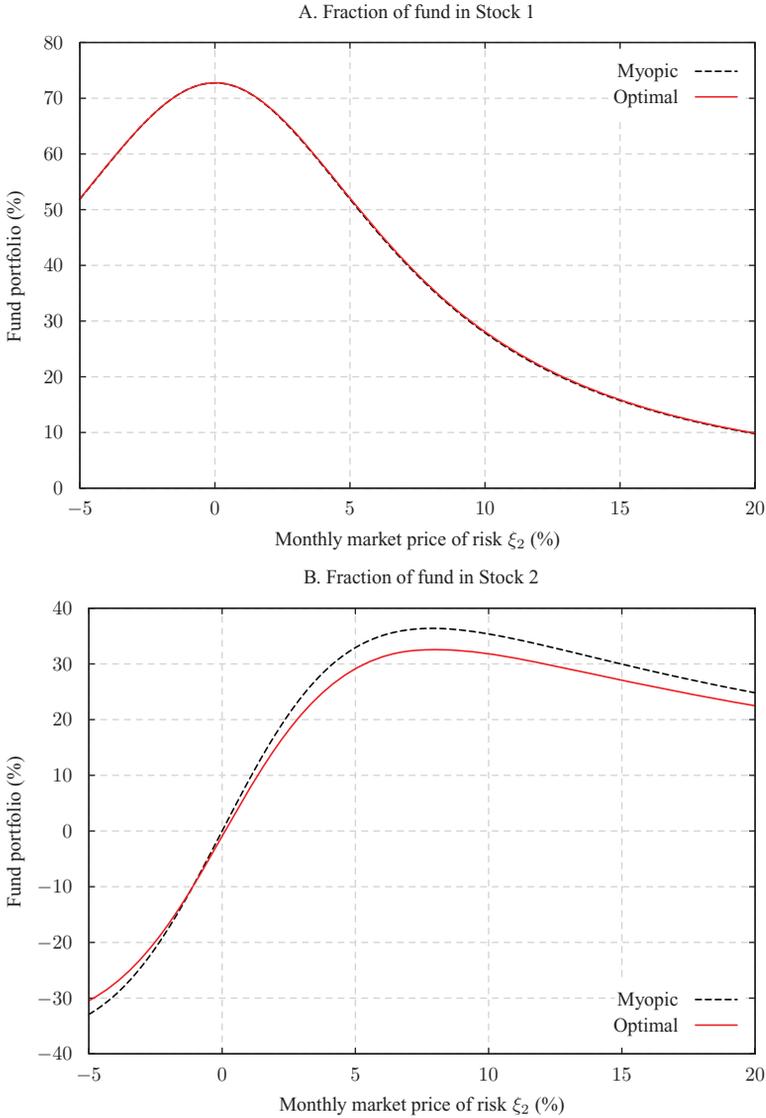
FIGURE 4.1.   Fund portfolio as a function of the market price of risk, ξ_2. The parameters of the model are reported in Table 4.1.

Given that the two securities are uncorrelated, Proposition 4.4 implies that the equilibrium net-of-fees Sharpe ratio is given by

$$\hat{\vartheta}_t = \frac{||\xi_t||^2 - \eta_t^2}{4\gamma}.$$

On the other hand, it follows from equation (3.14) that the maximal net-of-fees Sharpe ratio is given by

$$\tilde{\vartheta}_t = \frac{||\xi_t||^2}{4\gamma}.$$
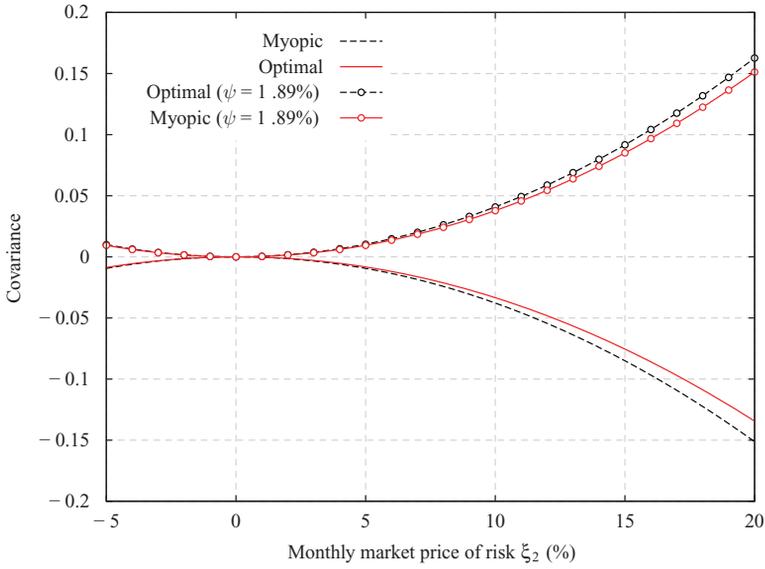
FIGURE 4.2. Instantaneous covariance between the net-of-fees Sharpe ratio and the investor's wealth as a function of the market price of risk, ξ_2. The parameters of the model are reported in Table 4.1.

A direct calculation then shows that the instantaneous covariance between changes in net-of-fees Sharpe ratios and changes in the investor's wealth for the optimal and myopic fund strategies, taking into account the investor's best response to each fund strategy, are respectively given by

$$\frac{\psi(\xi_{2t} - 4\psi A_1(t)\eta_t)(\xi_{2t} - \eta_t)W_t^i}{4\gamma},$$

and

$$\frac{\psi\xi_{2t}^2 W_t^i}{4\gamma}.$$

Figure 4.2 plots these two covariances as functions of the risk premium. As expected, the covariance associated with the optimal strategy is always higher than that associated with the myopic strategy.

Figure 4.3 shows the allocations to the two stocks as functions of the mean reversion parameter, λ. When the mean reversion parameter is very high, the market price of risk of the second stock is going to stay fairly close to its long-term mean. As a result, the optimal strategy will be fairly close to its myopic counterpart. On the other hand, for low levels of the mean reversion parameter the market price of risk of stock 2 becomes highly persistent, leading to larger flow hedging distortions.

Figure 4.4 shows the allocations to the stocks as functions of the market price of risk volatility, ψ. While for stock 2 the allocation increases as a function of ψ, for stock 1 the relationship is U-shaped. When the market price of risk volatility is negative and large in absolute value, the allocation to both stocks can differ significantly from the allocation under the myopic strategy.
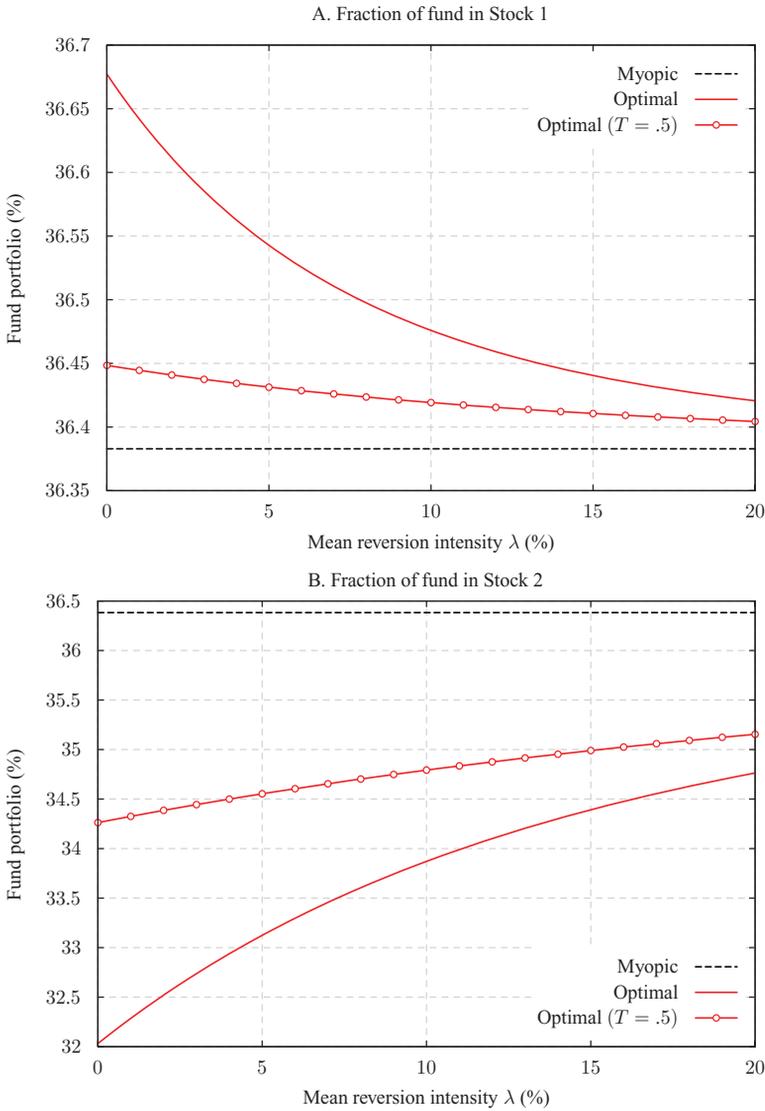
FIGURE 4.3. Fund portfolio as a function of the mean reversion parameter, λ. The parameters of the model are reported in Table 4.1.

Computing the ratio of the holdings in the two stocks shows that the fraction of equity invested in stock 2 is increasing as a function of ψ. For a given level of the absolute value of ψ, the magnitude of the distortion relative to the myopic case is larger when ψ is negative. This follows from combining the following two facts. First, when ψ is negative (positive) flow hedging considerations decrease (increase) the allocation to stock 2 as ψ decreases (increases). Second, a more volatile market price of risk increases uncertainty, and as such will tend to decrease the allocation to stock 2.

We next demonstrate that our model can produce an increasing relation between fund flows and lagged returns, as has been documented, for example, by Chevalier and Ellison (1997) and Sirri and Tufano (1998).

A. Fraction of fund in Stock 1

B. Fraction of fund in Stock 2
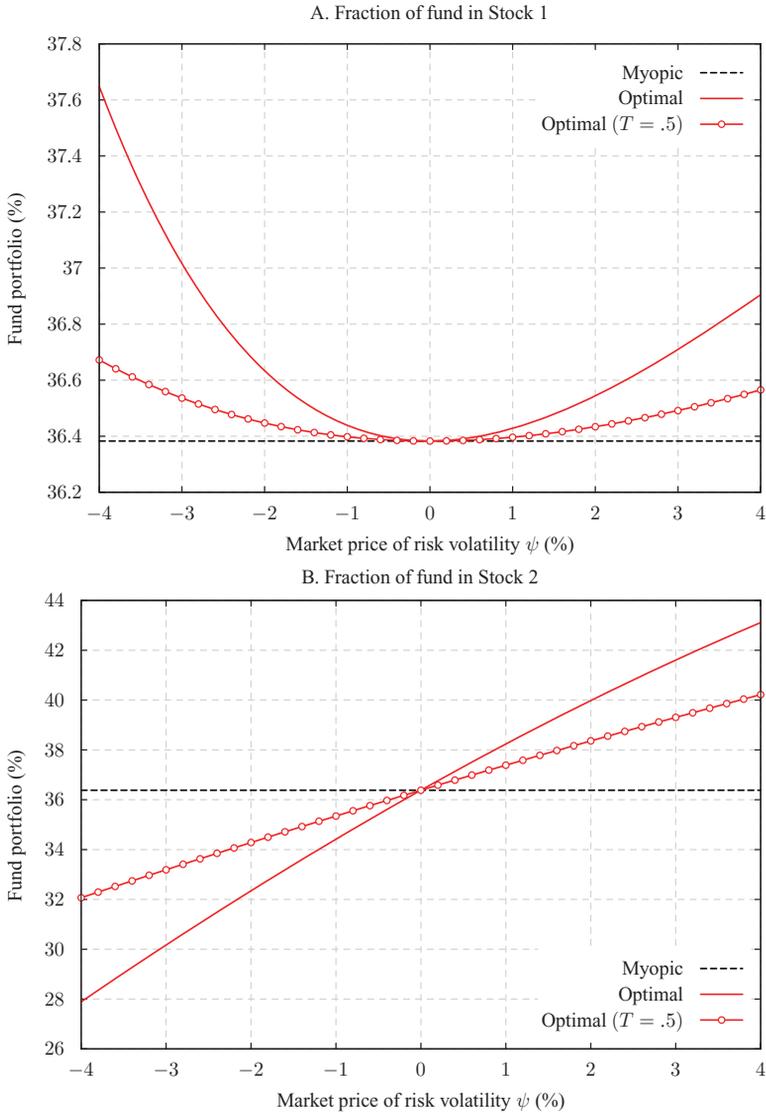
FIGURE 4.4. Fund portfolio as a function of volatility of the market price of risk, ψ.
The parameters of the model are reported in Table 4.1.

The equilibrium dynamics of the investor's wealth and of the fund returns are explicitly
given by

$$dW_t = W_t \left[\frac{1}{2}\xi_1 dB_{1t} + \frac{1}{2}(\xi_{2t} - \eta_t)dB_{2t} + \frac{1}{4}\|\xi_t - \mathbf{e}_2\eta_t\|^2 dt \right],$$

$$dF_t = \frac{F_t}{P_t} \left[\frac{1}{2}\xi_1 dB_{1t} + \frac{1}{2}(\xi_{2t} - \eta_t)dB_{2t} + \frac{1}{4}\|\xi_t - \mathbf{e}_2\eta_t\|^2 dt \right],$$

where the process η_t is defined in equation (4.9) and we have set

$$P_t := \frac{1}{4\gamma}[\xi_1^2 + \xi_{2t}^2 - \eta_t^2].$$

To analyze the relationship between flows and lagged returns, we simulate 100 sets of 500 trajectories of the processes (ξ_2, W, F), where each trajectory is created from a partition of 1000 subintervals. The first half of each sample path is used to compute the lagged returns and the second half to compute the subsequent fund flows according to the measure ρ of equation (4.5). For each of the 100 sets of trajectories, we then estimate the following regression on the simulated data

$$(4.10) \qquad \rho_t(\Delta) = \alpha + \beta[R_{t-\Delta}(\Delta) - 1] + \kappa[R_{t-\Delta}(\Delta) - 1]^2 + \epsilon_t.$$

Table 4.2 gives the quantiles of the distribution of the T-statistics for the parameter estimates across the 100 regressions. The base case corresponds to the parameters in Table 4.1 with a horizon of 2 years, so that past returns are measured over the first year and flows are measured over the second.

As shown in the table, flows are typically, significantly, positively related to past returns, the only exception being when shocks to ξ_2 and S_2 are positively correlated. However, the relationship is not convex, whereas the empirical evidence suggests that it is (see Chevalier and Ellison 1997; Sirri and Tufano 1998). When shocks to ξ_2 and S_2 are negatively correlated, high past fund returns will tend to be associated with current values of ξ_2 that are below its long-term mean and which will increase in the future due to mean reversion. Because the fund's net-of-fees Sharpe ratio is positively related to ξ_2, this leads to a positive observed relationship between lagged returns and fund flows. In addition, comparing the magnitude of the regression coefficients shows that larger fees and a shorter time horizon both imply a less sensitive relationship.[15]

Finally, we note that the source of the positive relationship that we identify is different from the one in Berk and Green (2004) and Dangl, Wu, and Zechner (2006). In these models, the information about managerial ability is incomplete, and both the investors and the manager are learning about it from observing fund returns. In such a setting, high past returns lead investors to increase their estimate of the manager's ability and thus increase their investment in the fund.

4.2. Restricting the Fund to Holding Only Equity

Given that the fund has an advantage in accessing equity relative to the investor but does not have an advantage in accessing the bond, it is natural to also consider the case in which the fund is restricted to holding only equity. For simplicity of exposition, we restrict the analysis of this section to the case with constant market coefficients.

For a given fee rate, the fund's optimal portfolio is still the one that maximizes the net-of-fees Sharpe ratio, with the additional restriction that it can trade only equity. Solving

$$(4.11) \qquad \sup_{\pi \in \mathbb{R}^n} \left\{ \frac{[\pi^* \sigma \xi - \gamma]^+}{\|\sigma^* \pi\|^2} \quad \text{s.t.} \quad \pi^* \mathbf{1} = 1 \right\}$$

and plugging the result into the formula of Proposition 3.1, we obtain the following counterpart to Corollary 4.2.

---

[15] Huang, Wei, and Yan (2005) document differences in performance flow sensitivity for different levels of total fees, where total fee is the expense ratio plus one-seventh of the up-front load.

TABLE 4.2
Quantiles of of the T-Statistics for the Estimates of the Parameters of the Regression
of Fund Flows on Lagged Returns (equation [4.10])

	(A) Quantiles of the T-statistics for β						
	5%	10%	25%	50%	75%	90%	95%
Base case	2.66	3.00	3.47	4.07	4.83	5.53	5.71
$\gamma = 0.000625$	2.54	2.86	3.46	4.05	4.57	5.21	5.48
$\gamma = 0.0025$	2.44	2.62	3.22	3.86	4.44	4.93	5.52
$\sigma_{11} = \sigma_{22} = 0.02$	2.23	2.61	3.50	4.03	4.77	5.43	5.83
$\sigma_{11} = \sigma_{22} = 0.09$	2.62	2.80	3.44	3.98	4.58	5.01	5.42
$\lambda = 0.01$	2.02	2.20	2.67	3.36	4.01	4.38	4.86
$\lambda = 0.05$	3.87	4.11	4.67	5.45	6.16	6.97	7.22
$\psi = -0.04$	1.65	1.97	2.50	3.32	4.00	4.33	4.65
$\psi = 0.0189$	-3.29	-3.09	-2.29	-1.60	-0.96	-0.39	-0.14
$\sigma_{22} = 0.09$	3.03	3.19	3.61	4.25	4.72	5.32	5.76
$\bar{\xi}_2 = \xi_1 = 0.04$	1.64	2.55	3.13	3.84	4.58	5.04	5.22
$T = 12$ months	1.73	1.95	2.62	3.18	3.81	4.50	4.66

	(B) Quantiles of the T-statistics for κ						
	5%	10%	25%	50%	75%	90%	95%
Base case	-2.31	-2.04	-1.49	-1.00	-0.23	0.63	1.04
$\gamma = 0.000625$	-2.35	-1.96	-1.43	-0.52	0.23	0.81	0.95
$\gamma = 0.0025$	-2.81	-2.34	-1.90	-1.25	-0.35	0.54	0.90
$\sigma_{11} = \sigma_{22} = 0.02$	-2.68	-2.24	-1.60	-0.89	-0.08	0.56	0.87
$\sigma_{11} = \sigma_{22} = 0.09$	-2.26	-1.87	-1.48	-0.71	-0.06	0.64	1.13
$\lambda = 0.01$	-2.00	-1.81	-1.23	-0.66	-0.03	0.78	1.39
$\lambda = 0.05$	-3.08	-2.65	-1.82	-1.20	-0.48	0.19	0.61
$\psi = -0.04$	-1.93	-1.33	-0.88	-0.26	0.62	1.30	1.62
$\psi = 0.0189$	-1.58	-1.26	-0.45	0.14	0.68	1.61	2.08
$\sigma_{22} = 0.09$	-2.43	-1.80	-1.40	-0.88	-0.15	0.60	0.79
$\bar{\xi}_2 = \xi_1 = 0.04$	-2.41	-2.15	-1.74	-0.96	-0.06	0.76	1.11
$T = 12$ months	-2.32	-1.83	-1.08	-0.65	0.07	0.97	1.36

Note: The base case corresponds to $t = T = 24$ months, $\Delta = T/2$, and the parameters
in Table 4.1. For each of the other models, one parameter is varied relative to the base
case.

PROPOSITION 4.6. *Assume that the coefficients are constant and that the fund is restricted
from investing in the bond. Then the equilibrium fund portfolio and the investor's trading
strategy are given by*

$$\hat{\pi} = \pi^{(\gamma)} := [\sigma^*]^{-1} \frac{\xi - \lambda^{(\gamma)} A}{\xi^* A - \lambda^{(\gamma)} \|A\|^2},$$

$$\phi_t^{\hat{\pi}} = (\xi^* A - \lambda^{(\gamma)} \|A\|^2) \left[\frac{W_t^{i\hat{\pi}}}{2} \right],$$

where $\lambda^{(\gamma)} := \gamma - \|\xi - \gamma A\| / \|A\|$ and $A := [\sigma^{-1}\mathbf{1}]$. The initial market value of the fees and the investor's expected utility are given by

$$(4.12) \qquad \hat{v}_{0(\gamma)} = W_0^i \left[1 - \exp\left(-\frac{\gamma}{2} [\xi - \lambda^{(\gamma)} A]^* AT \right) \right],$$

$$(4.13) \qquad V_i = \log\left(W_0^i S_T^0 \right) + \frac{1}{8} \|\xi - \lambda^{(\gamma)} A\|^2 T,$$

where W_0^i denotes the investor's strictly positive initial capital.

The fact that the fund is restricted to trading only equity implies that the instantaneous net-of-fees Sharpe ratio can be written as

$$\vartheta_t^\pi = \frac{\pi_t^* a}{\|\sigma^* \pi_t\|^2} - \frac{r + \gamma}{\|\sigma^* \pi_t\|^2},$$

where a is the drift of the stocks. Thus, higher fee rates will induce the manager to shift the fund's portfolio toward more volatile portfolios, similar to the case where the fund is allowed to trade also in the bond. The difference between the two is the mechanism used to increase the volatility. When the manager is allowed to also trade in the bond, the higher volatility is obtained by increasing leverage without changing the composition of the equity portion of the portfolio. When the manager is restricted to trading only equity, he cannot change the fund's leverage, so he increases volatility by changing the composition of the equity portfolio. As a result, it is no longer the case that the investor's effective equity portfolio is independent of the fee rate, as can be seen by multiplying the equilibrium fund portfolio and the investor's equilibrium trading strategy.

With fixed market coefficients, the optimal fund portfolio is the one that maximizes the relevant net-of-fees Sharpe ratio, regardless of whether the fund is allowed to invest in the bond or not. This Sharpe ratio will be at least as high in the case where it is allowed to trade also in the bond, because holding only equity is a feasible strategy. As a result, when the fund is restricted to trading only equity the manager's welfare should be at most the one obtained when the fund is unrestricted.

Maximization of the market value of the fees received by the manager, as one varies the instantaneous fee rate yields[16]:

COROLLARY 4.7.   *Assume that the coefficients are constant and that the fund is restricted from investing in the bond. Then the optimal fee rate is*

$$\hat{\gamma} := \arg\max_{\gamma > 0} \hat{v}_0(\gamma) = \frac{\|\xi\|^2}{2\xi^* A} = \frac{\|\xi\|^2}{2\sigma^{-1}\mathbf{1}}.$$

The corresponding equilibrium fund portfolio and investor's trading strategy are given by

$$\hat{\pi} = \pi^{(\hat{\gamma})} = [\sigma^*]^{-1} \frac{\xi}{(\xi^* A)}, \qquad \phi_t^{\hat{\pi}} = (\xi^* A) \left[\frac{W_t^{i\hat{\pi}}}{2} \right].$$

---

[16] This corollary could also be obtained directly from Corollary 4.2 by solving for the fee rate that implies that the fund is fully invested in equity.

Finally, the equilibrium expected utility of the investor and the initial market value of the fees are as in Corollary 4.2.

The welfare of both agents under this optimal fee rate are identical to that in Corollary 4.2. Therefore, an intuitive interpretation of the fee rate irrelevance result of Section 4 is as follows. The fund has an advantage in that it is able to access the risky assets whereas the investor cannot. On the other hand, it does not have any advantage in accessing the bond. As such, the "optimal equity access tax" that it can extract from the investor can be found by restricting the fund to trading only equity, but allowing it to set the fee rate. Allowing the fund to trade also in the bond enables it to extract this "optimal equity access tax" irrespective of the fee rate. It does so by varying its bond-stock mix so that the investor's effective exposure to equity, as well as the fees paid, remain the same as under the strategy that implements the "optimal equity access tax." For higher (lower) fee rates, it tilts its portfolio toward equity (the bond), taking into account the fact that a higher (lower) fee rate implies that the investor will be willing to invest less (more) in the fund. Thus, for every fee rate there is a corresponding fund portfolio that implements the optimal tax. Theorem 4.1 shows that this intuition also holds in the case of a stochastic opportunity set.

Given that the welfare of both agents under the fee rate of Corollary 4.7 is the same as in Corollary 4.2, when the fund manager is the one setting the fee rate, both agents are indifferent whether the fund is restricted to trading only equity or not. Differentiating equation (4.13) with respect to the fee rate shows that the investor's welfare is a decreasing function of the fee rate. Furthermore, equation (4.12) shows that the market value of the fees is a concave function of the fee rate, with a maximum at the fee rate of Corollary 4.7. Thus, if the investor gets to set the fee rate, the two agents have opposing views on whether to allow the fund to trade only equity or enable it to also trade in the bond.

4.3. Relaxing the Inaccessibility Assumption

We now extend the model by allowing the investor to have direct access to some of the stocks traded on the market, and show that the qualitative results derived earlier still hold in this more general setting.

To simplify the exposition, assume that the coefficients are constant. Let $m \leq n$ be the number of stocks that the investor can trade, and assume that these stocks are the first m components of the stock price process. Let $\bar{\sigma}$ denote the volatility matrix of these stocks, and define

$$\Omega := \mathbb{I}_n - \Omega^0 = \mathbb{I}_n - \bar{\sigma}^*[\bar{\sigma}\bar{\sigma}^*]^{-1}\bar{\sigma} \in \mathbb{R}^{n \times n},$$

where $\mathbb{I}_n$ denotes the n-dimensional unit matrix. The following proposition is the counterpart of Corollary 4.2 for the case where the investor can trade directly in a subset of the stocks.

PROPOSITION 4.8. *Assume that the market coefficients are constant. Then the equilibrium fund portfolio and the investor's trading strategy are given by*

$$\hat{\pi}_t = [\sigma^*]^{-1} \left[L_t + \frac{2\gamma\xi}{\xi^*\Omega\xi} \right],$$

$$\phi_t^{\hat{\pi}} = W_t^{i\hat{\pi}} \left(\frac{\xi^*\Omega\xi}{4\gamma} \right),$$

where L_t is any process with values in the null space of Ω. The initial market value of the fees and the investor's expected utility are given by

(4.14)
$$\hat{v}_0 := W_0^i \left[1 - \exp\left(-\frac{1}{4}\xi^*\Omega\xi\, T \right) \right],$$

$$V_i = \log\left(W_0^i S_T^0 \right) + \frac{\|\xi\|^2}{8}T + \frac{3\xi^*\Omega^0\xi}{8}T,$$

where W_0^i denotes the investor's strictly positive initial capital.

Since the volatility matrix $\bar{\sigma}$ has rank m, we have that the null space of the matrix Ω is an m-dimensional subspace of $\mathbb{R}^n$, and it follows that the manager has m degrees of freedom in determining the fund portfolio. Given this finding, it is natural to ask whether there exists an optimal fund portfolio which entails the fund to trade only in the $m - n$ stocks that the investor cannot trade. The following result shows that this is the case and provides an explicit characterization of this fund portfolio.

COROLLARY 4.9. *Assume that the market coefficients are constant. Then the fund portfolio process defined by*

$$\hat{\pi}^k := 1_{\{k>m\}} \frac{2\gamma[(\sigma^*)^{-1}\xi]^k}{\xi^*\Omega\xi}, \quad 1 \le k \le n,$$

is the unique optimal fund portfolio with the property that the fund does not invest in any of the stocks that the investor can access directly.

The results of this section are consistent with the ones obtained under the assumption that the investor is restricted from trading equity directly. First, the investor's holding of the fund is inversely related to the fee rate. Second, and as can be seen by multiplying $\phi^{\hat{\pi}}$ and $\hat{\pi}$, the investor's effective equity portfolio, as well as the fraction of his wealth invested in the bond, are independent of the fee rate. Third, increasing the fee rate results in the investor holding a larger part of his bond position directly versus indirectly through his holding of the fund. Fourth, the cumulative fee process $\Phi^{\hat{\pi}}$ is independent of the fee rate. Fifth, combining the second and the fourth earlier implies that the investor's wealth process, his value function, and the market value of the fees are independent of the fee rate.

As can be expected, the fact that the investor can trade in a subset of the risky assets allows him to improve his expected utility. Comparing equations (4.4) and (4.14) shows that direct access to m stocks increases the investor's welfare by the quantity

$$u(m) := \frac{3\xi^*\Omega^0\xi}{8}T,$$

which is nonnegative because the matrix Ω^0 is positive semi-definite. Intuitively, the increase in the investor's welfare compared to the base case of Section 4 should be increasing in the number of stocks that the investor can trade in. The following lemma confirms this intuition.

LEMMA 4.10. *The function $u : \{1, \ldots, n\} \to \mathbb{R}_+$ is increasing.*

Specifically, when the investor cannot trade any of the stocks directly, the matrix Ω^0 vanishes, and we recover the results of Section 4. On the other hand, when the investor has direct access to all of the risky assets, the matrix Ω^0 becomes the identity matrix, and the investor's expected utility becomes

$$V_i = \log \left(W_0^i \, S_T^0 \right) + \frac{\|\xi\|^2}{2} \, T,$$

which is the expected utility of a logarithmic investor who has complete access to the financial market. In this case, the fund becomes a redundant asset which the investor does not want to trade because of the fees, and, as a result, the manager is not able to extract anything from the investor.

5. MANAGERIAL EFFORT

In this section, we show that our main results are robust to the introduction of managerial ability. Specifically, we extend the model of Section 2 by allowing the manager to improve the returns of the fund by exerting costly effort and show that the main features of our solution remain valid in this more general setup.

5.1. The Model

Assume that, in addition to $B \in \mathbb{R}^n$, the probability space also supports a Brownian motion $W \in \mathbb{R}^d$, which is independent from B. As earlier, let $\mathbb{F}$ be the usual augmentation of the filtration generated by the Brownian motion B, and assume that the market parameters (r, σ, ξ) are bounded processes which are adapted with respect to $\mathbb{F}$.

Throughout this section we assume that, given a fund portfolio π, the fund value process evolves according to

$$\frac{dF_t}{F_t} = [r_t + e_t - \gamma]dt + \pi_t^* \sigma_t [dB_t + \xi_t dt],$$

where $e_t \geq 0$ represents the manager's instantaneous effort level at time t. Furthermore, we assume that in exerting effort the manager incurs a cost which is given by

$$c_t(e, \phi) = \phi \bar{c}_t(e) = \phi[\delta_t + \alpha_t e + \kappa_t e^2],$$

where (α, δ, κ) are nonnegative, bounded processes which are adapted with respect to the filtration $\mathbb{H} = (\mathcal{H}_t)_{0 \leq t \leq T}$ generated by the pair (B, W). The fact that the cost function is increasing in the size of the fund, ϕ, is intuitively appealing. It implies that it is much more costly for the manager to improve the return of a large fund than that of a small fund, and is thus akin to an assumption of decreasing returns to scale.

To guarantee the existence of a well-behaved equilibrium, we need to make sure that the cost of offsetting the management fee through effort is larger than the fee rate. Therefore, we assume that

(5.1) $$\bar{c}_t(\gamma) = \delta_t + \gamma \alpha_t + \gamma^2 \kappa_t > \gamma.$$

If this condition did not hold, then the optimal fund portfolio would be riskless, and as a result, the investor would choose not to invest at all.

5.2. Equilibrium

Using arguments similar to those of Sections 3 and 4, we obtain a complete description of the equilibrium for the model with managerial effort.

THEOREM 5.1. *Assume that equation (5.1) holds, define a pair of nonnegative processes by setting $\Gamma_t = \gamma - \hat{e}_t$ and $\Phi_t = \gamma - \bar{c}_t(\hat{e}_t)$ where*

$$(5.2) \qquad \hat{e}_t = \gamma - \left[\gamma^2 - \frac{(\gamma(1-\alpha_t) - \delta_t)^+}{\kappa_t} \right]^{1/2},$$

and let $\Psi_t(x) = \Gamma_t / (x\Gamma_t - \Phi_t)$. In equilibrium, the manager's effort level, the fund portfolio, and the investor's strategy are given by $\hat{e}_t$ and

$$(5.3) \qquad \hat{\pi}_t = 1_{\{\mathcal{B}\}} [\sigma_t^*]^{-1} \frac{2\Gamma_t(\xi_t - \Psi_t(P_t)Y_t)}{\|\xi_t\|^2 - \|\Psi_t(P_t)Y_t\|^2},$$

$$\phi_t^{\hat{\pi}} = 1_{\{\mathcal{B}\}} (\|\xi_t\|^2 - \|\Psi_t(P_t)Y_t\|^2) \frac{W_t^{i\hat{\pi}}}{4\Gamma_t},$$

where the triple (P, Y, K) is the maximal solution to the backward stochastic differential equation

$$(5.4) \qquad dP_t = 1_{\{\mathcal{B}\}} \frac{\|\xi_t - \Psi_t(P_t)Y_t\|^2}{4\Psi_t(P_t)} dt + Y_t^*[dB_t + \xi_t dt] + K_t^* dW_t$$

with terminal condition equal to zero and $\mathcal{B} \subset \Omega \times [0, T]$ denotes the set of events on which $\|\Psi(P)Y\| \le \|\xi\|$.

Equation (5.2) shows that, as expected, the equilibrium effort level is lower when the cost parameters α, δ, and κ are higher. Furthermore, taking the derivative of equation (5.2) with respect to γ shows that a higher fee rate induces the manager to exert more effort. The intuition for this result is as follows. A higher fee rate implies that the manager is receiving a larger fraction of the assets under management. Therefore, it creates a stronger incentive for the manager to work in order to increase the investor's wealth, and hence the amount of assets under management.

When the cost parameters α, δ, and κ are constant, the solution simplifies significantly because in that case the optimal effort level, the processes (Γ, Φ) and the mapping Ψ are time independent. This implies that equation (5.4) no longer depends on W, and it follows that the pair

$$(X, D) = \left(\frac{\Psi(0)}{\Psi(P_t)}, \Psi(0)Y \right)$$

is the minimal solution to the backward stochastic differential equation

$$-dX_t = \hat{g}(t, X_t, D_t)dt - D_t^* dZ_t$$

with terminal condition equal to one where the mapping $\hat{g}$ is defined as in Section 3.2. As is easily seen, this BSDE is the same as the one in equation (3.12), and because the minimal solution to this equation is uniquely defined, we obtain the following:

COROLLARY 5.2. *Assume that (α, δ, κ) are nonnegative constants such that equation (5.1) holds, and let the pair (X, D) be the minimal solution to equation (3.12). In*

equilibrium, the fund portfolio and the investor's strategy are given by equations (5.2) and (5.3) with $\Psi(P)Y = X^{-1}D$.

Comparing the earlier characterization to that provided in Theorem 4.1 shows that, with constant cost parameters, the equilibrium with managerial effort is isomorphic to a no-effort equilibrium with a modified fee rate given by

$$\Gamma = \gamma - \hat{e} = \left[\gamma^2 - \frac{(\gamma(1-\alpha)-\delta)^+}{\kappa}\right]^{\frac{1}{2}} \le \gamma.$$

In equilibrium, the costlier it is for the manager to exert effort, that is, the larger δ, α, and κ are, the more extreme will be his portfolio. Therefore, exerting effort and tilting the portfolio toward equity are partial substitutes for the manager. As the cost parameters increase (resp. decrease), the manager reduces (resp. augments) his effort level, and at the same time increases (resp. decreases) the volatility of the fund.

A direct computation shows that Γ is increasing in the fee rate if and only if $\gamma > (1-\alpha)/2\kappa$ or $\gamma(1-\alpha) - \delta \le 0$. As a result, it is no longer the case that the tilt toward equity in the fund portfolio is necessarily monotone in the fee rate, as was the case in the model without managerial effort. Furthermore, in the parameter region where the implicit fee rate, Γ, is increasing in γ, the attenuation ratio Γ/γ may not be monotone.

In equilibrium, all the benefits of exerting effort accrue to the manager. Given the isomorphism to the case with no effort and the fee-independence results of Section 4, we have that the investor's value function is exactly the same as in the no-effort case. On the other hand, the manager's value function (i.e., the market value of fees minus effort costs) is given by

$$P_0 = W_0^i(1 - X_0) + W_0^i\left[\frac{\Phi}{\Gamma} - 1\right].$$

The first term on the right-hand side corresponds to the manager's value function in the model without effort. The second part constitutes the gains to the manager and is strictly positive since $\Phi > \Gamma$.

6. CONCLUSION

In this paper, we have analyzed the implications of dynamic flows on a fund's portfolio decisions. To the best of our knowledge, this is the first paper to theoretically address this issue in a dynamic setting in which both the fund's portfolio decisions and the flows are determined endogenously.

With a constant opportunity set, the fund portfolio is chosen to maximize the fund's net-of-fees Sharpe ratio. A stochastic investment opportunity set, however, induces "flow hedging" on the part of the fund. Specifically, the manager distorts the fund's portfolio to increase the correlation between the investor's wealth and future fund Sharpe ratios.

Our model predicts a positive relationship between a fund's fee rate and its equity exposure or, alternatively, its volatility. However, allowing the manager to exert costly effort to improve fund returns breaks down this monotonicity, as exerting effort and tilting the portfolio toward equity become partial substitutes. In addition, the combination of the fund's relative advantage in accessing equity with the investors' ability to dynamically move money in and out of the fund implies the existence of a fee rate independent "equity access tax" that funds can extract from the investors.

While we view this paper as an important first step in understanding the dynamic portfolio decisions made by fund managers, there are some important extensions that are left for future work. First, to focus on a fund's portfolio choice decision, we made the simplifying assumption that investors are myopic. Allowing for utility functions other than log for the investor introduces hedging considerations also on their part. In such a setting, there will be both usual hedging considerations on the part of the investor and "flow hedging" considerations on the part of the fund. This would make the problem analytically intractable. For deterministic coefficients, Cetin (2006) considers the case in which the investor and the manager have mean-variance preferences, and solves the equilibrium numerically. He, too, finds that the value functions are independent of the fee rate.

Second, allowing for competition among mutual funds is an important and challenging extension of the model. As discussed in Section 4, the case of constant coefficients has been recently analyzed in Breton, Hugonnier, and Masmoudi (2010) and Cetin (2006), with results that are consistent with ours. Extending these results to the case in which funds are restricted to holding only equity would potentially lead to interesting insights as to how funds compete by jointly determining the fees they charge and the composition of their portfolios. In addition, allowing investors to access some assets directly while allowing different funds to have a relative advantage in accessing different assets, can help provide insights into fund competition through specialization.

Third, because some mutual funds use symmetric fulcrum performance fees and hedge funds utilize contracts with asymmetric performance fees, it would be interesting to allow for these types of fee structures. Finally, explicitly incorporating asymmetric information would be very interesting. Unfortunately, this extension of the model seems currently out of reach because it would introduce an inference problem on the part of the investors, which the manager can strategically manipulate.

APPENDIX A: PROOFS

Proof of Proposition 3.1. Let $\pi \in \Pi_f$ be an admissible fund portfolio process, ϕ^π be the candidate optimal strategy defined by equation (3.2) and $W^{i\pi}$ denote the corresponding wealth process. The fact that ϕ^π is an admissible trading strategy for the investor follows from the result of Lemma A.1 which also shows that the corresponding wealth process is strictly positive.

To verify that ϕ^π identifies the investor's best response, let ϕ be an arbitrary admissible trading strategy for the investor given the fund portfolio process and denote by W^i the corresponding wealth process. Using Itô's product rule in conjunction with equations (2.2) and (3.2), we have

$$
d\left[\frac{W_t^i}{W_t^{i\pi}}\right] = \frac{dW_t^i}{W_t^{i\pi}} + W_t^i d\left[\frac{1}{W_t^{i\pi}}\right] + d\left\langle W^i, \frac{1}{W^{i\pi}}\right\rangle_t
$$

$$
= \left(\phi_t - [\vartheta_t^\pi]^+ W_t^i\right)\frac{\pi_t^* \sigma_t}{W_t^{i\pi}} dB_t - \frac{\|\sigma_t^* \pi_t\|^2}{W_t^{i\pi}}\left([\vartheta_t^\pi]^- \phi_t\right)dt,
$$

where ϑ^π is the process defined by equation (3.1). Observing that both ϕ and W_i/W_i^π are nonnegative processes and applying Fatou's lemma we conclude that the latter process is a supermartingale. Using this in conjunction with the concavity and increase of the

logarithm we obtain

$$E\left[\log\left(W_T^i\right)\right] - E\left[\log\left(W_T^{i\pi}\right)\right] \leq \log E\left[\frac{W_T^i}{W_T^{i\pi}}\right] \leq \log(1) = 0$$

and it now only remains to observe that the trading strategy $\phi \in \Theta_i^\pi$ was arbitrary in order to complete the proof. $\qquad\square$

LEMMA A.1. *For every fund portfolio $\pi \in \Pi_f$, the cumulative fee process Φ^π is integrable under the risk-neutral probability measure.*

Proof. The process $W^{i\pi}$ being nonnegative, we deduce from equation (3.3) that Φ^π is increasing and since its initial value is equal to zero, it is also nonnegative. Let f be the progressively measurable mapping defined by

$$(A.1) \qquad f(t, \pi_t) := [\vartheta_t^\pi]^+ = \frac{[\pi_t^* \sigma_t \xi_t - \gamma]^+}{\|\sigma_t^* \pi_t\|^2}.$$

By definition, we have $f(t, \pi_t) \leq \sup_{\pi \in \mathbb{R}^n} f(t, \pi)$ for every $\pi \in \Pi_f$. A simple computation shows that the supremum is uniquely attained by

$$\tilde{\pi}_t := \arg\max_{\pi \in \mathbb{R}^n} f(t, \pi) = [\sigma_t^*]^{-1} \frac{2\gamma \xi_t}{\|\xi_t\|^2}$$

and that the corresponding value is $(4\gamma)^{-1}\|\xi_t\|^2$. Let now $\Theta \in \mathbb{R}_+$ denote a uniform bound on the norm of the market price of risk process. The increase of Φ^π and the boundedness of the discount factor imply that we have

$$E_Q\left[\sup_{0 \leq t \leq T} \Phi_t^\pi\right] = E_Q[\Phi_T^\pi] \leq c_1 \times E_Q\left[\int_0^T \frac{d\Phi_\tau^\pi}{S_\tau^0}\right]$$

$$\leq c_2 \times E_Q\left[\int_0^T \frac{W_\tau^{i\pi}\|\xi_\tau\|^2}{S_\tau^0}d\tau\right]$$

$$\leq c_2 \times E_Q\left[\int_0^T \frac{W_\tau^{i\pi}\Theta^2}{S_\tau^0}d\tau\right] \leq c_3 \times W_0^i$$

for some nonnegative constants $(c_i)_{i=1}^3$ where the third inequality follows from Fubini's theorem and the fact that the investor's discounted wealth process is a nonnegative supermartingale under Q. $\qquad\square$

Proof of Lemma 3.2. Applying Itô's lemma to the investor's best response wealth process and using equation (3.5), we obtain that the nonnegative process M^π is the stochastic exponential of the local martingale

$$N_t^\pi := \int_0^t f(\tau, \pi_\tau)\pi_\tau^* \sigma_\tau dZ_\tau,$$

hence a local martingale itself. Coming back to the definition of f in equation (A.1), it is easily seen that we have

$$\langle N^\pi \rangle_t = \int_0^t \|f(\tau, \pi_\tau)\sigma_\tau^* \pi_\tau\|^2 d\tau \leq \int_0^t \|\xi_\tau\|^2 d\tau \leq \Theta^2 T$$

for some $\Theta \in \mathbb{R}_+$, where the last inequality follows from the boundedness of the market price of risk. The above expression readily implies that the Novikov condition holds. □

Proof of Proposition 3.3. Let $\Theta \in \mathbb{R}_+$ denote a uniform bound on the norm of the risk premium. Using the definition of the set $\mathcal{B}$ in conjunction with the Cauchy–Schwartz inequality, we obtain

$$|\hat{g}(t, x, y)| \leq \Theta(1 + \Theta|x| + \|y\|^2)$$

and the existence of a minimal solution (X, Y) to equation (3.12) now follows from Theorem 2.3 in Kobylanski (2000). Because this solution is continuous, the process $\hat{\pi}$ is locally bounded. On the other hand, applying Itô's lemma we have that the solution $F = F^{\hat{\pi}}$ to equation (2.1) is given by

$$F_t = e^{-\gamma t} S_t^0 \exp\left[-\frac{1}{2} \int_0^t \left\| \sigma_\tau^* \hat{\pi}_\tau \right\|^2 ds + \int_0^t \hat{\pi}_\tau^* \sigma_\tau dZ_\tau \right].$$

The process $e^{\gamma t} F_t / S_t^0$ is thus a nonnegative local martingale under the risk-neutral measure and it follows that $\hat{\pi}$ is an admissible fund portfolio process. Using equation (3.11) in conjunction with the uniqueness of the solution to equation (3.8), we obtain

$$(A.2) \qquad \qquad \operatorname*{ess\,inf}_{\pi \in \Pi_f} X_t^\pi \leq X_t^{\hat{\pi}} = X_t.$$

On the other hand, the definition of $\hat{g}$ and Theorem 2.3 in Kobylanski (2000) show that we have $X \leq X^\pi$ for every admissible fund portfolio process and taking essential infimum over π on both sides we conclude that equality holds in equation (A.2). Using this fact in conjunction with equations (3.4) and (3.6), it is now easily seen that

$$v_0^{\hat{\pi}} = W_0^i\left(1 - X_0^{\hat{\pi}}\right) = W_0^i\left(1 - \operatorname*{ess\,inf}_{\pi \in \Pi_f} X_0^\pi\right) \geq W_0^i\left(1 - X_0^\pi\right) = v_0^\pi$$

for every admissible fund portfolio process. This implies that $\hat{\pi}$ constitutes the optimal fund portfolio process. □

Proof of Corollary 3.4. The deterministic function $\bar{X} : [0, T] \to (0, 1]$ satisfies the terminal condition $\bar{X}_T = 1$. On the other hand, differentiating (3.15) and comparing the result with equation (3.11) we obtain

$$-d\bar{X}_t = -\frac{1}{4} \bar{X}_t \|\xi_t\|^2 dt = (\hat{g}[t, \bar{X}_t, Y_t]dt - Y_t^* dZ_t)|_{Y_t=0}$$

and it follows that $(\bar{X}, 0)$ belongs to the set of solutions to equation (3.12). Now let (X, Y) denote another solution to this equation. Applying Itô's lemma and using the fact that X is uniformly bounded we deduce that

$$e^{-\frac{1}{4}\|\xi\|^2 t} X_t - X_0$$

is a martingale with initial value zero under the risk-neutral measure. This readily implies that $(X, Y) = (\bar{X}, 0)$. □

Before proceeding with the proof of Proposition 4.4, we start by identifying the unique risk neutral measure for the model of Section 4.1.

LEMMA A.2. *The density of the unique risk-neutral probability measure for the model of equations (4.6)–(4.8) is given by*

$$(\text{A.3}) \qquad \frac{dQ}{dP} = M_T := M_{1T} M_{2T},$$

where the nonnegative processes M_i are defined by

$$M_{1t} = \exp\left[-\frac{1}{2} |\xi_1|^2 t - \xi_1 B_{1t} \right],$$

$$M_{2t} = \exp\left[-\frac{1}{2} \int_0^t |\xi_{2s}|^2 ds - \int_0^t \xi_{2s} dB_{2s} \right].$$

Proof. Since ξ_1 is constant, we have that M_1 is a martingale on $[0, T]$. As a result, it follows from Girsanov theorem that the dynamics of ξ_2 remain the same under the equivalent probability measure defined by

$$\frac{dQ'}{dP} := M_{1T}$$

and the only thing left to prove in order to establish that equation (A.3) defines an equivalent probability measure is that M_2 is a martingale on the interval $[0, T]$ under Q'. This follows from the fact that an Ornstein–Uhlenbeck process is Gaussian, see Revuz and Yor (1994, p. 323). □

Proof of Proposition 4.4. To compute the optimal fund portfolio, we need to solve the backward stochastic differential equation

$$(\text{A.4}) \qquad dX_t = 1_{\{(t, X_t, Y_t) \in \mathcal{B}\}} \frac{X_t}{4} \left\| \xi_t - \frac{Y_t}{X_t} \right\|^2 dt + Y_t^* dZ_t, \quad X_T = 1,$$

where the set in the indicator function is defined by

$$\mathcal{B} := \left\{ (t, \omega, x, y) \in [0, T] \times \Omega \times \mathbb{R} \times \mathbb{R}^2 : \left\| \frac{y}{x} \right\|^2 \leq \|\xi_t\|^2 \right\}.$$

The argument we use to construct a solution to this nonlinear backward equation relies on two observations:

(a) If its drift did not contain an indicator function, then equation (A.4) could be solved explicitly by using a change of probability measure and a power transformation.
(b) For the dynamics of ξ_2 in equation (4.8), this equation can be solved in closed form as a function of time and the market price of risk. As a result, if we can show that this solution satisfies

$$(\text{A.5}) \qquad \left\| \frac{Y_t}{X_t} \right\|^2 \leq \|\xi_t\|^2$$

then it also solves the original BSDE and our fund portfolio choice problem is solved in closed form.

Applying an argument similar to that we used in the proof of Lemma A.2, we have that the formula

$$\frac{dR}{dQ}\bigg|_{\mathcal{F}_t} = \exp\left[-\frac{1}{8}\int_0^t \|\xi_s\|^2 ds + \frac{1}{2}\int_0^t \xi_s^* dZ_s\right]$$

defines an equivalent probability measure. Using this change of probability measure conjunction with Itô's lemma it is easily deduced that the unique solution to the modified BSDE

$$dX_t = \frac{X_t}{4}\left\|\xi_t - \frac{Y_t}{X_t}\right\|^2 dt + Y_t^* dZ_t, \quad X_T = 1,$$

is explicitly given by

$$(A.6) \qquad X_t = E_R\left[\exp\left(-\frac{1}{8}\int_t^T \|\xi_s\|^2 ds\right)\bigg|\mathcal{F}_t\right]^2.$$

Applying Girsanov theorem, we have that the dynamics of the risk premium under the probability measure R are given by

$$d\xi_{2t} = \lambda\left[\overline{\xi_2} - \left(1 + \frac{\psi}{2\lambda}\right)\xi_{2t}\right]dt + \psi dW_{1t},$$

where the process W is a two-dimensional Brownian motion under R. In particular, ξ_2 is still a mean reverting Ornstein–Uhlenbeck process under the new measure. This implies that the conditional expectation on the right-hand side of equation (A.6) can be computed as

$$X_t = \exp\{A_1(t)|\xi_{2t}|^2 + A_2(t)\xi_{2t} + A_3(t)\}^2,$$

where the deterministic functions $(A_i)_{i=1}^3$ are defined by

$$(A.7) \qquad A_1(t) = \frac{1 - e^{-\Omega(T-t)}}{4(\Omega - 2\lambda_R)(1 - e^{-\Omega(T-t)}) - 8\Omega},$$

$$(A.8) \qquad A_2(t) = \frac{\lambda\overline{\xi_2}(1 - e^{-\frac{1}{2}\Omega(T-t)})^2}{\Omega((\Omega - 2\lambda_R)(1 - e^{-\Omega(T-t)}) - 2\Omega)},$$

$$A_3(t) = \int_t^T\left[\lambda\overline{\xi_2}A_2(s) + \psi^2\left(A_1(s) + \frac{1}{2}(A_2(s))^2\right) - \frac{1}{8}|\xi_1|^2\right]ds$$

and we have set

$$\Omega := \sqrt{\psi^2 + 4\lambda_R^2}.$$

Applying Itô's formula to X and using the dynamics of ξ_2 in conjunction with the fact that volatility is invariant under a change of probability measure, we obtain that the volatility of X is given by $Y_{1t} = 0$ and

$$Y_{2t} = 2\psi X_t(2A_1(t)\xi_{2t} + A_2(t)).$$

Since ξ_2 is a Gaussian process, we have that equation (A.5) holds almost everywhere if and only if the parameters are such that

(A.9) $$\sup_{(\tau,x)\in[0,T]\times\mathbb{R}} \left\{ 4\psi^2 (2A_1(\tau)x + A_2(\tau))^2 - x^2 - |\xi_1|^2 \right\} \leq 0.$$

If this is indeed the case, then the pair (X, Y) is a solution to the original backward equation (A.4) and our proof is complete. $\square$

Proof of Proposition 4.6. Consider the optimization problem

(A.10) $$\sup_{\pi\in\mathbb{R}^n} \frac{\pi^*\sigma\xi - \gamma}{\|\sigma^*\pi\|^2} \quad \text{s.t.} \quad \pi^*\mathbf{1} = 1,$$

and observe that if this problem admits a unique solution which is such that the corresponding value is strictly positive, then it must be that this solution coincides with that of equation (4.11). Solving the first-order conditions associated with equation (A.10) and checking the associated Hessian constraint, we find that the unique maximizer is

$$\pi^{(\gamma)} := [\sigma^*]^{-1} \frac{\xi - \lambda^{(\gamma)} A}{\xi^* A - \lambda^{(\gamma)} \|A\|^2},$$

where $\lambda^{(\gamma)} := \gamma - \|\xi - \gamma A\| / \|A\|$ and $A := [\sigma^{-1}\mathbf{1}]$. Plugging this back into the objective function, we find that the corresponding value is

$$\frac{1}{2}\{(\xi - \gamma A)^* A + \|A\| \cdot \|\xi - \gamma A\|\}$$

and is nonnegative by application of the Cauchy–Schwartz inequality. It follows that the optimal fund portfolio is given by $\hat{\pi} = \pi^{(\gamma)}$. The remaining claims in the statement are obtained by plugging the new expression for $\hat{\pi}$ into the formulae of Sections 3 and 4, we omit the details. $\square$

Proof of Corollary 4.7. Using the results of Proposition 4.6 in conjunction with an argument similar to that which led to equation (3.7), it is easily seen that the optimal fee rate solves

$$\inf_{\gamma>0} z(\gamma) := \inf_{\gamma>0} \exp\left[-\frac{\gamma}{2} \left(\xi - \left\{ \gamma - \frac{\|\xi - \gamma A\|}{\|A\|} \right\} A \right)^* AT \right].$$

Solving the first-order conditions associated with this problem, we find that the unique candidate is given by $\hat{\gamma} = \|\xi\|^2 / 2(\xi^* A)$ and since $z''(\hat{\gamma}) \geq 0$, we conclude that $\hat{\gamma}$ is indeed the optimal fee rate. The remaining claims in the statement follow by setting $\gamma = \hat{\gamma}$ in Proposition 4.6. $\square$

Proof of Proposition 4.8. The result follows from arguments similar to those used in the proof of Proposition 3.1 and Corollary 4.2. $\square$

Proof of Corollary 4.9. From Proposition 4.8, we have that the optimal fund portfolio is any vector of the form

$$\pi_t = [\sigma^*]^{-1} \left[L_t + \frac{2\gamma\xi}{\xi^*\Omega\xi} \right]$$

for some $L_t \in \Omega^\perp$. Thus, all there is to prove is that for the candidate optimal fund portfolio process $\hat{\pi}$ defined in the statement we have

$$\Omega \left[\sigma^* \hat{\pi} - \frac{2\gamma\xi}{\xi^*\Omega\xi} \right] = \mathbf{0}_n.$$

This easily follows from the definition of Ω and the fact that the rows of σ are linearly independent. $\qquad\square$

Proof of Lemma 4.10. Assume that we extend the investor's opportunity set from m to $m + 1$ stocks. Let v^* be the volatility vector of the additional stock, denote by $\tilde{\sigma}$ the volatility matrix of the traded stocks and set

$$\tilde{\Omega}^0 := \tilde{\sigma}^* [\tilde{\sigma}\tilde{\sigma}^*]^{-1} \tilde{\sigma}.$$

Using the definition of u in conjunction with the standard formulae for the inversion of block matrices, we find that

$$\begin{aligned} u(m+1) &= \frac{3\xi^*\tilde{\Omega}^0\xi}{8} T = \frac{3\xi^*\Omega^0\xi}{8} T + \frac{(\xi^*\Omega v^*)^2}{v\Omega v^*} T \\ &= u(m) + \frac{(\xi^*\Omega v^*)^2}{v\Omega v^*} T. \end{aligned}$$

Observing that the matrix Ω is positive semidefinite by construction we deduce from the above equation that $u(m) \leq u(m+1)$ and the desired result now follows from an induction argument. $\qquad\square$

Proof of Proposition 4.3. The equilibrium dynamics of the fund value and of the investor's wealth are given by

$$(\mathrm{A.11}) \qquad\qquad dF_t = F_t \left[(r+\gamma)dt + \frac{2\gamma}{\|\xi\|^2} \xi^* dB_t \right],$$

$$(\mathrm{A.12}) \qquad\qquad dW_t^i = W_t^i \left[\left(r + \frac{\|\xi\|^2}{4} \right) dt + \frac{1}{2}\xi^* dB_t \right].$$

Using equation (A.11) to express the differential of $\xi^* B$ as a function of that of F and plugging the result back into equation (A.12), we obtain that the equilibrium dynamics of the investor's wealth are given by

$$\frac{dW_t^i}{W_t^i} = \left[r\left(1 - \frac{\|\xi\|^2}{4\gamma} \right) + \frac{\gamma}{2} \right] dt + \frac{\|\xi\|^2}{4\gamma} d\log F_t.$$

Using the definition of the total return $R(\Delta)$ and integrating this stochastic differential equation on $[t - \Delta, t]$, we obtain

$$W_t^i = e^{\alpha\Delta} W_{t-\Delta}^i R_t(\Delta)^{\frac{\|\xi\|^2}{4\gamma}},$$

where the constant α is given by

$$\alpha := \left(r + \frac{\gamma}{2} \right) \left[1 - \frac{\|\xi\|^2}{4\gamma} \right].$$

Finally, combining this expression with the result of Corollary 4.2 and using the definition of the flow measure we obtain

$$\rho_t(\Delta) = R_t(\Delta) \left[e^{\alpha\Delta} R_t(\Delta)^{\frac{\|\xi\|^2}{4\gamma} - 1} - 1 \right].$$

The comparative statics in the statement follow from the differentiation of this expression with respect to γ and $R_t(\Delta)$. □

Proof of Theorem 5.1. In accordance with the model set forth in Section 5.1, the instantaneous net-of-fees Sharpe ratio of the fund is given by

$$b_t^\pi(e_t) := \frac{\pi_t^* \sigma_t \xi_t + e_t - \gamma}{\|\sigma_t^* \pi_t\|^2}.$$

Using an argument similar to that of Section 3.1 yields that, for a given fund portfolio and effort strategy, the investor's best response is

$$\phi_t^\pi(e_t) = b_t^\pi(e_t)^+ W_t^i.$$

Therefore, we can write the manager's optimization problem as

$$\hat{P}_0 = \sup_{(\pi, e)} E_Q \left[\int_0^T \frac{1}{S_t^0} \left(\gamma \phi_t^\pi(e_t) - c_t(e, \phi_t^\pi(e_t)) \right) dt \right]$$

$$= \sup_{(\pi, e)} E_Q \left[\int_0^T (\gamma - \bar{c}_t(e)) b_t^\pi(e_t)^+ \frac{W_t^i}{S_t^0} dt \right],$$

where E_Q is the expectation operator under the risk-neutral probability measure. Now, assuming wlog that the investor's initial wealth is $W_0^i = 1$, and using arguments similar to those of Section 3.2 we obtain that

(A.13)
$$\hat{P}_0 = \sup_{(\pi, e)} P_0^{(\pi, e)},$$

where $P = P^{(\pi, e)}$ is the trajectory of the solution to the backward stochastic differential equation with dynamics

$$-dP_t = b_t^\pi(e_t)^+ [\gamma - \bar{c}_t(e_t) + (e_t - \gamma) P_t + \pi_t^* \sigma_t Y_t] dt - Y_t^* dZ_t - K_t^* dW_t$$

$$= q[t, \pi_t, e_t, P_t, Y_t] dt - Y_t^* dZ_t - K_t^* dW_t$$

and terminal condition zero. Here, the n-dimensional process

$$Z_t = W_t + \int_0^t \xi_s ds$$

is a standard Brownian motion under the risk-neutral probability measure. Note also that, since $\langle W, B \rangle = 0$, the process W remains a Brownian motion under the risk-neutral probability measure.

As in Section 3.2, the candidate optimal control $(\hat{e}, \hat{\pi})$ is obtained by maximizing the driver of the backward equation. This gives

$$(A.14) \qquad \hat{e}_t = \gamma - \left[\gamma^2 - \frac{(\gamma(1 - \alpha_t) - \delta_t)^+}{\kappa_t} \right]^{1/2},$$

$$(A.15) \qquad \hat{\pi}_t = 1_{\{B\}} [\sigma_t^*]^{-1} \frac{2\Gamma_t [\xi_t - \Psi_t(P_t) Y_t]}{\|\xi_t\|^2 - \|\Psi_t(P_t) Y_t\|^2}.$$

In the earlier equations, the strictly positive, bounded processes (Γ, Φ) and the mapping $\Psi_t(\cdot)$ are defined as in the statement, the progressively measurable set $\mathcal{B}$ is defined by

$$\mathcal{B} := \{(t, \omega) \in [0, T] \times \Omega : \|\Psi_t(P_t) Y_t\| \leq \|\xi_t\|\},$$

and the triple $(P, Y, K) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^d$ is a solution to the backward stochastic differential equation with driver

$$\hat{q}(t, P, Y, K) := q(t, \hat{e}_t, \hat{\pi}_t, P, Y) = 1_{\{B\}} \frac{(\Phi_t - P\Gamma_t)}{4\Gamma_t} \left\| \xi_t - \frac{Y\Gamma_t}{P\Gamma_t - \Phi_t} \right\|^2$$

and terminal condition zero. To legitimate the earlier approach, we need to prove that this BSDE admits a solution whose initial value coincides with the supremum defined in equation (A.13). This follows from Lemma A.3 below and it now suffices to plug the manager's optimal strategy $(\hat{\pi}, \hat{e})$ into the investor's best response $\phi_t^\pi(e_t)$ to obtain the desired result. □

LEMMA A.3. *Assume that equation (5.1) holds true. Then the backward stochastic differential equation with dynamics*

$$(A.16) \qquad -dP_t = \hat{q}(t, P_t, Y_t, K_t)\, dt - Y_t^* [dB_t + \xi_t dt] - K_t^* dW_t$$

and terminal condition zero admits a maximal solution. Furthermore, the pair $(\hat{e}, \hat{\pi})$ defined in terms of this solution by equations (A.14) and (A.15) is admissible and attains the supremum in equation (A.13).

Proof. Let $\Theta \in \mathbb{R}_+$ denote a uniform upper bound on the norm of the market price of risk. Using the definition of $\hat{q}$ in conjunction with the Cauchy–Schwartz inequality, we obtain

$$|\hat{q}(t, P, Y, K)| \leq \Theta(1 + \Phi_t/\Gamma_t + \Theta|P| + \|Y\|^2)$$
$$\leq \Theta(3 + \Theta|P| + \|Y\|^2),$$

where the last inequality follows from that fact that the ratio Φ_t/Γ_t takes its values in (1,2]. The existence of a maximal solution to equation (A.16) now follows directly from Theorem 2.3 in Kobylanski (2000).

The remaining claims in the statement follow from arguments similar to those used in the proof of Proposition 4.4, we omit the details. □

APPENDIX B: EX-ANTE VERSUS EX-POST FEES

In this appendix, we show that in continuous time the distinction between the case where the fees are computed on an ex-ante and the case where the they are computed an ex-post basis does not matter.

Consider a discrete time model. If the fees are computed on an ex-ante basis, then the fees due for the period running from date t to date $t + \Delta$ are given by

$$\gamma \phi_t \Delta,$$

where ϕ_t is the amount invested in the fund at time t. On the other hand, if the fees are computed on an ex-post basis, then the fees due for the period running from date t to date $t + \Delta$ are given by

$$\gamma \phi_t \frac{F_{t+\Delta}}{F_t} \Delta,$$

where $F_{t+\Delta}/F_t$ is the total return on the fund over the interval. As is easily see, these two amounts differ almost surely.

To see that this distinction does not matter in continuous time, let $n \in \mathbb{N}$ and partition the interval $[0, T]$ in n periods of equal length. The total fees computed on an ex-post basis are given by

$$\sum_{k=0}^{n-1} \gamma \phi_{t_k} \Delta_n \frac{F_{t_{k+1}}}{F_{t_k}} = \sum_{k=0}^{n-1} \gamma \phi_{t_k} \Delta_n + \sum_{k=0}^{n-1} \gamma \frac{\phi_{t_k}}{F_{t_k}} \Delta_n \left[F_{t_{k+1}} - F_{t_k} \right],$$

where we have set $\Delta_n := n^{-1} T$ and $t_k = k\Delta_n$. As the number of partitions increases to infinity the first term converges to

$$\lim_{n \to \infty} \sum_{k=0}^{n-1} \gamma \phi_{t_k} \Delta_n = \int_0^T \gamma \phi_t \, dt,$$

which coincides with the total fees computed on an ex-ante basis. On the other hand, the second term converges to zero since F is an infinite variation process and $\langle F, A \rangle = 0$ for all finite variation processes.

REFERENCES

ADMATI, A., and P. PFLEIDERER (1997): Does It All Add Up? Benchmarks and the Compensation of Active Portfolio Managers, *J. Bus.* 70, 323–350.

BAGCHI, A. (1984): *Stackelberg Differential Games in Economic Models*, Volume 64 of *Lecture Notes in Control and Information Sciences*, New York: Springer Verlag.

BASAK, S., A. PAVLOVA, and A. SHAPIRO (2007): Optimal Asset Allocation and Risk Shifting in Money Management, *Rev. Financ. Stud.* 20, 1683–1721.

BERK, J., and R. GREEN (2004): Mutual Fund Flows and Performance in Rational Markets, *J. Polit. Econ.* 112, 1269–1295.

CARHART, M. (1997): On the Persistence in Mutual Fund Performance, *J. Finance* 52, 57–82.

CARPENTER, J. (2000): The Optimal Dynamic Investment Policy for a Fund Manager Compensated with an Incentive Fee, *J. Finance* 55, 2311–2331.

CETIN, C. (2006): Delegated Dynamic Portfolio Management Under Mean-Variance Preferences, *J. Appl. Math. Dec. Sci.* 10, 1–22.

CHEN, H., and G. PENNACCHI (1999): Does Prior Performance Affect a Mutual Fund's Choice of Risk? Theory and Further Empirical Evidence, Working Paper, University of Illinois.

CHEVALIER, J., and G. ELLISON (1997): Risk Taking by Mutual Funds as a Response to Incentives, *J. Polit. Econ.* 105, 1167–1200.

CREMERS, M., and A. PETAJISTO (2007): How Active is Your Fund Manager? A New Measure that Predicts Performance, Working Paper, Yale School of Management.

CUOCO, D., and R. KANIEL (2007): Equilibrium Prices in the Presence of Delegated Portfolio Management, Working Paper, Duke University.

CVITANIĊ, J., and I. KARATZAS (1992): Convex Duality in Convex Portfolio Optimization, *Ann. Appl. Probab.* 2, 767–818.

DANGL, T., Y. WU, and J. ZECHNER (2006): Market Discipline and Internal Governance in the Mutual Fund Industry, *Rev. Financ. Stud.* 21(5), 2307–2343.

DAS, S., and R. SUNDARAM (2002a): Fee Speech: Signalling, Risk-Sharing, and the Impact of Fee Structures on Investor Welfare, *Rev. Financ. Stud.* 15, 1465–1497.

DAS, S., and R. SUNDARAM (2002b): On the Regulation of Mutual Fund Fee Structures; in *Quantitative Analysis in Financial Markets*, Vol. III, M. Avellaneda ed. World Scientific.

DROMS, W., and D. WALKER (1996): Mutual Fund Investment Performance, *Q. Rev. Econ. Finance* 36, 347–363.

DYBVIG, P., H. FARNSWORTH, and J. CARPENTER (1999): Portfolio Performance and Agency, Working Paper, Washington University.

EDELEN, R., and J. WARNER (2001): Aggregate Price Effects of Institutional Trading: A Study of Mutual Fund Flows and Market Returns, *J. Financ. Econ.* 59, 195–220.

ELTON, E., M. GRUBER, S. DAS, and M. HLVAKA (1993): Efficiency with Costly Information: A Reinterpretation of Evidence from Managed Portfolios, *Rev. Financ. Stud.* 6, 1–22.

GIL-BAZO, J., and P. RUIZ-VERDÙ (2008): The Relation Between Price and Performance in the Mutual Fund Industry, Working Paper, Universidad Carlos III de Madrid.

GOETZMANN, W., J. INGERSOLL, and S. ROSS (2003): High Water Marks and Hedge Fund Management Contracts, *J. Finance* 58(4), 1685–1717.

GRINBLATT, M., and S. TITMAN (1989): Adverse Risk Incentives and the Design of Performance Based Contracts, *Manage. Sci.* 35, 807–822.

HEINKEL, R., and N. STOUGHTON (1994): The Dynamics of Portfolio Management Contracts, *Rev. Financ. Stud.* 7, 351–387.

HU, Y., and P. IMKELLER, M. M. (2005): Utility Maximization in Incomplete Markets, *Ann. Appl. Prob.* 15(3), 1691–1712.

HUANG, J., K. WEI, and H. YAN (2005): Participation Costs and Sensitivity of Fund Flows to Past Performance, Working Paper, University of Texas at Austin, USA.

IPPOLITO, R. (1989): Efficiency with Costly Information: A Study of Mutual Fund Performance, 1965–1984, *Q. J. Econ.* 104, 1–23.

KARATZAS, I., J. P. LEHOCZKY, S. E. SHREVE, and G. L. XU (1991): Martingale and Duality Methods for Utility Maximization in Incomplete Markets, *SIAM J. Control Opt.* 29, 702–730.

KAROUI, N. E. (1981): *Les aspects probabilistes du contrôle stochastique*, Volume 876 of *Lecture Notes in Mathematics*. Berlin: Springer-Verlag.

KAROUI, N. E., M.-C. QUENEZ, and S. PENG (1997): Backward Stochastic Differential Equations in Finance, *Math. Finance* 7, 1–71.

KAROUI, N. E., and R. ROUGE (2000): Pricing Via Utility Maximization and Entropy, *Math. Finance* 10(2), 259–276.

KOBYLANSKI, M. (2000): Backward Stochastic Differential Equations and Partial Differential Equations with Quadratic Growth, *Ann. Prob.* 28, 558–602.

LIN, W., J. LU, and I. CHUANG (2004): Dynamic Mutual Fund Investment Decision in the Presence of Labor Income, *J. Financ. Stud.* 12(2), 143–170.

LYNCH, A., and D. MUSTO (1997): Understanding Fee Structures in the Asset Management Business, Working Paper, New York University, USA.

MA, J., and J. YONG (1999): Forward-Backward Stochastic Differential Equations and Their Applications, *Lecture Notes in Mathematics*, Vol. 1702, Berlin: Springer-Verlag.

MANIA, M., and M. SCHWEITZER (2005): Dynamic Exponential Utility Indifference Valuation, *Ann. Appl. Prob.* 15, 2113–2143.

BRETON, M., J. HUGONNIER, and T. MASMOUDI (2010): Mutual Fund Competition in the Presence of Dynamic Flows, *Automatica*, Forthcoming.

OU-YANG, H. (2003): Optimal Contracts in a Continuous-Time Principal-Agent Problem, *Rev. Financ. Stud.* 16(1), 173–208.

PATEL, J., D. HENDRIKS, and R. ZECKHAUSER (1991): The Rationality Struggle: Illustrations from Financial Markets, *Amer. Econ. Rev.* 81, 232–236.

REVUZ, D., and M. YOR (1994): *Continuous Martingales and Brownian Motion*, Volume 293 of *Grundlehren der mathematischen wissenschaften*, Heidelberg: Springer-Verlag.

ROLL, R. (1992): A Mean/Variance Analysis of Tracking Error, *J. Portfolio Manage.* 18(4), 13–22.

SIRRI, E., and P. TUFANO (1998): Costly Search and Mutual Fund Flows, *J. Finance* 53, 1589–1622.

STARKS, L. (1987): Performance Incentive Fees: An Agency Theoretic Approach, *J. Financ. Quant. Anal.* 22, 17–32.

WACHTER, J. (2002): Portfolio and Consumption Decisions Under Mean-Reverting Returns: An Exact Solution, *J. Financ. Quant. Anal.* 37, 63–91.

WARTHER, V. (1995): Aggregate Mutual Fund Flows and Security Returns, *J. Financ. Econ.* 39, 209–235.

YONG, J., and X. ZHOU (1999): *Stochastic Controls: Hamiltonian Systems and HJB Equations*, New York: Springer.