CFR (draft), XXXX, XX: 1–31

Mutual Fund Flows and Performance in Rational Markets (Revisited)

Heber Farnsworth

[1] *Michael F Price College of Business*
University of Oklahoma

ABSTRACT

In Berk and Green (2004) diseconomies of scale in asset management combined with performance-driven flows drive expected excess returns to zero at the optimal fund size. However this model incorrectly accounts for variations away from the optimal size due to investment returns. In this paper I explain the errors in Berk and Green (2004) and present an alternative model in which managers profess to have skill at identifying expected returns based on public information. Updating about whether managers posses this skill leads to contracts and performance/flow relationships which conform with those we observe. The model allows competition among managers and so, in general, updating depends not only on performance relative to a benchmark but relative a peer group also.

Keywords: Mutual Fund Flows, Active and Passive Management

JEL Codes: G10, G11, G12

The author thanks Ivo Welch (the editor), Jonathan Berk, Joel Vanden, Alexander David, Kerry Back, Phil Dybvig, Monika Piazzesi, as well as an anonymous referee for useful comments. All remaining errors are my own.

ISSN 2164-5744 ; DOI 10.1561/104.XXXXXXXX

1 Introduction

Flows in and out of actively managed funds have been found to be related to recent fund excess returns over passive benchmarks. Since mutual fund management fees are a fraction of assets under management, such flows imply that managers who outperform their benchmarks are rewarded with higher fee income. The existence of such flows is something of a mystery because there is little evidence that performance is predictable or that active managers can, on average, deliver superior performance.

Berk and Green (2004) present a model which attempts to explain this mystery. In the model, investors update about the ability of managers to outperform their benchmarks in the future. However there are decreasing returns to scale in asset management such that flows into a fund decrease the fund's ability to produce such outperformance so that high average performance, commonly known as “alpha", is not observed in equilibrium. In this way skilled managers are rewarded with higher fees but fund investors don't earn superior returns on average.

In this paper I point out that Berk and Green (2004) incorrectly accounts for variations in the size of a fund due to fund returns. Correctly accounting for fund returns in the model has the unfortunate outcome that managers who outperform will often be penalized with outflows rather than inflows and underperformers will often be rewarded with inflows. I shall also present an alternative model of active management and flows.

In my model I adopt a different definition of skill than that used by Berk and Green (2004). In Berk and Green (2004), a condition for an active manager to have any assets under management is that investors believe that the manager can produce a positive alpha (before fees and costs associated with scale) and the ability to produce such an alpha is the definition of skill.[1] But if we consider all portfolio strategies which would produce zero alphas it is certainly not the case and investors would be indifferent among them. An investor would rank such strategies according to their preferences and the distribution which they believe governs asset returns. But now suppose there is a manager who claims to have a better

[1] To be precise, a skilled manager can produce a positive alpha but, as pointed out in Berk and Van Binsbergen (2015), the alpha does not measure skill. Higher alpha does not necessarily mean higher skill. Berk and Van Binsbergen (2015) define skill as the ability to “extract money from markets" and assert that only irrational investors would pay for the services of managers who cannot add value in this way.

estimate of this distribution and so, knowing the investor's preferences, would choose a different zero alpha strategy for the investor. If the investor believes that the manager has skill, defined in this way, then they would certainly be willing to pay the manager to choose the investment strategy for them.

In my model a manager claims to have skill at processing publicly available information to identify return distributions of assets better than the investor can. However the investor is only partially convinced, at the outset, that the manager has such skill and so may not be willing to invest all their wealth with the manager, choosing to leave some wealth in a passive strategy (the passive strategy involves no fees). Over time the investor updates their belief about the skill of the manager and reallocates between the passive strategy and the manager's portfolio.

The restriction that managers may only use information which is also known to other market participants is sufficient to guarantee that alpha is zero. This explains why many, including the authors of most textbooks, discuss alphas associated with active management in the context of violations of market efficiency. The relation between public information and alphas first came to the fore in the debate between Roll (1977) and Mayers and Rice (1979) and is discussed in Ferson (2012), which also contains a proof that all strategies which involve only publicly available information must earn zero alphas, at least relative to an asset pricing model which correctly prices all the primitive assets which the a manager may choose from to form a portfolio.

It is important to note that there is not perfect overlap between the goals of the model presented in this paper and that of Berk and Green (2004), even though they both address the issue of performance and flows. Berk and Green (2004) also consider the entry and exit of funds and the cross-sectional distribution of managerial talent but neither of these are contemplated in this paper. The model of this paper also addresses competition among managers, which is not considered in Berk and Green (2004).

An important metric for how well the model I present reflects the economic realities of mutual funds will be the extent to which the optimal contract arising in the model is like those which we observe in practice. Investment management agreements almost always specify that the manager receives a fee which is a fixed fraction of assets under management. These contracts also allow investors to purchase new fund shares or redeem existing shares at any time. Without making any assumptions on the form of the contract in the model I find that the optimal fee structure does seem to take this form.

Similarly I will compare the implications of flows from my model to those documented in the empirical literature. Chevalier and Ellison (1997) was the first of many studies to document that the performance/flow relationship appears to be convex for mutual funds. I find that in the context of the model of this paper the performance/flow relationship may be convex or concave, depending on the prior beliefs of investors about the skill of the manager.

2 Berk and Green (2004)

Berk and Green (2004) begins with an assumption that the active manager's excess returns over a passive benchmark satisfy

$$R_t = \alpha + \varepsilon_t$$

where ε_t is normally distributed and α is unknown. Investors have a normally distributed prior over α and over time investors update their beliefs based on the history of excess returns.

Assets under management are denoted q_t. Active funds charge a fixed fee, f, as a fraction of assets under management, and experience diseconomies of scale $C(q_t)$ such that the excess return received by fund investors is

$$r_{t+1} = R_{t+1} - \frac{C(q_t)}{q_t} - f$$

and the diseconomies of scale are such that $C(q_t)/q_t$ is increasing in q_t.

The allocation decision to active vs passive managers is not explicitly modeled but it is assumed that there will be flows into the fund if $\mathbb{E}_t[r_{t+1}] > 0$ and flows out of the fund if $\mathbb{E}_t[r_{t+1}] < 0$.

If in period t the excess return is high then investors' posterior estimate of α will increase. This will result in flows into the fund, which will increase the diseconomies of scale until $\mathbb{E}_t[r_{t+1}] = 0$, at which point flows will cease. Similarly a negative excess return, leading to a lower posterior estimate α, results in flows out of the fund until, once again $\mathbb{E}_t[r_{t+1}] = 0$. This means that for each possible estimate of α there is a size, q_t, that is consistent with it. This size is the *efficient size* of the fund.

Since fund size fluctuates with realized excess returns it would seem that this gives us the performance related flows that we observe. However this is incorrect. Notice that what Berk and Green (2004) have given us is a model in which changes in assets under management, q_t depend ONLY on realized excess returns, not realized total returns. In this model positive excess returns can lead to inflows, outflows, or no flows at all.

To see this, note that the percent change in assets under management in the Berk and Green (2004) model, under the additional assumption that $C(q)$ is quadratic, is given in equation (30) of that paper as

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

where t is the age of the fund, and γ and ω are the precisions of the prior and the excess return distributions respectively. From the above we can clearly see that if r_t is zero then this should result in no change in assets under management. This is because when $r_t = 0$ investors' beliefs about the manager's skill are unchanged and so the efficient size of the fund is unchanged.[2]

The problem is that zero excess returns can occur when the fund and benchmark returns are both positive or both negative. Suppose the fund and the benchmark both had a return of 10%. Then the fund size will have changed by 10% before any flows. But since the efficient size of the fund hasn't changed there must be outflows of 10% to return the fund to it's efficient size.

This implication of the model was missed by Berk and Green (2004) who defined flows in equation (33) of that paper.

> Empirical studies, however, commonly consider the flow of new funds, that is, the percentage change in new assets, which is typically defined to be
>
> $$n_t(r_t, q_t) \equiv \frac{q_t - q_{t-1}(1 + r_t)}{q_{t-1}}$$

The problem is that this is not the definition of flows of new funds used by empirical studies. Let R_t^B be the benchmark return such that $r_t + R_t^B$ is

[2] Berk and Green (2004) go on to present a slightly modified version of the above formula to allow for endogenous exit of funds but this version is sufficient for the present purposes.

the investment return of the fund. The correct expression for flows of new funds used in the empirical literature is

$$n_t = \frac{q_t - q_{t-1}(1 + r_t + R_t^B)}{q_{t-1}} = \frac{q_t - q_{t-1}}{q_{t-1}} - r_t - R_t^B.$$

Substituting in from the expression for the percent change in assets under management gives the definition for flows which would be implied by the Berk and Green (2004) model.

$$n_t = \left[\frac{1}{f}\left(\frac{\omega}{\gamma + t\omega}\right) - 1\right] r_t + \left(\frac{\omega}{\gamma + t\omega}\right)^2 \frac{r_t^2}{4f^2} - R_t^B.$$

In Berk and Green (2004) it is claimed that “The flow of funds is zero at an excess return of zero” but from the above we can see that at zero excess return flows will be equal to $-R_t^B$.

Since, in this model, expected excess returns are zero for all active funds, one would expect that average excess returns across all managers would be close to zero in any given period. It follows that flows in and out of the actively managed fund industry should be negatively correlated with the stock market return, as proxied by the benchmark return R_t^B. There is no evidence that such a negative correlation exists.

Besides the lack of empirical support, the above expression for flows is troubling because of what it implies about the incentives for managers to achieve high excess returns, and so increase their perceived skill. In terms of flows, the manager will be rewarded for outperformance, but only if R_t^B is not too large. Similarly the manager will be punished for underperformance, but only if R_t^B is not too negative. Perversely, a manager who outperforms in up markets is punished with outflows and a manager who underperforms in down markets is rewarded with inflows.

3 An Alternative Model

I will now present a model which is an alternative to the model of Berk and Green (2004). Market participants in this model trade in a frictionless security market with continuous trading. All market participants are able to trade in all the assets. All market participants are symmetrically informed and believe that all assets are fairly priced and earn expected returns commensurate with their risk exposures. So they would expect all portfolios, whether passive or active, would earn zero alphas.

There are N non-redundant, risky assets traded in the market and a locally risk-free asset with return $r(t)$. Because of the continuous-time framework all returns, expected returns, and volatilities are to be understood as “instantaneous”. Let $dR(t)$ denote the N-vector of instantaneous returns with instantaneous volatility matrix $\sigma(t)$.

An inconvenient truth that all financial econometricians are familiar with is that it is much easier to obtain accurate estimates of volatilities than to obtain accurate estimates of expected returns. If we increase the frequency of the data, even over a finite time horizon, then our estimates of volatilities converge almost surely to the truth, but not so with our estimates of expected return. Accordingly I will assume that two observers, like the manager and the investor, cannot disagree about the volatility matrix but they certainly can disagree about expected returns of the risky assets.

There exists a benchmark portfolio which charges no fees and which the investor, without the input of manager, would believe produces the maximum Sharpe ratio available in the market. In other words, at each date t there is an expected return vector $\mu_B(t)$ such that

$$w_B(t) = \left(\sigma(t)\sigma(t)^\top\right)^{-1}(\mu_B(t) - r(t)\mathbf{1}) \tag{1}$$

gives the weights in the risky assets of the benchmark portfolio at time t. Note that a portfolio with these weights attains the maximum Sharpe ratio if $\mu_B(t)$ is the vector of expected returns at time t. But the manager claims to be able to deliver a more accurate vector of expected returns. Let $\mu_M(t)$ be the expected return vector the manager identifies at time t. The manager is willing, for a fee, to manage a portfolio with weights

$$w_M(t) = \left(\sigma(t)\sigma(t)^\top\right)^{-1}(\mu_M(t) - r(t)\mathbf{1}) \tag{2}$$

which delivers the maximum Sharpe ratio according to the manager’s beliefs.

All market participants observe the returns on the underlying assets but $\mu_M(t)$ is private information of the manager. This assumption deserves a bit of explanation. While it is true that all market participants observe the same data it does not follow that they would identify the same expected returns on the basis of this data. The situation is the same as two econometricians who use the same macro data to forecast GDP. Their forecasts will be different if they are using different models. If the forecasters don't know each other's models then they won't know each other's forecasts. So saying that $\mu_M(t)$ is private information of the manager is equivalent to saying that the investor has access to same data as the manager but doesn't know the mapping from data to $\mu_M(t)$.

The investor is only partially convinced that the manager has the skill they claim they have. Initially the probability the investor puts on the investor having skill is $p(0)$, though the investor may change this assessment over time. So at time t the investor believes the managed portfolio has the highest Sharpe ratio with probability $p(t)$ and believes the benchmark has the highest Sharpe ratio with probability $1-p(t)$ such that at time t the investor believes the instantaneous expected return vector to be

$$\mu_I(t) = p(t)\mu_M(t) + (1-p(t))\mu_B(t). \tag{3}$$

The model is inherently partial equilibrium in that I will not be concerned with price formation or interest rate determination. Instead the main object of interest will be the nature of the optimal contract that arises between investor and manager. The mathematical details associated with deriving the optimal contract are relegated to the appendix. In the remainder of this section I will lay out the main results and the intuition behind them.

3.1 Allocation to actively managed fund

The first result is that the investor's allocation to the actively managed portfolio at time t is determined by $p(t)$, the investor's belief that that manager has skill. This allocation decision is part of the dynamic portfolio choice problem of the investor. We have known since Merton (1973) that dynamic portfolio strategies have two components: an allocation to a maximum Sharpe ratio portfolio and an allocation to a hedge portfolio. Log investors have no hedging demand so it's easiest to consider log investors first.

For a log investor the result is that the time t fraction of wealth that a log investor will invest with the manager is $p(t)$, allocating the rest to the benchmark. This result follows from the linearity of the weights of the maximum Sharpe ratio portfolio in the expected excess return vector, $\mu_I(t)$

and equation (3). Letting $dR_B(t)$, $dR_M(t)$, and $dR_I(r)$ denote the instantaneous returns to the benchmark portfolio, the manager's recommended portfolio, and the investor's optimal portfolio we have

$$dR_I(t) = p(t)dR_M(t) + (1 - p(t))dR_B(t).$$

Investors with non-log utility functions split their wealth between a maximum Sharpe ratio portfolio and a hedge portfolio and so $p(t)$ is the fraction, not of total wealth, but of the wealth that the investor would allocate to the maximum Sharpe ratio portfolio at time t.

3.2 Bayesian updating

Over time the investor will update their belief in the manager's skill. The form of the updating rule will play a role in contract design as well as in determining flows.

Since the investor is updating about the expected return vector of the available risky assets it is certainly possible that their updating rule would depend on realized returns of all these assets. However the restriction in (3) that the investor's expected return vector lies on a line connecting $\mu_B(t)$ and $\mu_M(t)$ leads to an updating rule which depends only on excess return of the managed portfolio over the passive benchmark.

This result is proven in the appendix but it can motivated formally by treating $dR(t)$ as a normally distributed vector with known covariance matrix $\sigma(t)\sigma(t)^\top dt$ and mean vector of the form (3). Using the definitions of $w_B(t)$ and $w_M(t)$ in (1) and (2) respectively we can write the score (the derivative of the log likelihood with respect to $p(t)$) as

$$dR_M(t) - dR_B(t) - (w_M(t) - w_B(t))^\top(\mu_B(t) + p(t)\mu_M(t) - \mu_B(t))dt$$

but note that this can be written as

$$dR_M(t) - dR_B(t) - \mathbb{E}_t^I[dR_M - dR_B(t)] \tag{4}$$

where the superscript on the expectation operator indicates we are using the investor's beliefs.

In Berk and Green (2004) updating also depends on excess returns over a benchmark but that was an assumption of the model. Here it is a result.

Having motivated the notion that the updating rule should depend on excess returns via (4) we turn to the form of the updating rule itself. In the appendix it is proven that

$$dp(t) = p(t)(1-p(t))\left(dR_M(t) - dR_B(t) - \mathbb{E}_t^I[dR_M(t) - dR_B(t)]\right) \quad (5)$$

from which we see that $p(t)$ is a martingale from the perspective of the investor. The martingale convergence theorem guarantees that the $p(t)$ process will converge to either 1 or 0, at which point either all or none of the investor's money will be in the managed portfolio. However notice that the responsiveness of $dp(t)$ to a given excess return depends on the value of $p(t)$ prior to the excess return being realized. As $p(t)$ approaches either 1 or 0 additional moves toward these endpoints become more difficult to achieve in terms of excess return. So although convergence is guaranteed, the rate of convergence is very slow.[3]

3.3 The Optimal Contract

From the perspective of contract theory, the most peculiar thing about money management contracts is that they allow investors to put more money in the fund or withdraw money from the fund at any time, without triggering a renegotiation. What makes this surprising is that the management fee is a fraction of the assets that the investor chooses to allocate to the actively managed portfolio. One would hardly expect the same of a contract for an executive at a corporation. If the shareholders voted to spin off half the company it seems unlikely that the CEO's compensation would automatically become half of what it was. But this is exactly what money managers agree to in the mutual fund industry, as well as in this model.

There is no moral hazard in this model. I assume that there is no costly effort involved in managing the portfolio. The optimal contract is obtained by maximizing the investor's utility subject to a budget constraint and participation by the manager.[4] I assume that both agents have time-separable expected utilities with a common rate of time preference, ρ.

[3] An interesting extension of this model would be to allow exit (through retirement or death) of skilled managers. In this case it seems likely there would be discontinuous change in $p(t)$ as the fund moves to new management. While I think might be a fruitful direction for future research it is beyond the scope of the current paper.

[4] The same contract would be obtained by maximizing manager utility subject to a budget constraint and participation by the investor.

Let $c(t)$ be the consumption of the investor at time t and $\phi(t)$ be the consumption of the manager (the manager's fee). From the first-order conditions for this problem (derived in the appendix) the marginal utilities of the contracting parties satisfy

$$\frac{u_I'(c(t))}{u_M'(\phi(t))} = \lambda \frac{p(t)}{p(0)}$$

where λ is the multiplier on the participation constraint.[5]

Since both manager and investor are assumed to have concave utility functions, a decrease in marginal utility is associated with an increase in consumption (or fee, for the manager). So the above first-order condition tells us that the manager's fee, $\phi(t)$, rises faster than the investor's consumption when $p(t)$ increases, and vice-versa for decreases in $p(t)$. The intuition for this stems from optimal risk sharing. From the investor's perspective the probability that the manager is skilled is a martingale, i.e. they feel it is as just as likely they will revise this probability upward as downward in the future. But from the manager's perspective this is not the case, the manager believes that the data will eventually convince the investor of the manager's skill. So the cheapest way for the investor to compensate the manager is to let the manager bet on the future evolution of $p(t)$ by agreeing to a pay schedule which fluctuates in $p(t)$.[6] Since the investor's allocation to the managed portfolio increases linearly in $p(t)$, this can be accomplished by having the manager agree to be paid according to assets under management.

Obviously, one easy way to formulate such a contract is to pay the manager a fixed-fraction of assets under management. This is not always exactly the case, but it is the case when both investor and manager have log utility. I will show later that the manager's fee is approximately a fixed-fraction of assets under management even for non-log utility.

---

[5]Notice that the so-called Borch rule, which states that for an optimal risk-sharing contract the ratio of marginal utilities should be constant, does not hold in this case. The reason for this is that the two parties have different beliefs. Instead the condition that holds here is that the ratio of the investor's marginal utility to the manager's marginal utility is a martingale, because $p(t)/p(0)$ is a martingale, at least from the perspective of the investor.

[6]In Adrian and Westerfield (2009) it was shown that an optimal contract between parties with differing beliefs will involve a "side bet" involving the relative likelihoods of the outcome under the two sets of beliefs. This feature of the model is essentially that side bet.

If both investor and manager have log utility then from the above first-order condition consumption and fees satisfy

$$\phi(t) = c(t)\lambda \frac{p(t)}{p(0)}. \tag{6}$$

Let asset under management at t be denoted by $A(t)$. In the appendix it is shown that

$$A(t) = \frac{c(t)p(t) + \phi(t)}{\rho}$$

Now consider the fee received by the manager as a proportion of assets under management. Using (6) gives

$$\frac{\phi(t)}{A(t)} = \frac{\rho}{\frac{c(t)}{\phi(t)}p(t) + 1} = \frac{\rho\lambda}{p(0) + \lambda} \tag{7}$$

which says that the manager's fee is a fixed fraction of assets under management.

This has an appealing similarity to the contracts that we actually observe. In practice, fund managers do not change their fees, relying instead on the growth of assets under management to reap the reward, as they suppose, of their skill.[7]

3.4 Competition among managers

One could argue that a world in which investors choose how to split their investments between an index portfolio and a single active manager is not very interesting because in the real world there are many active portfolio managers. The optimal contract can also be obtained in this case under the assumption that the investor believes that at most one of the managers could be skilled. So the investor's inference problem is to find which (if any) of the managers has skill.

The updating is somewhat more interesting in this multiple-manager case because simple excess returns over the benchmark are not sufficient. Instead the excess returns of all other managers matter. Let M denote the number of active managers and let $p_j(t)$ denote the probability, assessed

[7] Actually this is a bit of an overstatement. Warner and Wu (2011) document that when mutual fund advisory contracts are renewed by the fund board there is occasionally a small change of fee.

by the investor at time t, that the jth manager is skilled. The return realized by the investor is given by

$$dR_I(t) = \sum_{j=1}^{M} p_j(t) dR_j(t) + \left(1 - \sum_{j=1}^{M} p_j(t)\right) dR_B(t).$$

Using this expression it is straightforward to show that the updating rule for $p_j(t)$ analogous to (5) is

$$\begin{aligned}\frac{dp_j(t)}{p_j(t)} &= \left(dR_j(t) - dR_B(t) - \mathbb{E}_t^I\left[dR_j(t) - dR_B(t)\right]\right) \\ &\qquad - \sum_{k=1}^{M} p_k(t)\left(dR_k(t) - dR_B(t) - \mathbb{E}_t^I\left[dR_k(t) - dR_B(t)\right]\right).\end{aligned}$$

Now the updating depends on the excess return of a given managed portfolio over the benchmark and also on the weighted average of the excess returns of all the funds in the manager's peer group over the same benchmark. Whether this corresponds to industry practice is hard to say. Certainly managers are compared to their peers as well as to their benchmarks. Whether such comparisons take the form above is an open question and a fruitful direction for future research.

If we let $\phi_j(t)$ denote the fee of the $j$the manager then the expression for assets under management for the jth manager is

$$A_j(t) = \frac{c(t)p_j(t) + \phi_j(t)}{\rho}.$$

An expression analogous to (6) holds for each manager which implies that the fee paid to each manager is a constant fraction of the assets under their own management.

4 Flows and Power Utility

Applying the Ito formula to the expression for $A(t)$ from the one-manager case with log utility gives us the following.

$$dA(t) = -(c(t)p(t) + \phi(t))dt + A(t)dR_M(t)$$

Recall that flows are defined as the percentage changes in assets under management net of investment returns. From the above we can see that in this model the only flows are consumption and fee withdrawals.[8] But since these are independent of performance there are *no performance related flows* in this model, at least not as defined in the empirical literature.

There is an argument to be made that the definition used in the empirical literature can be misleading. From the above we see that there are no flows in or out of the managed portfolio. However this does not mean that the investor is not reacting to excess returns by updating and changing the allocation to the managed portfolio. In the one-manager case the initial allocation is between the benchmark and the managed portfolio. If the manager beats the benchmark then the investor updates $p(t)$ in the positive direction and now desires a higher allocation to the managed portfolio compared to the benchmark. But since the manager beat the benchmark the allocation has already changed in that direction. For log utility this exactly matches the desired change.[9]

Similar logic works for the multiple manager case. In that case the initial allocation is to the benchmark and each of the several managers. As returns are realized there are no reallocating trades. Better performing managers end up with greater assets under management and managers who underperform see their assets under management shrink.

Now compare with an investment problem in which there is no updating. If one asset outperforms the others then the investor will reallocate back to their ideal weights. Here the lack of reallocation can be considered a flow; it is a deliberate change to a new allocation in response to updating. But it is not a flow that we would observe using the definition from the empirical literature.

The fact that we do observe performance driven flows using the definition adopted in the empirical literature is evidence against the log utility version of the model I have presented above. However it turns out that a slight generalization of the model, to power utility, does result in flows using this definition while still retaining many of the desirable features of

[8]Interestingly this is also true in the multiple manager model. There are no flows between competing managers or between any manager and the benchmark. The investor makes an initial allocation among managers and the benchmark depending on $p_j(0)$ and then never reallocates.

[9]As I will show below this is due to the fact that log investors have unit relative risk aversion.

the log utility case. However it is more difficult to show because, in general, closed-form solutions are not available, which is why I started with the log utility case.

In the log utility case there was no need to specify the processes which risk prices and interest rates followed because the model could be solved in general. For power utility we must solve a PDE to obtain optimal contracts and portfolios and the form of the PDE depends on the assumptions we make about these processes. Suppose for instance that in addition to updating risk there are changes in the investment opportunity set that are driven by some state variable $x(t)$ such that $p(t)$ and $x(t)$ are jointly Markovian. Let $G(p(t), x(t))$ be the wealth/consumption ratio for a power utility investor. In the one-manager case where manager and investor both have power utility with relative risk aversion γ the present value of investor consumption is $c(t)G(p(t), x(t))$. Because of the manager's beliefs, the function G must be evaluated at $p(t) = 1$ such that the present value of manager fees is $\phi(t)G(1, x(t))$. Log utility is a special case of power utility with $\gamma = 1$ and in this case G reduces to $1/\rho$. For $\gamma \neq 1$ the function G solves a PDE derived in the appendix.

Recall that power investors allocate their wealth between the maximum Sharpe ratio portfolio, the risk-free asset, and a hedge portfolio, to hedge against changes in the investment opportunity set. The size of this hedging demand depend on the shape of the function G and it turns out that for the specification considered in this paper the hedging demand is very small so I will ignore it in what follows.

We have the following expression for assets under management

$$A(t) = c(t)G(p(t), x(t))p(t) + \phi(t)G(1, x(t)).$$

To derive the dynamics of $A(t)$ one must specify the dynamics of $x(t)$ and how it affects the investment opportunity set so that the PDE for G can be solved. A convenient specification is to imagine that μ_B is constant and that the manager's risk price vector fluctuates around this vector.[10] The appendix gives the exact specification and describes how the PDE which G must satisfy can be solved.

Given a numerical solution for G we can examine the manager's fee as a fraction of assets under management. In figure 1 this fraction is plotted

[10]Technically the assumption is that the risk prices associated with μ_B are constant I assume that r and σ are constant these two statements are equivalent.

as a function of $p(t)$ for several different values of risk aversion. Notice that the fraction is very nearly constant except at very low values of $p(t)$. The deviation from a constant fee can be extreme if $p(t)$ is very low and the risk aversion is very high. The fact that we almost always observe constant fee fractions in practice would then suggest that there are very few manager's whose reputation is so extremely low and that the typical investor's relative risk aversion is moderate.

To examine the nature of the performance/flow relationship consider applying the Ito formula to the expression for $A(t)$ above. Solving the PDE which G satisfies numerically shows that it is close to flat in $x(t)$ and $p(t)$ (implying that hedging demand is small) so to simplify the resulting expression we will ignore any terms that involve derivatives of G. It is shown in the appendix that we obtain the following.

$$
\begin{aligned}
dA(t) = & -\frac{c(t)p(t)+\phi(t)}{\gamma}dt + \\
& A(t)\left[\frac{\gamma-1}{\gamma}r\,dt + \frac{1}{\gamma}dR_M(t)\right] + \frac{\gamma-1}{\gamma^2}c(t)G(1,x(t))dp(t)
\end{aligned}
$$

If we let $\gamma = 1$ then we obtain the expression for log utility. The last term is the performance driven flows. Recall from (5) that $dp(t)$ depends on the excess return of the manager. If $\gamma > 1$ then positive excess returns result in flows into the managed fund.

The above expression is for instantaneous changes. Typically, empirical researchers use much longer horizons over which to examine flows. It is straightforward to simulate from this model to obtain histories of assets under management for funds in order to examine the nature of the performance/flow relationship over long horizons. For my simulation I take the horizon to be 1-year, the same as Chevalier and Ellison (1997). To compute the benchmark and managed return over this horizon I compute ending portfolio values $V_B(T)$ and $V_M(T)$ according to

$$
\frac{dV_B(t)}{V_B(t)} = dR_B(t), \quad \text{and} \quad \frac{dV_M(t)}{V_M(t)} = dR_M(t)
$$

up to time T (one year). Consistent with the literature I then define flows as

$$
Flows = \frac{A(T)-A(0)}{A(0)} - \frac{V_M(T)-V_M(0)}{V_M(0)}
$$

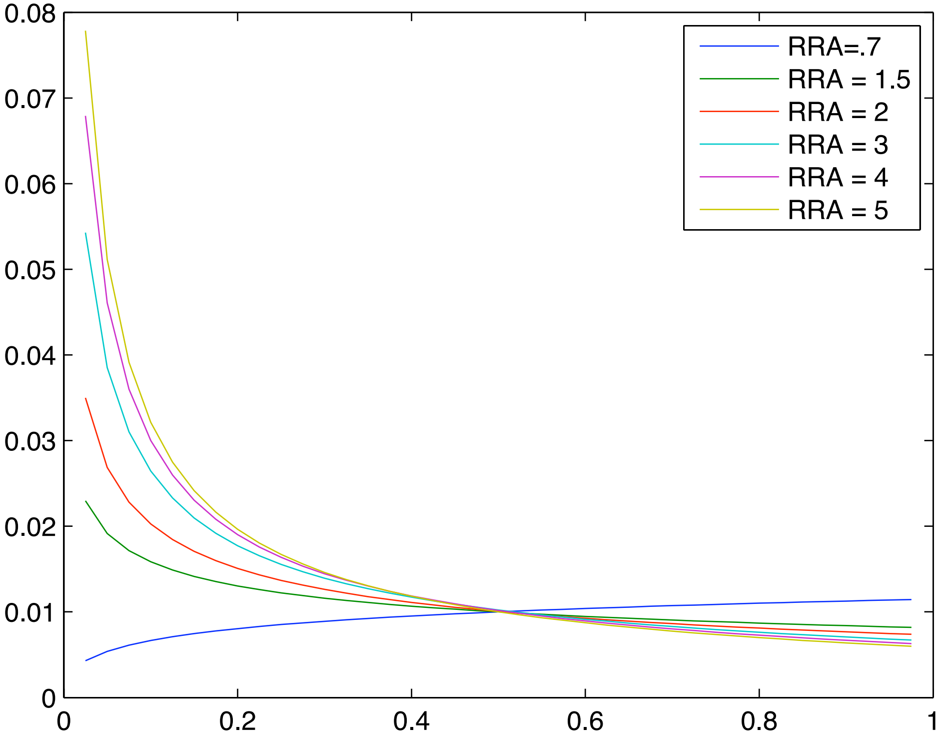

Figure 1: Fee Fractions

Description: The manager's fee expressed as a fraction of assets under management plotted against $p(t)$. Each curve is for a different relative risk aversion (RRA).

Interpretation: For a log investor the optimal contract pays the manager a fixed fraction of assets under management. For power utility the fee fraction is close to constant except for managers with very low $p(t)$.

and the excess return over the horizon as

$$Excess = \frac{V_M(T) - V_M(0)}{V_M(0)} - \frac{V_B(T) - V_B(0)}{V_B(0)}.$$

Using this simulated data I then estimate the performance/flow relationship

$$Flows = a + b_1 Excess + b_2 Excess * (Excess > 0) + e$$

where $(Excess > 0)$ is a dummy variable which is 1 when the excess return of the fund is positive and zero otherwise. The dummy variable allows me to test the significance of any convexity or concavity in the performance/flow relationship.

For the simulation I assume that manager and investor have a common coefficient of risk aversion, γ which is equal to 3, and a common subjective rate of time preference, ρ, equal to 0.06. The riskless rate is constant at 0.04 and the expected return vector, μ_B, is also assumed to be constant for simplicity. The full specification and the assumed parameter values are given in the appendix. I repeat the regression for various starting distributions of manager reputation. Each starting distribution is uniform on the interval indicated at the top of the column. Each regression uses 3000 simulated fund-year observations.

Table 1: Simulated Regression Results

Coefficient	(0,0.2)	(0.2,0.4)	(0.4,0.6)	(0.6,0.8)	(0.8,1.0)
a	−0.0222	−0.0256	−0.0288	−0.0323	−0.0359
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
b_1	1.3499	1.0650	0.7756	0.4845	0.1827
	(0.0059)	(0.0062)	(0.0066)	(0.0074)	(0.0076)
b_2	0.2388	0.1311	0.0106	−0.0949	−0.1212
	(0.0158)	(0.0144)	(0.0150)	(0.0173)	(0.0191)

Note: Estimates from the regression of simulated flows on simulated excess returns and simulated excess returns interacted with a dummy which is 1 if the simulated excess returns are positive. Standard errors are in parentheses. The coefficient b_2 measures the difference in the response of flows when excess returns are positive compared to when excess returns are negative. A significantly positive b_2 indicates a convex performance/flow relationship. Each column is from a different simulation in which each simulated fund has a starting reputation, $p(0)$, drawn uniformly from the range indicated at the head of the column.

There are several things to note about the table. As might be expected, the coefficient b_1 is always positive and significant so that flows respond positively to performance. However the strength of the response is greater when the reputation is lower. The reason for this is simple. When $p(t)$ is very high the investor has already allocated most of their wealth to the managed portfolio. This means that the investor has only a little more wealth to move in response to an upward revision. And since assets under management are already high for managers with high reputations the percentage change in assets will be smaller. The estimate of b_2 is significant and positive for low beginning reputations, meaning that the performance/flow relationship is convex. When reputations are high the relationship becomes concave.

Although there is no evidence of concave performance/flow relationships in mutual fund data Goetzmann *et al.* (2003) presents evidence that such a concave relationship may exist for hedge funds. Some may argue that hedge fund managers tend to have higher reputations than mutual fund managers so it is tempting to interpret this as being consistent with the model of this paper. However, other features of hedge funds, such as lock-up periods and high-water marks, are not consistent with this model so I hesitate to put too much weight on this.

5 Conclusion

In in this paper I have tried to explain the errors in the Berk and Green (2004) model. This is important because it has been the leading candidate to explain the simultaneous lack of evidence of alphas for mutual funds and the clear evidence that mutual fund investors chase performance, as measured by recent excess returns over passive benchmarks. The Berk and Green (2004) model is based on an assumption of diseconomies of scale in asset management. While diseconomies of scale may exist (there is mixed evidence in the literature) the Berk and Green (2004) model, correctly understood, incorrectly implies flows will be negatively correlated with the benchmark return such that outperformance will only result in inflows when the benchmark return is low or negative.

I have also presented an alternative model. In this model investors and managers are both equally informed and so neither expects that managers can produce a positive alpha. Despite this, investors still decide to hire active managers because they think it is possible that the managers have a better estimate of expected returns than what investors could come up with on their own. Investors update on this possibility based on the excess returns over a benchmark which, absent the manager, the investor would deem to have the maximum Sharpe ratio available. The optimal contract pays each manager a fee which is a fixed-fraction (or very nearly so) of assets under management, just as we observe in practice. If there are multiple active managers to choose from then investors update based on comparing each manager to the passive benchmark as well as to the performance of their peers.

The performance/flow relationship depends on the risk aversion of investors. Existing evidence on flows is consistent with investors having relative risk-aversion greater than 1. Simulation evidence based on a simple model of time-varying expected returns suggests that the relationship between performance and flows will be convex if the typical manager has a relatively low reputation but the relationship can be concave for managers with very high reputations.

References

Adrian, T. and M. M. Westerfield. 2009. "Disagreement and Learning in a Dynamic Contracting Model". *The Review of Financial Studies*. 22(10): 3873–3906.

Berk, J. B. and J. H. Van Binsbergen. 2015. "Measuring skill in the mutual fund industry". *Journal of Financial Economics*. 118(1): 1–20.

Berk, J. B. and R. C. Green. 2004. "Mutual Fund Flows and Performance in Rational Markets". *The Journal of Political Economy*. 112: 1269–1295.

Borch, K. 1962. "Equilibrium in a Reinsurance Market". *Econometrica*. 30(3): 424–444.

Chevalier, J. and G. Ellison. 1997. "Risk Taking by Mutual Funds as a Response to Incentives". *The Journal of Political Economy*. 105(6): 1167–1200.

Cox, J. C. and C. Huang. 1989. "Optimal consumption and portfolio choices when asset prices follow a diffusion process". *Journal of Economic Theory*. 49: 33–83.

Ferson, W. 2012. "Investment Performance: A Review and Synthesis". In: *Handbook of Economics and Finance, Volume 2B*. Ed. by G. Constantinides, M. Harris, and R. Stulz. Elsevier. 969–1010.

Goetzmann, W. N., J. E. Ingersoll Jr, and S. A. Ross. 2003. "High-water marks and hedge fund management contracts". *The Journal of Finance*. 58(4): 1685–1718.

Karatzas, I., J. P. Lehoczky, and S. E. Shreve. 1987. "Optimal Portfolio and Consumption Decisions for a "Small Investor" on a Finite Horizon". *SIAM Journal on Control and Optimization*. 25(6): 1557–1586.

Li, T. 2007. "Heterogeneous beliefs, asset prices, and volatility in a pure exchange economy". *Journal of Economic Dynamics and Control*. 31: 1697–1727.

Mayers, D. and E. M. Rice. 1979. "Measuring Portfolio Performance and the Empirical Content of Asset Pricing Models". *Journal of Financial Economics*. 7: 3–28.

Merton, R. C. 1973. "An intertemporal capital asset pricing model". *Econometrica*: 867–887.

Roll, R. 1977. "A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory". *Journal of financial economics*. 4(2): 129–176.

Warner, J. B. and J. S. Wu. 2011. “Why Do Mutual Fund Advisory Contracts Change? Performance, Growth, and Spillover Effects”. *Journal of Finance*. 66(1): 271–306.

A The Market

Since the investor and manager observe the same returns on all assets but disagree about the expected returns they must also disagree about the innovations in asset returns. The sources of these innovations are N-dimensional standard Brownian motions which we will denote by $Z_I(t)$ and $Z_M(t)$. The following are two equivalent descriptions of the vector of instantaneous returns of the N risky assets.

$$dR(t) = \mu_I(t)dt + \sigma(t)dZ_I(t)$$

$$dR(t) = \mu_M(t)dt + \sigma(t)dZ_M(t)$$

Provided that $\mu_I(t)$ or $\mu_M(t)$ are such that neither investor or manager perceive that there are arbitrage opportunities available in this market, there exist corresponding vectors of risk prices

$$\theta_I(t) = \sigma(t)^{-1}(\mu_I(t) - r(t)\mathbf{1}), \text{ and } \theta_M(t) = \sigma(t)^{-1}(\mu_M(t) - r(t)\mathbf{1})$$

and we can rewrite the return dynamics as

$$dR(t) = r(t)\mathbf{1}dt + \sigma(t)[\theta_I(t)dt + dZ_I(t)] \tag{A1}$$

or

$$dR(t) = r(t)\mathbf{1}dt + \sigma(t)[\theta_M(t)dt + dZ_M(t)]. \tag{A2}$$

Equating (A1) and (A2) gives

$$\theta_M(t)dt + dZ_M(t) = \theta_I(t)dt + dZ_I(t) \tag{A3}$$

the well-known *filtering equation*[11] which will be of use late

B Bayesian Updating

Since the manager and investor each have their own risk price vector and innovation process they also have their own stochastic discount factor with dynamics

$$\frac{dH_I(t)}{H_I(t)} = -r(t)dt - \theta_I(t)^\top dZ_I(t) \text{ and}$$

[11]The filtering equation is normally written as $dZ_M = (\theta_I - \theta_M)dt + dZ_I$ but the above expression is more useful in the present context.

$$\frac{dH_M(t)}{H_M(t)} = -r(t)dt - \theta_M(t)^\top dZ_M(t).$$

The ratio of these SDFs is a martingale under the investor's beliefs.

$$d\left(\frac{H_I(t)}{H_M(t)}\right) = \frac{H_I(t)}{H_M(t)}\left(\theta_M(t) - \theta_I(t)\right)^\top dZ_I(t) \tag{B4}$$

Recall that stochastic discount factors can be described as

$$H(t) = e^{-\int_0^t r(s)ds}\frac{d\mathbb{Q}(t)}{d\mathbb{P}(t)}$$

where $\mathbb{Q}$ is the risk-neutral measure and $\mathbb{P}$ is the probability measure corresponding to the beliefs of some agent. The measure $\mathbb{Q}$ does not depend on which agent's perspective we take. Now we can see that the ratio of the stochastic discount factor of the investor to that of the manager is also the likelihood ratio of the observed data using the manager's beliefs and the investor's beliefs

$$\frac{H_I(t)}{H_M(t)} = \frac{d\mathbb{P}_M(t)}{d\mathbb{P}_I(t)} \tag{B5}$$

I will now show that the updating rule for $p(t)$ depends on this ratio as

$$p(t) = p(0)\frac{H_I(t)}{H_M(t)}$$

Proof. The proof below largely follows the proof found in Li (2007). Let $p(0)$ be the prior probability that the manager is skilled and let $p(t)$ be the posterior probability, after observing returns up to time t. Bayes' theorem says that

$$p(t) = p(0)\frac{dP_M(t)}{p(0)dP_M(t) + (1-p(0))dP_B(t)}$$

which I will rewrite as

$$p(t) = \frac{\ell(t)}{\ell(t) + \frac{1-p(0)}{p(0)}}$$

where $\ell(t)$ is defined to be the likelihood ratio $\frac{dP_M(t)}{dP_B(t)}$. Now note that the function $g(x) = \frac{x}{x+k}$ has as its first and second derivatives

$$g'(x) = \frac{g(x)(1-g(x))}{x} \quad \text{and} \quad g''(x) = -2\frac{g^2(x)(1-g(x))}{x^2}.$$

With this in mind Ito's formula gives

$$\frac{dp(t)}{p(t)} = (1-p(t))\left[\frac{d\ell(t)}{\ell(t)} - p(t)\frac{d\langle\ell(t)\rangle}{\ell^2(t)}\right]. \tag{B6}$$

Now recall that $\ell(t) = \frac{dP_M(t)}{dP_B(t)} = \frac{H_B(t)}{H_M(t)}$. In terms of $dZ_I(t)$ we have that

$$\frac{dH_i(t)}{H_i(t)} = -r\,dt - \theta_i(t)^\top\left[(\theta_I(t) - \theta_i(t))dt + dZ_I(t)\right] \quad i = M, B$$

where

$$\theta_I(t) = p(t)\theta_M(t) + (1-p(t))\theta_B(t). \tag{B7}$$

The Ito formula gives us that

$$\frac{d\ell(t)}{\ell(t)} = (\theta_M(t) - \theta_B(t))dZ_I(t) + (\theta_M(t) - \theta_B(t))^\top(\theta_I(t) - \theta_B(t))dt$$

so that

$$\frac{\langle d\ell(t)\rangle}{\ell^2(t)} = (\theta_M(t) - \theta_B(t))^\top(\theta_M(t) - \theta_B(t))dt$$

With this observation we can substitute in to (B6) to obtain

$$\frac{dp(t)}{p(t)} = (1-p(t))(\theta_M(t) - \theta_B(t))^\top dZ_I(t).$$

But now note that

$$\theta_M(t) - \theta_I(t) = (1-p(t))(\theta_M(t) - \theta_B(t))$$

so that

$$\frac{dp(t)}{p(t)} = (\theta_M(t) - \theta_I(t))^\top dZ_I(t).$$

Comparing with (B4) we see that $p(t)$ and $H_I(t)/H_M(t)$ have the same dynamics. Since $H_I(0) = H_M(0) = 1$ this gives us

$$\frac{p(t)}{p(0)} = \frac{H_I(t)}{H_M(t)} \tag{B8}$$

as was desired.

C Optimal Fee

I will adopt the so-called *martingale approach* of Karatzas *et al.* (1987) and Cox and Huang (1989). The optimal contracting problem is

$$\max_{c,\phi} \mathbb{E}^I \int_0^\infty e^{-\rho t} u_I(c(t))dt$$

subject to participation by the manager

$$\mathbb{E}^M \int_0^\infty e^{-\rho t} u_M(\phi(t))dt \geq \bar{u}$$

and the budget constraint

$$\mathbb{E}^I \int_0^\infty H_I(t)(c(t)+\phi(t))\,dt \leq W(0)$$

where $\bar{u}$ is the manager's reservation utility and $W(0)$ is initial wealth. The superscripts on the expectation operators indicate whose probability beliefs are being used.

In order to solve the problem we need to be able to express the Lagrangian as an expectation using a single probability measure. Since the ratio of stochastic discount factors is a likelihood ratio we can use (B5) to change probability measure in the manager's participation constraint.

The first-order conditions are

$$u_I'(c(t)) = e^{\rho t} y H_I(t)$$

$$u_M'(\phi(t)) = e^{\rho t} \frac{y}{\lambda} H_M(t)$$

where y is the multiplier on the budget constraint and λ is the multiplier on the manager's participation constraint. The ratio of marginal utilities is

$$\frac{u_I'(c(t))}{u_M'(\phi(t))} = \lambda \frac{H_I(t)}{H_M(t)}$$

or, using, (B8)

$$\frac{u_I'(c(t))}{u_M'(\phi(t))} = \lambda \frac{p(t)}{p(0)}$$

which is the expression in the text.

D Power Utility

Suppose both investor and manager have power utility with relative risk aversion coefficient γ. We can solve the first-order conditions for the optimal consumption and fee.

$$c(t) = \left(e^{\rho t} y H_I(t)\right)^{-1/\gamma} \tag{D9}$$

$$\phi(t) = \left(e^{\rho t} \frac{y}{\lambda} H_m(t)\right)^{-1/\gamma}$$

By the Ito formula we also have the following dynamics for $c(t)$

$$\frac{dc(t)}{c(t)} = \left(\frac{r-\rho}{\gamma} + \frac{1}{2}\frac{1}{\gamma}\left(\frac{1+\gamma}{\gamma}\right)\theta_I(t)^\top \theta_I(t)\right)dt + \frac{1}{\gamma}\theta_I(t)^\top dZ_I(t).$$

For simplicity in what follows I will write this as

$$\frac{dc(t)}{c(t)} = \mu_c(t)dt + \frac{1}{\gamma}\theta_I(t)^\top dZ_I(t).$$

Optimal wealth has two parts: the present value of investor consumption and the present value of manager fees. The present value of consumption is

$$\frac{y^{-1/\gamma}}{H_I(t)} E_t^I \int_t^\infty e^{-\rho u/\gamma} H_I(u)^{(\gamma-1)/\gamma} du$$

Define $F(u,t)$ such that

$$\mathbb{E}_t^I H_I(u)^{(\gamma-1)/\gamma} = H_I(t)^{(\gamma-1)/\gamma} F(u,t).$$

and define $G(t)$ as

$$G(t) = \int_t^\infty e^{-\rho u/\gamma} F(t,u) du.$$

Plugging this in above and using (D9) we have that the present value of investor consumption is $c(t)G(t)$.

Now write the dynamics of $G(t)$ as

$$\frac{dG(t)}{G(t)} = \mu_G(t)dt + \sigma_G(t)dZ_I(t)$$

where the drift rate and volatility are, for the moment, unspecified. From the Ito formula we have

$$\frac{d(c(t)G(t))}{c(t)G(t)} = \left[\mu_G(t) + \mu_c(t) + \frac{1}{\gamma}\theta_I(t)^\top \sigma_G(t)\right] dt + \left[\sigma_G(t) + \frac{1}{\gamma}\theta_I(t)\right]^\top dZ_I(t)$$

But we also know that $c(t)G(t)$ is a wealth process which finances the consumption stream $c(t)$. To prevent arbitrage then we must have the following.

$$\frac{d(c(t)G(t))}{c(t)G(t)} = \left(r - \frac{1}{G(t)}\right) dt + \left[\sigma_G(t) + \frac{1}{\gamma}\theta_I(t)\right]^\top [\theta_I(t)dt + dZ_I(t)]$$

Equating the dt portions of both these SDEs gives the following no-arbitrage condition on $G(t)$.

$$\mu_G(t) + \mu_c(t) + \frac{1-\gamma}{\gamma}\theta_I(t)^\top \sigma_G(t) - \frac{1}{\gamma}\theta_I(t)^\top \theta_I(t) = r - \frac{1}{G(t)} \qquad \text{(D10)}$$

Consider a state variable x which solves the following SDE

$$dx(t) = x(t)(1 - x(t))\sigma_x^\top dZ_I(t)$$

so that x is restricted to values between 0 and 1. At time t we have

$$\theta_M(t) = \theta_B + (2x(t) - 1)D$$

where D is a constant vector. Notice that for $x(t) = 1$ we have $\theta_M(t) = \theta_B + D$, for $x = 0.5$ we have $\theta_M(t) = \theta_B$, and for $x = 0$ we have $\theta_M(t) = \theta_B - D$.

Using this specification for $x(t)$ and $\theta_M(t)$ we can apply the Ito formula to $G(t)$ to obtain

$$\begin{aligned}\mu_G(t) \quad &= \quad \frac{1}{2}\frac{G_{xx}}{G}x^2(1-x)^2\sigma_x^\top\sigma_x + \frac{1}{2}\frac{G_{pp}}{G}p^2(1-p)^2(2x-1)^2D^\top D \\ &\quad + \frac{G_{xp}}{G}x(1-x)p(1-p)(2x-1)\sigma_x^\top D\end{aligned}$$

and

$$\sigma_G(t) = \frac{G_x}{G}x(1-x)\sigma_x + \frac{G_p}{G}p(1-p)(2x-1)D.$$

Plugging these and the definition of $\mu_c(t)$ into (D10) gives us the following.

$$\frac{1}{2}G_{xx}x^2(1-x)^2\sigma_x^\top\sigma_x+\frac{1}{2}G_{pp}p^2(1-p)^2(2x-1)^2D^\top D$$

$$+G_{xp}x(1-x)p(1-p)(2x-1)\sigma_x^\top D$$

$$-G\left(r\frac{1-\gamma}{\gamma}-\frac{\rho}{\gamma}+\frac{1}{2}\frac{1-\gamma}{\gamma^2}\theta_I(t)^\top\theta_I(t)\right)$$

$$+\frac{\gamma-1}{\gamma}G_x x(1-x)\sigma_x^\top\theta_I(t)+\frac{\gamma-1}{\gamma}G_p p(1-p)(2x-1)D^\top\theta_I(t)=-1$$

Now substituting in the definition of $\theta_I(t)$ and noting that

$$\theta_I(t)^\top\theta_I(t)=\left(\theta_B^\top\theta_B+2p(2x-1)\theta_B^\top D+p^2(2x-1)^2D^\top D\right)$$

gives us the PDE which must hold on $[0,1]\times[0,1]$.

In order to solve this PDE numerically we need to know the behavior of the solution on the boundary of this region. At each of the four boundaries ($p=1$, $p=0$, $x=1$, and $x=0$) the PDE reduces to an ODE, each of which is straightforward to solve provided we know the solution at the corners.

At each corner all the derivatives disappear from the PDE and so we have the closed form solution

$$G=\frac{\gamma}{\rho-r(1-\gamma)-\frac{1}{2}\left(\frac{1-\gamma}{\gamma}\right)\theta_I(t)^\top\theta_I(t)}$$

where $\theta_I(t)^\top\theta_I(t)$ is equal to $\theta_B^\top\theta_B$ at both corners where $p=0$, and when $p=1$ we have $\theta_I(t)^\top\theta_I(t)=\theta_B^\top\theta_B+2D^\top\theta_B+D^\top D$ when $x=1$ and $\theta_I(t)^\top\theta_I(t)=\theta_B^\top\theta_B-2D^\top\theta_B+D^\top D$ when $x=0$.

Knowing all four corners means we can solve the equation on all four boundaries and then solve the PDE on the interior using standard techniques.

The following parameter values are used in the simulation: $\theta_B=[.25\ .25]^\top$, $D=[.05\ -.05]^\top$, and $\sigma_x=[-.03\ .03]^\top$. I take the initial distribution of x to be uniform on $[0,\ 1]$.

E Flows

Recall from the text that

$$A(t) = c(t)G(p(t))p(t) + \phi(t)G(1).$$

We wish to obtain the approximate dynamics for $A(t)$ by ignoring the hedging demand. This means we set σ_G to zero.

The present value of all future fee payments then has dynamics

$$\frac{d(\phi(t)G(1))}{\phi(t)G(1)} = \left(r - \frac{1}{G(1)}\right)dt + \frac{1}{\gamma}\theta_M(t)^\top[\theta_I(t)dt + dZ_I(t)]$$

or, more conveniently

$$d(\phi(t)G(1)) = -\phi(t)dt + \phi(t)G(1)r\,dt + \phi(t)G(1)\frac{1}{\gamma}dR_M(t). \quad \text{(E11)}$$

The dynamics of the present value of consumption is given by

$$\frac{d\left(c(t)G(p(t))\right)}{c(t)G(p(t))} = \left(r - \frac{1}{G(p(t))}\right)dt + \frac{1}{\gamma}\theta_I(t)^\top[\theta_I(t)dt + dZ_I(t)]$$

from which we obtain

$$\begin{aligned}
\frac{d(c(t)G(p(t))p(t))}{c(t)G(p(t))p(t)} &= \left(r - \frac{1}{G(p(t))}\right)dt + \frac{1}{\gamma}\theta_I(t)^\top[\theta_I(t)dt + dZ_I(t)] \\
&\quad + (\theta_M - \theta_I(t))^\top dZ_I(t) + \frac{1}{\gamma}\theta_I(t)^\top[\theta_M(t) - \theta_I(t)]dt \\
&= \left(r - \frac{1}{G(p(t))}\right)dt + \frac{1}{\gamma}\theta_M(t)^\top[\theta_I(t)dt + dZ_I(t)] \\
&\quad + \frac{\gamma-1}{\gamma}(\theta_M(t) - \theta_I(t))^\top dZ_I(t)
\end{aligned}$$

which, after some cancellation, yields

$$\begin{aligned}
d(c(t)G(p(t))p(t)) &= -c(t)p(t)dt + \frac{\gamma-1}{\gamma}c(t)G(p(t))p(t)r\,dt \\
&\quad + c(t)G(p(t))p(t)\frac{1}{\gamma}dR_M(t) + \frac{\gamma-1}{\gamma}c(t)G(p(t))dp(t). \quad \text{(E12)}
\end{aligned}$$

Combining (E12) with (E11) gives the following expression for the dynamics of $A(t)$.

$$dA(t) = -(c(t)p(t) + \phi(t))dt + \\ A(t)\left[\frac{\gamma-1}{\gamma} r\,dt + \frac{1}{\gamma} dR_M(t)\right] + \frac{\gamma-1}{\gamma} c(t)G(p(t))dp(t).$$

The first term is consumption and fee withdrawals, the second term is growth in assets due to investment return, and the last term is flows.