The life cycle of investment management when “today’s alpha is tomorrow’s beta”

Georgios Magkotsios*
Marshall School of Business, University of Southern California
November 5, 2017

Abstract

I present a model where competition in the fund management industry has positive and negative effects. When funds have increasing (decreasing) returns to scale at the industry level, the flow-performance relation is concave (convex). Active funds outperform passive benchmarks initially, but the average returns from active investing are not persistent. A growing number of competing funds and rising trading costs gradually deplete the profitable opportunities in the aggregate. The total investment surplus declines to zero and most active managers underperform relative to passive funds. Aggregate risk is reduced through “closet indexing” over time, until all active funds form a large and scalable pool of passively invested capital.

Keywords: investment management; alpha; returns to scale; flow-performance relation; network externality.

JEL Classification Numbers: G11, G12, G23

*Georgios.Magkotsios.2017@marshall.usc.edu. This paper was the recipient of the John Doukas best PhD paper award at the 2017 EFMA meeting. I am grateful to my supervisors Wayne Ferson, Kevin Murphy, Kenneth Ahern, Odilon Câmara, Arthur Korteweg, and other faculty at USC. I have received valuable feedback from Thierry Foucault, Raman Uppal, and the participants of the 2016 Trans-Atlantic Doctoral Conference in London, the 2017 FMA European meeting in Portugal, the 2017 EFMA meeting in Greece, and the 2017 FMA meeting in Boston.
I. Introduction

This paper compares the benefits to the investor from active and passive fund management. The bulk of the empirical research on this important topic has focused on equity mutual funds. The average actively managed equity fund underperforms relative to passive indexes, but a minority of talented managers outperform their benchmark and receive more capital from investors.\(^1\) More recent evidence reveal a heterogeneity in the relations between scale, performance, and investor flows of capital that is puzzling for existing theories. For instance, Pástor et al. (2015) show decreasing returns to scale at the industry level for equity mutual funds, while Magkotsios (2017) finds increasing returns to scale for hedge funds and fixed income funds.\(^2\) I provide a theoretical model of the competition among active managers that can explain these empirical results.

The main predictions include the conditions for successful active asset management, the endogenous emergence of increasing and decreasing returns to scale at the industry level, and an empirically testable connection between returns to scale and investor flows across funds. Active funds can outperform passive funds and operate with increasing returns to scale at the industry level, or be less profitable than passive funds and have diminishing returns to scale. The effect of scale on fund performance and the subsequent flows by the investor have been studied separately in the empirical literature. My model combines these two literatures under a common framework. I also provide a potential explanation for the popularity of equity mutual funds among investors.


\(^2\)Harvey and Liu (2017) propose instead that the performance of the average equity fund declines mostly with the fund’s own assets. Kaplan and Schoar (2005), Getmansky (2012), and Goldstein et al. (2017) find concave flow-performance relations for private equity, hedge funds, and corporate bond funds respectively. Magkotsios (2017) also shows that the returns of index funds and ETFs are inert to variations in fund or industry scale.
The model introduces a life cycle for asset management with time-varying R&D and trading costs for every manager. The key feature is that competition has positive and negative effects on fund performance, where costs and investor flows of capital determine which effect dominates. Trading costs are low when competition is moderate, and the average active fund outperforms passive benchmarks. On the other hand, strong competition increases trading costs and erodes performance until the net returns of the average active fund fall behind the returns of passive indexes. Despite the poor performance in risk-adjusted returns (alpha), the investor doesn’t withdraw all his capital from active funds. Instead, he diversifies optimally across managers on the efficient frontier.

During the early stage of the life cycle, the supply of managers and aggregate demand expand simultaneously in equilibrium. The rising number of funds reduces the R&D costs. This allows the average manager to identify profitable opportunities more efficiently. In addition, the investor filters out the worst performing managers and augments with capital inflows the top performers by observing their track record. This process raises the average level of managerial talent in the cross-section, and it creates a barrier to fund entry for future cohorts. The declining R&D costs and rising level of average talent express the positive effect of competition. Active management is more profitable to the investor than passive investing during this stage, and the funds have increasing returns to scale at the industry level. Investor flows are more sensitive to bad-performing funds and less sensitive to top-performing ones. This implies a concave flow-performance relation.

During the late stage of the life cycle, the empirical predictions of the model are the opposite. Although the incumbent managers are on average more talented over time, they also become more homogeneous in their strategies and crowd into a diminishing opportunity set. Investing in the same direction increases trading costs and reduces the value of investment in the aggregate. Crowded trading and increasing costs is the negative effect of competition. The average active fund underperforms relative to passive indexes. Funds have diminishing returns to scale at the industry level, because the increased trading costs limit the opportunities to obtain alpha. Investor flows are more sensitive to good-performing funds, because it is more strenuous for managers to outperform their
rivals when they are equally talented. The flows are less sensitive to bad-performing funds, because expected losses are small. This implies a convex flow-performance relation.

Fund managers may choose to index a fraction of their assets to mitigate their costs. Over the life cycle of asset management, I show that active funds transform from a risky investment vehicle that is rich in opportunities for alpha to a set of funds whose performance and risk are similar to those of passive funds. This is the concept of “today’s alpha is tomorrow’s beta”.\(^3\) It provides a potential explanation for the high investor demand for mature asset classes such as equity mutual funds. Assets in passive funds are subject to market risk only, while active management late in its life cycle provides diversification across multiple funds on the efficient frontier.

Previous theoretical work has assumed various relations between scale and performance. Berk and Green (2004) assume diminishing returns with fund size for a monopolist manager. The manager can extract all surplus from investment by optimally increasing his fee in equilibrium. Their model implies positive returns before fees and zero net-of-fee returns in excess of passive benchmarks. In reality, managers compete for investor capital and superior performance. With competition, it is not obvious that managers can extract all the surplus. Pástor and Stambaugh (2012) and Feldman et al. (2016) assume that fund returns decline with industry scale to explain the persistence of poor track records for active managers.\(^4\) My model derives endogenously increasing and decreasing returns to scale at the industry level, depending on the level of competition among managers with symmetric information.

The competition among managers with asymmetric information in Garcia and Vanden (2009) and Gârleanu and Pedersen (2016) gives rise to diseconomies of scale in equilibrium. Diseconomies of scale at the industry level arise in their models as the market becomes more efficient and attenuates the comparative advantage of informed managers. My model is different, because it also derives increasing returns to scale at the industry level for moderate competition. In addition, the

\(^3\)This expression was coined by Andrew Lo. See also Cho (2017) for a costly arbitrage-based argument where alphas turn to betas.

\(^4\)Theoretical models in the international trade literature also make similar assumptions. For instance, Krugman (1980, 1991) assumes a fixed cost of labor to generate increasing returns to scale at the firm level, while Grossman and Rossi-Hansberg (2010) assume external economies of scale.
unique link that I show between returns to scale and the sensitivity of investor flows to performance has not been discussed before in the theoretical and empirical literature.

II. The model

The model describes the competition among active managers for superior performance and investor flows of capital. The markets make rational expectations, and there is no moral hazard or adverse selection. Funds have a dual cost structure from active investing. The first type involves R&D costs in identifying profitable investment opportunities within capital markets. The second type involves the fund’s trading costs to implement the strategies that are the outcome of the R&D process. For instance, funds can recruit quantitative financial analysts to search for mispriced assets in the market and recommend new strategies to the portfolio manager. The compensation of the analysts is part of the R&D costs. The portfolio manager will then execute trades that incorporate the new strategies and are subject to transaction costs.5

I show that competition has positive and negative effects on the performance of the average fund. The positive effect arises from economies of agglomeration. The discovery of a new profitable opportunity increases the average performance in the cross section, because it attracts more talented managers that attempt to exploit it. The investor augments the positive effect through an assortative matching between talent and fund assets in equilibrium. The negative effect is related to crowded trading, when a large number of managers cluster around a small amount of profitable opportunities. This effect increases the cost of trading for every manager and reduces the net value of investment in the aggregate. The balance between R&D and trading costs determines in equilibrium whether the positive or negative effect of competition dominates.

5Klepper (1996) also introduces a dual cost structure during the life cycle of industries in the real economy. This includes product R&D costs for innovation and process R&D costs for more efficient production.
A. Main setup

A manager’s alpha reflects his ability in exploiting mispriced investment opportunities and achieving risk-adjusted returns in excess of a benchmark. Let $\tau_i$ be an exogenous level of talent for manager $i$ that defines his alpha, $q_{it}$ the size of his assets under management at time $t$, and $Q_t$ the aggregate size for a total number $N_t$ of competing funds. The gross risk-adjusted return for fund $i$ that is realized at time $t + 1$ is

$$R_{it+1} = 1 + \tau_i + \varepsilon_{it+1}.$$  \hspace{1cm} (1)

The noise terms $\varepsilon_{it+1}$ are jointly distributed over time and across managers with zero mean. These shocks reflect the component of luck in the realized return. Passive funds (labeled with $i = 0$ hereafter) require no talent, implying the zero-alpha definition $E_t[R_{0t+1}] \equiv 1$. Depending on the distribution of talent among active managers, the average fund may have either positive or negative alpha.

The stylized cost function for manager $i$ has two components

$$C(q_{it}, HH_t) = \frac{cq_{it}^2(1 - HH_t)}{2HH_t(1 + \tau_i)} + \frac{hq_{it}HH_t}{1 + \tau_i},$$ \hspace{1cm} (2)

where $c$ and $h$ are constants. The first term describes the trading cost for the fund. This cost arises from the impact on security prices from the manager’s trading activity, and it is most sensitive to the volume of assets that the fund trades in the market. The second term is the fund’s R&D cost, reflecting the manager’s effort in discovering mispriced opportunities for alpha in the market. Both terms are inversely proportional to the manager’s talent, implying that managers have differential ability in reducing trading and R&D costs to maximize their return.
The Herfindahl–Hirschman index $HH_t \in [0, 1]$ is a measure of concentration at the industry level. It is defined as

$$HH_t \equiv \sum_{i=1}^{N_t} \left( \frac{q_{it}}{Q_t} \right)^2,$$

and it reflects the positive and negative effects of competition on fund costs. The positive effect reflects economies of agglomeration. A decreasing $HH_t$ implies less concentration within the industry. The agglomeration of talented managers reduces the R&D costs for all incumbents, assuming knowledge spillovers among funds. The negative effect is the price impact of crowded trades. Crowded trading emerges when a large number of similarly talented managers trade in the same direction to compete for an investment opportunity. The intuition is similar to Foster and Viswanathan (1996), who show that trading is less profitable when many agents chase the same information signal.

Let $\mu_{\tau,t}$ and $\sigma_{\tau,t}$ be the cross-sectional weighted average and dispersion of talent respectively. Within a homogeneous group of very talented competitors, every manager has the potential to identify an investment opportunity that is available. In equilibrium, the combination of large $N_t$, large $\mu_{\tau,t}$, and small $\sigma_{\tau,t}$ decreases $HH_t$, and induces increased transaction costs from crowded trades. This increases prices and reduces the number of profitable trades. Therefore, increased trading costs destroy net investment value and make alphas more elusive in the aggregate.

The fund managers participate in a monopolistic competition. Managers are vertically differentiated based on talent in obtaining alpha from arbitrage opportunities. They are also horizontally differentiated based on their risk-return tradeoff on the efficient frontier. For instance, some managers provide high-risk and high-return strategies, while others provide low-risk and low-return strategies. Monopolistic competition implies that perturbations to the expected return and fees of a single fund cannot affect the aggregate indices of returns $SR \equiv \sum_j E_t[R_{jt+1}]$ and fees $Sf \equiv \sum_j f_{jt}$ respectively. The manager of fund $i$ sets his fee $f_{it}$ at time $t$ to maximize his ex-
pected profits from active investing

$$\max_{q_{it}} E_t[\Pi_{it+1}] = E_t \left[ f_{it} q_{it} - \frac{cq_{it}^2 (1 - HH_t)}{2HH_t(1 + \tau_i)} - \frac{hq_{it}HH_t}{1 + \tau_i} \right], \quad (4)$$

where $q_{it}$ is the investor demand that clears the market. The quadratic term for the transaction costs implies diminishing returns to scale at the fund level.\(^6\) The fund fee represents the price that the investor pays for asset management. I do not make a distinction between management and performance fees.

The investor has finite wealth and mean-variance preferences. His portfolio of funds at time $t$ involves $N_t$ active managers and one passive fund. The investor’s problem is the following

$$\max_{q_t} U = E_t[q_t'(r_{t+1} - f_t)] - \frac{\gamma}{2} q_t' V_t q_t \quad s.t. \quad q_t' 1 = W_t \quad , \quad (5)$$

where $\gamma$ is a parameter related to the investor’s risk aversion, and $W_t$ is the exogenous total wealth invested among the active funds and passive index at time $t$. All vectors have $N_t + 1$ elements, where $q_t$, $f_t$, and $r_{t+1}$ are the vectors for fund sizes, fees, and nominal returns realized at $t + 1$. In addition, $1$ is a vector of ones, and $V_t$ is the covariance matrix for the returns of the passive index and incumbent active managers.

**B. Equilibrium**

Talent is unobservable and imperfectly known to the market, including the managers themselves. All market participants observe the realized returns over time and update their estimates about every manager’s talent. The learning process involves Bayesian updating under symmetric information. Potential entrants draw their exogenous talent from a normal distribution $\mathcal{H}(\tau_i) \sim N(\mu, \sigma^2)$ that is common to all entry cohorts over time. The conditional estimates for the cross-sectional mean $\hat{\mu}_{\tau,t}$ and dispersion $\hat{\sigma}_{\tau,t}$ among incumbent managers are $\hat{\mu}_{\tau,t}$ and $\hat{\sigma}_{\tau,t}$ respectively.

\(^6\)Edelen et al. (2007) show that transaction costs are an important determinant of diminishing returns to scale for equity mutual funds.
The Appendix discusses the details of the stochastic learning process and derives expressions for the conditional estimates on the mean and dispersion of talent. As the uncertainty about each manager’s talent is resolved over time, the estimates converge to the population mean and dispersion.

Define the vector of expected risk-adjusted returns that are net of fees and costs

$$E_t[R_{t+1} - f_t] \equiv V_t^{-1} E_t[r_t - f_t] .$$

(6)

Each element of this vector corresponds to a fund’s net alpha, and $R_{it+1}$ is the gross return given by equation (1). The solution to the investor’s problem in equation (5) is the following:

**LEMMA 1.** The equilibrium size for fund $i$ at time $t$ is given by

$$q_{it}^* = \gamma W_t b_{it} / N_t + 1 + b_1 E_t[R_{it+1} - f_{it}] - \frac{b_{it}}{N_t + 1} \sum_{j=0}^{N_t} E_t[R_{jt+1} - f_{jt}] ,$$

(7)

where $E_t[R_{jt+1} - f_{jt}]$ is the expected net alpha for fund $j$, while $b_1$ and $b_{it}$ are

$$b_1 \equiv \frac{1}{\gamma} , \quad b_{it} \equiv \frac{1}{\gamma} \frac{V^{-1}1}{VV^{-1}1} .$$

(8)

The investor values returns net of fees and costs. The demand for a fund $i$ depends on the fund’s own net-of-fee alpha, but also includes the average of net-of-fee alphas from every competing fund. The term $b_{it}$ embeds the correlations in performance between fund $i$ and its rival funds (including the passive fund). The relation $b_1 > b_{it}/(N_t + 1)$ in equation (8) implies that investor flows are always more sensitive to the fund’s own performance than the average performance among the competing funds. Lemma 1 implies that any arbitrary attempt by a manager to raise his own fees may trigger outflows that are redistributed to rival funds in equilibrium.

The Nash equilibrium for equations (4) and (7) gives the fund’s response function

$$f_{it}^* = [b_1 (N_t + 1)(h HH_t + \hat{\sigma}_{\tau,t}^2/2) HH_t - c b_{it} (1 - HH_t) (SR - Sf - \gamma W_t) +$$

(8)
\[ E_t[R_{it+1} + (b_1 c(N_t + 1)(1 - HH_t) + b_1(N_t + 1)HH_t - b_{it}HH_t \cdot (SR - Sf - \gamma W_t))] \cdot \left[ b_1(N_t + 1)\left(b_1 c(HH_t - 1) + 2E_t[R_{it+1}HH_t]\right)\right]^{-1}, \] (9)

where \( SR \) and \( Sf \) the aggregate indices of returns and fees respectively. Equation (9) involves the approximation \( R^2_{it+1} \approx R_{it+1} + \tilde{\sigma}^2_{\tau_{it}}/2 \). The fund’s response function depends on the sum of equilibrium fees for all incumbent managers. Adding up equations (9) for all funds gives the solution for the sum of fees. Then the fee for each fund can be retrieved by substitution to equation (9). The equilibrium fund size is given by equation (7). The equilibrium industry size \( Q^*_t \) is the sum of the fund sizes for the \( N_t \) active managers.

Potential entrant managers enter if their expected profits are positive, while incumbents with negative expected profits sustain capital outflows until they liquidate and exit. In equilibrium, the marginal incumbent at time \( t \) is the manager who has zero expected profits. As a result, the number of incumbent managers \( N_t \) at time \( t \) is specified by the following condition

\[ E_t[\Pi_{mt+1}] = E_t \left[ f^*_m q^*_m - \frac{c(q^*_m)^2 (1 - HH^*_t)}{2HH^*_t(1 + \tau_t)} - \frac{hq^*_mHH^*_t}{1 + \tau_t} \right] = 0, \] (10)

where \( m \) is the marginal incumbent fund, and the asterisks denote equilibrium values that are specified by the solution to equations (4) and (5). The solution to equation (10) gives \( N^*_t \), the equilibrium number of funds at time \( t \).

## III. The life cycle of investment management

The comparative statics analysis on the endogenous variables \( f^*_t, q^*_t, \) and \( Q^*_t \) reveals a life cycle for active management. I show below that the two driving forces for the evolution of this life cycle are the performance-based flows of capital across funds by the investor, and the balance between R&D and trading costs (see equation (2)).
A. Investor flows and competition

The equilibrium allocations of capital to funds are assortative, with more talented managers administering larger portions of the assets. The following lemma shows a monotonically increasing relation between investor flows and expected fund performance.

**LEMMA 2.** Investor flows to each fund are monotonically increasing in performance, but decline to zero as \( HH_t \) diminishes. Specifically,

\[
\frac{\partial q_{it}^*}{\partial E_t[R_{it+1}]} \geq 0 \quad \forall i, t , \quad (11)
\]

with the equality satisfied when \( HH_t \to 0 \).

The investor forms estimates about every manager’s talent and expected returns by observing their track record. The fund flows are his response to innovations about performance, where an innovation is the difference between expected and realized returns. He supplies inflows to funds with positive realizations in returns, in anticipation of larger returns in the future. On the contrary, the investor removes capital from funds when realized returns are lower than expected. Therefore, Lemma 2 implies that learning and investor flows filter out over time the bad managers and favor the most talented ones in equilibrium.

Competition and the investor’s response to performance influence the cross-sectional distribution of talent. The less talented managers experience outflows from investors until they liquidate the fund and exit. Their capital is reallocated to other incumbents or new cohorts of more talented managers. The transfer of capital increases the estimate for the weighted average talent in the cross section \( \hat{\mu}_{\tau,t} \) (equation (B.20)), and also the population average talent \( \mu_{\tau,t} \) as the uncertainty about talent is gradually resolved. The rising average talent suggests that many incumbents who attain positive abnormal returns at a certain time will eventually be surpassed by more talented rivals in the future.\(^7\) This feature of competition is similar to a “creative destruction” (Schumpeter, 1942).

\(^7\)Pástor et al. (2015) document empirically a rising average talent over time among equity mutual funds, and explain it in terms of changes to the population of managers.
Furthermore, the increasing $\mu_{\tau,t}$ and the exogenous prior distribution $H(\tau_i)$ for the talent of potential entrants imply that the cross-sectional dispersion $\sigma_{\tau,t}$ for incumbent managers must decrease over time. The effects of managerial turnover during the life cycle of active management are summarized below.

**PROPOSITION 1 (Creative destruction).** The competition among managers and investor flows concentrate the cross-sectional distribution of talent at larger values, i.e.

$$
\frac{d\mu_{\tau,t}}{dt} > 0 \quad \text{and} \quad \frac{d\sigma_{\tau,t}}{dt} < 0 .
$$

(12)

Proposition 1 implies that the barrier to fund entry is talent, and this barrier is raised over time. This equilibrium mechanism induces fund exits. Incumbent managers whose exogenous talent is above average and harvest positive alphas may be forced to exit later in the life cycle if their talent becomes much lower than the rising average talent. Therefore, the creative destruction mechanism implies that returns from active management are not persistent.

The dynamic evolution of talent in the cross section affects the investor’s flows in equilibrium. Intuitively, the uncertainty about talent in the cross section is larger early in the life cycle. For two funds of the same talent $\tau_i$, the one that enters earlier will experience larger and more volatile flows. As $\mu_{\tau,t}$ and $\sigma_{\tau,t}$ evolve, the talent range for successful entrants narrows. The prior for subsequent entrant cohorts has smaller dispersion, and the investor resolves the uncertainty about talent faster and with smaller flows in absolute magnitude.

**COROLLARY 1.** Funds in earlier entry cohorts receive comparatively larger flows than funds in subsequent entry cohorts. Specifically, between two managers of identical talent and realized returns who entered at different cohorts, the manager who entered first will receive larger flows over time.

The following lemma and Figure 1 show some comparative statics. The independent variable in Figure 1 is the Herfindahl index. The green curve is the correlation between the industry size
$Q_t^*$ and expected fund performance. When this curve is positive, the funds operate under increasing returns to scale at the industry level and vice versa. The blue curve describes the sensitivity of investor flows to performance. When this curve is positive, the flow-performance relation is convex. This means that investor flows are more sensitive to good performers and less sensitive to bad performers. The opposite is true when the blue curve is negative.

**Lemma 3.** The following equations

$$
\frac{\partial f_{it}^*}{\partial E_t[R_{it+1}]} = 0 , \quad \frac{\partial Q_t^*}{\partial E_t[R_{it+1}]} = 0 , \quad \text{and} \quad \frac{\partial^2 q_{it}^*}{\partial E_t[R_{it+1}]^2} = 0
$$

have a common and unique positive root $HH_{thr}$ in terms of the Herfindahl index.

Lemma 3 and Figure 1 suggest that the life cycle of active management may be separated in two stages, depending on the value of the Herfindahl index relative to its threshold value $HH_{thr}$. Each stage has different implications for returns to scale, the sensitivity of flows to performance, and the performance of the average active fund relative to passive benchmarks. I discuss these implications below for each stage of the life cycle.

**B. The early stage of the life cycle**

Let’s assume a small number of funds early in the life cycle. This implies a highly concentrated asset class. The Herfindahl index is near 1, and it satisfies this relation: $HH_t > HH_{thr}$. The combination of small $N_t$ and relatively large $\hat{\sigma}_{\tau,t}$ (Proposition 1) implies a less competitive landscape for potential entrants, and the cost of trading is low.

The first cohort of managers discover unexploited investment opportunities and obtain alphas. Their performance attracts more managers subsequently, and the supply of funds is expanding. The investor benefits from the agglomeration of labor, because he can diversify his wealth across funds to mitigate risk. He is learning about every manager’s talent over time, and distributes his capital among the best performing managers (Lemma 2). The concurrent increase in the supply and demand for active management improves the performance of the average fund (Proposition 1). The
Figure 1: Comparative statics of fees on performance (red solid line), returns to scale at the industry level (green dashed line), and curvature of the flow-performance relation (blue dot-dashed line) as functions of the Herfindahl index. The vertical dashed line marks the threshold value $HH_{thr}$ where the returns to scale change from increasing to decreasing, and the curvature of the flow-performance relation switches from concave to convex. The graph also shows limiting values for all curves when $HH_t \to 0$ and $HH_t \to 1$. These are:

(Ping A) $\lim_{HH_t \to 0} \frac{\partial f^*_i}{\partial E_t[R_{it+1}]} = + \infty$

(Passing B) $\lim_{HH_t \to 1} \frac{\partial f^*_i}{\partial E_t[R_{it+1}]} = - \frac{(N_t + 1)b_1 + b_{it}(h + \hat{\sigma}^2_{\tau,t}/2)}{2(N_t + 1)b_1 E_t[R_{it+1}]^2}$

(Passing C) $\lim_{HH_t \to 1} \frac{\partial Q^*_t}{\partial E_t[R_{it+1}]} = \frac{b_{0t}(h + \hat{\sigma}^2_{\tau,t}/2)}{(N_t + 1)E_t[R_{it+1}]^2}$

(Passing D) $\lim_{HH_t \to 0} \frac{\partial Q^*_t}{\partial E_t[R_{it+1}]} = - \infty$

(Passing E) $\lim_{HH_t \to 1} \frac{\partial^2 q^*_i}{\partial E_t[R_{it+1}]^2} = - \frac{(h + \hat{\sigma}^2_{\tau,t}/2)((N_t + 1)b_1 - b_{it})}{(N_t + 1)E_t[R_{it+1}]^3}$
improved performance at the industry level further amplifies the expansion of supply and demand through a positive feedback loop. As a result, the early stage of the life cycle features a positive network externality similar to Katz and Shapiro (1985). The agglomeration of funds and investor flows create suitable conditions for the growth of active management and its success relative to passive investing. The following lemma shows the conditions for the outset of the life cycle and the dominance of active asset management against passive investing.

**Lemma 4.** The number of funds $N_t^*$ increases over time, the Herfindahl index $HH_t^*$ decreases, and the average active fund outperforms passive benchmarks when its return satisfies the following conditions

$$1 + \frac{1}{2} \left( h + \frac{\hat{\sigma}^2_{\tau,t}}{2} \right) + \frac{\gamma W_t}{2(N_t + 1)} < \hat{\mu}_{\tau,t} < \frac{1}{2} \left( h + \frac{\hat{\sigma}^2_{\tau,t}}{2} \right) + \frac{\gamma W_t}{N_t + 1}$$  \hspace{1cm} (14)

Figure 2 illustrates the conditions for the initiation of the network effect. The contour plot between the average performance in the cross-section $\mu_{\tau,t}$ and investor total wealth $W_t$ is divided in three areas. Area (A) corresponds to the case where a small number of managers have positive alphas, but investor demand is small. The investment opportunities are relatively undiscovered by the market. The network effect does not start, because the aggregate capital invested is insufficient. Area (C) corresponds to the case where the available capital for investment is large, but the average performance of active funds is below that of passive indexes. The poor performance among a small number of funds does not attract any new managers. Area (B) corresponds to the case where the conditions in Lemma 4 hold true. The number of funds $N_t$ increases when there is sufficient investor capital available. The network effect initiates the life cycle, and the rising number of funds coincides with an expansion in the aggregate investor demand.

The network externality, also known as demand-side economies of scale, is the positive effect from the competition among managers. The following proposition summarizes its implications about the life cycle of active management.
Figure 2: Average performance and investor demand conditions for network effect. The coefficients for the curves marking different areas in the contour plot are \( a_1 = 1/2 \left( h + \hat{\sigma}_{\tau,t}^2/2 \right) + 1 \), \( a_2 = 1/2 \left( h + \hat{\sigma}_{\tau,t}^2/2 \right) \), \( d_1 = \gamma W_t/(2(N_t + 1)) \), and \( d_2 = \gamma W_t/(N_t + 1) \).

PROPOSITION 2 (Network effect). During the early stage of the life cycle, (a) the net-of-fee returns of the average active fund outperform passive benchmarks, (b) the funds have increasing returns to scale at the industry level, (c) the flow-performance relation is concave, (d) fund fees are negatively correlated with expected returns, and (e) the investor’s surplus from active management increases.

The green curve in Figure 1 shows that for \( HH_t > HH_{thr} \) the correlation between aggregate size \( Q_t^* \) and the expected return of a fund is positive. Although each manager has diminishing returns to scale from R&D and trading costs, the industry as a whole operates under increasing returns to scale because of the network effect. This result highlights the difference between fund-level and industry-level returns to scale, and demonstrates the benefit to the investor from competition. The performance of a monopolist manager deteriorates from diseconomies of scale. However, competition and the investor’s response to performance create conditions where the average performance and aggregate demand increase simultaneously in a positive feedback loop.
The flow-performance relation during the early stage of the life cycle is concave. Figure 1 shows that the second derivative of fund size with expected returns (blue curve) is negative when $HH_t > HH_{thr}$. Large values of $\hat{\sigma}_{\tau,t}$ during this stage of the life cycle (Proposition 1) imply the possibility of large realized losses from managers on the left tail of the distribution. A concave flow-performance relation means that investor flows are more sensitive to bad performance and less sensitive to good performance. In equilibrium, the investor anticipates more talented managers to enter the industry in the future, because of the dynamic network effect and increasing returns to scale. Therefore, his optimal response is to remove the worst-performing managers from his portfolio, rather than invest more capital into the current top performers.

The red curve in Figure 1 suggests that fund fees are negatively correlated with expected returns during the early stage of the life cycle. The negative correlation stems from the increase in returns because of the network effect, and the downward pressure on fees from competition. Specifically, the competition among managers provides a way for the investor to counteract rising fees in equilibrium. The investor diversifies his wealth across multiple funds with allocations proportional to net-of-fee returns. If a manager attempts to increase his fees, the investor reacts by redirecting flows to rival managers of similar talent. A manager cannot affect the aggregate indices of returns and fees under monopolistic competition. If he decreases his fee he can attract larger capital inflows by the investor. The discount of fees from incumbent managers is optimal, because the market anticipates the entry of more talented managers in the future (Proposition 1).

The results of Proposition 2 are based on monopolistic competition. Appendix C shows an extension of the model with differentiated Bertrand competition that has the same empirical predictions for the early stage of the life cycle.

C. The late stage of the life cycle

The growth of active management during the early stage of its life cycle is unsustainable. As the competition intensifies and funds harvest alphas, the managers gradually crowd into a diminishing

---

8Christoffersen (2001) shows that money market funds relinquish fees to increase future investor flows.
set of investment opportunities. Crowded trading emerges from the combination of rising \(N_t\), rising \(\hat{\mu}_{\tau,t}\), and decreasing \(\hat{\sigma}_{\tau,t}\). The declining Herfindahl index \(HH_t^*\) reduces the R&D costs over time, and facilitates the discovery of profitable investment opportunities by managers. However, it increases trading costs increase and affect prices in the aggregate. The rising trading costs attenuate the network effect, diminish the net value of the available investment opportunities, and make alphas more elusive at the industry level. When the Herfindahl index reduces below the critical threshold \(HH_{thr}\), the life cycle of active management transitions to its late stage. Managers who outperform their peers and benchmark net of costs by a wide margin are rare during this stage, because the incumbents are alike in talent.

The rising trading costs and crowded investing reflect the negative aspect of competition. Investment opportunities deplete and value is destroyed as it becomes more strenuous to outperform rivals over time. The following proposition summarizes the empirical implications from active management for this stage of the life cycle.

**PROPOSITION 3 (Depletion of alpha).** During the late stage of the life cycle, (a) the net-of-fee returns of the average active fund fall behind passive benchmarks, (b) the funds have diminishing returns to scale at the industry level, (c) the flow-performance relation is convex, (d) fund fees are positively correlated with expected gross returns, and (e) the total surplus from active investing decreases asymptotically to zero.

Proposition 3 suggests that large and very competitive groups of funds will have only few managers who achieve positive net alpha. The lack of profitable opportunities and managerial competition impose negative net alphas for most incumbents. This result is consistent with evidence for equity mutual funds in Kosowski et al. (2006) and Fama and French (2010), who find that only a small fraction of managers at the right tail of the distribution attain risk-adjusted returns in excess of their benchmark. It is also consistent with a voluminous empirical literature that documents the poor average performance of equity funds relative to passive funds.

The green curve in Figure 1 shows that for \(HH_t < HH_{thr}\) the correlation between the total actively managed assets \(Q_t^*\) and the expected return of a fund is negative. This implies that
managers have diminishing returns to scale at the industry level, related to the lack of profitable investment opportunities. An increase in aggregate investor demand destroys investment value, because more managers chase the remaining opportunities for alpha at higher trading costs. The correlation between fund fees and performance becomes positive. Although managers are very talented on average, those who can obtain positive alphas after costs are rare and increase their fees in equilibrium.

The flow-performance relation during the late stage of the life cycle is convex. The blue curve in Figure 1 shows that the second derivative of fund size with expected returns is positive for $0 < HH_t < HH_{thr}$. Convexity implies that investor flows are more sensitive to good performance, and less sensitive to bad performance. The investor rewards with inflows the few managers who earn positive alphas when the opportunities in the aggregate are limited.$^9$ In addition, investor outflows from poor-performing managers are moderate, because all incumbents are alike in talent and realized losses are smaller compared to the early stage of the life cycle.

### IV. Today’s alpha is tomorrow’s beta

#### A. Closet indexing and endogenous benchmark

The main setup shows that active management is more profitable than passive investing during the early stage, but less profitable during the late stage of the life cycle. Area (C) in the contour plot of Figure 2 shows that poor average performance when $N_t$ is small would prevent the network effect from starting. However, the number of funds $N_t$ is large during the late stage of the life cycle. Why is active management popular during the late stage? For instance, why are equity mutual funds popular, even though they underperform passive benchmarks on average? To answer these questions, I extend the main setup by introducing an endogenous benchmark and allowing active managers to invest a fraction of their assets into passive strategies.

---

$^9$Berk and Green (2004) also derive a convex flow-performance relation when managerial talent is scarce.
The gross risk-adjusted return for fund $i$ is

$$R_{it+1} = 1 + \tau_i - \left[(1 - HH_t)\mu_{\tau,t} + HH_t \cdot R_{Mt+1}\right] + \varepsilon_{it+1}.$$  

(15)

The term in brackets is the endogenous benchmark, which is a weighted average between the market return $R_{Mt+1}$ and the average talent $\mu_{\tau,t}$ in the cross-section of active managers. The weight is the Herfindahl index $HH_t$, and it decreases over the life cycle. Intuitively, the benchmark during the early stage is an exogenous factor, such as the market return. The asset class could in principle grow until the managers hold the full underlying market of securities. In this case, the active managers would be the benchmark themselves and the average gross alpha would be zero. The endogenous benchmark at time $t$ is a weighted average between these two extremes, with $HH_t$ as the weight.

Equation (15) implies that competition and investor learning gradually assimilate alpha-earning strategies into the benchmark. For instance, equity fund managers could obtain alphas when their benchmark is $R_{Mt+1}$ by trading on small stocks. This strategy would qualify as alpha, because only a few managers use the strategy at low trading cost. As the Herfindahl index decreases over time and more managers adopt the strategy, net alpha is competed away. The average talent $\mu_{\tau,t}$ in equation (15) captures this effect, and size-related strategies become a risk factor in the benchmark.

The cross-sectional distribution for managerial talent affects investment value. A very talented manager can attain large alphas when he competes against mediocre rivals with low $\mu_{\tau,t}$. However, when the same manager competes against similarly talented rivals and $\mu_{\tau,t}$ is large, then everyone has the potential to exploit the same opportunities for alpha, and this affects adversely the performance of all incumbents. Therefore, a manager’s performance before fees and costs is affected by his talent relative to the environment where he competes, rather than his individual level of talent only.

The asymptotic limit where funds manage the full market portfolio is consistent with Sharpe’s arithmetic of active management (Sharpe, 1991). The active managers compete in a zero-sum game, and for every winner there is an offsetting loser. Therefore, the average gross alpha is zero.
by definition. However, this identity is valid if the market portfolio does not change and no new securities are issued (Pedersen, 2017), or active managers maintain full control of the market portfolio for every period. For any other case, the average gross alpha is not zero.

Each manager may reduce his trading and R&D costs if he invests a fraction of his assets into passive or previously assimilated strategies. This is termed “closet indexing”, because the investor cannot monitor whether all capital is actively traded or not. The drawback of closet indexing for the manager is that the investor observes a smaller realized alpha and reduces the fund’s assets in equilibrium. The manager trades actively \( x_{it}q_{it} \) of his assets under management, and invests the rest of his capital \((1 - x_{it})q_{it}\) in passive strategies that have zero alpha and \(E_t[R_{0t+1}] = 1\) by definition. The optimal fraction \(x^*_{it} < 1\) is found by maximizing the following objective:

\[
\max_{x_{it}} \left\{ f^*_{it}q^*_{it} \left( x_{it}E_t[R_{it+1}] + (1 - x_{it})E_t[R_{0t+1}] \right) - \frac{c(x_{it}q^*_{it})^2(1 - HH^*_t)}{2HH^*_t} - hx_{it}q^*_it HH^*_t \right\}.
\] (16)

The asterisks in equation (16) denote equilibrium values for capital and fees that are determined by equations (4) and (7). The investor allocates capital to the fund based on its expected return \(x^*_{it}E_t[R_{it+1}] + (1 - x^*_{it})E_t[R_{0t+1}]\) in equilibrium.

**Lemma 5.** In equilibrium, each manager \(i\) trades actively at time \(t\) only a fraction of his assets under management given by

\[
x^*_{it} = \frac{f^*_{it}(E_t[R_{it+1}] - 1) - hHH^*_t}{cq^*_it(1 - HH^*_t)}.
\] (17)

Lemma 5 shows that the managers with the largest alphas are more active, and tend to increase the fraction of actively managed assets when fees and the Herfindahl index are large. However, managers tend to become more passive for increasing fund size or average R&D cost \(h\). In equilibrium, a declining \(HH_t\) implies a rising number of funds, increasing talent among competitors.
on average, and decreasing dispersion of talent. As the competition intensifies and opportunities to obtain alpha deplete, active managers become more passive. Indexing is less risky than active management, suggesting that aggregate risk is reduced as $x_t^*$ decreases for every fund. The funds transform over the life cycle from investment vehicles that offer positive surplus from active trading to a large pool of capital where risk and performance are similar to indexing at the margin. This is the intuition of “today’s alpha” becoming inevitably “tomorrow’s beta”, where the opportunities for alpha are gradually competed away and assimilated into the benchmark.\(^{10}\)

PROPOSITION 4 (**Today’s alpha is tomorrow’s beta**). Manager gross alphas decline in absolute magnitude, and the average active fund becomes the benchmark as the funds hold an increasing fraction of the market portfolio. The aggregate risk is reduced over time. At the end of the life cycle, the asset class becomes a perfectly competitive market and net entry is zero as $HH_t^* \rightarrow 0$.

The investor in my model represents the aggregate demand for active management. The aggregate demand remains large during the late stage of the life cycle despite the poor track records on average. Although a very competitive group of funds is deficient in alpha, it provides a platform to investors for large managed pools of capital that are scalable and well diversified. The scalability stems from indexing the bulk of managed assets, because indexed capital is not subject to diminishing returns to scale. The investor diversifies his wealth optimally across managers on the efficient frontier, instead of allocating all capital to the passive fund and be subject to a single source of market risk. In principle, the life cycle will evolve as long as there are investors who chase alpha, until alpha is fully depleted in the aggregate. However, Proposition 4 implies that most investors value the benefits of diversification more than alpha. This is a potential explanation for the popularity of equity mutual funds, despite the lack of superior returns on average relative to passive funds.

\(^{10}\)For instance, Stulz (2007) predicts that the performance gap between hedge funds and mutual funds will narrow, and hedge funds will become more regulated and less risky in the future.
B. Closed-end funds and improving skill

The quadratic trading cost in the main setup implies diminishing returns to scale at the fund level. This implies that capital inflows following large realized returns may erode the fund’s future performance. Some managers may choose to close their fund to new investment and cap their assets, so that they can maintain large returns. This practice does not change the dynamics of the model. The closed-ended funds are still subject to diminishing returns to scale at the industry level, because the trading cost includes the Herfindahl index. The performance for every fund depends on the environment where managers compete, in addition to the fund’s own size. A closed-end fund that invests its assets actively may still be forced to exit as the average talent increases over time and $HH_t^*$ declines. As a result, enforcing a fund size cap is not a dominant strategy in the long term, because it is optimal to be open to new investment and charge fees while indexing a fraction of the assets under management.

Another possible extension is to allow improvements in managerial skill with experience. It is likely that managers of long-lived funds have improved their ability in identifying profitable opportunities or reducing costs. This can be modeled by extending the exogenous talent level $\tau_i$ with the sum of a fixed component and a linear trend. This would extend the manager’s lifetime, but it doesn’t change the dynamics of the life cycle at the industry level. The results in Propositions 1 to 4 involve aggregated variables, and are invariant to the identity of incumbent managers. Other extensions to the functional form of talent, such as models with manager effort and career concerns (Holmström, 1999) cannot change the evolution of the life cycle either for the same reason.

V. Empirical implications

Recent evidence by Pástor et al. (2015) show that equity mutual funds have diminishing returns to scale at the industry level. My model provides a potential theoretical explanation for the origin of this result, which is the depletion of the opportunities for alpha (Propositions 1 and 3). Gårleanu and Pedersen (2016) provide an alternative explanation for diseconomies of scale, based on the
competition among managers and informational inefficiency in investment management. However, their model does not predict increasing returns to scale at the industry level. Proposition 2 of this paper suggests that the equity mutual fund industry must have operated under increasing returns to scale during the early growth phase of its life cycle. The same prediction applies for other fund classes too, such as bond mutual funds, hedge funds, and other alternative investment funds.

Within mature fund classes like equity mutual funds, the empirical literature shows insignificant net abnormal returns and low gross abnormal returns on average. These results have been interpreted through rent-seeking managers that successfully absorb all surplus from investment (e.g. Berk and Green (2004)). I argue that mature fund classes have little alpha to offer. This should make it hard to measure alpha with statistical precision. This is consistent with the empirical results of Kosowski et al. (2006) and Fama and French (2010), who show that most equity mutual funds have zero alpha and managers at the tails have small positive and negative alphas.

My model shows that fund performance is not persistent at the industry level, and predicts a non-monotonic evolution over time for the return of the average fund. It increases initially, but eventually declines. Depending on the choice of sample period, average performance may seem persistent due to the effect of the network externality. However, the model predicts that performance must eventually erode, along with the profitable investment opportunities at the industry level.

The model predicts that the aggregate size of equity mutual funds will continue to grow in the long run even with poor abnormal returns, as they have evolved into a relatively safe investment vehicle that can manage large pools of capital at low cost. This result is similar to Glode (2011), who justifies the negative expected performance as an insurance premium that investors pay to protect themselves against bad states of the economy. The reduced aggregate risk in my model during strong competition stems from a large fraction of indexed assets. This is consistent with Cremers and Petajisto (2009), who show an increase by 30% for the fraction of closet indexers among equity mutual fund managers since 1980.

The competition among managers and the availability of investment opportunities link the returns to scale at the industry level with the curvature of the flow-performance relation. Funds with
limited opportunities have diminishing returns to scale and convex flows. Equity mutual funds are an asset class that fits this description. The literature on the convexity of flows within equity mutual funds is extensive.\textsuperscript{11} The model predicts that this fund class has diminishing returns to scale at the industry level, and Pástor et al. (2015) verify this prediction. However, when managers have sufficient investment opportunities they should have increasing returns to scale and concave flow-performance relations. Goldstein et al. (2017), Kaplan and Schoar (2005), and Getmansky (2012) find concave flows among corporate bond funds, private equity, and hedge funds respectively. These fund classes are good candidates to test for increasing returns to scale at the industry level. Sensoy et al. (2014) show a maturing of the private equity industry that resembles the life cycle evolution of my model.

The mechanism for the life cycle of active management allows the creation of “mega funds”, namely funds with significantly larger assets under management than rivals. The most successful managers over the life cycle will be those who entered early with a talent level that is deep in the right tail of the prior distribution $\mathcal{H}(\tau_i)$. These managers receive the largest inflows over time (Corollary 1), and their long tenure allows them to amass large amounts of capital. The existence of mega funds does not affect the final number of funds.

\section*{VI. Conclusion}

This paper provides a model for the life cycle of investment management. Fund managers compete for investor flows and profitable opportunities. The markets learn about managerial talent from the fund’s track record. The driving forces for the life cycle are the competition among funds and the investor’s response to performance.

During the early stage of the life cycle, the competition among managers triggers a network externality for the investor and funds operate under increasing returns to industry scale. The flow-performance relation is concave. The investor surplus from alpha increases during this stage. As the competition among managers intensifies, the opportunities for abnormal returns are curtailed

\textsuperscript{11}See also a survey for this literature by Christoffersen et al. (2014).
and the funds operate under diminishing returns to industry scale. The flow-performance relation becomes convex, and the total surplus from active investing is depleted by the end of the life cycle. Active investing outperforms passive strategies during the early stage, although managers become more passive over time to mitigate trading costs. By the end of the life cycle, the managers index all their assets.

Appendix

A. Proofs

Proof of LEMMA 1. To simplify the notation for this proof, I omit the time subscript $t$ from vectors. The Lagrangian for the investor’s problem (see equation (5)) is

$$\mathcal{L} = E_t[q'(r - f)] - \frac{\gamma}{2} q'V q - \lambda(q'1 - W_t) , \tag{A.1}$$

where $q = (q_{0t}, q_{1t}, \ldots, q_{Nt})'$ is the fund size vector, $r = (r_{0t+1}, r_{1t+1}, \ldots, r_{Nt+1})'$ is the fund nominal return vector, $f = (f_{0t}, f_{1t}, \ldots, f_{Nt})'$ is the fund fee vector, $1$ is a $(N + 1) \times 1$ vector of ones, $V$ is the $(N + 1) \times (N + 1)$ covariance matrix for the returns of the passive index and the incumbent active managers, while $\gamma, W_t, \text{ and } \lambda$ are constants.

The first-order condition for $q'$ is

$$E_t[r - f] - \gamma V q = \lambda 1 \tag{A.2}$$

and the optimal fund sizes are given by

$$q^* = \frac{1}{\gamma} V^{-1} [E_t[r - f] - \lambda 1] \equiv \frac{1}{\gamma} \left[ E_t[R - f] - \lambda V^{-1} 1 \right] , \tag{A.3}$$
where the net-of-fee alpha is defined as the net-of-fee Sharpe ratio

\[ E_t[R - f] \equiv V^{-1} E_t[r - f] \quad (A.4) \]

with \( R = (1, R_{1t+1}, \ldots, R_{Nt+1}) \) the vector of risk-adjusted fund returns in excess of the benchmark, i.e. the fund gross alphas. The first element of \( R \) corresponds to the passive index, and it has zero alpha by definition. Multiplying equation (A.3) by \( 1' \) allows to reconstruct the budget constraint and solve for \( \lambda \)

\[ 1' q^* = \frac{1}{\gamma} \left[ 1' E_t[R - f] - \lambda \left( 1' V^{-1} 1 \right) \right] = W_t \Rightarrow \]

\[ \lambda = \frac{1' E_t[R - f]}{\left( 1' V^{-1} 1 \right)} - \frac{W_t \gamma}{\left( 1' V^{-1} 1 \right)} . \quad (A.6) \]

Substituting for \( \lambda \) in equation (A.3) gives the equilibrium demand function

\[ q^* = \frac{W_t}{\left( 1' V^{-1} 1 \right)} V^{-1} 1 + \frac{1}{\gamma} E_t[R - f] - \frac{1' E_t[R - f]}{\gamma \left( 1' V^{-1} 1 \right)} V^{-1} 1 . \quad (A.7) \]

As a result, the demand function for fund \( i \) at time \( t \) is given by

\[ q^*_{it} = \frac{W_t}{\gamma} \sum_{j=0}^{N_t} \omega_{ij} + \frac{1}{\gamma} E_t[R_{it+1} - f_{it}] - \frac{\sum_{j=0}^{N_t} \omega_{ij}}{\gamma} \sum_{k=0}^{N_t} \sum_{j \geq k} \omega_{kj} E_t[R_{jt+1} - f_{jt}] , \quad (A.8) \]

where \( \omega_{ij} \) the matrix element of \( V^{-1} \) at row \( i \) and column \( j \). The following ratio has order of magnitude

\[ \frac{\sum_{j=0}^{N_t} \omega_{ij}}{\sum_{k=0}^{N_t} \sum_{j \geq k} \omega_{kj}} = O \left( \frac{1}{N_t + 1} \right) . \quad (A.9) \]
As a result, the optimal fund size may be written as

$$q_{it}^* = \frac{\gamma W_t b_{it}}{N_t + 1} + b_1 E_t[R_{it+1} - f_{it}] - \frac{b_{it}}{N_t + 1} \sum_{j=0}^{N_t} E_t[R_{jt+1} - f_{jt}]$$  \hspace{1cm} (A.10)

with $b_1 > b_{it}/(N_t+1)$ for all $N_t \geq 1$. The covariance matrix $V$ is positive definite, implying that its inverse is positive definite too. Therefore, the coefficient $b_{it}$ is positive for all funds $i$, because it involves summations of matrix elements along a dimension of $V^{-1}$.

Proof of LEMMA 2. The correlation between size and expected performance describes the flow-performance relation. The derivative

$$\frac{\partial q_{it}^*}{\partial E_t[R_{it+1}]}$$  \hspace{1cm} (A.11)

has no root for positive values of $HH_t$, and its limiting values are

$$\lim_{HH_t \to 0} \left[ \frac{\partial q_{it}^*}{\partial E_t[R_{it+1}]} \right] = 0 \quad (A.12)$$

$$\lim_{HH_t \to 1} \left[ \frac{\partial q_{it}^*}{\partial E_t[R_{it+1}]} \right] = \frac{b_1^2 (N_t + 1) \left( 2E_t[R_{it+1}]^2 + \hat{\sigma}_{\tau,t}^2/2 + h \right)}{2b_1 (N_t + 1) E_t[R_{it+1}]^2} - \frac{b_1 b_{it} \left( \hat{\sigma}_{\tau,t}^2/2 + h \right)}{2b_1 (N_t + 1) E_t[R_{it+1}]^2} > 0 \quad , (A.13)$$

because $b_1 (N_t + 1) > b_{it}$. As a result,

$$\frac{\partial q_{it}^*}{\partial E_t[R_{it+1}]} > 0 \quad \text{for} \quad HH_t > 0 \quad , (A.14)$$

which implies a monotonically increasing flow-performance. The equality is satisfied when $HH_t$ declines to zero.

$\Box$
Proof of PROPOSITION 1. The weighted average talent among the incumbent managers $\mu_{\tau,t}$ at time $t$ is

$$
\mu_{\tau,t} \equiv \frac{1}{Q_t} \sum_{i=1}^{N_t} q_{it} \tau_i,
$$

(A.15)

using fund assets under management as weights. Lemma 2 shows that managers with larger expected returns are assigned more capital. The worst-performing managers lose capital, while the best ones receive more capital. This implies that the investor flows gradually decrease the weights in equation (A.15) from the left tail of the distribution for talent. On the other hand, the flows increase the weights on the right tail of the distribution over time. As the uncertainty about talent decreases over time, the estimate $\hat{\mu}_{\tau,t}$ converges to $\mu_{\tau,t}$. Therefore, the weighted average talent shifts upward

$$
\frac{d\mu_{\tau,t}}{dt} > 0.
$$

(A.16)

Appendix B shows that the average talent creates a barrier to entry, and the cross-sectional distribution of talent among incumbents is truncated at the left tail. The barrier to entry during a period is the talent value for the marginal incumbent with zero profits. The rising $\mu_{\tau,t}$ implies that the marginal incumbent with talent $\hat{\tau}_{min,t}$ at time $t$ is also more talented over time. All incumbents before time $t$ with talent $\tau_i < \hat{\tau}_{min,t}$ will start losing capital from time $t$ and beyond. In the long run, these managers are forced to exit by liquidation, and the distribution of talent among the incumbents is truncated within the range $\tau_i \in [\hat{\tau}_{min,t}, \infty)$. Since the lower bound increases over time, the cross-sectional dispersion $\sigma_{\tau,t}$ among the incumbent managers must decline, i.e.

$$
\frac{d\sigma_{\tau,t}}{dt} < 0.
$$

(A.17)
Proof of \textit{COROLLARY 1}. Let $t_1 < t_2$. Proposition 1 shows that $\sigma_{\tau,t_1} > \sigma_{\tau,t_2}$. Therefore, the dispersion of the prior distribution for successful entrants at $t_1$ is larger than the corresponding prior for the cohort at $t_2$. Equations (B.16) to (B.18) in the Appendix show that a smaller variance in the prior ($\sigma^2$ in that notation) reduces the Kalman gain, and the market puts more weight on the previous estimate rather than the observations from realized returns. This implies that the smaller uncertainty for every fund in the cohort at $t_2$ gets resolved faster. The fund flows by the investor are his response to the innovations about performance. Since the weight on observations is smaller for the cohort at time $t_2$, the flows to every manager will be smaller too. The smaller Kalman gain implies that two managers with the same talent and identical track records of realized returns will receive flows of different magnitude if they entered at different time periods.

\[\square\]

\textit{Proof of \textit{LEMMA 3}.} The equation

$$\frac{\partial f^*_{it}}{\partial E_t[R_{it+1}]} = 0 \quad \text{(A.18)}$$

is a polynomial of degree six. According to the Abel–Ruffini theorem, polynomials with abstract coefficients of degree five or higher may lack a closed-form solution. Depending on parameter values, equation (A.18) has either of the following forms

$$-a_0 - a_1 HH_t - a_2 HH_t^2 + a_3 HH_t^3 + a_4 HH_t^4 + a_5 HH_t^5 + a_6 HH_t^6 = 0 \quad \text{(A.19)}$$

$$-a_0 - a_1 HH_t - a_2 HH_t^2 - a_3 HH_t^3 - a_4 HH_t^4 + a_5 HH_t^5 + a_6 HH_t^6 = 0 \quad \text{(A.20)}$$

where the coefficients $a_0$ to $a_6$ are positive. Descartes’ rule of signs implies that this polynomial has a unique positive root $HH_{thr}$, and five negative or complex roots.

Define the following auxiliary variables

$$I_t \equiv \sum_{j=0}^{N_t} \frac{b_{jt}E_t[R_{jt+1}]}{b_1 c(1 - HH_t) + 2E_t[R_{jt+1}]HH_t} \quad \text{(A.21)}$$
\[
J_t \equiv \sum_{j=0}^{N_t} \frac{b_{jt}}{b_{0t}} \frac{b_{jt}}{b_{0t}} \left(1 - HH_t \right) + 2E_t[R_{jt+1}]HH_t .
\]

(A.22)

The correlations for industry-level returns to scale and the curvature of the flow performance relation are

\[
\frac{\partial Q^*_t}{\partial E_t[R_{it+1}]} = - \left[ b_1 b_{0t} (b_{1t} + 2E_t[R_{it+1}]y_t) \right] \cdot \\
\left[ b_1^2 c((N_t + 1) - cJ_t) - y_t(2I_t y_t - b_{it})E_t[R_{it+1}] + \\
b_1 \left( 2(N_t + 1)E_t[R_{it+1}]y_t - c(I_t y_t - b_{it} + \\
2J_t E_t[R_{it+1}]y_t) \right)^{-1} \cdot \frac{\partial f^*_t}{\partial E_t[R_{it+1}]} \right]^{\frac{1}{2}}.
\]

(A.23)

\[
\frac{\partial^2 q^*_t}{\partial E_t[R_{it+1}]} = 2b_1 y_t \left[ 2b_1^4 c^2 (N_t + 1 - cJ_t)^2 + \\
4I_t y_t^3 (2I_t y_t - b_{it})E_t[R_{it+1}]^2 + b_1 \left( b_1^2 c(N_t + 1 - cJ_t) \cdot \\
(8(N_t + 1)E_t[R_{it+1}]y_t - c(4I_t y_t - b_{it} + 8J_t E_t[R_{it+1}]y_t) + \\
b_1^2 y_t(8(N_t + 1)^2 E_t[R_{it+1}]y_t^2 - 16c(N_t + 1)E_t[R_{it+1}]y_t(I_t + \\
J_t E_t[R_{it+1}]) + c^2(2I_t^2 y_t - I_t b_{it} + \\
16J_t I_t E_t[R_{it+1}]y_t + 8J_t E_t[R_{it+1}]y_t^2)) - b_1 E_t[R_{it+1}]y_t \cdot \\
\left( 4(N_t + 1)y_t(4I_t y_t + b_{it})E_t[R_{it+1}] - c(8I_t y_t^2 + 16J_t I_t y_t^2 E_t[R_{it+1}] \\
- b_{it}(b_{it} - 4J_t E_t[R_{it+1}]y_t)) \right) \right] \cdot \frac{\partial f^*_t}{\partial E_t[R_{it+1}]},
\]

(A.24)

where \(I_t\) and \(J_t\) are defined in equations (A.21) and (A.22) respectively, and \(b_{0t}\) is the coefficient \(b_{it}\) for the passive fund \((i = 0)\). The equations above show that these two correlations are proportional to the sensitivity of fund fees to performance. As a result, they have the same positive root \(HH_{thr}\).
Proof of LEMMA 4. The number of funds in equilibrium is given by the zero-profit condition in equation (10). The solution of this equation for large $\hat{\sigma}_{\tau,t}$ is

$$N_t^* = \left[ E_t[R_{mt+1} \left(1 + \gamma b_{mt} (-1 + 2SR - (h + \hat{\sigma}^2_{\tau,t}/2)SRinv - 2\gamma W_t) + h + \hat{\sigma}^2_{\tau,t}/2 \right. \right.$$

$$- 2E_t[R_{mt+1}^2] \cdot \left. \left( E_t[R_{mt+1}] (2E_t[R_{mt+1}] + \gamma b_{mt} - 1) - h - \hat{\sigma}^2_{\tau,t}/2 \right) \right]^{-1} , \quad \text{(A.25)}$$

where $m$ is the marginal incumbent fund, $SR \equiv \sum_{j=0}^{N_t} E_t[R_{jt+1}]$ is the aggregate index of fund returns, and $SRinv \equiv \sum_{j=0}^{N_t} 1/E_t[R_{jt+1}]$. The first condition for the network effect to initiate is that the number of funds increases with performance in the aggregate, including the performance of the marginal incumbent

$$\frac{\partial N_t^*}{\partial E_t[R_{jt+1}]} = \gamma b_{mt} \left( -2SR + 2\gamma W_t + (h + \hat{\sigma}^2_{\tau,t}/2)SRinv \right) \left( h + \hat{\sigma}^2_{\tau,t}/2 \right.$$

$$\left. + E_t[R_{mt+1}]^2 \right) \cdot \left[ \left( E_t[R_{mt+1}] (2E_t[R_{mt+1}] + \gamma b_{mt} - 1) - h - \hat{\sigma}^2_{\tau,t}/2 \right) \right]^{-1} > 0 . \quad \text{(A.26)}$$

Using the approximation $SRinv \approx (N_t + 1)$, the condition becomes

$$\frac{SR}{N_t + 1} \approx \frac{\hat{\mu}_{\tau,t}}{N_t + 1} < \frac{1}{2} \left( h + \frac{\hat{\sigma}^2_{\tau,t}}{2} \right) + \frac{\gamma W_t}{N_t + 1} . \quad \text{(A.27)}$$

The return of the benchmark has zero alpha by definition, implying $E_t[R_{0t+1}] = 1$ for the passive fund $i = 0$ (see equation (1)). The average active fund will outperform net-of-fees the passive benchmark if

$$\frac{\sum_{j=0}^{N_t} (E_t[R_{jt+1}] - f_{jt}^*)}{N_t + 1} > 1 , \quad \text{(A.28)}$$
which gives
\[
\frac{SR}{N_t + 1} \approx \mu_{\tau, t} + \frac{1}{2} \left( h + \frac{\sigma_{\tau, t}^2}{2} \right) + \frac{\gamma W_t}{2(N_t + 1)}. \tag{A.29}
\]

The Herfindahl index is negatively correlated with the number of funds
\[
\frac{\partial HH_t^*}{\partial N_t} = \gamma W_t b_1 b_{0t} \left[ 2 \left( \frac{SR}{N_t + 1} - \frac{\gamma W_t}{N_t + 1} \right) - \left( 1 + h + \frac{\sigma_{\tau, t}^2}{2} \right) \right] < 0, \tag{A.30}
\]
because of condition (A.27). As a result, the number of funds increases and the Herfindahl index decreases during the early stage of the life cycle.

\[\Box\]

Proof of PROPOSITION 2. Lemma 4 shows that the average active fund has positive alpha after the network effect begins. The returns to scale at the industry level, the curvature of the flow-performance relation, and the sensitivity of a fund’s fee relative to its performance at the asymptotic limit for $HH_t \to 1$ are given by
\[
\lim_{HH_t \to 1} \frac{\partial f^*_{it}}{\partial E_t[R_{it+1}]} = \left( \frac{(N_t + 1)b_1 + b_{it}}{2(N_t + 1)b_1 E_t[R_{it+1}]} \right) < 0, \tag{A.31}
\]
\[
\lim_{HH_t \to 1} \frac{\partial Q^*_{it}}{\partial E_t[R_{it+1}]} = \left( \frac{b_{0t}(h + \frac{\sigma_{\tau, t}^2}{2})}{(N_t + 1)E_t[R_{it+1}]} \right) > 0, \tag{A.32}
\]
\[
\lim_{HH_t \to 1} \frac{\partial^2 q^*_{it}}{\partial E_t[R_{it+1}]^2} = \left( \frac{(h + \frac{\sigma_{\tau, t}^2}{2})(N_t + 1)b_1 - b_{it}}{(N_t + 1)E_t[R_{it+1}]^3} \right) > 0, \tag{A.33}
\]
for all funds $i$. As a result, the returns to scale are increasing at the industry level, the flow-performance relation is concave, and the fees are negatively correlated with expected returns within the range $HH_t \in (HH_{thr}, 1]$. The investor’s surplus increases over time, because the network effect increases gross returns and competition decreases fund fees. As a result, the net-of-fee returns increase in the aggregate.

\[\Box\]
Proof of PROPOSITION 3. Let’s assume that the average fund has negative net alpha. This implies the following inequality

\[
\frac{\sum_{j=0}^{N_t} (E_t[R_{jt+1}] - f^*_j)}{N_t + 1} < 1.
\]  

(A.34)

The net-of-fees return for the average fund at small \( HH_t \) is

\[
\frac{\sum_{j=0}^{N_t} (E_t[R_{jt+1}] - f^*_j)}{N_t + 1} = \frac{\sum_{j=0}^{N_t} (E_t[R_{jt+1}]x - cW_t)}{(N_t + 1)x},
\]

(A.35)

where \( x \equiv 2HH_tE_tE_t[R_{it+1}]/(b_1c(1 - HH_t)) \) a small quantity. Solving for the average gross return gives

\[
\frac{\sum_{j=0}^{N_t} E_t[R_{jt+1}]}{N_t + 1} < 1 + \frac{cW_t}{x},
\]

(A.36)

which is always true because \( x \) is small. As a result, the assumption that the average active fund underperforms relative to the benchmark holds true.

Lemma 3 shows that \( HH_{thr} \) is the unique root for the industry-level returns to scale, the curvature of the flow-performance relation, and the correlation between fees and gross returns. The Herfindahl index is \( HH_t < HH_{thr} \) during the late stage of the life cycle. Proposition 2 shows that

\[
\frac{\partial Q^*_t}{\partial E_t[R_{it+1}]} > 0, \quad \frac{\partial^2 q^*_it}{\partial E_t[R_{it+1}]^2} < 0 \quad \text{and} \quad \frac{\partial f^*_it}{\partial E_t[R_{it+1}]} < 0
\]

(A.37)

for all funds \( i \) when \( HH_t > HH_{thr} \). As a result, each of these correlations has the opposite sign within the range \( HH_t \in (0, HH_{thr}) \). Therefore, the funds have diminishing returns to scale at the industry level, the flow-performance is convex, and fees are positively correlated with gross returns during the late stage of the life cycle.
The equilibrium demand for an active manager is zero when the profitable opportunities are depleted near $HH_t \to 0$. As a result, the total surplus from active investing

$$TS_t \equiv \sum_{j=0}^{N_t} \left( q_{jt}^* E_t[R_{jt+1}] - \frac{cq_{jt}^2(1 - HH_t)}{2HH_t} - hq_{jt} \right)$$  \hspace{1cm} (A.38)

also declines asymptotically to zero.

\[ \square \]

*Proof of Lemma 5.* The first-order condition from equation (16) for fund $i$ at time $t$ is

$$f_{it}^* q_{it}^* (E_t[R_{it+1}] - 1) - \frac{cx_{it}(q_{it}^*)^2(1 - HH_t^*)}{HH_t^*} - hq_{it}^* = 0$$  \hspace{1cm} (A.39)

which implies a fraction of assets under management

$$x_{it}^* = \frac{f_{it}^* (E_t[R_{it+1}] - 1) - h}{cq_{it}^*(1 - HH_t^*)}$$  \hspace{1cm} (A.40)

that is actively invested in equilibrium.

\[ \square \]

*Proof of Proposition 4.* Equation (15) shows that every manager’s alpha declines as $HH_t^*$ decreases and $\mu_{\tau,t}$ increases. The weighted average return for the endogenous benchmark shifts progressively toward the average talent in the cross-section. The aggregate risk is reduced over the life cycle, because managers index progressively larger portions of their assets (Lemma 5). The profits from active investing decline asymptotically to zero as $HH_t^* \to 0$. At this limit, the net entry is zero because managers are indifferent between entering or exiting. Since all managers have the same talent as $\sigma_{\tau,t} \to 0$ and expected profits are zero, they have a common fee $f_t^*$. Therefore, the funds participate in a perfect competition.

\[ \square \]
B. Learning with a truncated distribution

The learning process is identical to all funds, and it is independent from the time of entry. The prior probability density for the talent of potential entrants is a normal distribution $\mathcal{H}(\tau_i) \sim N(\mu, \sigma^2)$. The investor updates his estimates for the talent of every fund by observing their returns. During any period, the marginal incumbent whose profits satisfy the zero-profit condition (equation (10)) is the least talented among the competing managers. Therefore, the conditional distribution for the cross section of talent among incumbents $\mathcal{H}_{t_1}(\tau_i)$ at time $t_1$ is a normal distribution that is truncated at its left tail. The truncation point $\hat{\tau}_{\text{min},t_1}$ is defined as the talent of the marginal fund at time $t_1$. Any potential entrant who draws at time $t_1$ from the unrestricted prior distribution $\mathcal{H}(\tau_i)$ a value of talent $\tau_i < \hat{\tau}_{\text{min},t_1}$ would not enter successfully. The investor considers this effect in equilibrium, and the prior distribution for a cohort of managers who enter successfully at $t_1$ will also be truncated at $\hat{\tau}_{\text{min},t_1}$.

In order to get a closed-form solution for learning, the noise terms $\varepsilon_{it+1}$ only need to be jointly distributed over time and across managers with zero mean. I simplify the notation by assuming independence among those shocks, i.e. $\varepsilon_{it+1} \sim N\left(0, \sigma^2_{\varepsilon_i}\right)$ in equation (1). This is a strong assumption, but the formulas are scalar rather than in matrix format. The results from learning are unaffected by correlated returns. I focus on a single cohort and reset the time of entry at $t_1 = 1$. As a result, all quantities that are indexed by time $t$ refer to $t$ periods after the fund’s entry.

The normality and independence assumptions for the entrant talent and noise terms allow the use of a simple Kalman filter for learning. What is unusual about the learning process in this model, is that the prior distribution of talent among incumbents is truncated. The standard Kalman filter solution gives estimates for every period that are unbiased but suboptimal in terms of variance. On the other hand, the constrained Kalman filter for truncated normal priors by Simon and Simon (2010) gives estimates that are optimal in terms of variance, but potentially biased for every period. This extension to the standard filter optimally includes the hard inequality constraint $\tau_i \geq \hat{\tau}_{\text{min},t_1}$.
Alternative versions of constrained Kalman filters use projection methods and give estimators that are both optimal and unbiased. Their disadvantage is the lack of a closed-form solution.

Once the manager enters, his talent is fixed over time. This simplification allows to write the learning problem in the following state space notation

\[ \tau_{it+1} = \tau_{it} \] 
\[ y_{it} = \tau_{it} + \varepsilon_{it} \]  

where (B.1) is the state transition equation, and (B.2) is the observation equation. The investor observes over time the process \( y_{it} \), which is defined as

\[ y_{it} \equiv R_{it} - 1 \]  

The recursive equations from the unconstrained Kalman filter are

\[ K_t = \frac{\sigma^2}{\sigma^2 + t\sigma^2} \]  
\[ \tilde{\tau}_{it+1} = (1 - K_t)\tilde{\tau}_{it} + K_t y_{it} \]  
\[ \tilde{P}_{it+1}^2 = \frac{\sigma^2 K_t}{\sigma^2 + t\sigma^2} \]  

where \( K_t \) is the Kalman gain, and \( \tilde{\tau}_{it} \) is the conditional expectation for talent during the period \( t \) to \( t + 1 \), using the full history of observations until time \( t \). Also, \( \tilde{P}_{it}^2 \) is the conditional variance of the talent estimator \( \tilde{\tau}_{it} \).

The solutions for the posterior mean and variance in talent are

\[ \tilde{\tau}_{it} = \frac{\sigma^2 \mu_t + t \sigma^2 \bar{y}_i}{\sigma^2 + t \sigma^2} \]  
\[ \tilde{P}_{it}^2 = \frac{(\sigma \sigma_{\bar{y}_i})^2}{\sigma^2 + t \sigma^2} \]  

36
where $\bar{R}_t$ is the sample average of all innovations on $R_{it}$ for fund $i$ until time $t$. The estimator $\tilde{\tau}_{it}$ is optimal when the mean and variance of an unconstrained normal distribution are used. However, it can also be used for the constrained problem. The estimator $\tilde{\tau}_{it}$ is unbiased for the posterior, but as I show below it is not efficient in terms of variance.

The constrained Kalman filter of Simon and Simon (2010) uses the posterior mean and variance for truncated normal priors. Their algorithm involves transforming the unconstrained state variable into a standard normal variable, then enforce the constraint with a renormalization of the density function, and apply the reverse transformation to retrieve the constrained posterior for talent. The resulting estimate is the mean of the truncated posterior with a variance that is smaller than the one from the unconstrained Kalman filter. However, the estimator can be biased, and the direction of the bias is hard to assess.

The prior for the talent of potential entrants at time $t_1$ is distributed as $\mathcal{H}_{t_1}(\tau_i) \sim N(\mu, \sigma^2)$ but for $\tau_i \geq \hat{\tau}_{\text{min},t_1}$. Ignoring the constraint for now, define $z_i \equiv (\tau_i - \mu)/\sigma$ as the auxiliary transformation of talent to a standard normal variable. Applying the constraint implies that the minimum $z_i$ is

$$z_{\text{min}} = \frac{\hat{\tau}_{\text{min},t_1} - \mu}{\sigma} .$$  \hspace{1cm} (B.9)

The normalization constant of the truncated normal is

$$\frac{1}{M} \int_{z_{\text{min}}}^{\infty} e^{-\zeta^2/2} d\zeta = 1 \iff M = \sqrt{\frac{\pi}{2}} \left[ 1 - \text{erf} \left( \frac{z_{\text{min}}}{\sqrt{2}} \right) \right] ,$$  \hspace{1cm} (B.10)

where $\text{erf}(\cdot)$ is the error function. The mean and variance of the truncated distribution for $z_i$ are

$$E[z_i] = \frac{1}{M} \int_{z_{\text{min}}}^{\infty} \zeta e^{-\zeta^2/2} d\zeta = \frac{1}{M} e^{-\frac{z_{\text{min}}^2}{2}} > 0$$  \hspace{1cm} (B.11)

$$\text{Var}(z_i) = \frac{1}{M} \int_{z_{\text{min}}}^{\infty} (\zeta - E[z_i])^2 e^{-\zeta^2/2} d\zeta = 1 - E[z_i]^2 - z_{\text{min}} E[z_i] < 1 .$$  \hspace{1cm} (B.12)
The reverse transformation on the constrained $z_i$ gives the constrained Kalman filter estimator for the manager’s talent with a truncated prior. The reverse transformation is simply

$$
\tau_i = \sigma z_i + \mu ,
$$

where $\mu$ and $\sigma$ are the mean and standard deviation respectively of the unconstrained density function, implying that their posterior values could be retrieved from the standard Kalman filter solution (equations (B.7) and (B.8) respectively). As a result, the posterior estimates for the constrained state variable at time $t$ are

$$
\hat{\tau}_{it} = \sqrt{\hat{P}_{it}^2} \ E[z_i] + \tau_{it}
$$

$$
\hat{P}_{it} = Var(z_i) \hat{P}_{it}^2 ,
$$

where the conditional estimates use all history of observations until time $t$. Interestingly, $\hat{P}_{it}^2 < \hat{P}_{it}^2$ which makes the constrained estimator more efficient in terms of variance. In the asymptotic limit $t \to \infty$ the variances tend to zero and the two estimators are identical. As a result, the uncertainty about the manager’s true talent is resolved over time through the learning process.

On the other hand, the posterior estimate for talent can be biased for every period, since the estimator’s mean value is not necessarily equal to the true mean of the population for potential entrants. The direction of this bias is hard to assess. The potential bias in the posterior estimates does not stem from the existence of the lower bound for talent, but rather is an artifact of the specific algorithm for the learning process. Alternative algorithms for constrained Kalman filters provide both efficient and unbiased estimators, but lack a closed-form solution (Simon and Simon, 2010). However, the artificial bias does not affect any features of the model, because it is related to the true mean of the distribution for the potential entrants $\mathcal{H}_{t_1}(\tau_i)$, not the distribution of talent among the incumbents who have successfully entered the industry.
For the purpose of clarity, the recursive equations for the constrained Kalman filter are

\[ K_{t-1} = \frac{\sigma^2}{\sigma_{\xi_i}^2 + (t-1)\sigma^2} \]  
\[ \hat{\tau}_{it} = \sigma_{\xi_i} E[z_i] K_{t-1}^{3/2} + (1 - K_{t-1}) \hat{\tau}_{it-1} + K_{t-1} y_{it-1} \]  
\[ \hat{P}_{it}^2 = \text{Var}(z_i) \sigma_{\xi_i}^2 K_{t-1} \]  
\[ (B.16) \]  
\[ (B.17) \]  
\[ (B.18) \]

The true value of talent \( \tau_i \) for fund \( i \) is subject to estimation errors \( u_{it} \). This implies that the true value of talent is equal to

\[ \tau_i = \hat{\tau}_{it} + u_{it} \]  
\[ (B.19) \]

with the conditional variance \( \text{Var}_t(u_{it}) = \hat{P}_{it}^2 \) from equation (B.18).

The estimates for the cross-sectional weighted mean and dispersion of talent among the incumbent managers are

\[ \hat{\mu}_{\tau,t} = \frac{1}{Q_t} \sum_{i=1}^{N_t} q_{it} \hat{\tau}_{it} \]  
\[ (B.20) \]

\[ \hat{\sigma}_{\tau,t} = \left[ \frac{1}{Q_t} \sum_{i=1}^{N_t} q_{it} (\hat{\tau}_{it} - \hat{\mu}_{\tau,t})^2 \right]^{1/2} \]  
\[ (B.21) \]

From equation (B.19), the cross-sectional dispersion of talent \( \sigma_{\tau,t} \) is equal to

\[ \sigma_{\tau,t} = \left[ \hat{\sigma}_{\tau,t}^2 + \frac{1}{Q_t} \sum_{i=1}^{N_t} q_{it} \hat{P}_{it}^2 \right]^{1/2} \]  
\[ (B.22) \]

As the uncertainty about each value of talent is resolved over time and \( \hat{P}_{it}^2 \) tends asymptotically to zero for all funds \( i \), the cross-sectional estimates \( \hat{\mu}_{\tau,t} \) and \( \hat{\sigma}_{\tau,t} \) converge to the population mean \( \mu_{\tau,t} \) and dispersion \( \sigma_{\tau,t} \) respectively.
C. Differentiated Bertrand competition

The sensitivity of fund fees to performance for a differentiated Bertrand competition and large values of \( \hat{\sigma}_{\tau,t} \) is
\[
\lim_{\hat{\sigma}_{\tau,t} \to \infty} \frac{\partial f^*_it}{\partial E_t[R_{it+1}]} = \frac{((N_t + 1)b_1 - b_{it})(E_t[R_{it+1}]^2 - h)}{(2(N_t + 1)b_1 - b_{it})E_t[R_{it+1}]^2} > 0. \tag{C.1}
\]

The first and second derivatives of fund size on performance are
\[
\lim_{\hat{\sigma}_{\tau,t} \to \infty} \frac{\partial q^*_it}{\partial E_t[R_{it+1}]} = \left[ ((N_t + 1)b_1 - b_{it})\left((N_t + 1)b_1 \left(E_t[R_{it+1}]^2 + h\right) - b_{it}E_t[R_{it+1}]^2\right) \right] \cdot \left[ (N_t + 1)(2(N_t + 1)b_1 - b_{it})E_t[R_{it+1}]^2\right]^{-1} > 0 \tag{C.2}
\]
\[
\lim_{\hat{\sigma}_{\tau,t} \to \infty} \frac{\partial^2 q^*_it}{\partial E_t[R_{it+1}]^2} = -\frac{2b_1h((N_t + 1)b_1 - b_{it})}{(2(N_t + 1)b_1 - b_{it})E_t[R_{it+1}]^3} < 0, \tag{C.3}
\]
which imply that the flow-performance relation is monotonically increasing and concave. The correlation between aggregate demand and fund performance is
\[
\lim_{\hat{\sigma}_{\tau,t} \to \infty} \frac{\partial Q^*_t}{\partial E_t[R_{it+1}]} = b_1 - \frac{1}{N_t + 1} \sum_{j=1}^{N_t} b_{jt} > 0, \tag{C.4}
\]
and the funds operate under increasing returns to aggregate scale. A fund’s fee must be non-negative, giving the following constraint for the index of returns
\[
SR \equiv \sum_{j=0}^{N_t} E_t[R_{jt+1}] \geq \frac{\gamma W_t b_{it} E_t[R_{it+1}] + b_1(N_t + 1)(E_t[R_{it+1}]^2 + h) - b_{it}h}{b_{it}E_t[R_{it+1}]}, > \frac{b_1(N_t + 1)}{b_{it}} > N_t + 1 \tag{C.5}
\]
proving that the aggregate gross alpha is positive, and the average active fund outperforms passive funds. As a result, the results from a differentiated Bertrand competition are consistent with those from monopolistic competition during the early stage of the life cycle.
References


Chen, Yong, and Wayne Ferson, 2016, How many good and bad funds are there, really?, *Working paper*.


Feldman, David, Konark Saxena, and Jingrui Xu, 2016, Is the active fund management industry concentrated enough?, *Working paper*.


42


