Market Frictions, Investor Heterogeneity, and Persistence in Mutual Fund Performance[*]

Ariadna Dumitrescu

ESADE Business School and University Ramon Llull

Javier Gil-Bazo

University Pompeu Fabra and Barcelona GSE

Abstract

If there are diseconomies of scale in asset management, any predictability in mutual fund performance will be arbitraged away by rational investors seeking funds with the highest expected performance (Berk and Green, 2004). In contrast, the performance of equity mutual funds persists through time. In this paper, we show how market frictions can reconcile the assumptions of investor rationality and diseconomies of scale with the empirical evidence. More specifically, we extend the model of Berk and Green (2004) to account for financial constraints and heterogeneity in investors' reservation returns reflecting the idea that less financially sophisticated investors face higher search costs. In our model, both negative and positive expected fund performance are possible in equilibrium. The model also predicts that expected fund performance increases with managerial ability and explains why predictable differences in performance across funds are more prevalent in markets populated by less sophisticated investors.

JEL codes: G2; G23.

Keywords: mutual fund performance persistence; market frictions; investor sophistication.

---

1   Introduction

Like investors in other retail financial markets, mutual fund investors face non-negligible search costs, entry costs, and switching costs, and are likely to be financially constrained. While the role of market frictions on investor choices has received some attention in the mutual fund literature, the implications of frictions for the determination of mutual fund performance are still not well understood. In this paper, we investigate how market frictions impact investors' investment and disinvestment decisions and the determination of mutual fund performance in equilibrium.

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 performance 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 zero. In equilibrium, all funds deliver zero net expected performance. Therefore, fund performance is not predictable from fund characteristics or past performance.

BG's influential work has changed the prevalent view on mutual fund performance persistence by showing that lack of predictability in mutual fund performance is consistent with a market populated by competing rational investors, even if fund managers possess skill. However, there exists abundant empirical evidence that underperforming US equity funds continue to underperform in the long term (e.g., Carhart, 1997). The model cannot explain, either, why performance persists for winners in the short term (Bollen and Busse, 2005). Ferreira et al. (2013) show that fund performance persistence is a widespread phenomenom throughout the world. Under the framework of BG, such persistence in mutual fund performance is an anomaly that needs to be explained.

One possible explanation for the discrepancy between the model's implication of performance unpredictability and the empirical evidence on performance persistence is that the assumption of diseconomies of scale in asset management is not a good characterization of the mutual fund industry. However, the available empirical evidence suggests that US equity fund performance

decreases with size. Chen et al. (2004) show that, conditional on other fund characteristics, performance decreases with lagged assets under management, especially for funds investing in small-cap growth stocks, suggesting that liquidity is a source of diseconomies of scale portfolio management. Yan et al. (2008) confirm these findings using more direct measures of portfolio liquidity.

An alternative explanation is that market frictions such as search costs, switching costs, and liquidity constraints, distort investor decisions and affect mutual fund equilibrium performance. Understanding the effects of frictions on the determination of equilibrium in the mutual fund market is precisely the purpose of our study. More specifically, we develop a model of performance determination that retains the key features of the model of BG, namely diseconomies of scale and competition among investors, but extends it in several directions. First, we assume that investors' reservation risk-adjusted returns are negative, not zero, for many investors. The idea that mutual fund investors have negative reservation risk-adjusted returns is indeed consistent with the abundant empirical evidence that the average actively managed equity fund under-performs passive benchmarks after fees and trading costs. Negative reservation risk-adjusted returns can arise as a consequence of search costs. For instance, BG assume that the investment alternative to actively managed funds is an index fund. In the presence of search costs, the risk-adjusted return of investing in the index fund net of search costs is negative. Consistently with this view, Hortaçsu and Syverson (2004) attribute the large dispersion of fees across index mutual funds tracking the same index to search costs. Since search costs are likely to vary across investors due to heterogeneity in financial sophistication, in our model we assume that reservation returns are lower for unsophisticated investors. We believe that regulatory attempts to make low-cost passively managed funds a mandatory default option give further credence to the notion that for a substantial fraction of investors, finding a cheap passive alternatively to actively managed funds is not costless.

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. Moreover, investors face the risk of a liquidity shock, which would prevent them from investing in a mutual fund. We assume

that this risk is higher for unsophisticated investors.

Like in the model of BG, each period investors must choose between an actively managed fund and an index fund, an alternative investment opportunity available to all investors with the same risk as the managed portfolio. We assume that while the fund's current investors can reinvest their last period's wealth as well as invest their current endowment in the fund, new investors can only invest their current period's endowment.

If investors were not financially constrained, any fund's expected risk-adjusted net return would be equal in equilibrium to the reservation return of the most unsophisticated investor in the market. Otherwise, there would be excess demand by the most unsophisticated investors for any fund with a higher level of performance. An increase in expected managerial ability would not lead to an increase in the fund's net performance, it would simply result in more flows from unsophisticated investors. However, when there is a limit on the amount of money each investor can invest, inflows from the least sophisticated investors do not drive fund performance down to their reservation return, so the fund can still attract more sophisticated investors. In this different setup, more sophisticated investors decide to invest in the fund as long as the fund's expected performance exceeds their reservation return. In equilibrium, any 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 risk-adjusted returns in the market. As managerial ability increases, the fund's equilibrium expected performance increases and the fund attracts more sophisticated investors. A fund can offer a positive expected risk-adjusted net return provided that investors' inflows are not sufficient to drive the fund's performance down to zero. In sum, in our model both negative and positive expected fund performance are possible in equilibrium.[1] Moreover, expected fund performance increases with managerial ability. To the extent that managerial ability is persistent through time, so is fund performance. Therefore, heterogeneity in investors' reservation returns together with financial constraints can rationalize

---

[1]Note, however, that if positive managerial ability is scarce, positive expected performance, although compatible with equilibrium, will be rarely observed in the data.

the evidence on fund performance persistence.

The model developed in this paper also predicts that differences in performance between high- and low-ability managers increase with the amount of unsophisticated investors in the market. The intuition for this result is as follows. If managerial ability is low, a fund can operate only if the investors that populate the market are unsophisticated enough. Moreover, other things equal, a fund captures more assets if there are more unsophisticated investors in the market, which in turn, hampers its performance. Further, the performance of the fund improves faster with managerial ability in a market with more unsophisticated investors. The reason is that the fund's current investors are less sophisticated and have less money to invest, so their decision to enter the fund as managerial ability improves is less harmful to fund performance. Also, current investors require a lower expected risk-adjusted return in order to decide to reinvest with a fund in a less sophisticated market, so it takes a lower level of managerial ability for all current investors to reinvest with the fund. Once all current investors have decided to reinvest, new investors may enter the fund, but since new investors only have their current endowment to invest, the effect of their entry on fund performance is limited. In sum, other things equal, the expected performance of a fund in a market with less sophisticated investors is lower, but rises faster with managerial ability. Holding the cross-sectional distribution of managerial ability constant, the model predicts that performance differences will be more likely to survive in markets with less sophisticated investors.

The effect of market frictions on investor decisions has been previously investigated in the mutual fund literature in the context of studies of mutual fund flows. Sirri and Tufano (1998) are the first to show that search costs affect investor decisions. In particular, they find that the flow-performance relation is less steep for funds associated with higher search costs. Huang et al. (2007) propose a model in which search costs combined with Bayesian learning from past returns lead investors to consider only funds with the highest recent performance since the costs of researching a new fund with less than top recent performance outweigh the expected benefits. More recently, Navone (2012) shows that the sensitivity of flows to past performance decreases with past performance but increases with different proxies for fund visibility.

Our paper is related to the empirical study of Glode et al. (2011), who investigate time variation in performance persistence. Glode et al. (2011) find evidence of more persistence after up-markets, which they attribute to a larger presence of unsophisticated investors in the market. Our model provides a precise mechanism through which the entry of less sophisticated investors in the market results in more persistence in performance.

The rest of the paper is organized as follows. In section 2, we describe the model's setup. Section 3 characterizes the equilibrium and discusses the model's predictions. Section 4 concludes. The Appendix contains all the proofs.

2   The model

BG consider a fund that can generate returns in excess of a passive benchmark due to its manager's ability. Let R_t denote the fund's return in excess of a passive benchmark before fees and expenses, $R_t = \alpha + \varepsilon_t$, where α reflects managerial ability and ε_t is an idiosyncratic shock that is normally distributed with mean 0 and variance σ^2. For simplicity, throughout the paper we refer to risk-adjusted return as return. Managerial ability, α, is not known to managers or investors, who estimate it using the information contained in past returns.

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: $C(0) = 0$, $\lim_{q\to\infty} C'(q) = \infty$ 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 $r_t \equiv R_t - \dfrac{C(q_t)}{q_t} - f$, where q_t is the $t-1$ investment in the fund and f is the fund's fee, which is exogenously given.[2] If the revenues

---

[2] It can be shown that when the fund's fee is endogenous, it increases with managerial skill although less than proportionally, so our results are not substantially changed if we do comparative statics with respect to managerial skill holding the fee constant or with respect to managerial skill net of the endogenous fee.

collected by the manager at time t, fq_t, cover the fixed costs of the fund, the fund continues its activity, otherwise the fund closes down. We assume for simplicity and without loss of generality that fixed costs are zero.

We depart from BG in that the fund's potential investors have limited funds to invest and exhibit different degrees of financial sophistication. To model different degrees of sophistication we allow for reservation returns to vary across investors. Like BG, we assume that each investor i has a specific search cost γ_i that reflects her ability to find an alternative fund. For simplicity, we assume that the alternative for all investors is an index fund with zero expected risk-adjusted return. Net of search costs, the reservation expected risk-adjusted return (henceforth reservation return) of the $i - th$ investor is $-\gamma_i$. Therefore, unlike in the model of BG, the investor's reservation return is different from zero and is also different across investors. Note that γ_i is the search cost for investor i, per dollar invested. If search costs are fixed, a wealthier investor faces a lower γ. 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 γ uniformly distributed over the interval $\left[0, \gamma^{MAX}\right]$, with $\gamma^{MAX} \leq 1$. Therefore, we assume that all investors in the market have negative reservation returns net of search costs. Alternatively we could allow some investors to have positive reservation returns without altering the conclusions. The parameter γ^{MAX} determines the overall level of sophistication in the fund in which the market is offered.

Finally, we also allow for the possibility that new investors who enter the fund at date t must pay an entry cost K, reflecting either an explicit fee charged by the fund (front-end loads) or other costs (e.g., switching costs).

The timing of the events is the following:

Date $t - 1$:

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

Date t:

7

- The fund's return at date t is realized and current investors obtain its net return.

- 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 wealth of $m(1 - \gamma_i)$. This assumption captures the idea that less sophisticated investors face more severe financial constraints. For instance, investor i could be exposed at time t to the possibility of a liquidity shock with probability γ_i.

Date $t + 1$:

- The fund's return at date $t + 1$ is realized and the fund's investors obtain its net return.

3 Equilibrium

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 $\{R_s\}_{s=1}^{s=t}$ and update their beliefs about the fund manager's ability through Bayesian updating:

$$\phi_{t+1} = E\left(R_{t+1} \mid R_1, ..., R_t\right).$$

Investor i demands shares of the fund if the fund's expected return net of trading costs and fees (Total Performance, TP) exceeds her reservation return $-\gamma_i$. The fund's expected net return in period t equals

$$
\begin{aligned}
TP_{t+1}\left(q_{t+1}\right) &= E\left[r_{t+1} | R_1, ..., R_t\right] \\
&= E\left[R_{t+1} - \frac{C\left(q_{t+1}\right)}{q_{t+1}} - f \middle| R_1, ..., R_t\right].
\end{aligned}
$$

A current investor will either withdraw her date $t - 1$ 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.

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 $TP_{t+1}\left(q_{t+1}^*\right) = \phi_{t+1} - \dfrac{C\left(q_{t+1}^*\right)}{q_{t+1}^*} - f.$

- All investors who withdraw their money from the fund have reservation returns higher than $TP_{t+1}\left(q_{t+1}^*\right).$

- All investors who invest new money in the fund have reservation returns less than or equal to $TP_{t+1}\left(q_{t+1}^*\right).$

- The equilibrium amount of assets q_{t+1}^* is such that $0 \leq q_{t+1}^* \leq v_t + M$, where $v_t \equiv m\left(\gamma_{MAX} - \overline{\gamma}\right)$ is the value at t of current investors' investment at $t - 1$ and M denotes the maximum inflow possible in this period: $m\left(\gamma_{MAX} - \frac{1}{2}\gamma_{MAX}^2\right).$

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

$$
\begin{aligned}
TP_{t+1}\left(q_{t+1}^C\right) &= -\gamma^C, \\
q_{t+1}^C &= 2m\left(\gamma_{MAX} - \gamma^C\right) - \frac{m}{2}(\gamma_{MAX}^2 - (\gamma^C)^2).
\end{aligned}
$$

Depending on the value of the solution γ^C, there are three possible alternatives:

Case 1: $\gamma_{MAX} \leq \gamma^C$. Even if all existing 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: $\overline{\gamma} \leq \gamma^C < \gamma_{MAX}$. Current investors with reservation returns higher than $-\gamma^C$ exit the fund and those with reservation returns lower than $-\gamma^C$ reinvest with the fund. The fund's

9

expected net return equals $E\left(r_{t+1}\right) = -\gamma^C < 0$ and the fund's assets $q_{t+1}^* = q_{t+1}^C$.

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

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

- The fund's expected return equals $TP_{t+1}\left(q_{t+1}^*\right)$.

- New investors who invest in the fund have reservation returns less than or equal to $TP_{t+1}\left(q_{t+1}^*\right) - K$.

- New investors who decide not to invest in the fund have reservation returns higher than $TP_{t+1}\left(q_{t+1}^*\right) - K$.

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:

$$
\begin{aligned}
TP_{t+1}\left(q_{t+1}^N\right) - K &= -\gamma^N, \\
q_{t+1}^N &= v_t + m\left((\gamma_{MAX} - \gamma^N) - \frac{1}{2}(\gamma_{MAX}^2 - \left(\gamma^N\right)^2)\right).
\end{aligned}
$$

We now distinguish two cases depending on whether the solution γ^N is higher or smaller than $\overline{\gamma}$. When $\gamma^N \geq \overline{\gamma}$, no new investors want to enter the fund. Even if only current investors reinvested with the fund, the fund's performance would not be enough to convince investors to pay the entry cost. As a result, only current investors invest in the fund and the amount invested in the fund at $t+1$ is $q_{t+1}^* = 2v_t - \frac{m}{2}(\gamma_{MAX}^2 - \overline{\gamma}^2) \equiv \overline{q}_{t+1}$. The expected return in this case is $E\left(r_{t+1}\right) = TP_{t+1}\left(\overline{q}_{t+1}\right)$.

On the other hand, when $\gamma^N < \overline{\gamma}$, new investors enter the fund. The last investor i to enter the fund in the period t has $\gamma_i = \gamma^N$, and the quantity invested in the fund is $q_{t+1}^* =$

$v_t + m\left(\left(\gamma_{MAX} - \gamma^N\right) - \frac{1}{2}(\gamma_{MAX}^2 - \left(\gamma^N\right)^2)\right)$. If $\gamma^N < 0$, then all potential investors enter the fund and the quantity invested in the fund is $q_{t+1}^* = v_t + m\gamma_{MAX}\left(1 - \frac{\gamma_{MAX}}{2}\right) = v_t + M$. Consequently, the fund's expected net return is $E\left(r_{t+1}\right) = K - \gamma^N$, if $\overline{\gamma} > \gamma^N > 0$ and $E\left(r_{t+1}\right) = TP_{t+1}\left(v_t + M\right)$, if $\gamma^N \leq 0$.

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

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

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

where Φ_j, $j = \overline{1,3}$ are defined in the Appendix and γ^C, γ^N, equal:

$$\begin{aligned} \gamma^C &= \frac{1}{cm}\left(1 + 2cm - A^{1/2}\right), \quad where \\ A &\equiv 1 + 2cm\left(2 + \phi - f\right) + c^2 m^2\left(2 - \gamma_{MAX}\right)^2 \end{aligned}$$

and

$$\begin{aligned} \gamma^N &= \frac{1}{cm}\left(1 + cm - B^{1/2}\right), \quad where \\ B &\equiv 1 + 2cm\left(1 + \phi - f - K\right) + c^2 m^2\left(\left(1 - \gamma_{MAX}\right)^2 - \frac{2}{m}v_t\right), \end{aligned}$$

respectively.

Figure 1 shows graphically the fund's expected net return as a function of expected managerial ability holding the fund's fee constant and assuming that there are no entry costs for new investors $\left(K = 0\right)$. 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
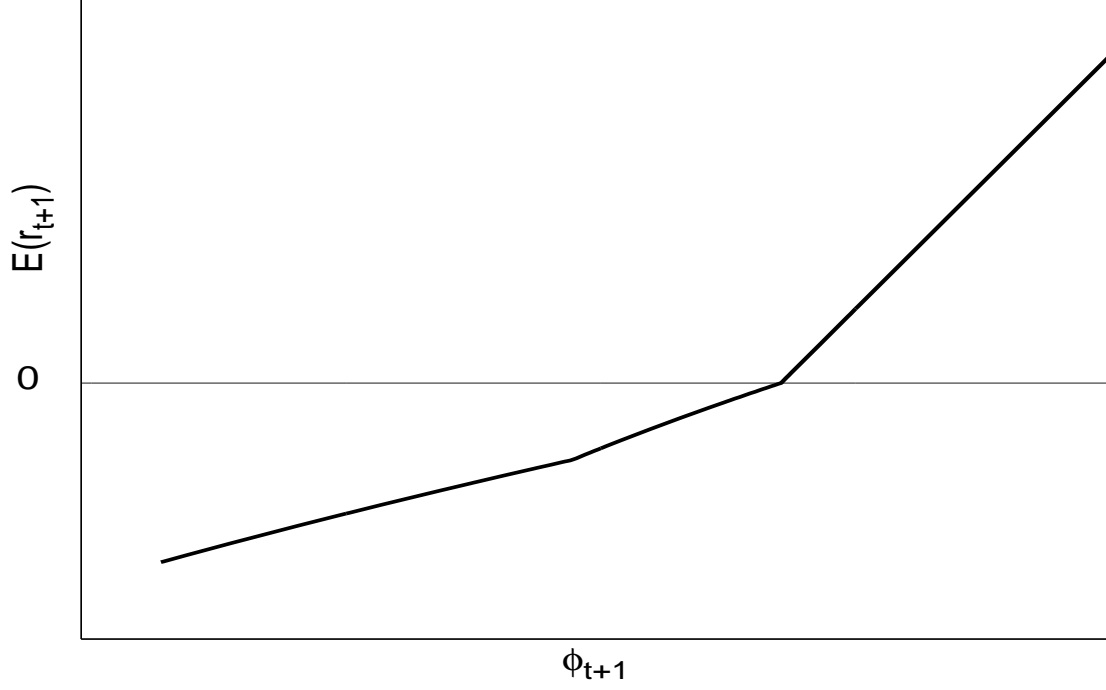
Figure 1: Expected net return as a function of expected managerial skill. Parameter values: $m = 200$, $c = 0.01$, $K = 0$, $\gamma_{MAX} = 0.7$, $f = 0.01$.

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 increase less rapidly with increases in managerial ability, so the fund's expected return increases faster. Once all potential investors are in the fund, fund performance increases one-to-one with managerial skill.

Proposition 1 shows that the fund's expected net return in equilibrium can be different from zero. Equilibrium expected net returns 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. Positive equilibrium expected net returns can be obtained when managerial ability increases and either entry costs prevent new investors from entering the fund and eroding funds' performance or all potential investors have invested with

12

the fund. Therefore, the interaction of financial constraints and negative reservation returns prevents investors' money from flowing freely into and out of the fund and eliminating non-zero performance.

Note that in order to observe dispersion in expected performance in the data, other than that induced by differences in γ_{MAX} or $\overline{\gamma}$, we need to have dispersion in managerial ability. If managerial ability persists through time, then fund performance also persists.

Note that the necessary conditions to obtain expected net returns different form zero are: heterogeneity of investors' reservation returns and limited capital to invest. Investor heterogeneity ensures that in equilibrium we have expected returns different from zero, but also different expected returns for different levels of managerial ability. 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 in the market. If investors were 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, it would simply attract more flows from unsophisticated investors.

As we can see from Proposition 1, expected net return of a given fund depends on the level of sophistication of the investors in the market for that fund, which is given by γ_{MAX}. Notice that both γ^C and γ^N increase with γ_{MAX} and this is due to the fact that when the potential investors are less sophisticated, there is a larger amount available for reinvestment in the fund, and therefore, the fund performance is eroded faster by money inflows. As a result, if investors are more sophisticated, the fund may earn a higher expected net return in equilibrium. However, this does not guarantee that an increase in sophistication always increases expected performance. To see this, let us consider two otherwise identical funds being offered in two different markets, each one corresponding to a different value of γ_{MAX}. Henceforth, we refer to the first market as the unsophisticated market (U), with $\gamma_{MAX} = \gamma_{MAX}^U$, and to the second market as the sophisticated market (S), with $\gamma_{MAX} = \gamma_{MAX}^S$, and $\gamma_{MAX}^U > \gamma_{MAX}^S$. We assume that the total amount currently invested in both cases is the same, v_t. We denote by Φ_j^U, Φ_j^S the cut-off points

13

for the unsophisticated and sophisticated markets, respectively.

Proposition 2 *There exist K_1 and K_2 as defined in the Appendix such that:*

1. If $K < K_1$, then $E^S\left(r_{t+1}\left(\phi_{t+1}\right)\right) > E^U\left(r_{t+1}\left(\phi_{t+1}\right)\right)$, for any ϕ_{t+1}.

2. If $K \in [K_1, K_2]$ then there exist $\phi_1 \in \left(\Phi_2^U, \Phi_2^U + K\right)$ and $\phi_2 \in \left(\Phi_2^S, \Phi_2^S + K\right), \phi_2 > \Phi_2^U + K$ such that $E^S\left(r_{t+1}\left(\phi_j\right)\right) = E^U\left(r_{t+1}\left(\phi_j\right)\right), j = 1, 2$. Then, for any $\phi_{t+1} < \phi_1$ and $\phi_{t+1} > \phi_2$, we have that $E^S\left(r_{t+1}\left(\phi_{t+1}\right)\right) > E^U\left(r_{t+1}\left(\phi_{t+1}\right)\right)$ and for $\phi_{t+1} \in (\phi_1, \phi_2)$, $E^S\left(r_{t+1}\left(\phi_{t+1}\right)\right) > E^U\left(r_{t+1}\left(\phi_{t+1}\right)\right)$.

3. If $K > K_2$, then there exists $\phi_1 \in \left(\Phi_2^U, \Phi_2^U + K\right)$ such that $E^S\left(r_{t+1}\left(\phi_1\right)\right) = E^U\left(r_{t+1}\left(\phi_1\right)\right)$. Then, for any $\phi_{t+1} < \phi_1$, $E^S\left(r_{t+1}\left(\phi_{t+1}\right)\right) > E^U\left(r_{t+1}\left(\phi_{t+1}\right)\right)$ and for $\phi_{t+1} > \phi_1$, $E^S\left(r_{t+1}\left(\phi_{t+1}\right)\right) < E^U\left(r_{t+1}\left(\phi_{t+1}\right)\right)$.

Proposition 2 characterizes the conditions under which a fund in the unsophisticated market underperforms a fund in the sophisticated market. When entry costs are small, $K < K_1$ (see Figure 2 for the case $K = 0$), the fund in the unsophisticated market underperforms an otherwise identical fund in a market with more sophisticated investors for any level of managerial ability. A fund targeted to more unsophisticated investors captures more investors for any level of managerial ability, which reduces its performance. As can be seen in Figure 2, the performance gap between funds targeted to sophisticated investors and funds targeted to unsophisticated investors narrows as managerial ability increases. This is because it takes a low level of managerial ability for all current investors of the latter to decide to reinvest with the fund: They have lower reservation returns and, because they have less money to invest (on average), their decision to reinvest is not as harmful for fund performance. Once all current investors have decided to reinvest, new investors enter the fund but entry of new investors has a less detrimental effect on fund performance than reinvestment by current investors. Therefore, for low entry costs, differences in expected performance between both funds are more apparent in the lower end of managerial ability.

Figure 2 suggests that, holding the distribution of managerial ability constant, there will be more cross-sectional dispersion in fund performance as investor sophistication decreases.
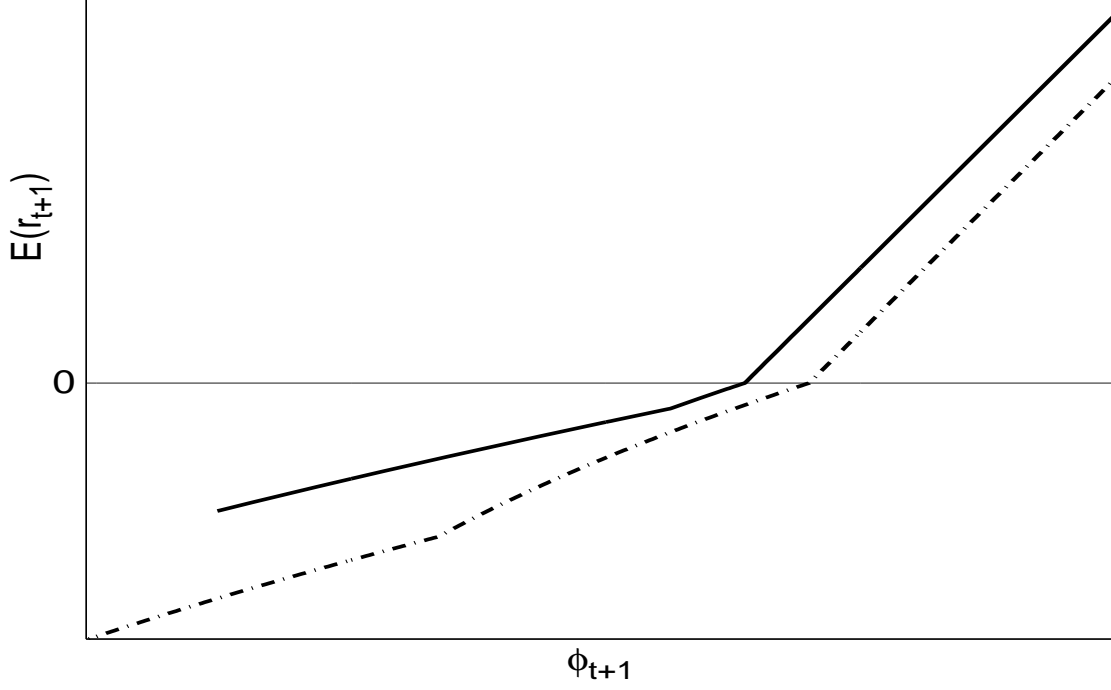
14

Figure 2: Expected net return as a function of expected managerial skill and market sophistication, with no entry costs. The solid (dotted) line corresponds to the sophisticated (unsophisticated) market, i.e., low (high) γ_{MAX}. Parameter values: $m = 200$, $c = 0.01$, $K = 0$, $\gamma_{MAX}^U = 1$, $\gamma_{MAX}^S = 0.5$, $f = 0.01$.

Figure 3 shows the expected performance of both types of funds when $K \in [K_1, K_2]$. In this case, there exists an interval, (ϕ_1, ϕ_2) in which the fund in the unsophisticated market outperforms the fund in the sophisticated market. When all current investors have decided to reinvest in the former, no new investors are willing to enter the fund as long as its expected performance does not exceed the reservation return of the least sophisticated new investor plus the entry cost. In that interval, the fund's expected performance increases one-to-one with managerial ability. The fund in the sophisticated market, however, continues to retain its current investors' money and attract their $t-$date endowment, so its expected performance increases slowly with ability. The lower bound of the entry cost interval, K_1, guarantees that the expected performance of both types funds cross in the interval $\left(\Phi_2^U, \Phi_2^U + K\right)$. Existence of the intersection
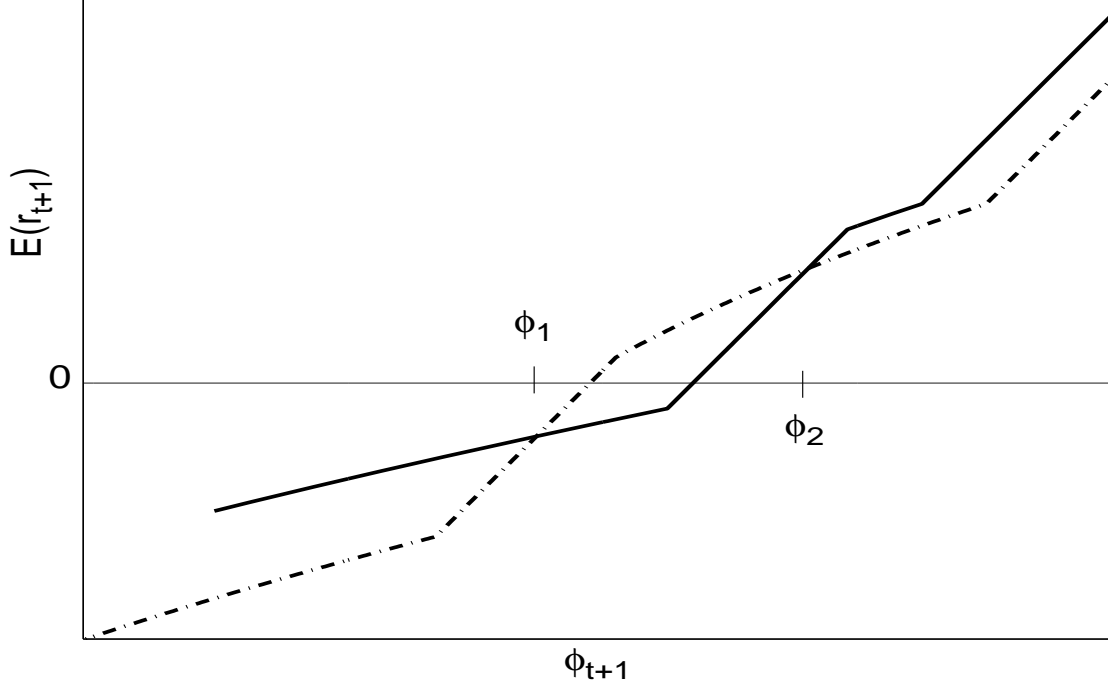
15

Figure 3: Expected net return as a function of expected managerial skill and market sophistication, with a positive entry cost. The solid (dotted) line corresponds to the sophisticated (unsophisticated) market, i.e., low (high) γ_{MAX}. Parameter values: $m = 200$, $c = 0.01$, $K = 0.75$, $\gamma_{MAX}^{U} = 1$, $\gamma_{MAX}^{S} = 0.5$, $f = 0.01$.

is guaranteed by the fact that unsophisticated investors have less money to invest, which gives funds in less sophisticated markets a performance advantage over funds in more sophisticated markets when current investors have reinvested with both funds and no new investors wish to enter. For higher levels of ability, new investors start to enter the fund. Since new investors in the less sophisticated market enter for lower levels of ability, fund performance deteriorates sooner as ability improves. In the limit, all possible investors decide to invest. Since the fund in the market with more unsophisticated investors attracts a larger set of investors, it is larger and necessarily underperform.

Finally, when entry costs are very high, i.e., when $K > K_2$, there will be no new investors willing to enter the fund for the range of managerial ability considered. In this case, the fund in the less sophisticated market outperforms the fund in the more sophisticated market for levels

of expected managerial ability that are above a minimum level, ϕ_1.

Our model suggests that both negative and positive expected performance are possible in equilibrium in a market with frictions. It also predicts that expected fund performance increases with managerial ability, which explains the evidence that cross-sectional differences in observed performance persist through time.

Finally, the model also delivers a new prediction: Investor sophistication decreases cross-sectional dispersion in fund performance. Therefore, the model not only explains why and how performance persistence arises in equilibrium, but it also provides a mechanism through which the degree of investor sophistication impacts the persistence of performance differences across funds.

4 Conclusions

Previous studies have noted that price competition alone may not be sufficient to eliminate differences in net performance across funds when investors fail to react to differences in expected performance and management companies react strategically (Christoffersen and Musto, 2002; Gil-Bazo and Ruiz-Verdú, 2008). In this paper, we argue that even if investors react rationally to differences in expected performance, market frictions distort their choices with respect to what would be expected in a friction-less market, such as the one described by Berk and Green (2004), and can generate predictability in fund performance both in the time series and in the cross section.

The model not only explains persistence in fund performance, but it also delivers a new prediction: Predictable differences in performance across funds decrease with the presence of more unsophisticated investors in the market.

An important implication of our results is that policies aimed at improving the efficiency of the market for mutual funds should focus on eliminating frictions and, particularly, facilitating product comparisons both within active funds and between active funds and passive alternatives.

Appendix

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 γ^C is such that $TP_{t+1}\left(q_{t+1}^*\right) = -\gamma^C$. The quantity invested in the fund is

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

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 $TP_{t+1}\left(q_{t+1}^*\right) = -\gamma^C$ can be rewritten as

$$\phi - cm\left(2(\gamma_{MAX} - \gamma^C) - \frac{1}{2}(\gamma_{MAX}^2 - (\gamma^C)^2)\right) - f = -\gamma^C. \tag{1}$$

Solving for γ^C, we obtain

$$\gamma^C = \frac{1}{cm}\left(1 + 2cm - A^{1/2}\right), \text{ where}$$
$$A \equiv 1 + 2cm\left(2 + \phi - f\right) + c^2m^2\left(2 - \gamma_{MAX}\right)^2.$$

γ^C is a real solution of equation (1) if $A > 0$ and a sufficient condition for $A > 0$ is $2 + \phi > f$, which is a reasonable assumption.

Notice that if $\gamma^C < \overline{\gamma}$ all current investors re-entry and we have also possible entry of new investors. The new investors have to pay the cost K to enter the fund and therefore, their cutoff reservation return, $-\gamma^N$, is obtained from:

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

We solve for γ^N from the equilibrium condition

$$\phi - cm\left(2\gamma_{MAX} - \gamma^N - \overline{\gamma} - \frac{1}{2}\left(\gamma_{MAX}^2 - (\gamma^N)^2\right)\right) - f - K = -\gamma^N, \qquad (2)$$

and obtain

$$
\begin{aligned}
\gamma^N &= \frac{1}{cm}\left(1 + cm - B^{1/2}\right), \text{ where} \\
B &\equiv 1 + 2cm\left(1 + \phi - f - K\right) + c^2m^2\left(1 + 2\overline{\gamma} + \gamma_{MAX}^2 - 4\gamma_{MAX}\right) \\
&= 1 + 2cm\left(1 + \phi - f - K\right) + c^2m^2\left((1 - \gamma_{MAX})^2 - \frac{2}{m}v_t\right).
\end{aligned}
$$

γ^N is a real solution of equation (2) if $B \geq 0$. For B to be higher or equal than 0 we need to have $K < \overline{K}(\gamma_{MAX}) \equiv \frac{1}{2cm}\left(1 + 2cm\left(1 + \phi - f\right) + c^2m^2\left((1 - \gamma_{MAX})^2 - \frac{2}{m}v_t\right)\right)$. So if $K < \overline{K}(\gamma_{MAX})$ there is a solution to equation (2), otherwise there is no real solution (and therefore no new investors enter the fund). When there is a real solution, we distinguish two cases depending on whether the solution γ^N is higher or smaller than $\overline{\gamma}$. When $\gamma^N \geq \overline{\gamma}$, no new investors want to enter the fund because the performance of the fund is lower than the sum of their reservation return and the entry cost. The expected return in this case equals $TP_{t+1}\left(\overline{q}_{t+1}\right) > -\overline{\gamma}$. On the other hand, when $0 \leq \gamma^N < \overline{\gamma}$, new investors enter the fund. Since the last new investor that entered has reservation return $-\gamma^N$, the expected return in this case is $K - \gamma^N$.

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

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

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

where $\Phi_1 \equiv f - \gamma_{MAX}$, $\Phi_2 \equiv f + 2cv_t - \overline{\gamma} - \frac{1}{2}cm\left(\gamma_{MAX}^2 - \overline{\gamma}^2\right)$ and $\Phi_3 \equiv f + cv_t + cm\gamma_{MAX}\left(1 - \frac{\gamma_{MAX}}{2}\right)$.

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

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

■

Proof of Proposition 2. Notice that, since $\gamma_{MAX}^U - \gamma_{MAX}^S > 0$, we have that $\Phi_1^S > \Phi_1^U$, $\Phi_2^S > \Phi_2^U$ but $\Phi_3^S < \Phi_3^U$.

We search for $\phi_1 \in \left(\Phi_2^U, \Phi_2^U + K\right)$ such that

$$E^S\left(r_{t+1}\right) = E^U\left(r_{t+1}\right)$$
$$i.e. - \gamma^C = \phi_1 - cq_{t+1}^* - f.$$

Notice that $\phi_1 - cq_{t+1}^* - f = \phi_1 - \Phi_2^U + \Phi_2^U - cv_t - f = \phi_1 - \Phi_2^U - \overline{\gamma}^U$, and $-\gamma^C = -\left(1 + 2a - A^{1/2}\right)$. We define a by $a \equiv cm$.

Solving for A we obtain $A = \left(2a + a\left(\phi_1 - \Phi_2^U - \overline{\gamma}^U\right) + 1\right)^2$, if $2a + a\left(\phi_1 - \Phi_2^U - \overline{\gamma}^U\right) + 1 > 0$ i.e. $\phi_1 > \Phi_2^U + \overline{\gamma}^U - 2 + \frac{1}{a}$ and this is satisfied for $\phi_1 > \Phi_2^U$. Since on the other hand $A = 1 + 2a\left(2 + \phi_1 - f\right) + a^2\left(2 - \gamma_{MAX}^S\right)^2$ we have that

$$\left(2a + a\left(\phi_1 - \Phi_2^U - \overline{\gamma}^U\right) + 1\right)^2 = 1 + 2a\left(2 + \phi_1 - f\right) + a^2\left(2 - \gamma_{MAX}^S\right)^2$$
$$\left(2a + a\left(\phi_1 - \Phi_2^U - K - \overline{\gamma}^U + K\right) + 1\right)^2 = 1 + 2a\left(2 + \phi_1 - \left(\Phi_2^U + K\right) + \Phi_2^U + K - f\right)$$
$$+ a^2\left(2 - \gamma_{MAX}^S\right)^2$$
$$\left(2a + a\left(-x - \overline{\gamma}^U + K\right) + 1\right)^2 = 1 + 2a\left(2 - x + \Phi_2^U + K - f\right) + a^2\left(2 - \gamma_{MAX}^S\right)^2,$$

where by definition $x \equiv \Phi_2^U + K - \phi_1$.

We define

$$T \equiv a^2 \left(2 - \gamma_{MAX}^S\right)^2 \ and$$

$$and \ V \equiv a \left(4cv_t - \overline{\gamma}^U - cv_t \left(\gamma_{MAX}^U + \overline{\gamma}^U\right)\right)$$

$$= av_t \left(4c + \frac{1}{m} + \frac{cv_t}{m}\right) - \gamma_{MAX}^U \left(1 + 2cv_t\right).$$

We obtain two solutions $x_{1,2}^* = \left(2 - \overline{\gamma}^U + K \pm \frac{1}{a}\sqrt{T + V}\right)$. If $\gamma_{MAX}^U < 2$ and $K > K_1 \equiv \frac{1}{a}\left(\sqrt{T + V} - a\left(2 - \overline{\gamma}^U\right)\right)$, the solution $x_1^* = \left(2 - \overline{\gamma}^U + K - \frac{1}{a}\sqrt{T + V}\right) \in (0, K)$. Consequently, $\phi_1 = \Phi_2^U + K - x_1^* \in \left(\Phi_2^U, \Phi_2^U + K\right)$.

Notice that $x_2^* = \left(2 - \overline{\gamma}^U + K + \frac{1}{a}\sqrt{T + V}\right)$ is always a positive solution but is also higher than K, so it cannot be solution of our problem.

We have shown in Proposition 1 that if $K \geq \overline{K}\left(\gamma_{MAX}^U\right) \equiv K_2$ then no new investors will enter the fund that targets the unsophisticated investors. The expected return of this fund increases one to one with ϕ_{t+1}, and since $\Phi_2^S > \Phi_2^U$ it implies that $E^U\left(r_{t+1}\left(\Phi_2^S + K\right)\right) > E^S\left(r_{t+1}\left(\Phi_2^S + K\right)\right)$ for any $\phi_{t+1} > \phi_1$.

Let us then consider the case when $K < \overline{K}\left(\gamma_{MAX}^U\right)$. To prove next that there is $\phi_2 \in \left(\Phi_2^S, \Phi_2^S + K\right)$ such that $E^S\left(r_{t+1}\right) = E^U\left(r_{t+1}\right)$ is enough to prove that the following two conditions are true: $E^U\left(r_{t+1}\left(\Phi_2^U + K\right)\right) > E^S\left(r_{t+1}\left(\Phi_2^U + K\right)\right)$ and $E^U\left(r_{t+1}\left(\Phi_2^S + K\right)\right) < E^S\left(r_{t+1}\left(\Phi_2^S + K\right)\right)$.

We have shown that when $K > K_1$ it exists $\phi_1 \in \left(\Phi_2^U, \Phi_2^U + K\right)$ such that $E^S\left(r_{t+1}\right) = E^U\left(r_{t+1}\right)$ and this implies that $E^S\left(r_{t+1}\left(\Phi_2^U + K\right)\right)$ could be either $-\gamma^C$ or $\Phi_2^U + K - c\overline{q}_{t+1} - f$. In the first case, it is straightforward that since the return does not change the slope in that interval, $E^U\left(r_{t+1}\left(\Phi_2^U + K\right)\right) > E^S\left(r_{t+1}\left(\Phi_2^U + K\right)\right)$. In the second case, if $E^S\left(r_{t+1}\left(\Phi_2^U + K\right)\right) = \Phi_2^U + K - c\overline{q}_{t+1} - f$ we have then that

$$E^U\left(r_{t+1}\left(\Phi_2^U + K\right)\right) = \Phi_2^U + K - c\overline{q}_{t+1}\left(\gamma_{MAX}^U\right) - f >$$

$$E^S\left(r_{t+1}\left(\Phi_2^U + K\right)\right) = \Phi_2^U + K - c\overline{q}_{t+1}\left(\gamma_{MAX}^S\right) - f \Leftrightarrow$$

$$\overline{q}_{t+1}\left(\gamma_{MAX}^U\right) < \overline{q}_{t+1}\left(\gamma_{MAX}^S\right).$$

Notice that $\overline{q}_{t+1}\left(\gamma_{MAX}\right) = 2v_t - \dfrac{m}{2}(\gamma_{MAX}^2 - \overline{\gamma}^2) = 2v_t - \dfrac{v_t}{2}(\gamma_{MAX} + \overline{\gamma}) = v_t\left(2 - \dfrac{1}{2}(2\gamma_{MAX} - \frac{v_t}{m})\right).$
Since $\overline{q}_{t+1}\left(\gamma_{MAX}\right)$ decreases with γ_{MAX} it results that $\overline{q}_{t+1}\left(\gamma_{MAX}^U\right) < \overline{q}_{t+1}\left(\gamma_{MAX}^S\right).$

To prove that $E^U\left(r_{t+1}\left(\Phi_2^S + K\right)\right) < E^S\left(r_{t+1}\left(\Phi_2^S + K\right)\right)$ we calculate both the expected adjusted returns evaluated in $\Phi_2^S + K$. Notice that in this range both returns equal to $K - \gamma^N$ and γ^N increase with γ_{MAX} if $\gamma_{MAX} < 1$. Since $\gamma_{MAX}^S < \gamma_{MAX}^U$ it implies $K - \gamma^{S,N} > K - \gamma^{U,N}$ and therefore $E^U\left(r_{t+1}\left(\Phi_2^S + K\right)\right) < E^S\left(r_{t+1}\left(\Phi_2^S + K\right)\right)$, with $\gamma^{S,N}$, $\gamma^{U,N}$ denoting the value of γ^N in the sophisticated and the unsophisticated market, respectively. If $K > \overline{K}\left(\gamma_{MAX}\right)$, the return for the sophisticated equals $E^S\left(r_{t+1}\left(\Phi_2^S + K\right)\right) = \Phi_2^S + K - c\overline{q}_{t+1}\left(\gamma_{MAX}^U\right) - f > K - \gamma^{S,N}$ and $E^U\left(r_{t+1}\left(\Phi_2^S + K\right)\right) = K - \gamma^{U,N}$, so again $E^U\left(r_{t+1}\left(\Phi_2^S + K\right)\right) < E^S\left(r_{t+1}\left(\Phi_2^S + K\right)\right)$ q.e.d. ∎

References

Berk, J. and R. Green (2004). Mutual fund flows and performance in rational markets. *Journal of Political Economy 112*(6), 1269–1295.

Bollen, N. and J. Busse (2005). Short-term persistence in mutual fund performance. *Review of Financial Studies 18*(2), 569–597.

Carhart, M. (1997). On persistence in mutual fund performance. *Journal of Finance 52*(1), 57–82.

Chen, J., H. Hong, M. Huang, and J. Kubik (2004). Does fund size erode mutual fund performance? The role of liquidity and organization. *The American Economic Review 94*(5), 1276–1302.

Christoffersen, S. and D. Musto (2002). Demand curves and the pricing of money management. *Review of Financial Studies 15*(5), 1499–1524.

Ferreira, M. A., A. Keswani, A. F. Miguel, and S. Ramos (2013). Testing the Berk and Green model around the world. *Working paper*.

Gil-Bazo, J. and P. Ruiz-Verdú (2008). When cheaper is better: Fee determination in the market for equity mutual funds. *Journal of Economic Behavior & Organization 67*(3), 871–885.

Glode, V., B. Hollifield, M. Kacperczyk, and S. Kogan (2011). Time-varying predictability in mutual fund returns. *Working paper*.

Hortaçsu, A. and C. Syverson (2004). Search costs, product differentiation, and welfare effects of entry: A case study of SP 500 Index Funds. *Quarterly Journal of Economics 119*(4), 403–456.

Huang, J., K. Wei, and H. Yan (2007). Participation costs and the sensitivity of fund flows to past performance. *The Journal of Finance 62*(3), 1273–1311.

Navone, M. (2012). Investors distraction and strategic repricing decisions. *Journal of Banking & Finance 36*(5), 1291–1303.

Sirri, E. and P. Tufano (1998). Costly search and mutual fund flows. *The Journal of Finance 53*(5), 1589–1622.

Yan, X. et al. (2008). Liquidity, investment style, and the relation between fund size and fund performance. *Journal of Financial and Quantitative Analysis 43*(3), 741–768.