



# Market frictions, investor sophistication, and persistence in mutual fund performance



Ariadna Dumitrescu <sup>a,\*</sup>, Javier Gil-Bazo <sup>b</sup>

<sup>a</sup> ESAD Business School and University Ramon Llull, Av. Pedralbes, 60-62, 08034 Barcelona, Spain

<sup>b</sup> Universitat Pompeu Fabra and Barcelona GSE, Ramon Trias Fargas, 25-27, 08006 Barcelona, Spain

## ARTICLE INFO

### Article history:

Received 17 October 2016

Revised 5 January 2018

Accepted 9 January 2018

Available online 26 April 2018

### JEL classification:

G2

G23

### Keywords:

Mutual fund performance persistence

Market frictions

Investor sophistication

## ABSTRACT

We extend the model of Berk and Green (2004) to investigate the impact of two frictions, search costs and financial constraints, on mutual fund performance. Our model predicts the survival of underperforming funds and delivers two new predictions: 1) differences in performance across funds are likely to persist; and 2) mutual funds targeted to less sophisticated investors should exhibit higher dispersion in expected performance. Using data on U.S. domestic diversified equity mutual funds we find empirical support for these predictions.

© 2018 Elsevier B.V. All rights reserved.

## 1. Introduction

Like investors in other retail financial markets, mutual fund investors face non-negligible costs of gathering and processing information as well as financial constraints, which can affect both their decision to participate in the market and their fund choices. While we understand well how fund performance is determined in equilibrium in an ideal frictionless setting, the consequences of frictions for the determination of mutual fund performance are a priori unclear. In this paper, we investigate how the presence of frictions leads to market outcomes that differ from those of the ideal setting.

The starting point of our analysis is the model of Berk and Green (BG) (2004), who characterize the competitive provision of capital to mutual funds. In their model, investors learn about managerial ability from past returns and demand shares of all funds with positive expected risk-adjusted returns net of fees and other costs. If there are diseconomies of scale in portfolio management, the flows of money into (out of) outperforming (underperforming) funds drive their performance down (up) to investors' reservation return, which is assumed to be zero. In equilibrium, all funds deliver zero net expected performance, so fund performance is not predictable from past performance.

Consistent with the model of BG, empirical studies have traditionally found little evidence that active mutual fund managers have the ability to consistently outperform the market after fees, expenses, and transaction costs (e.g., Sharpe, 1966). However, this consensus has been challenged by studies that consider time-variation in managerial skill (Glode, 2011; Kosowski, 2011; Kacperczyk et al., 2014) or identify observable fund characteristics that predict outperformance (Cremers and Petajisto, 2009). Moreover, there is abundant empirical evidence that underperforming U.S. equity funds continue to underperform even in the

\* Corresponding author.

E-mail addresses: [ariadna.dumitrescu@esade.edu](mailto:ariadna.dumitrescu@esade.edu) (A. Dumitrescu), [javier.gil-bazo@upf.edu](mailto:javier.gil-bazo@upf.edu) (J. Gil-Bazo).

long run (e.g., [Carhart, 1997](#)). One possible explanation is that market frictions distort investor decisions and affect mutual fund equilibrium performance. To investigate the plausibility of this explanation, we develop a model of performance determination that retains the key features of the BG model, namely diseconomies of scale and competition among investors, but extends it in two directions. Like in the BG model, each period investors must choose between an actively managed fund and an index fund, the alternative investment opportunity available, which offers a zero expected risk-adjusted return. However, we depart from BG by accounting for two different types of frictions. First, we assume that when searching for suitable alternative investments, investors bear an information collection and processing cost. As a consequence, investors' reservation return, defined as the expected risk-adjusted return on their outside investment option, may be lower than zero. This assumption is consistent with [Müller and Weber \(2010\)](#), who find that individuals who score lower in financial literacy tests, less educated people, and those working outside of the financial services sector, are less likely to be aware of the existence of low-cost, passively managed funds, such as index funds and ETFs. It is also consistent with regulatory proposals to make low-cost, passively managed funds a mandatory default option in 401(k) plans, which suggest that finding a cheap passive alternative to actively managed funds is not costless for many investors ([Zingales, 2009](#)). We further assume that while collecting and processing relevant information is costly for all investors, less sophisticated investors are more adversely affected by this friction so their reservation returns are lower than those of more sophisticated investors.<sup>1</sup>

Second, we assume that investors are financially constrained, i.e., they face a limit on the amount of money they can invest in a mutual fund each period, so the supply of capital is limited. More specifically, the fund's current investors can reinvest their last period's wealth and invest their current endowment in the fund, whereas new investors can invest only their current period's endowment. We believe that financial constraints are an important feature of the mutual fund market. Our belief is founded on two facts. First, households are the predominant investor type in the mutual fund market and many of the households owning mutual funds have low or moderate incomes.<sup>2</sup> Moreover, financial constraints have been found to play an important role in explaining the portfolio decisions of many households (see, e.g., [Brown et al., 2017](#)). Second, the prevalence of minimum initial investments among mutual funds supports the idea that financial constraints are binding for many investors, including smaller institutions ([James and Karceski, 2006](#)). In our model, all investors are financially constrained. However, we allow financial constraints to vary across investors. More specifically, we assume that investors' endowment is increasing in their level of sophistication.<sup>3</sup>

In equilibrium, an actively managed fund offers an expected risk-adjusted net return at least as high as the reservation return of the most sophisticated investor who decides to invest with the fund. When managerial ability is low, a fund can survive offering a negative risk-adjusted expected net return if there are investors with low enough reservation returns in the market. Therefore, our model explains the survival of funds that can be expected to underperform.

But the model also delivers new predictions. First, the fund's equilibrium expected performance increases with managerial ability, despite the fact that the fund attracts more investors. This result is the consequence of the interaction of heterogeneity in investors' reservation returns with financial constraints. To the extent that managerial ability is persistent through time, so is fund performance. Therefore, the model predicts a monotonic relation between past performance and future performance.<sup>4</sup>

Second, our model predicts that fund performance exhibits higher sensitivity to managerial ability in markets with less sophisticated investors (e.g., funds that cater to retail investors) than in markets populated by more sophisticated investors (e.g., institutional funds). The higher sensitivity of expected performance to managerial ability in turn induces larger dispersion in expected performance (holding the distribution of ability constant). Moreover, flows of new money to mutual funds exhibit more sensitivity to past performance in markets with less sophisticated investors, consistent with empirical studies comparing different market segments ([Del Guercio and Tkac, 2002](#); [James and Karceski, 2006](#); [Del Guercio and Reuter, 2014](#)). Finally, we show that performance predictability in our model holds regardless of whether the fee is exogenously determined or the manager is allowed to set the fee to maximize profits.

We test our two new predictions using data on U.S. actively managed domestic equity funds in the 1962–2015 period. We document that short-term fund performance increases monotonically with fund performance ranking in the previous three years. We also show that the performance of funds targeted to retail investors exhibits significantly more dispersion than the performance of institutional funds, and this difference in dispersion is not due to differences in fees, investment style, or idiosyncratic risk.

To the best of our knowledge, we are the first to extend the BG model to account for frictions. In a different setup, [Garleanu and Pedersen \(2015\)](#) study the effect of frictions on mutual fund performance. In particular, they characterize equilibrium in both the asset market and the market for delegated portfolio management when asset managers face no diseconomies of scale.

<sup>1</sup> Heterogeneity in investor sophistication is consistent with ample empirical evidence that individuals differ in their ability to make optimal financial decisions. For instance, [Calvet et al. \(2009\)](#) report that families with high financial wealth and high education make fewer investment mistakes. Also, financially illiterate individuals (those lacking basic financial knowledge) are less likely to plan for retirement and participate in the stock market, and more likely to use high-costs methods of borrowing, take too much debt, and pay excessive mortgage rates ([Lusardi and Mitchell, 2014](#)).

<sup>2</sup> As of 2016, households in the U.S. held 89% of mutual fund assets; 17% of the households owning mutual funds had annual income of less than \$50,000 and 18% had between \$50,000 and \$75,000. (Investment Company Institute's 2017 Mutual Fund Factbook, [https://www.ici.org/pdf/2017\\_factbook.pdf](https://www.ici.org/pdf/2017_factbook.pdf)).

<sup>3</sup> Allowing for financial sophistication and wealth to be positively correlated is consistent with the empirical findings of [Calvet et al. \(2009\)](#). [Garleanu and Pedersen \(2015\)](#) make a similar assumption when they model heterogeneity across investors.

<sup>4</sup> Note that persistence in performance differences does not imply that some funds will outperform the market. Throughout the paper, we refer to persistence as the ability of past performance to predict future performance, even if funds are never expected to outperform the market.

In their model, managers invest in an asset market that is inefficient due to information acquisition costs. Uninformed investors must pay a cost to find and vet active asset managers. In equilibrium, informed asset managers must outperform passive investment to compensate investors for that cost. Also, uninformed managers survive despite underperforming passive investment after fees due to the presence of investors who pick managers randomly. Interestingly, in [Garleanu and Pedersen's \(2015\)](#) model, diseconomies of scale arise endogeneously at the industry level: As investors allocate more capital to active funds, asset prices become more efficient. Like their model, ours explains persistent differences in expected performance across funds. Unlike in their model, we retain BG's assumptions: asset prices are exogenous; there is no asymmetric information about managerial performance; investors are rational; and we assume diseconomies of scale at the fund level.

Our paper is also related to two empirical studies. [Glode et al. \(2017\)](#) investigate time variation in performance persistence and find evidence of a monotonic relation between past and future performance following months in which the stock market experiences large positive returns, but not after down-markets. They attribute their results to a larger presence of unsophisticated investors in the market following up-markets. Our model provides a precise mechanism through which the presence of unsophisticated investors in the market generates a monotonic relation between past and future performance.

[Berk and Tonks \(2007\)](#) argue that differences in the speed of learning across investors cause the composition of a fund's investor base to change with performance, since the first investors to leave or enter a fund are those who update their beliefs the fastest. As a consequence, the remaining investors of a fund that has underperformed have a lower flow-to-performance sensitivity, which prevents the fund's assets from shrinking should the fund continue to underperform. The authors report evidence consistent with this hypothesis.

Finally, our paper is related to empirical studies that investigate how market frictions affect mutual fund flows (see, e.g., [Sirri and Tufano, 1998](#); [Huang et al., 2007](#); and [Navone, 2012](#)).

The rest of the paper is organized as follows. In Section 2, we present the theoretical framework of our analysis and derive the main results when fund fees are exogenous. In Section 3, we compare two markets characterized by different levels of investor sophistication. In Section 4, we extend the model to allow for endogeneously determined fees. In Section 5, we present the empirical evidence and conclude in Section 6. [Appendix 1](#) and [Appendix 2](#) contain all the proofs.

## 2. The model

BG assume that actively managed funds can generate expected returns in excess of a passive benchmark due to managerial ability. Investors can also invest in an index fund that generates a zero risk-adjusted expected return. In the BG model, the equilibrium amount of assets under management of each active fund is such that the fund's net performance is zero (investors' reservation return) given the manager's ability and the fund's fee, defined as a fraction of the fund's assets. This is a consequence of diseconomies of scale and an unlimited supply of capital. Therefore, a manager's profits do not depend on other managers' actions, so each manager behaves as a monopolist and chooses the fee that maximizes his profits subject to the zero-performance constraint. However, when the supply of capital is limited, as is the case in our model, flows of money do not necessarily drive net performance to zero. Consequently, a manager may have incentives to outperform other funds by setting a lower fee and attract all assets. In order to abstract from the consequences of strategic competition among actively managed funds and focus on the direct impact of frictions and investor heterogeneity on fund equilibrium performance, we assume that there is a single actively managed fund competing against a passive alternative investment. As in BG, we assume that the passive alternative for all investors is an index fund with zero expected risk-adjusted return.

### 2.1. Setup

Let  $R_t$  denote the active fund's return in excess of a passive benchmark before fees and expenses,  $R_t = \alpha + \varepsilon_t$ , where  $\alpha$  reflects managerial ability and  $\varepsilon_t$  is an idiosyncratic shock that is normally distributed with mean 0 and variance  $\sigma^2$ .<sup>5</sup> We denote the precision of the error  $\omega = \frac{1}{\sigma^2}$ . For simplicity, throughout the paper we refer to expected risk-adjusted return as return. Like in BG, managerial ability,  $\alpha$ , is not known to managers or investors, who estimate it using the information contained in past returns. Similar to BG, we model the distribution of the manager's ability,  $\alpha$ , as normal with mean  $\phi_0$  and variance  $\sigma_\alpha^2$  and we denote by  $\delta = \frac{1}{\sigma_\alpha^2}$  its precision.

The cost of managing the portfolio is denoted by  $C(q)$ , where  $q$  is the dollar value of assets under management.  $C(q)$  is common knowledge and it satisfies the following properties:  $C(0) = 0$ ,  $\lim_{q \rightarrow \infty} C'(q) = \infty$  and for all  $q \geq 0$ ,  $C(q) \geq 0$ ,  $C'(q) > 0$ ,  $C''(q) > 0$ . The last assumption, increasing marginal costs, captures diseconomies in scale in asset trading and is key to the model's implications.

Similarly to BG, we model a fund that began operating at time 0 and study the investors' decisions at time  $t$ . Since we do not study fund dynamics, our model analyzes a single-period's decision. The fund's net return at time  $t$  is defined as

<sup>5</sup> Note that the return process in our model is exogenous and therefore not affected by the asset management market. Since our goal is to investigate whether frictions can reconcile the empirical evidence with the equilibrating mechanism of BG, our model retains the main assumptions of BG, including the exogeneity of fund returns. Several authors have investigated the equilibrium implications of delegated portfolio management for the asset market, including [Cuoco and Kaniel \(2011\)](#), [Kaniel and Kondor \(2012\)](#), [Basak and Pavlova \(2013\)](#), [He and Krishnamurthy \(2013\)](#), [Garleanu and Pedersen \(2015\)](#), and [Sato \(2016\)](#), among others.

$r_t \equiv R_t - \frac{C(q_t)}{q_t} - f_t$ , where  $q_t$  is the  $t - 1$  investment in the fund and  $f_t$  is the fund's fee. If the revenues collected by the manager at time  $t$ ,  $f_t q_t$ , cover the fixed costs of the fund, the fund continues its activity, otherwise the fund closes down. We assume without loss of generality that fixed costs are zero. We analyze first the case in which fees are exogenously determined. This enables us to abstract from the manager's actions and focus on the consequences of investors' decisions on fund performance. In Section 4, we allow the manager to set the fund's fee every period in order to maximize profits. We show that our main conclusions hold with endogenously determined fees, too.

We depart from BG in that the fund's potential investors exhibit different degrees of financial sophistication and have limited funds to invest. As explained in the Introduction, to model different degrees of sophistication in the simplest possible way, we allow for reservation returns to vary across investors. More specifically, we assume that each investor  $i$  is associated with a specific cost  $\gamma_i$  of investing in the passively managed fund, which captures lack of knowledge or understanding of financial products, inability to find them, or difficulty in making financial choices. Net of that cost, the reservation expected risk-adjusted return (henceforth reservation return) of the  $i$ -th investor is  $-\gamma_i$ . Therefore, unlike in the BG model, the investor's reservation return is different from zero and is also different across investors.<sup>6</sup> We assume that there is a continuum of investors in the market in which the fund is offered with absolute value of the reservation return  $\gamma$  uniformly distributed over the interval  $[0, \gamma_{MAX}]$ , with  $\gamma_{MAX} \leq 1$ . Therefore, we assume that all investors in the market have negative reservation returns net of costs. Alternatively, we could allow some investors to have positive reservation returns without altering the conclusions. The parameter  $\gamma_{MAX}$  determines the overall level of sophistication in the market in which the fund is offered.

The timing of the events is as follows:

Date  $t - 1$ :

- Investors enter the fund. We denote by  $\bar{\gamma}$  the absolute value of the reservation return of the most sophisticated investor who enters the fund. Therefore, all investors with  $\gamma$  in the  $[\bar{\gamma}, \gamma_{MAX}]$  are invested in the fund.

Date  $t$ :

- The fund's return at date  $t$  is realized and earned by current investors.
- After observing the return at date  $t$ , the fund's current investors decide whether to reinvest with the fund or withdraw their current investment.
- New investors decide whether they want to invest with the fund.
- We assume that each current investor holds an investment in the fund that is worth  $m$  dollars at  $t$ . Also, each investor is endowed with a quantity  $m(1 - \gamma_i)$  of new money to invest at  $t$ . This assumption captures the idea that less sophisticated investors face more severe financial constraints.

Date  $t + 1$ :

- The fund's return at date  $t + 1$  is realized and earned by all current and new investors.

We study equilibrium at  $t$ .

Upon observing the series of net returns and total assets under management from 1 to  $t$ ,  $\{r_s, q_s\}_{s=1}^{s=t}$ , investors can infer the series of returns before transaction costs and fees  $\{R_s\}_{s=1}^{s=t}$  and form an expectation of the fund manager's ability through Bayesian updating:

$$\phi_{t+1} = E(R_{t+1} | R_1, \dots, R_t).$$

Investor  $i$  demands shares of the fund if the fund's expected net return exceeds her reservation return,  $-\gamma_i$ . The fund's expected return net of trading costs and fees (total performance,  $TP$ ) in period  $t + 1$  conditional on information up to period  $t$ , equals

$$TP_{t+1}(q_{t+1}) = E[r_{t+1} | R_1, \dots, R_t] = E\left[R_{t+1} - \frac{C(q_{t+1})}{q_{t+1}} - f_{t+1} | R_1, \dots, R_t\right]$$

A current investor will either withdraw her investment from the fund or keep her current investment and invest her date  $t$  endowment in the fund depending on whether the fund's expected net return at date  $t$  is below or above her reservation return.

## 2.2. Equilibrium

An equilibrium at  $t$  is defined as the amount of assets under management,  $q_{t+1}^*$ , such that investors maximize their expected risk-adjusted return. In an equilibrium in which only current investors enter the fund, the following conditions must hold:

- The fund's expected net return is given by  $TP_{t+1}(q_{t+1}^*) = \phi_{t+1} - \frac{C(q_{t+1}^*)}{q_{t+1}^*} - f_{t+1}$ .
- All investors who withdraw their money from the fund have reservation returns higher than  $TP_{t+1}(q_{t+1}^*)$ .

<sup>6</sup> Note that  $\gamma_i$  is the cost for investor  $i$ , per dollar invested. If the costs of investing in passive funds are fixed, wealthier investors face lower  $\gamma$ .

- All investors who invest new money in the fund have reservation returns less than or equal to  $TP_{t+1}(q_{t+1}^*)$ .
- The equilibrium amount of assets  $q_{t+1}^*$  is such that  $0 \leq q_{t+1}^* \leq v_t + M$ , where  $v_t \equiv m(\gamma_{\text{MAX}} - \bar{\gamma})$  is the value at  $t$  of current investors' investment at  $t - 1$  and  $M$  denotes the maximum inflow possible in this period:  $m(\gamma_{\text{MAX}} - \frac{1}{2}\gamma_{\text{MAX}}^2)$ .

To find the cutoff reservation return,  $-\gamma^C$ , such that all current investors with reservation returns lower than  $-\gamma^C$  reinvest in the fund and all current investors with reservation returns higher than  $-\gamma^C$  leave the fund, we solve the system:

$$TP_{t+1}(q_{t+1}^C) = -\gamma^C,$$

$$q_{t+1}^C = 2m(\gamma_{\text{MAX}} - \gamma^C) - \frac{m}{2}(\gamma_{\text{MAX}}^2 - (\gamma^C)^2).$$

Depending on the value of the solution  $\gamma^C$ , there are three possible alternatives:

Case 1:  $\gamma_{\text{MAX}} \leq \gamma^C$ . Even if all investors left the fund, so  $q_{t+1} = 0$  and  $C(q_{t+1}) = 0$ , the fund's expected net return would be lower than the reservation return of the market's most unsophisticated investor. Therefore, the fund must close down and  $q_{t+1}^* = 0$ .

Case 2:  $\bar{\gamma} \leq \gamma^C < \gamma_{\text{MAX}}$ . Current investors with reservation returns higher than  $-\gamma^C$  exit the fund and those with reservation returns lower than  $-\gamma^C$  reinvest in the fund. The fund's expected net return equals  $E(r_{t+1}) = -\gamma^C < 0$  and the fund's assets will be  $q_{t+1}^* = q_{t+1}^C$ .

Case 3:  $\gamma^C < \bar{\gamma}$ . Even if all current investors reinvested in  $\bar{\gamma}$ , the fund's expected net return would be higher than the reservation return of the fund's most sophisticated target investor, so some new and more sophisticated investors will want to enter the fund. Therefore,  $q_{t+1}^* \geq 2m(\gamma_{\text{MAX}} - \bar{\gamma}) - \frac{m}{2}(\gamma_{\text{MAX}}^2 - \bar{\gamma}^2)$ . In this case, we are interested in knowing how many new investors would enter the fund.

In an equilibrium in which new investors enter the fund, the following conditions must hold:

- The fund's expected return equals  $TP_{t+1}(q_{t+1}^*)$ .
- New investors who invest in the fund have reservation returns less than or equal to  $TP_{t+1}(q_{t+1}^*)$ .
- Potential new investors who decide not to invest in the fund have reservation returns higher than  $TP_{t+1}(q_{t+1}^*)$ .

To find the cutoff reservation return,  $-\gamma^N$ , such that all current investors reinvest with the fund, new investors with reservation returns lower than  $-\gamma^N$  enter the fund, and new investors with reservation returns higher than  $-\gamma^N$  do not invest with the fund, we solve the system:

$$TP_{t+1}(q_{t+1}^N) = -\gamma^N,$$

$$q_{t+1}^N = v_t + m(\gamma_{\text{MAX}} - \gamma^N) - \frac{1}{2}(\gamma_{\text{MAX}}^2 - (\gamma^N)^2).$$

Since  $\gamma^C < \bar{\gamma}$ , we have that  $\gamma^N < \bar{\gamma}$ , too, and therefore at least some new investors enter the fund. The last investor  $i$  to enter the fund in period  $t$  has  $\gamma_i = \gamma^N$ , and the quantity invested in the fund is  $q_{t+1}^* = v_t + m(\gamma_{\text{MAX}} - \gamma^N) - \frac{1}{2}(\gamma_{\text{MAX}}^2 - (\gamma^N)^2)$ . If  $\gamma^N < 0$ , then all potential investors enter the fund and the quantity invested in the fund is  $q_{t+1}^* = v_t + m\gamma_{\text{MAX}}(1 - \frac{\gamma_{\text{MAX}}}{2}) = v_t + M$ . Consequently, the fund's expected net return is  $E(r_{t+1}) = -\gamma^N$ , if  $\bar{\gamma} > \gamma^N > 0$  and  $E(r_{t+1}) = TP_{t+1}(v_t + M)$ , if  $\gamma^N \leq 0$ .

Henceforth, we assume for simplicity that  $C(q) = cq^2$ .

**Proposition 1.** *The expected net return of a fund offered in a market with investors in the interval  $[0, \gamma_{\text{MAX}}]$  equals:*

$$E(r_{t+1}(\phi_{t+1})) = \begin{cases} -\gamma^C, & \text{if } \Phi_1 \leq \phi_{t+1} < \Phi_2 \\ -\gamma^N, & \text{if } \Phi_2 \leq \phi_{t+1} < \Phi_3 \\ TP_{t+1}(v_t + M), & \text{if } \Phi_3 \leq \phi_{t+1}, \end{cases}$$

where  $\Phi_j, j = \overline{1, 3}$  are defined in [Appendix 1](#) and  $\gamma^C, \gamma^N$ , equal:

$$\gamma^C = \frac{1}{cm} (1 + 2cm - A^{1/2}), \quad \text{where}$$

$$A \equiv 1 + 2cm (2 + \phi_{t+1} - f_{t+1}) + c^2 m^2 (2 - \gamma_{\text{MAX}})^2,$$

$$\gamma^N = \frac{1}{cm} (1 + cm - B^{1/2}), \quad \text{where}$$

$$B \equiv 1 + 2cm (1 + \phi_{t+1} - f_{t+1}) + c^2 m^2 \left( (1 - \gamma_{\text{MAX}})^2 - \frac{2}{m} v_t \right),$$

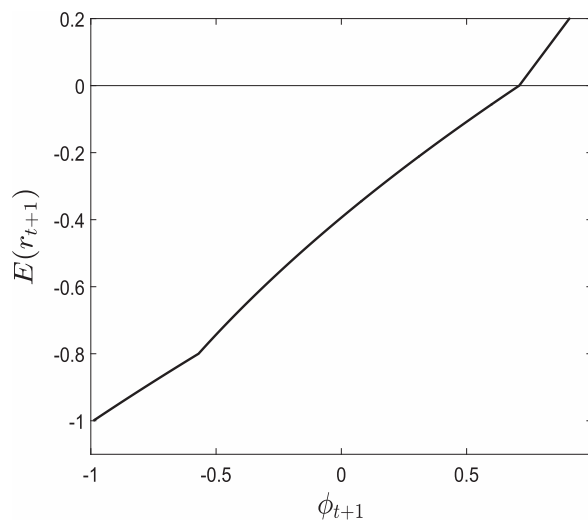


Fig. 1. Expected net return as a function of expected managerial ability. Parameter values:  $m = 100$ ,  $c = 0.01$ ,  $\gamma_{\text{MAX}} = 1$ ,  $\bar{\gamma} = 0.8$ ,  $f = 0.01$ .

respectively.

Proposition 1 shows that the fund's expected net return can be different from zero. The equilibrium expected net return may be negative in our setup when investors prefer to keep their investment in the fund despite earning a negative return because this return is still higher than their reservation return. Therefore, the model can be used to rationalize the well-documented survival of mutual funds that are expected to underperform.

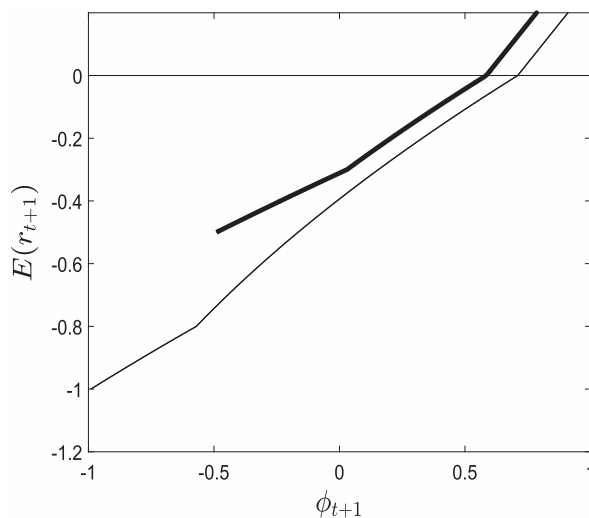
However, another prediction emerges from Proposition 1: The fund's equilibrium expected performance increases with managerial ability,  $\phi_{t+1}$ . This finding is important because if ability persists through time, as one would expect, then fund performance persists, too. Therefore, past performance and future performance should be positively related. As noted above, persistence by itself does not imply that funds can outperform the market, but simply that future expected performance can be predicted from past performance. In other words, expected fund performance increases monotonically with past recent performance, even if it never becomes positive. This is the first new prediction of our model.

The sufficient conditions for this result are: heterogeneity of investors' reservation returns and limited capital to invest. The assumption that investors are financially constrained prevents funds from having a risk-adjusted expected net return equal to the reservation return of the most unsophisticated investor among the fund's target investors. If investors are not constrained, there would be excess demand from the most unsophisticated investors for any fund with a higher level of performance. Therefore, an increase in managerial ability would not lead to an increase in the fund's net performance, and the fund would simply attract more unsophisticated investors. Heterogeneity in investors' reservation returns ensures that in equilibrium we have different expected returns for different levels of managerial ability. Therefore, it is not the case that each one of these conditions by itself generates persistence in fund performance.

Fig. 1 shows the fund's expected net return as a function of expected managerial ability holding the fund's fee constant. If managerial ability is too low, the fund must close down. As managerial ability increases, the fund starts to operate with the most unsophisticated investors among all its potential investors. Investors' limited capital allows fund performance to increase with managerial ability. If managerial ability is high enough, all current investors reinvest with the fund and new more sophisticated investors start to invest. Because new investors invest only their current endowment, the fund's assets grow less rapidly as managerial ability increases, so the fund's expected return increases faster. Once all potential investors are in the fund, fund performance increases one-to-one with managerial ability.<sup>7</sup>

Finally, note that all investors in our model are financially constrained. However, in reality, sophisticated arbitrageurs, such as high net worth individuals, large institutions, hedge funds, and funds-of-funds, would want to invest in mutual funds if there were positive abnormal returns to be earned. If we allowed for the presence of a sophisticated unconstrained investor in our model, our conclusions would hardly change. The reason is that such an investor would enter the market only for high levels of fund performance. Once in the fund, flows of money from this investor would drive fund performance to her reservation return regardless of managerial ability, just like in BG. However, for lower levels of managerial ability, only less sophisticated constrained investors would invest in the fund, so our results regarding the relation between expected performance and managerial ability would be the same in that region.

<sup>7</sup> Positive fund performance is, of course, a consequence of the exogeneity of the fund fee. As we show in Section 4, when fees are endogenous, the manager extracts all the rents and ensures that net performance does not exceed the reservation return of the most sophisticated investor in the market.



**Fig. 2.** Expected net return as a function of expected managerial ability and market sophistication, with heterogeneous financial constraints across investors. The thick (thin) line corresponds to the sophisticated (unsophisticated) market, i.e., low (high)  $\gamma_{\text{MAX}}$ . Parameter values:  $m = 100$ ,  $c = 0.01$ ,  $\gamma_{\text{MAX}}^U = 1$ ,  $\gamma_{\text{MAX}}^S = 0.5$ ,  $\bar{\gamma}^U = 0.8$ ,  $\bar{\gamma}^S = 0.3$ ,  $f = 0.01$ .

### 3. Market segmentation and investor sophistication

Having shown that performance persistence can arise as a consequence of heterogeneity in the effects of frictions across investors, in this section we compare two different market segments characterized by different degrees of investor sophistication. Our analysis is motivated by empirical studies suggesting that the market for delegated portfolio management is segmented on the basis of investor sophistication. More specifically, [Del Guercio and Tkac \(2002\)](#) find that in contrast to mutual fund investors, pension fund investors, who can be argued to be more sophisticated, punish poorly performing managers. [James and Karceski \(2006\)](#) report that larger minimum investment requirements among institutional funds are associated with higher fund performance, higher sensitivity of flows to risk-adjusted performance, and lower soft-money expenses. Finally, [Del Guercio and Reuter \(2014\)](#) show that within retail funds, the broker-sold segment exhibits less sensitivity of flows to performance, higher fees, and lower alphas than in the directly-sold segment of the market, consistent with the idea that investors in that segment are willing to pay extra fees and sacrifice performance in exchange for advice.

Although we do not model market segmentation explicitly, several theoretical studies have shown how segmentation in the market for delegated portfolio management may arise in equilibrium when funds compete for investors' money and investors differ along dimensions, such as their valuation of quality, their liquidity needs, their ability to interpret fees, or their trust in asset managers ([Metrick and Zeckhauser, 1998](#); [Nanda et al., 2000](#); [Gil-Bazo and Ruiz-Verdú, 2008](#); [Gennaioli et al., 2015](#)).

In our model, the overall level of sophistication of the fund's potential investors is determined by  $\gamma_{\text{MAX}}$ . We first show (Proof of Proposition 1) that a fund targeting less sophisticated investors always underperforms the fund with more sophisticated investors conditional on the level of managerial ability, since both  $\gamma^C$  and  $\gamma^N$  increase with  $\gamma_{\text{MAX}}$ . This is simply the consequence of the fund in the less sophisticated market always capturing more assets.

In [Fig. 2](#), we plot the expected net return of two otherwise identical funds whose target investors are characterized by two different values of  $\gamma_{\text{MAX}}$ . The values of  $\bar{\gamma}$  are such that the mass of current investors is the same for both funds. The figure shows that the fund targeted to less sophisticated investors starts to operate at a lower level of managerial ability. Moreover, it takes a lower level of performance for all current investors to reinvest with the fund in the less sophisticated market since its investors have lower reservation returns. Once all current investors have decided to reinvest, new investors enter the fund. However, the entry of new investors has a less detrimental effect on fund performance than reinvestment by current investors, so fund performance rises faster with managerial ability.

Although expected performance increases with managerial ability for both funds, the rate of increase is higher for the fund in the less sophisticated market. The reason is that for the same level of ability, investors in this fund are less sophisticated and have less money to invest. Consequently, identical differences in managerial ability generate larger differences in expected performance for the fund operating in a market populated by less sophisticated investors. Holding the distribution of managerial ability constant, this finding implies that funds will exhibit more dispersion in expected performance in markets with less sophisticated investors. This is the second new prediction of our model.

The assumption that less sophisticated investors are endowed with less money to invest plays an important role in differences in dispersion between the fund in the sophisticated market and the fund in the unsophisticated market. To see this, in [Appendix 1](#) we also describe the equilibrium when all investors receive the same endowment,  $m$ . In particular, we show that:

**Remark 1.** The slope of the expected return with respect to managerial ability for a fund offered in a market with investors in the interval  $[0, \gamma_{MAX}]$  does not depend on  $\gamma_{MAX}$  if the endowment at  $t$  is constant across investors.

Motivated by the aforementioned empirical studies that compare different segments of the asset management market, in the rest of this section we study the relation between flows of money and past performance and compare it across market segments characterized by different degrees of investor sophistication.

Investors update their beliefs about managerial ability upon observing past fund returns using Bayes' rule:

$$\begin{aligned} \phi_{t+1} &= E(R_{t+1} | R_1, \dots, R_t) = E(\alpha | R_{t+1}, R_t, \dots, R_1) \\ &= (1 - \Delta) \phi_t + \Delta R_{t+1}, \end{aligned}$$

where  $\Delta \equiv \frac{\omega}{\delta + (t + 1)\omega}$ .

Similarly to BG, we define flows of money as the percentage change in the fund's assets under management:

$$\begin{aligned} Flows_{t+1} &= \frac{q_{t+1} - q_t}{q_t} \\ &= \begin{cases} -1, & \text{if } \phi_{t+1} < \Phi_1 \\ \frac{2m}{v_t} \left( (\gamma_{MAX} - \gamma^C) - \frac{1}{2}(\gamma_{MAX}^2 - (\gamma^C)^2) \right) - 1, & \text{if } \Phi_1 \leq \phi_{t+1} < \Phi_2 \\ \frac{m}{v_t} \left( (\gamma_{MAX} - \gamma^N) - \frac{1}{2}(\gamma_{MAX}^2 - (\gamma^N)^2) \right) - 1, & \text{if } \Phi_2 \leq \phi_{t+1} < \Phi_3 \\ \frac{M}{v_t}, & \text{if } \Phi_3 \leq \phi_{t+1}, \end{cases} \end{aligned}$$

where we replaced  $q_t = v_t$ , and  $q_{t+1}$  is given by equation (A.16) in Appendix 1. As shown in Proposition 1,  $\gamma^C$  and  $\gamma^N$  depend on expected managerial ability, which in turn is a function of past performance. Therefore, the equations above describe the flow-past performance relation.

We solve for  $\gamma^C$  and  $\gamma^N$  and compute fund flows as a function of past performance for two different levels of  $\gamma_{MAX}$ , assuming the same initial amount of assets under management,  $q_t$ . The results are shown graphically in Fig. 3. The shape of the flow-performance relation is similar to that shown by BG. Flows increase and are convex in recent performance. The main difference with respect to flows in BG is that when new investors enter the fund, the slope of the flow-performance relation declines. This is because new investors only invest their current period's endowment. We also notice that in the more sophisticated segment, investors are less tolerant to poor fund performance. However, the most interesting takeaway from Fig. 3 is that the flow-performance relation is flatter in the less sophisticated segment of the market conditional on the same level of past performance. The reason is that less sophisticated investors are more financially constrained, so flows are less responsive to increases in expected performance. Moreover, it takes a lower level of past performance for all current investors to reinvest with the fund and the flow-performance relation less steep when new investors enter the fund. To summarize, the presence of frictions does not qualitatively alter the shape of the flow-performance relation with respect to the ideal setting of BG. However, markets that are more afflicted by frictions exhibit less sensitivity of flows to performance, which is consistent with the findings of empirical studies comparing different segments of the market characterized by different degrees of investor sophistication (Del Guercio and Tkac, 2002; James and Karceski, 2006; Del Guercio and Reuter, 2014).

#### 4. Endogenous fees

So far, we have taken fees as exogenous and have focused on the impact of frictions for fund performance. In this section, we allow the manager to set the fee that maximizes fee revenues subject to the investors's participation constraint.

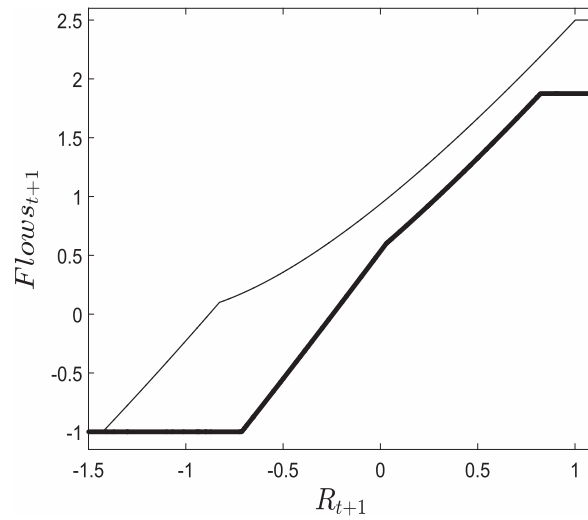
The participation constraint of an investor who decides to invest with the fund is that the expected return per dollar invested is higher than or equal to her reservation return from investing in the index. Thus, at  $t + 1$  the manager solves the following problem:

$$\begin{aligned} &\max_{f_{t+1} \geq 0} f_{t+1} q_{t+1} \\ &s.t. \quad TP_{t+1}(q_{t+1}) \geq -\gamma^* \\ &\quad q_{t+1} \geq 0, \end{aligned}$$

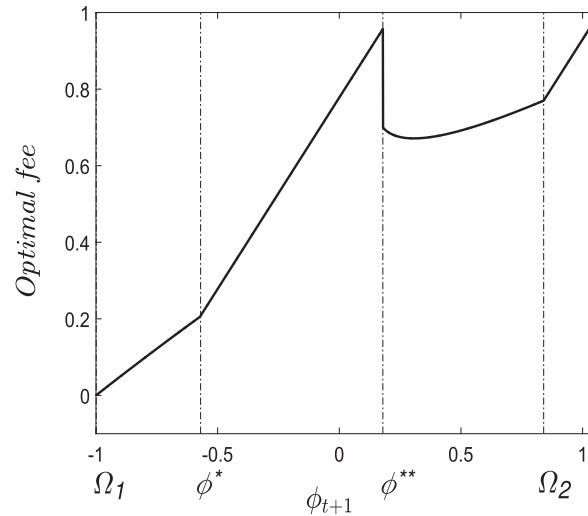
where  $f_{t+1}$  is the fee the manager charges,  $q_{t+1}$  is the total quantity invested in the fund at time  $t + 1$  and is given by (A.16) in Appendix 1, and  $\gamma^*$  is the reservation return of the marginal investor.

We solve numerically for the optimal fee  $f_{t+1}$  and plot it in Fig. 4 as a function of managerial ability,  $\phi_{t+1}$ . The figure reveals that when managerial ability is very low ( $\phi_{t+1} < \Omega_1$ , where  $\Omega_1 \equiv -\gamma_{MAX}$ ), there is no fee  $f_{t+1} \geq 0$  for which investors are willing to invest in the fund. As performance increases ( $\Omega_1 \leq \phi_{t+1} < \phi^*$ , where  $\phi^*$  is given by (B.13) in Appendix 2) there exists a fee  $f_{t+1} > 0$ , such that the fund operates with some investors. The current investors decide whether to stay in the fund depending





**Fig. 3.** Fund flows as a function of last-period returns. The thick (thin) line corresponds to the sophisticated (unsophisticated) market, i.e., low (high)  $\gamma_{MAX}$ . Parameter values:  $m = 100, c = 0.001, \gamma_{MAX}^U = 1, \gamma_{MAX}^S = 0.5, \bar{\gamma}^U = 0.8, \bar{\gamma}^S = 0.3, \Delta = 0.6, \phi_t = 0.03, f = 0.01$ .



**Fig. 4.** Optimal fee as a function of expected managerial ability. Parameter values:  $m = 100, c = 0.001, \gamma_{MAX} = 1, \bar{\gamma} = 0.8$ .

on whether their participation constraint is satisfied. In [Appendix 2](#), we show that the optimal fee set by the manager is strictly increasing in  $\phi_{t+1}$  when only current investors invest in the fund. However, once all the current investors have reinvested in the fund ( $\phi_{t+1} \geq \phi^*$ ), the manager faces a trade-off between increasing the fee and extracting rents from current investors or charging a lower fee and attracting assets from new investors. The manager initially chooses to increase the fee one-to-one with managerial ability, so new investors do not enter. Intuitively, since current investors have more money to invest (the previous period money and the current period money), the manager prefers them to new investors. When managerial ability increases substantially ( $\phi_{t+1} \geq \phi^{**}$ , derived in [Appendix 2](#)), the manager optimally allows new investors to participate. When new investors enter the fund, the negative effect of a fee increase on assets is factorized by  $\gamma^N - 1$ . For low levels of managerial ability,  $\gamma^N$  is high and close to  $\bar{\gamma}$  and  $\gamma^N - 1$  is relatively high so the negative effect of a fee increase on quantity dominates its direct positive impact on profits. In that case, the manager optimally decreases the fee and more new investors enter the fund. As managerial ability increases, more sophisticated investors enter the fund,  $\gamma^N - 1$  becomes smaller, and the negative effect of the fee increase on quantity is dominated. The manager then chooses to increase the fee as managerial ability increases. Finally, we define  $\Omega_2 \equiv f_{opt} + c(v_t + M)$ , where  $f_{opt}$  is the optimal fee set by the manager when the performance equals the minimum performance for which all investors enter the fund. When managerial ability is very high ( $\Omega_2 \leq \phi_{t+1}$ ), all current and new investors are invested in the fund. Since the fund is already at its maximum capacity, it is optimal for the manager to increase the fee one-to-one with managerial ability.

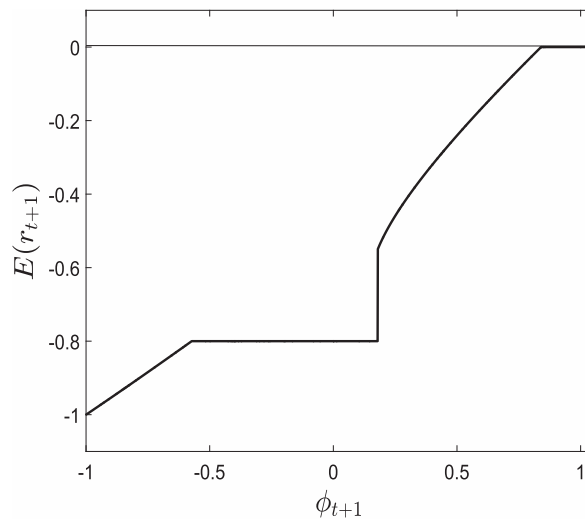


Fig. 5. Expected net return as a function of expected managerial ability. Parameter values:  $m = 100$ ,  $c = 0.001$ ,  $\gamma_{\text{MAX}} = 1$ ,  $\bar{\gamma} = 0.8$ .

It is also useful to understand the effect of managerial ability  $\phi_{t+1}$  on the fund's total performance  $TP_{t+1}(q_{t+1})$  in the presence of frictions when fees are endogenous. In Fig. 5, we can see that when the fee is set by the manager, the fund's total performance is non-decreasing in managerial ability  $\phi_{t+1}$  and is negative or equal to zero. So, unlike in the model of BG, in the presence of frictions, total performance can be negative. Moreover, for total performance to be equal to zero, as in BG, managerial ability  $\phi_{t+1}$  has to be large enough for all investors to decide to participate in the fund. Unlike the case in which fees are exogenously determined, the manager extracts all the rents and the active fund never outperforms passive investment after fees and trading costs. Nevertheless, Fig. 5 shows that the non-decreasing relation between expected performance and managerial ability is preserved even when the manager is allowed to set the fee every period.

## 5. Empirical evidence

In this section, we explore the consistency between the model and the data. In particular, we are interested in testing our two novel empirical predictions:

1. Expected fund performance increases monotonically with past recent performance.
2. Funds that are sold in markets populated by more sophisticated investors exhibit more dispersion in performance.

### 5.1. Data

We construct a data set of U.S. domestic diversified equity open-end mutual funds in the 1962–2015 period from the CRSP Survivor-Bias-Free U.S. Mutual Fund Database. The CRSP data are at the share class level. To aggregate data at the mutual fund level, we group share classes for the same fund using CRSP's class group identifier `crsp_cl_grp`. When this field is missing, we use the fund's name to identify share classes of the same fund. We then compute a fund's assets under management as the sum of net assets of all classes of the same fund. Fund age is the age of the fund's oldest share class. Fund returns and expense ratios are defined as asset-weighted averages of these variables across classes of the same fund. To avoid the effect of reporting biases, we exclude funds with assets under management below \$15 million and age less than three years (Evans, 2010; Elton et al., 2011). We also exclude index funds, which we identify by CRSP's identifier or by the fund's name, when name is missing. Our final data set consists of 451,029 fund-month observations corresponding to 3,206 unique funds.

Mutual funds catering to institutional investors are a distinct segment of the market. Such funds or share classes make it explicit in their names and prospectuses that their target clients are institutions. Other investors are restricted from entering by means of high minimum initial investment. It is plausible to think of institutional investors as sophisticated investors (see, e.g., Evans and Fahlenbrach, 2012). Therefore, to test our second prediction, we compare the dispersion in performance of retail funds with that of institutional funds. We identify institutional share classes based on CRSP's identifiers, complemented with fund name when identifiers are missing.<sup>8</sup> We then compute the fraction of assets of each fund in institutional share classes and

<sup>8</sup> More specifically, when CRSP institutional and retail flags are missing, we flag a share class as institutional if either the fund's name contains the string "inst" or the share class name is commonly associated with institutional classes ("I," "Y," and "X").

**Table 1**

Descriptive statistics. The table reports descriptive statistics for our sample of actively managed, U.S. domestic diversified equity mutual funds in the 1962–2015 period. Each observation corresponds to one mutual fund and one month. Fund assets are computed as the sum of the total net assets across share classes of the same fund. Expense ratio is the asset-weighted average of a fund's share class expense ratios. Fund age is the time since the first offer date of the fund's oldest share class. % Institutional is the percentage of the fund's assets under management in share classes sold to institutional investors. Alpha is the fund's monthly excess return (asset-weighted average return across share classes) minus the realized risk-premium, computed as the product of Fama-French-Carhart factor loadings and factor realizations. R-squared corresponds to the time-series regression of excess returns on the Fama-French-Carhart factors in the previous 36 months employed to estimate factor loadings.

	25th pctile.	Median	75th pctile.	Mean	SD	Obs.
Assets (USD millions)	73	222	793	1,207	4,633	451,029
Expense ratio (%)	0.90	1.14	1.42	1.16	0.46	443,688
Age (Years)	6.75	11.25	18.83	15.51	13.38	451,029
% Institutional	0	0	43.61	24.39	37.48	451,029
Alpha (%)	−0.95	−0.11	0.70	−0.12	1.83	325,086
R-squared (%)	89.19	93.47	96.29	91.20	8.93	325,440

define institutional (retail) funds as those in which the fraction of assets in institutional classes is 100% (0%).

We estimate risk-adjusted performance, alpha, using Carhart's (1997) four-factor model:

$$r_{it} = \alpha_i + \beta_{rmf,i} r_{rmf,t} + \beta_{smb,i} smb_t + \beta_{hml,i} hml_t + \beta_{pr1y} i^{pr1y}_t + \varepsilon_{it},$$

where  $r_{it}$  is fund  $i$ 's return in month  $t$  in excess of the 30-day risk-free interest rate, and  $rmf_t$ ,  $smb_t$ , and  $hml_t$  denote the return on portfolios that proxy for the market, size, and book-to-market risk factors, respectively. The term  $pr1y_t$  is the return difference between stocks with high and low returns in the previous year and is included to account for passive momentum strategies. Interest rates, the Fama-French factors, and momentum are downloaded from Kenneth French's website.

We construct a panel of monthly fund risk-adjusted returns, alphas, by first regressing the fund's excess return on the three Fama-French factors and momentum over the previous three years. We require at least 30 non-missing observations in the time-series regression. The fund's monthly alpha is then computed by subtracting the product of the estimated betas and factor realizations in each month from the fund's excess return.

Table 1 reports descriptive statistics for our data set. The average fund in our sample manages assets worth 1,207 million USD, but the distribution is heavily skewed so median assets under management are only 222 million USD. The mean fund in our sample charges an expense ratio of 1.16% and is 15.5 years old. Although the median fund has no institutional share classes, roughly one quarter of the average fund's assets are managed on behalf of institutional investors. Consistent with the literature, the average fund fails to earn risk-adjusted returns that offset expenses: the mean value of (after-fee) alpha is −0.12% per month (−1.44% annualized). Finally, the high values for the R-squared of the time-series regressions suggest that the Fama-French-Carhart factors explain most of the time-series variability of fund returns in our sample.

## 5.2. Performance persistence

In our model, higher skilled managers offer higher expected performance net of fees and transaction costs. If ability is persistent, differences in performance across funds will also persist through time. To test this hypothesis, we rank mutual funds each month  $t$  based on the intercept of the three-year time-series regression of excess returns on risk factors and momentum ending in month  $t - 1$ . We then split observations in ten decile groups and compute the average monthly alpha of each group.

We also look at performance persistence over horizons longer than one month. More specifically, we compute 3-, 6-, and 12-month risk-adjusted performance for each fund in month  $t$  as the average monthly alpha over the period from month  $t$  to month  $t + s$ , with  $s = 2, 5, \text{ and } 11$ , respectively. We group observations into decile bins based on performance estimated over the previous three years and compute the mean 3-, 6-, and 12-month performance for each bin using non-overlapping evaluation periods.

Table 2 reports the results. For each decile and horizon, the table displays the mean annualized alpha over that horizon for funds in each decile according to previous 3-year performance, the  $t$ -stat for the null hypothesis that the mean is equal to zero, as well as the number of observations. Consistent with a large body of research, funds in the top decile do not exhibit positive alphas. However, there is evidence of persistent differences in performance. For a 1-month horizon, mean alpha and its associated  $t$ -stat increase monotonically across recent performance deciles. All average alphas are statistically significant at conventional significance levels, except for funds in the top decile. Differences are also economically significant. Funds in the top decile beat funds in the bottom decile by 3.8% on an annualized basis. Looking at longer horizons, we still find strong evidence of relative performance persistence within the year as funds in the bottom (top) deciles underperform (outperform) other funds at all horizons. However, there is also evidence of mean reversion. Annualized alpha of funds in the lower deciles increases while annualized alpha of the top deciles decreases as the length of the evaluation period increases.

**Table 2**

Performance persistence. Each period we rank mutual funds based on the intercept of the time-series regression of excess returns on the Fama-French-Carhart factors over the previous three years and split funds into 10 decile groups. The table reports the mean risk-adjusted performance (in %) for each decile group. In column (1), risk-adjusted performance is the fund's monthly excess return minus the realized risk premium (the product of factor realizations and factor loadings estimated using the time-series regression). In column (2), performance is defined as the average of monthly alphas in each quarter. In column (3), performance is defined as the average of monthly alphas in each semester. In column (4), performance is defined as the average of monthly alphas in each year. To facilitate comparison across columns, risk-adjusted performance is annualized. The table reports for each decile and horizon the *t*-statistic (in parentheses) for the null hypotheses that performance is zero, as well as the number of observations.

Past-Performance Decile	Horizon			
	1 Month (1)	3 Months (2)	6 Months (3)	12 Months (4)
1 (Top)	-0.14 (-0.93) 32,783	-0.38 (-2.35) 10,696	-0.43 (-2.58) 5,223	-0.75 (-4.36) 2,484
2	-0.70 (-5.86) 32,468	-0.62 (-4.90) 10,631	-0.67 (-5.16) 5,199	-0.99 (-7.36) 2,502
3	-0.76 (-6.93) 32,555	-0.93 (-8.10) 10,674	-0.97 (-8.16) 5,226	-1.23 (-9.59) 2,504
4	-0.96 (-9.20) 32,475	-1.02 (-9.34) 10,644	-1.00 (-8.75) 5,191	-1.07 (-8.67) 2,489
5	-1.17 (-11.13) 32,427	-1.14 (-10.65) 10,611	-1.10 (-9.93) 5,182	-1.14 (-9.7351) 2,464
6	-1.45 (-13.68) 32,584	-1.43 (-13.09) 10,668	-1.37 (-11.99) 5,209	-1.33 (-10.70) 2,502
7	-1.52 (-14.22) 32,553	-1.55 (-14.00) 10,652	-1.72 (-15.14) 5,190	-1.70 (-13.86) 2,468
8	-1.82 (-16.12) 32,488	-1.74 (-14.71) 10,608	-1.70 (-13.80) 5,146	-1.47 (-10.97) 2,428
9	-2.34 (-19.36) 32,506	-2.28 (-17.86) 10,545	-2.23 (-16.48) 5,107	-1.80 (-12.35) 2,372
10 (Bottom)	-3.95 (-24.58) 32,247	-3.57 (-22.29) 10,357	-3.24 (-19.30) 4,928	-2.73 (-14.90) 2,263

Therefore, the data show that differences in risk-adjusted returns across mutual funds persist at least in the short term, which is consistent with our model and the assumption that managerial ability varies slowly through time.

### 5.3. Dispersion in performance

We now look at differences in dispersion in performance between retail and institutional funds. In Table 3, we report the standard deviation of alphas in both subsamples and test the null hypothesis of equal variance. More specifically, we report the F-test statistic for the null hypothesis that the ratio of the variance of retail funds' alphas to the variance of institutional fund's alphas equals 1. We also report Levene's test statistic, which is robust to departures from normality.

As reported in row (1) of Table 3, the standard deviation of monthly alphas is 53 bps higher for retail funds than for institutional funds (2.04% vs. 1.51%). The ratio of variances is 1.82, which is statistically different from 1 at conventional significance levels.

Institutional funds are relatively scarce. In particular, the number of observations of monthly alphas for retail funds is roughly six times that for institutional funds (168,540 vs. 27,312). A lower dispersion in alphas among institutional funds could be due to this segment of the market being less heterogeneous than the retail segment along some dimension. To increase the size of the institutional sample and potentially its heterogeneity, in row (2) we repeat the test considering as institutional (retail) funds those in which more (less) than 75% (25%) of the assets are in institutional classes. With this alternative definition, the institutional sample size doubles and the retail sample increases by 40%. Despite this large increase in the relative sample size of institutional funds, our conclusions hardly change. In unreported tests, we find almost identical results when using more

**Table 3**

Performance dispersion: retail vs. institutional funds. The table reports the standard deviation (SD) of monthly risk-adjusted performance (in %), defined as the fund's monthly excess return minus the realized risk-premium, computed as the product of Fama-French-Carhart factor loadings and factor realizations, for the subsample of retail funds and institutional funds. In row (2), monthly expense ratios are added back to alphas. In row (3), institutional (retail) funds are defined as those in which more (less) than 75% (25%) of the assets are in institutional classes. Funds are split according to their investment objective in rows (4)–(10), and according to the R-squared of the time-series regression employed to estimate factor loadings in rows (11)–(15). The table also reports the F-test statistic (ratio of variance of retail funds' risk-adjusted performance to variance of institutional funds' risk-adjusted performance) and Levene's test statistic for the null hypothesis of equal variance.

	Retail		Institutional		F-test	Levene's test
	SD	Obs.	SD	Obs.		
<i>Full Sample</i>						
(1) After-fee alphas	2.04	168,540	1.51	27,312	1.82 ( $p < 0.01$ )	1534.05 ( $p < 0.01$ )
(2) Before-fee alphas	2.05	166,969	1.52	26,774	1.81 ( $p < 0.01$ )	1498.72 ( $p < 0.01$ )
(3) 25%–75% cutoff	1.95	233,482	1.50	54,706	1.69 ( $p < 0.01$ )	2073.66 ( $p < 0.01$ )
<i>Inv. Objectives</i>						
(4) Micro Cap	2.33	1,015	1.72	661	1.84 ( $p < 0.01$ )	24.45 ( $p < 0.01$ )
(5) Small Cap	2.31	20,665	1.72	6,645	1.80 ( $p < 0.01$ )	338.91 ( $p < 0.01$ )
(6) Mid Cap	2.41	9,713	1.75	2,766	1.89 ( $p < 0.01$ )	128.52 ( $p < 0.01$ )
(7) Income	1.42	4,216	2.74	618	0.27 ( $p < 0.01$ )	13.18 ( $p < 0.01$ )
(8) Growth & Income	1.46	33,228	1.14	6,838	1.65 ( $p < 0.01$ )	264.80 ( $p < 0.01$ )
(9) Growth	2.11	97,336	1.40	9,628	2.27 ( $p < 0.01$ )	1003.54 ( $p < 0.01$ )
(10) Hedged	2.73	1,630	0.85	132	10.26 ( $p < 0.01$ )	17.02 ( $p < 0.01$ )
<i>R-squared Quintiles</i>						
(11) Quintile 1 (Top)	1.05	22,322	0.91	9,616	1.34 ( $p < 0.01$ )	193.11 ( $p < 0.01$ )
(12) Quintile 2	1.50	29,663	1.32	5,453	1.29 ( $p < 0.01$ )	71.36 ( $p < 0.01$ )
(13) Quintile 3	1.81	33,546	1.59	4,890	1.30 ( $p < 0.01$ )	80.01 ( $p < 0.01$ )
(14) Quintile 4	2.09	37,767	1.90	4,534	1.21 ( $p < 0.01$ )	42.40 ( $p < 0.01$ )
(15) Quintile 5 (Bottom)	2.72	45,242	2.42	2,819	1.27 ( $p < 0.01$ )	60.41 ( $p < 0.01$ )

comprehensive definitions of institutional and retail funds.

There are two possible concerns with our test. First, the higher dispersion in after-fee alphas among retail funds is consistent with our second prediction but it is also consistent with retail funds exhibiting higher dispersion than institutional funds in expense ratios. To discard this possibility, we repeat the test using before-fee alphas, as proxied by after-fee alphas plus annual expense ratios divided by 12, and report the results in row (3) in Table 3. The standard deviations of before alphas in both groups, however, are almost identical to those of after-fee alphas, and so are the test statistics. In other words, differences in dispersion are not driven by differences in expenses.

Second, differences in dispersion in alphas could be a consequence of the heterogeneity in the way retail and institutional funds invest. For instance, an omitted risk factor would generate variability in estimated alphas. If retail and institutional funds exhibit differences in their exposure to this risk factor, they will also exhibit differences in dispersion in alphas, even if true alphas have equal variance in both groups of funds. While we cannot fully discard this possibility, we can at least ensure that differences in dispersion are not due to differences in investment style. In particular, we repeat the tests for equality of variance within investment objectives, as provided by CRSP.<sup>9</sup> Results are shown in rows (4)–(10) in Table 3. In all investment objectives, except for one (“Income”), retail funds exhibit significantly more dispersion in alphas than institutional funds.

Finally, note that even if the performance attribution model is well specified, differences in idiosyncratic risk between institutional and retail funds would also generate differences in estimated alphas. We account for this possibility by grouping observations in quintiles based on the R-squared of the time series regression used to estimate factor loadings. As reported in rows (11)–(15), differences in dispersion between institutional and retail investors within each R-squared quintile become smaller than in the full sample, which suggests that differences in R-squared indeed can explain part of the differences in dispersion of alphas documented for the full sample. However, retail funds still exhibit significantly higher dispersion than institutional funds within each one of the five quintiles.

In sum, alphas are more disperse among retail funds than institutional funds, and such difference in dispersion cannot be explained away by differences in fees, investment objective, or idiosyncratic risk, between both segments of the mutual fund market.

## 6. Conclusions

It is now well understood that lack of frictions and competition among rational investors lead to performance unpredictability in asset management. However, the consequences of abandoning this paradigm are unclear. In this paper, we investigate how mutual fund performance is determined in equilibrium in the presence of two types of frictions that have a heterogenous impact across investors: a cost of finding investment alternatives and financial constraints.

<sup>9</sup> Investment objectives without observations of institutional or retail fund alphas are naturally excluded.

Our model can explain the puzzling survival of persistently underperforming funds. But the model also delivers two new predictions. First, if managerial ability is persistent, future performance should be directly related to past performance (regardless of whether some funds can outperform the market). Second, funds operating in markets with less sophisticated investors will exhibit higher dispersion in expected performance than otherwise identical funds in markets with more sophisticated investors. Using data on actively managed U.S. diversified equity funds, we report evidence that is consistent with both predictions.

Our study can be used to guide further empirical research on performance predictability. The conclusions also have important implication for policies aimed at improving the efficiency of the market for mutual funds, as they suggest that frictions, and in particular, the difficulties faced by the least sophisticated investors in finding low-cost investment alternatives and financial constraints, may explain the survival of persistently underperforming funds documented in the literature.

## Acknowledgements

The authors thank the editor, Tarun Chordia, and an anonymous referee, as well as Ioana Schiopu, Calin Arclean, Ramiro Losada, Bill Zu, Juan-Pedro Gomez, Francesco Consonni, Florinda Silva, and seminar participants at ESADE Business School, Universitat Pompeu Fabra, Universidad Carlos III, 2012 Annual Meeting of the Academy of Behavioral Finance and Economics, 2012 European Summer Symposium in Financial Markets (Gerzensee), VIII Workshop in Public Policy Design (Universitat de Girona), 2013 Finance Forum, 2013 Midwest Finance Association Conference, and 2014 European Financial Management Association Meeting, for helpful comments and suggestions. All remaining errors are the authors' responsibility. The authors also acknowledge financial support from the Government of Spain (grants ECO2011-24928, ECO2014-55488-P, PR2015-00645, and PR2015-00606), the Government of Catalonia (grants 2014-SGR-1079 and 2014-SGR-549), Bank Sabadell, and La Caixa Foundation.

## Appendix 1. Proof of Proposition 1

**Proof of Proposition 1.** The current investors exit or reinvest their wealth depending on whether their reservation return is lower or higher than  $-\gamma^C$ , where  $\gamma^C$  is such that  $TP_{t+1}(q_{t+1}^*) = -\gamma^C$ . The quantity invested in the fund is:

$$q_{t+1}^* = m(\gamma_{\text{MAX}} - \gamma^C) + m\left((\gamma_{\text{MAX}} - \gamma^C) - \frac{1}{2}(\gamma_{\text{MAX}}^2 - (\gamma^C)^2)\right),$$

where the first term corresponds to the period  $t - 1$  investment that is reinvested and the second term corresponds to the period  $t$  investment. The equilibrium condition  $TP_{t+1}(q_{t+1}^*) = -\gamma^C$  can be rewritten as:

$$\phi_{t+1} - cm\left(2(\gamma_{\text{MAX}} - \gamma^C) - \frac{1}{2}(\gamma_{\text{MAX}}^2 - (\gamma^C)^2)\right) - f_{t+1} = -\gamma^C. \quad (\text{A1.11})$$

Solving for  $\gamma^C$ , we obtain:

$$\gamma^C = \frac{1}{cm}(1 + 2cm - A^{1/2}), \text{ where} \quad (\text{A1.12})$$

$$A \equiv 1 + 2cm(2 + \phi_{t+1} - f_{t+1}) + c^2m^2(2 - \gamma_{\text{MAX}})^2. \quad (\text{A1.13})$$

$\gamma^C$  is a real solution of equation (A.11) if  $A > 0$  and a sufficient condition for  $A > 0$  is  $2 + \phi_{t+1} > f_{t+1}$ , which is a reasonable assumption.

Notice that if  $\gamma^C < \bar{\gamma}$ , all current investors remain in the funds and we also have possible entry of new investors. Since the new investors are more sophisticated than the current investors, they enter the fund only if all the current investors remain in the fund and, their cutoff reservation return,  $-\gamma^N$ , is obtained from:

$$TP_{t+1}(q_{t+1}^*) = -\gamma^N,$$

$$\text{where } q_{t+1}^* = v_t + m\left((\gamma_{\text{MAX}} - \gamma^N) - \frac{1}{2}(\gamma_{\text{MAX}}^2 - (\gamma^N)^2)\right).$$

We solve for  $\gamma^N$  from the equilibrium condition:

$$\phi_{t+1} - cm\left(2\gamma_{\text{MAX}} - \gamma^N - \bar{\gamma} - \frac{1}{2}(\gamma_{\text{MAX}}^2 - (\gamma^N)^2)\right) - f_{t+1} = -\gamma^N, \quad (\text{A1.14})$$

and obtain:

$$\gamma^N = \frac{1}{cm}(1 + cm - B^{1/2}), \text{ where}$$

$$\begin{aligned} B &\equiv 1 + 2cm(1 + \phi_{t+1} - f_{t+1}) + c^2m^2(1 + 2\bar{\gamma} + \gamma_{\text{MAX}}^2 - 4\gamma_{\text{MAX}}) \\ &= 1 + 2cm(1 + \phi_{t+1} - f_{t+1}) + c^2m^2\left((1 - \gamma_{\text{MAX}})^2 - \frac{2}{m}v_t\right). \end{aligned} \quad (\text{A1.15})$$

$\gamma^N$  is a real solution of equation (A.14) if  $B \geq 0$ . Notice that the entry of new investors takes place only if  $\gamma^C < \bar{\gamma}$ , and this implies that if  $A \geq 0$ , then  $B \geq 0$ . This also implies that  $\gamma^N < \bar{\gamma}$ . If in addition,  $\gamma^N \leq 0$ , then all new investors enter the fund.

Notice also that both  $\gamma^C$  and  $\gamma^N$  increase with  $\gamma_{\text{MAX}}$  if  $\gamma_{\text{MAX}} < 2$ .

Consequently, the amount invested in the fund at time  $t + 1$  is:

$$q_{t+1} = \begin{cases} 0, & \text{if } \phi_{t+1} < \Phi_1 \\ m \left( 2(\gamma_{\text{MAX}} - \gamma^C) - \frac{1}{2}(\gamma_{\text{MAX}}^2 - (\gamma^C)^2) \right), & \text{if } \Phi_1 \leq \phi_{t+1} < \Phi_2 \\ v_t + m \left( (\gamma_{\text{MAX}} - \gamma^N) - \frac{1}{2}(\gamma_{\text{MAX}}^2 - (\gamma^N)^2) \right), & \text{if } \Phi_2 \leq \phi_{t+1} < \Phi_3 \\ v_t + M, & \text{if } \Phi_3 \leq \phi_{t+1}, \end{cases} \quad (\text{A1.16})$$

where:

$$\Phi_1 \equiv f_{t+1} - \gamma_{\text{MAX}},$$

$$\Phi_2 \equiv f_{t+1} + 2cv_t - \bar{\gamma} - \frac{1}{2}cm \left( \gamma_{\text{MAX}}^2 - \bar{\gamma}^2 \right) = f_{t+1} + \frac{1}{2}a\bar{\gamma}^2 - \bar{\gamma}(2a + 1) - a\eta_1$$

$$\Phi_3 \equiv f_{t+1} + c \left( v_t + c\gamma_{\text{MAX}} \left( 1 - \frac{\gamma_{\text{MAX}}}{2} \right) \right) = f_{t+1} + c(v_t + M),$$

$$\eta_1 \equiv \left( 2\gamma_{\text{MAX}} - \frac{1}{2}\gamma_{\text{MAX}}^2 \right) \text{ and}$$

$$a \equiv cm.$$

Notice that if  $\phi_{t+1} < \Phi_1$ , the fund closes down. As a result, the expected return equals to:

$$E(r_{t+1}(\phi_{t+1})) = \begin{cases} -\gamma^C, & \text{if } \Phi_1 \leq \phi_{t+1} < \Phi_2 \\ -\gamma^N, & \text{if } \Phi_2 \leq \phi_{t+1} < \Phi_3 \\ TP_{t+1}(v_t + M), & \text{if } \Phi_3 \leq \phi_{t+1}. \end{cases}$$

When the new capital to be invested in the second period does not vary with the level of investor sophistication, and equals  $m$  for all investors,  $\gamma^C$  and  $\gamma^N$  have the following functional forms:

$$\gamma_i^C = \frac{1}{2a+1} (f_{t+1} - \phi_{t+1} + 2a\gamma_{\text{MAX}})$$

$$\gamma_i^N = \frac{1}{a+1} (f_{t+1} - \phi_{t+1} + a(2\gamma_{\text{MAX}} - \bar{\gamma})),$$

and the equilibrium quantity is:

$$q_{i,t+1} = \begin{cases} 0, & \text{if } \phi_{t+1} < \Phi_{i,1} \\ 2m(\gamma_{\text{MAX}} - \gamma_i^C), & \text{if } \Phi_{i,1} \leq \phi_{t+1} < \Phi_{i,2} \\ v_t + m(\gamma_{\text{MAX}} - \gamma_i^N), & \text{if } \Phi_{i,2} \leq \phi_{t+1} < \Phi_{i,3} \\ v_t + M, & \text{if } \Phi_{i,3} \leq \phi_{t+1}, \end{cases}$$

where:

$$\Phi_{i,1} \equiv f_{t+1} - \gamma_{\text{MAX}} = \Phi_1$$

$$\Phi_{i,2} \equiv f_{t+1} + 2cv_t - \bar{\gamma} \text{ and}$$

$$\Phi_{i,3} \equiv f_{t+1} + c(v_t + m\gamma_{\text{MAX}}).$$

The expected return is similar to the more general case in which less sophisticated investors have less money to invest. However, the equilibrium reservation returns  $\gamma_i^C, \gamma_i^N$  and the cut-off points change. This implies that the slope of the expected return of the fund with respect to managerial ability is  $\frac{1}{2a+1}$  when only some of the current investors enter the fund, and equals  $\frac{1}{a+1}$  when new investors enter the fund. Therefore, the slope does not depend on  $\gamma_{\text{MAX}}$ , in contrast to the case in which wealth increases with financial sophistication. ■

## Appendix 2. Endogenous fees

In this Appendix, we characterize the equilibrium when the manager sets the fee to maximize the fees revenues and study the implications for fund flows.

As we have seen, the manager solves the following problem:

$$\begin{aligned} & \max_{f_{t+1} \geq 0} f_{t+1} q_{t+1} \\ & \text{s.t. } TP_{t+1}(q_{t+1}) \geq -\gamma^* \\ & q_{t+1} \geq 0. \end{aligned}$$

Notice that the quantity  $q_{t+1}$  invested in the fund, given by (A.16), depends on the fee  $f_{t+1}$  and it takes different values depending on the managerial ability  $\phi_{t+1}$ . This quantity is the one that makes the constraint binding, so the problem is equivalent to:

$$\max_{f_{t+1} \geq 0} f_{t+1} q_{t+1} = \begin{cases} 0, & \text{if } \phi_{t+1} < \Phi_1 \\ f_{t+1} m \left( 2(\gamma_{\text{MAX}} - \gamma^C) - \frac{1}{2}(\gamma_{\text{MAX}}^2 - (\gamma^C)^2) \right), & \text{if } \Phi_1 \leq \phi_{t+1} < \Phi_2 \\ f_{t+1} \left( v_t + m \left( (\gamma_{\text{MAX}} - \gamma^N) - \frac{1}{2}(\gamma_{\text{MAX}}^2 - (\gamma^N)^2) \right) \right), & \text{if } \Phi_2 \leq \phi_{t+1} < \Phi_3 \\ f_{t+1} (v_t + M), & \text{if } \Phi_3 \leq \phi_{t+1}. \end{cases}$$

The first-order condition for this problem is:

$$\begin{aligned} \left( 2(\gamma_{\text{MAX}} - \gamma^C) - \frac{1}{2}(\gamma_{\text{MAX}}^2 - (\gamma^C)^2) \right) + f_{t+1} (-2 + \gamma^C) \frac{\partial \gamma^C}{\partial f_{t+1}} &= 0, & \text{if } \Phi_1 \leq \phi_{t+1} < \Phi_2 \\ \left( v_t + m \left( (\gamma_{\text{MAX}} - \gamma^N) - \frac{1}{2}(\gamma_{\text{MAX}}^2 - (\gamma^N)^2) \right) \right) & & \text{if } \Phi_2 \leq \phi_{t+1} < \Phi_3 \\ + f_{t+1} (-1 + \gamma^N) \frac{\partial \gamma^N}{\partial f_{t+1}} &= 0, & \\ v_t + M \geq 0, & & \text{if } \Phi_3 \leq \phi_{t+1}. \end{aligned}$$

Note that  $\Phi_i, i = 1, \dots, 3$  all depend on the fee  $f_{t+1}$ .

When the managerial ability is very low,  $\phi_{t+1} < \Phi_1$ , there exists no fee  $f_{t+1} \geq 0$  that can determine investors to enter. Note also, that when the managerial ability is very high, the quantity does not change with the fee (is maximum as all investors invested in the fund). However, if the fee is too high, new investors might decide not to invest in the fund. By increasing the fee one-to-one with managerial ability, the fund manager ensures that all investors are invested in the fund. So when  $\Phi_3 \leq \phi_{t+1}$ , the optimal fee is  $f_{t+1} = \phi_{t+1} - c\eta_2$ , where  $\eta_2 = v_t + M$ . For the other two regions (when only current investors enter and when new investors also enter), we have a solution in both cases. When the solutions are interior, we just have to compare these two solutions with the one obtained from the case when all investors enter to obtain the global optimum. However, when the solution we obtain in a certain case is not interior, the comparison is not straightforward anymore because the limits of the four intervals all vary with the fee,  $f_{t+1}$ . In what follows, we discuss these cases.

Let us consider first the case when only current investors enter. In this case, the first-order condition is:

$$\left( 2(\gamma_{\text{MAX}} - \gamma^C) - \frac{1}{2}(\gamma_{\text{MAX}}^2 - (\gamma^C)^2) \right) + f_{t+1} (-2 + \gamma^C) \frac{\partial \gamma^C}{\partial f_{t+1}} = 0. \tag{A2.11}$$

Solving for  $f_{t+1}$ , we obtain:

$$f_{t+1} = \frac{\left( \eta_1 - 2\gamma^C + \frac{1}{2}(\gamma^C)^2 \right)}{(2 - \gamma^C) \frac{\partial \gamma^C}{\partial f_{t+1}}}. \tag{A2.12}$$

Let us next define:

$$F(f_{t+1}, \phi_{t+1}) \equiv \left( 2(\gamma_{\text{MAX}} - \gamma^C) - \frac{1}{2}(\gamma_{\text{MAX}}^2 - (\gamma^C)^2) \right) + f_{t+1} (-2 + \gamma^C) \frac{\partial \gamma^C}{\partial f_{t+1}}.$$

Remember that the reservation return of the marginal current investor who enters the fund is:

$$\begin{aligned} \gamma^C &= \frac{1}{a} (1 + 2a - A^{1/2}), \text{ where} \\ A &= T + 2a (\phi_{t+1} - f_{t+1}), \\ T &= 1 + 4a + a^2 (2 - \gamma_{\text{MAX}})^2. \end{aligned}$$

Implicit function theorem implies that:



$$\begin{aligned} \frac{\partial f_{t+1}}{\partial \phi_{t+1}} &= -\frac{\frac{\partial F(f_{t+1}, \phi_{t+1})}{\partial \phi_{t+1}}}{\frac{\partial F(f_{t+1}, \phi_{t+1})}{\partial f_{t+1}}} = -\frac{\left(\gamma^C + \frac{1}{\sqrt{A}}f_{t+1} - 2\right) \left(-\frac{1}{\sqrt{A}}\right)}{\left(\gamma^C + \frac{1}{\sqrt{A}}f_{t+1} - 2 - 2 + \gamma^C\right) \frac{1}{\sqrt{A}}} = \\ &= \frac{\left(\gamma^C + \frac{1}{\sqrt{A}}f_{t+1} - 2\right)}{\left(2\gamma^C + \frac{1}{\sqrt{A}}f_{t+1} - 4\right)} > 0, \end{aligned}$$

since by replacing  $f_{t+1}$  given by (B.12) we have:

$$\gamma^C + \frac{1}{\sqrt{A}}f_{t+1} - 2 = \gamma^C + \frac{\left(\eta_1 - 2\gamma^C + \frac{1}{2}(\gamma^C)^2\right)}{(2 - \gamma^C)} - 2 = \frac{1}{2}(\gamma^C - 2) + \frac{2\gamma_{\text{MAX}} - \frac{1}{2}\gamma_{\text{MAX}}^2 - 2}{2 - \gamma^C} < 0.$$

Next, we calculate:

$$\frac{\partial^2 f_{t+1}}{\partial^2 \phi_{t+1}} = \frac{\partial}{\partial \phi_{t+1}} \left( \frac{\left(\gamma^C + \frac{1}{\sqrt{A}}f_{t+1} - 2\right)}{\left(2\gamma^C + \frac{1}{\sqrt{A}}f_{t+1} - 4\right)} \right) = \frac{8(\eta_1 - 2)(\gamma^C - 2)}{\left(-12\gamma^C - 2\eta_1 + 3(\gamma^C)^2 + 16\right)^2} > 0.$$

These results imply that for  $\Phi_1 \leq \phi_{t+1} < \Phi_2$ , the optimal fee increases with managerial ability  $\phi_{t+1}$  and is convex. Let us determine next the optimal fee when  $\gamma^C = \bar{\gamma}$ .

We define  $X = \phi_{t+1} - f_{t+1}$  and solve the system:

$$\begin{aligned} \frac{1}{a}(1 + 2a - A^{1/2}) &= \bar{\gamma} \\ T + 2aX &= A, \end{aligned}$$

it results that

$$\begin{aligned} A &= (2a - a\bar{\gamma} + 1)^2, \\ X &= \frac{1}{2a}(A - T) = \frac{1}{2a}(a^2\bar{\gamma}^2 - 2a\bar{\gamma}(2a + 1) - a^2\gamma_{\text{MAX}}(\gamma_{\text{MAX}} - 4)) \\ &= \frac{1}{2}(a\bar{\gamma}^2 - 2\bar{\gamma}(2a + 1) + a\gamma_{\text{MAX}}(4 - \gamma_{\text{MAX}})). \end{aligned}$$

Note that for  $\gamma^C = \bar{\gamma}$  to be an optimum, we need to have:

$$\left(\eta_1 - 2\bar{\gamma} + \frac{1}{2}\bar{\gamma}^2\right) + f_{t+1}(-2 + \bar{\gamma}) \frac{1}{(2a - a\bar{\gamma}^2 + 1)} = 0.$$

Since  $\frac{\bar{\gamma}-2}{2a-a\bar{\gamma}^2+1} \neq 0$ , the optimal solution is:

$$f^* = \frac{1}{2-\bar{\gamma}}(2a - a\bar{\gamma} + 1) \left(\eta_1 - 2\bar{\gamma} + \frac{1}{2}\bar{\gamma}^2\right)$$

and it is realized when:

$$\phi^* = X + f^*. \tag{A2.13}$$

Let us denote by  $\Pi_C(f_{t+1}, \phi_{t+1})$  the profit of the manager when only current investors are in the fund:

$$\Pi_C(f_{t+1}, \phi_{t+1}) = f_{t+1} \left( m \left( \eta_1 - 2\gamma^C + \frac{1}{2}(\gamma^C)^2 \right) \right).$$

Then for any  $\phi_{t+1} > \phi^*$ , we should have potential entry of new investors. However, we have to see whether the fee that maximizes the manager's profit when he allows new investors to enter is not dominated by a fee determined in such a way that the manager caters only to current investors. Notice that as  $\phi_{t+1}$  increases above  $\phi^*$ , the manager could increase the fee one-to-one with  $\phi_{t+1}$  and no current investor would have incentives to leave.

When new investors enter the market, the first-order condition for this problem is:

$$2\gamma_{\text{MAX}} - 2\bar{\gamma} - \frac{1}{2}\gamma_{\text{MAX}}^2 + \frac{1}{2}\bar{\gamma}^2 + \frac{f_{t+1}}{\sqrt{U - 2a(f_{t+1} - \phi_{t+1})}} (\bar{\gamma} - 1) = 0 \Leftrightarrow$$

$$x + \frac{f_{t+1}}{\sqrt{U - 2a(f_{t+1} - \phi_{t+1})}} (\bar{\gamma} - 1) = 0,$$

where:

$$U = 1 + 2a + a^2 \left( (1 - \gamma_{\text{MAX}})^2 - \frac{2}{m} v_t \right)$$

$$x \equiv 2\gamma_{\text{MAX}} - 2\bar{\gamma} - \frac{1}{2}\gamma_{\text{MAX}}^2 + \frac{1}{2}\bar{\gamma}^2.$$

Solving for  $f_{t+1}$ , we obtain that:

$$f_{t+1}(\phi_{t+1}) = x \frac{\pm \sqrt{U + 2a\phi_{t+1} + U\bar{\gamma}^2 + x^2a^2 - 2U\bar{\gamma} - 4a\bar{\gamma}\phi_{t+1} + 2a\bar{\gamma}^2\phi_{t+1} - xa}}{(\bar{\gamma} - 1)^2}.$$

However, since:

$$-x \frac{\sqrt{U + 2a\phi_{t+1} + U\bar{\gamma}^2 + x^2a^2 - 2U\bar{\gamma} - 4a\bar{\gamma}\phi_{t+1} + 2a\bar{\gamma}^2\phi_{t+1} + xa}}{(\bar{\gamma} - 1)^2} < 0,$$

the only feasible solution is:

$$f^{**}(\phi_{t+1}) = x \frac{\sqrt{U + 2a\phi_{t+1} + U\bar{\gamma}^2 + x^2a^2 - 2U\bar{\gamma} - 4a\bar{\gamma}\phi_{t+1} + 2a\bar{\gamma}^2\phi_{t+1} - xa}}{(\bar{\gamma} - 1)^2}. \tag{A2.14}$$

We need to check next that charging  $f^* = (\phi_{t+1} - X)$  and catering only to current investors does not dominate charging  $f^{**}(\phi_{t+1})$  and catering also to new investors. We find  $\phi^{**}$  such that  $\phi^{**} - X = f^{**}(\phi^{**})$ . When  $\phi_{t+1} = \phi^{**}$ ,  $\gamma^C = \gamma^N = \bar{\gamma}$ , so the manager is indifferent between charging  $f^*$  or  $f^{**}$  because the quantity is the same  $\bar{\gamma}$ . When  $\phi^* \leq \phi_{t+1} < \phi^{**}$  then  $f^* > f^{**}$  and this implies  $\Pi(f^*, \bar{\gamma}) > \Pi(f^{**}, \bar{\gamma})$ .<sup>10</sup>

Let us define  $\phi^{**}$  the solution of:

$$\phi_{t+1} - X = f^{**}(\phi_{t+1}). \tag{A2.15}$$

Using the solution in (B.14) we have that:

$$(\phi_{t+1} - X) = \frac{x}{(\bar{\gamma} - 1)^2} \left( \sqrt{U + 2a\phi_{t+1} + U\bar{\gamma}^2 + x^2a^2 - 2U\bar{\gamma} - 4a\bar{\gamma}\phi_{t+1} + 2a\bar{\gamma}^2\phi_{t+1} - xa} \right)$$

$$0 = \bar{\gamma}^2\phi_{t+1}^2 - 2X\phi_{t+1} + \phi_{t+1}^2 - x^2U - 2X^2\bar{\gamma} - 2\bar{\gamma}\phi_{t+1}^2 + X^2\bar{\gamma}^2 + X^2 + 4X\bar{\gamma}\phi_{t+1} - 2x^2Xa - 2X\bar{\gamma}^2\phi_{t+1}$$

We define:

$$H(\phi_{t+1}, \bar{\gamma}) = \bar{\gamma}^2\phi_{t+1}^2 - 2X\phi_{t+1} + \phi_{t+1}^2 - x^2U - 2X^2\bar{\gamma} - 2\bar{\gamma}\phi_{t+1}^2 + X^2\bar{\gamma}^2 + X^2 + 4X\bar{\gamma}\phi_{t+1} - 2x^2Xa - 2X\bar{\gamma}^2\phi_{t+1}.$$

Note that when  $\bar{\gamma} = 1$ ,

$$H(\phi_{t+1}, 1) = -x^2(U + 2Xa) < 0,$$

so always  $f^{**} < f^*$ .

When  $\bar{\gamma} \neq 1$ , the solution for equation (B.15) is:

$$\phi^{**} = X \pm x \frac{\sqrt{U + 2Xa}}{1 - \bar{\gamma}}.$$

Note that since  $\phi^* < \phi^{**}$ , and  $\phi^* = X + f^* \geq X$ , only one solution is feasible. Thus,

$$\phi^{**} = X + x \frac{\sqrt{U + 2Xa}}{1 - \bar{\gamma}} \quad \text{if } \bar{\gamma} < 1$$

$$\phi^{**} = X - x \frac{\sqrt{U + 2Xa}}{1 - \bar{\gamma}} \quad \text{if } \bar{\gamma} > 1.$$

<sup>10</sup> Note that both  $\phi_{t+1} - X$  and  $f^{**}(\phi_{t+1})$  are increasing in  $\phi_{t+1}$ .

Since we have:

$$H(X, \bar{\gamma}) = -\frac{1}{4}(\gamma_{\text{MAX}} - \bar{\gamma})^2(\gamma_{\text{MAX}} + \bar{\gamma} - 4)^2(U + 2Xa) < 0,$$

it implies that  $H(\phi_{t+1}, \bar{\gamma}) \leq 0$  for all  $\phi^* \leq X \leq \phi_{t+1} < \phi^{**}$ , and this is equivalent to  $f^{**} < f^*$  for all  $\phi^* \leq X \leq \phi_{t+1} < \phi^{**}$ .

Next, note that if  $\gamma^N(f^{**})$  is the reservation return of the last new investor who wants to enter the fund, we have that  $\gamma^N(f^{**}) \leq \bar{\gamma}$ . This implies that  $H(\phi_{t+1}, \bar{\gamma}) \leq 0$  for all  $\phi^* \leq X \leq \phi_{t+1} < \phi^{**}$  and this is equivalent to  $f^{**} < f^*$  for all  $\phi^* \leq X \leq \phi_{t+1} < \phi^{**}$ .

Note that for  $\phi^* \leq X \leq \phi_{t+1} < \phi^{**}$ , the manager's profit is:

$$\Pi_0(f_{t+1}, \phi_{t+1}) = \Pi_c(f_{t+1}, \phi_{t+1}) \Big|_{\gamma=\bar{\gamma}} = f_{t+1} \left( (\eta_1 - \bar{\gamma} + \frac{1}{2}(\bar{\gamma})^2) \right),$$

and the optimum fee is  $f_0 = \phi_{t+1} - X$ , so the fee increases one-to-one with performance.

Next, note that if  $\gamma^N(f^{**})$  is the reservation return of the last new investor who wants to enter the fund, we have that  $\gamma^N(f^{**}) \leq \bar{\gamma}$ .

We define the manager's profit as:

$$\Pi_N(f_{t+1}, \phi_{t+1}) = f_{t+1} \left( (\eta_2 - \gamma^N + \frac{1}{2}(\gamma^N)^2) \right).$$

The derivative of the manager's profit with respect to the fee  $f_{t+1}$  is:

$$\frac{\partial \Pi_N(f_{t+1}, \phi_{t+1})}{\partial f_{t+1}} = (\eta_2 - \gamma^N + \frac{1}{2}(\gamma^N)^2) + f_{t+1}(-1 + \gamma^N) \frac{1}{\sqrt{U + 2a(\phi_{t+1} - f_{t+1})}}. \quad (\text{A2.16})$$

Note that for  $\phi^* \leq X \leq \phi_{t+1} < \phi^{**}$ , we have that:

$$\frac{\partial \Pi_0(f_{t+1}, \phi_{t+1})}{\partial f_{t+1}} = 1 > 0.$$

Let us calculate:

$$\begin{aligned} \frac{\partial \Pi_N(f_{t+1}, \phi_{t+1})}{\partial f_{t+1}} \Big|_{\phi_{t+1}=\phi^{**}+\delta} &= (\eta_2 - \gamma^N + \frac{1}{2}(\gamma^N)^2) + f^{**}(-1 + \gamma^N) \frac{1}{\sqrt{U + 2a(\phi_{t+1} - f^{**})}} \Big|_{\phi_{t+1}=\phi^{**}+\delta} \\ &= (\eta_2 - \gamma^N + \frac{1}{2}(\gamma^N)^2) + \frac{f^{**}(-1 + \gamma^N)}{\sqrt{U + 2a(\phi^{**} + \delta - f^{**})}} \\ &= (\eta_2 - \gamma^N + \frac{1}{2}(\gamma^N)^2) + \frac{f^{**}(-1 + \gamma^N)}{\sqrt{U + 2a(X + \delta)}}, \end{aligned}$$

where  $\gamma^N = \frac{1}{a}(1 + a - \sqrt{U + 2a(X + \delta)})$  does not depend on  $f^{**}$  and  $\delta > 0$  is relatively small.

This implies that whenever  $\gamma^N - 1 < 0$ , we have that locally the derivative of the profit  $\frac{\partial \Pi_N(f_{t+1}, \phi_{t+1})}{\partial f_{t+1}} < 0$ , and therefore by decreasing the fee  $f^{**}$  the manager can increase his profit. As a result, the optimal fee may decrease when the managerial ability is such that new investors enter the fund.

## References

- Basak, S., Pavlova, A., 2013. Asset prices and institutional investors. *Am. Econ. Rev.* 103 (5), 1728–1758.
- Berk, J., Green, R., 2004. Mutual fund flows and performance in rational markets. *J. Polit. Econ.* 112 (6), 1269–1295.
- Berk, J., Tonks, I., 2007. Return Persistence and Fund Flows in the Worst Performing Mutual Funds. NBER Working Paper no. 13042.
- Brown, S.J., Veld, C., Veld-Merkoulova, Y.V., 2017. Why do individuals not participate in the stock market? In: Working Paper, 29th Australasian Finance and Banking Conference 2016 available at SSRN: <https://ssrn.com/abstract=2822094>.
- Calvet, L.E., Campbell, J.Y., Sodini, P., 2009. Measuring the financial sophistication of households. *Am. Econ. Rev.* 99 (2), 393–398.
- Carhart, M., 1997. On persistence in mutual fund performance. *J. Finance* 52 (1), 57–82.
- Cremers, K.M., Petajisto, A., 2009. How active is your fund manager? A new measure that predicts performance. *Rev. Financ. Stud.* 22 (9), 3329–3365.
- Cuoco, D., Kaniel, R., 2011. Equilibrium prices in the presence of delegated portfolio management. *J. Financ. Econ.* 101 (2), 264–296.
- Del Guercio, D., Reuter, J., 2014. Mutual fund performance and the incentive to generate alpha. *J. Finance* 69 (4), 1673–1704.
- Del Guercio, D., Tkac, P., 2002. The determinants of the flow of funds of managed portfolios: mutual funds vs. pension funds. *J. Financ. Quant. Anal.* 37 (4), 523–558.
- Elton, E., Gruber, M., Blake, C., 2011. Does Size Matter? the Relationship between Size and Performance. Fordham University School of Business Research Paper No. 1826406, available at SSRN: <https://ssrn.com/abstract=1826406>.
- Evans, R., 2010. Mutual fund incubation. *J. Finance* 65 (4), 1581–1611.

- Evans, R.B., Fahlenbrach, R., 2012. Institutional investors and mutual fund governance: evidence from retail–institutional fund twins. *Rev. Financ. Stud.* 25 (12), 3530–3571.
- Garleanu, N.B., Pedersen, L.H., 2015. Efficiently Inefficient Markets for Assets and Asset Management. NBER Working Paper No. 21563.
- Gennaioli, N., Shleifer, A., Vishny, R., 2015. Money doctors. *J. Finance* 70 (1), 91–114.
- Gil-Bazo, J., Ruiz-Verdú, P., 2008. When cheaper is better: Fee determination in the market for equity mutual funds. *J. Econ. Behav. Org.* 67 (3), 871–885.
- Glode, V., 2011. Why mutual funds underperform. *J. Financ. Econ.* 99 (3), 546–559.
- Glode, V., Hollifield, B., Kacperczyk, M., Kogan, S., 2017. Is investor rationality time varying? Evidence from the mutual fund industry. In: *Behavioral Finance: where Do Investors' Biases Come from?*. World Scientific, pp. 67–113.
- He, Z., Krishnamurthy, A., 2013. Intermediary asset pricing. *Am. Econ. Rev.* 103 (2), 732–770.
- Huang, J., Wei, K., Yan, H., 2007. Participation costs and the sensitivity of fund flows to past performance. *J. Finance* 62 (3), 1273–1311.
- James, C., Karceski, J., 2006. Investor monitoring and differences in mutual fund performance. *J. Bank. Finance* 30 (10), 2787–2808.
- Kacperczyk, M., Van Nieuwerburgh, S., Veldkamp, L., 2014. Time-varying fund manager skill. *J. Finance* 69 (4), 1455–1484.
- Kaniel, R., Kondor, P., 2012. The delegated Lucas tree. *Rev. Financ. Stud.* 26 (4), 929–984.
- Kosowski, R., 2011. Do mutual funds perform when it matters most to investors? US mutual fund performance and risk in recessions and expansions. *Q. J. Finance* 1 (03), 607–664.
- Lusardi, A., Mitchell, O.S., 2014. The economic importance of financial literacy: Theory and evidence. *J. Econ. Lit.* 52 (1), 5–44.
- Metrick, A., Zeckhauser, R., 1998. Price versus quantity: Market-clearing mechanisms when consumers are uncertain about quality. *J. Risk Uncertain.* 17 (3), 215–243.
- Müller, S., Weber, M., 2010. Financial literacy and mutual fund investments: who buys actively managed funds? *Schmalenbach Bus. Rev.* 62, 126–153.
- Nanda, V., Narayanan, M., Warther, V.A., 2000. Liquidity, investment ability, and mutual fund structure. *J. Financ. Econ.* 57 (3), 417–443.
- Navone, M., 2012. Investors distraction and strategic repricing decisions. *J. Bank. Finance* 36 (5), 1291–1303.
- Sato, Y., 2016. Delegated portfolio management, optimal fee contracts, and asset prices. *J. Econ. Theor.* 165, 360–389.
- Sharpe, W.F., 1966. Mutual fund performance. *J. Bus.* 39 (1), 119–138.
- Sirri, E., Tufano, P., 1998. Costly search and mutual fund flows. *J. Finance* 53 (5), 1589–1622.
- Zingales, L., 2009. The future of securities regulation. *J. Account. Res.* 47 (2), 391–425.