IS THE ACTIVE FUND MANAGEMENT INDUSTRY CONCENTRATED ENOUGH?

David Feldman, Konark Saxena and Jingrui Xu\textsuperscript{a,b}

Latest Revision October 7, 2016

\textbf{Abstract.} We study the effects of market concentration levels on the active fund management industry (AFMI). We introduce a model of an AFMI equilibria in which size, performance, and effort are endogenously determined under a continuum of exogenous market concentration levels. Higher market concentration (for a given number of funds) leaves more unexplored investment opportunities and allows managers to more efficiently use industry resources, making marginal managerial effort more productive in creating gross alphas. However, with higher market concentration, managers can get higher compensation for their effort, causing a higher opportunity cost of effort. We find that in equilibrium, higher market concentration levels induce higher net alphas and AFMI size (the ratio of assets under active management to total wealth) if and only if gains from better investment opportunities exceed the consequences of higher managerial costs. We specialize the model to allow endogenous concentration levels and, using the Herfindahl-Hirschman and other indices, empirically study its key predictions in the United States equity AFMI in the last four decades. We find that, on average, AFMI net alphas and AFMI size increase with market concentration. Given the current low market concentration in the U.S. AFMI and with no change in managerial productivity/effort opportunity costs, an increasing market concentration is likely to increase both AFMI net alphas and size. We also look at equilibria with colluding fund managers and examine AFMI’s direct benefits.

\textbf{JEL Codes:} G10, G20, L10

\textbf{Keywords:} Active management, Mutual funds, Effort, Performance, Market concentration, Competition, Herfindahl-Hirschman index (HHI), Industry size, Net alpha, Equilibrium, Returns to scale, Learning

\textsuperscript{a} All from Banking and Finance, UNSW Business School, UNSW Australia, UNSW Sydney, NSW 2052, Australia. Emails: d.feldman@unsw.edu.au, k.saxena@unsw.edu.au, jingrui.xu@unsw.edu.au.

1 Introduction

Two central underpinnings of free market economics are 1) competition leads to better outcomes and 2) agents earn economic rents if and only if they have a competitive advantage. Because the incentives of earning future economic rents are crucial in motivating people to act and because people need an environment where a competitive advantage can be created so they can earn future economic rents, it is natural to try to understand whether the level of competition (or concentration\(^1\)) in a given industry is optimal. This question is at the core of a central financial economic issue: the efficiency of the active fund management industry (AFMI) equilibrium. Extensive literature on the AFMI has focused on trying to understand the economic forces that can explain fund manager compensation, their ability to generate value, and the exponential growth of an industry where identifying economic value added seems elusive.\(^2\)

We introduce a model of AFMI competition, effort\(^3\), size, and performance, and provide a novel perspective of the U.S. active equity mutual AFMI. Specifically, we note that competition among asset management firms has grown dramatically over the past few decades with advancements in financial products and technology (see, for example, Gruber (1996) and Philippon and Reshef, (2012)). Worldwide, vast numbers of active fund managers are estimating the value of assets each day. These highly trained experts act to exploit any perceived differential—however small—between price and estimated asset value, hoping to be compensated for their efforts. This phenomenon raises important questions. Clearly, one needs some active management to ensure that security prices properly reflect relevant information, but do market concentration levels in the AFMI optimally balance opportunities and costs of gross alpha production? Our model provides economic insights regarding two opposing forces that influence economic outcomes when the concentration level of AFMI changes: available gross alpha-production opportunities and the corresponding effort costs.\(^4\)

We define market concentration to reflect the market competitiveness level.\(^5\) Two hypotheses about the effects of market concentration prevail in the banking literature: the efficient-structure hypothesis, which suggests a positive relation between market

---

\(^1\) We use concentration and competition as opposites.

\(^2\) See, for example, Jensen (1968), Daniel, Grinblatt, Titman and Wermers (1997), Wermers (2000), Berk and Green (2004), Chan, Covrig and Ng (2005), Khorana, Servaes and Tufano (2005), Khorana, Servaes and Tufano (2008), Pastor and Stambaugh (2012), Ferreira, Keswani, Miguel and Ramos (2012a) and (2012b), and Berk and Binsbergen (2015).

\(^3\) The “effort” in our model may be regarded as an effort-skill combination.

\(^4\) Managers’ efforts’ cost levels may be viewed, of course, as reflections of their skills.

\(^5\) Our results hold for any \(M, M > 1\). Determining \(M\) endogenously would not change our results.
concentration and firm efficiency, and the structure-conduct-performance hypothesis, which asserts a positive relation between market concentration and firms’ performance due to extractions of monopolistic rents. We note that these two hypotheses are not necessarily mutually exclusive. Following these hypotheses, we expect higher market concentration to

1) leave more unexplored investment opportunities and allow fund managers to more efficiently use industry resources, such as human capital, inducing higher marginal effort efficiency; and

2) facilitate an increase in the opportunity cost of effort, inducing fund managers to require more compensation for effort.

Our model allows incorporating these effects of market concentration, calibrating parameters with real-world data, and ascertaining which implications are consistent with market equilibrium.

Pastor and Stambaugh (2012), (PS), studied a framework with decreasing returns to scale (i.e., decreasing active managers’ marginal ability to outperform passive benchmarks), where interactions between active fund managers and investors determine expected net alphas (net of management fees). Within their world, we model a continuum of market concentration levels and allow active fund managers to (optimally) exert costly effort when competing over investment funds by producing net alphas. We study equilibria with four types of investors: a single risk-neutral investor, infinitely many risk-neutral investors, a single risk-averse investor, and infinitely many risk-averse investors.

Within an AFMI equilibrium, we study the impact of changes in market concentration on endogenous managerial costly effort, endogenous net alpha production, and endogenous AFMI size. As in PS, and without loss of generality, we fix the number of funds and define AFMI size as the ratio of assets under active management to total wealth. We show that with

---

6 These hypotheses and some related hypotheses, such as relative market power hypothesis, are discussed and tested, for example, in Berger and Hannan (1989), Berger (1995), and Goldberg and Rai (1996).

7 An example for increasing productivity with concentration is as follows. Suppose that gold has been found and a handful of diggers yield high returns with little effort. As the number of diggers increases, the area available to each digger, and with it diggers’ “productivity,” decreases, even under optimal effort. Fund managers are like gold diggers; they seek net alpha, but as their number increases, unexplored opportunities decline concomitant with their productivity.

8 In the literature, market concentration levels are exogenous, or endogenous (see, for example, Aguerrevere (2009) and Ambrose, Diop, and Yoshida (2014)); but, here, assuming exogenous concentration helps put a focus on what we study.

9 Each of these cases fits particular economies. The cases where investors are risk-neutral are consistent with the situation where households employ private banks or other financial institutions to manage their wealth. These institutions then invest in active funds. Acting as intermediaries, they mainly care about funds’ expected returns so may be regarded as risk-neutral investors. The cases where there is a single investor in the market is consistent with the situation where the societal wealth is centrally managed by the government, which determines the allocation of investments to active funds and other securities.
infinitely many mean-variance risk-averse investors, the effects of higher market concentration depend on a tradeoff. Higher market concentration not only increases opportunities in gross alpha production (the marginal efficiency of managerial effort in producing gross alpha) but also increases managerial effort costs due to increase in opportunity costs. We further show that if the former effect dominates (is dominated by) the latter one, higher market concentration induces higher (lower) equilibrium fund expected net alphas and larger (smaller) AFMI size.

We define the AFMI’s direct benefit (function) as the (equilibrium) increase in gross alpha induced by (optimal) efforts level minus the cost of these efforts, for given market concentration levels.\(^{10,11}\) We show that equilibrium fund expected net alphas and AFMI size increase (decrease) with market concentration if and only if the AFMI’s direct benefits increase (decrease) with market concentration. Also, we find that if equilibrium fund expected net alphas are concave in market concentration, then the AFMI’s direct benefits are concave in market concentration. Consequently, equilibrium AFMI size is also concave in market concentration. On the other hand,\(^{12}\) if equilibrium AFMI size is convex in market concentration, then the AFMI’s direct benefits are convex in market concentration and, consequently, equilibrium expected fund net alphas are convex in market concentration. Which of the cases prevail in the real world is an empirical question, which we study below.

Here, as in the literature, by construction, aggregate net alphas are zero-sum as they shift wealth between subsets of investors; see, for example, discussion in PS.\(^{13,14}\)

We specialize our model to allow endogenous concentration levels, and, using the Herfindahl-Hirschman and other indices, study it empirically. We study how the fund net alphas and size of the U.S. active equity mutual fund industry change with market concentration. We find that both fund net alphas and AFMI size, on average, increase with market concentration. Moreover, both fund net alphas and AFMI size are concave in concentration.

\(^{10}\) For brevity and simplicity, we use the term *benefits* in the general sense allowing for negative benefits.

\(^{11}\) The AFMI, however, may have indirect benefits which we do not model here. For example, monitoring, studying, and analyzing firms might incentivize managements to improve governance and productivity, and to reduce agency costs. Also, active fund management may induce transfer of wealth from less productive firms/investors to more productive ones, and even within endowment economies, may increase investors’ derived utilities by improving information processing and/or risk sharing.

\(^{12}\) Please note that the order of statements of this argument is different from that in the previous one, for reasons explained below.

\(^{13}\) Pages 748-750, including footnote 6, and references therein. As in PS, we do not model the “other” investors, who facilitate the zero-sum.

\(^{14}\) We note that this is also true at non-zero AFMI’s direct and/or indirect benefits, as we measure net alphas ex-post under the equilibrium active fund management level.
Our empirical methodology uses Pastor, Stambaugh, and Taylor’s (2015) (PST) recursive-demeaning estimator to address endogeneity and omitted-variable-related issues, and uses vector auto-regression (VAR) techniques to account for simultaneity in determination of AFMI size and market concentration. We control for survival bias by using the Morningstar U.S. mutual fund database, which contains both surviving and terminated funds. Our empirical results are robust to the use of alternative methods and measures.

Our empirical findings are consistent with our model’s theoretical implications under plausible parameter values, and have policy implications. Given the low market concentration in the current AFMI, and assuming no change in the tradeoff of managerial productivity and effort cost, increased market concentration is likely to increase both fund net alphas and AFMI size; under plausible parameter values, AFMI’s direct benefits also increase. The literature has shown multiple ways to increase fund market concentration. For example, Massa (2003) suggested that mutual fund families allow investors to move money across family funds of different categories at low costs, lowering effective fees and reducing competition among funds. Under our findings, formation of fund families in such markets is likely to be socially efficient.

In addition to the above topics we look at equilibria with colluding fund managers.

Section 2 develops the theoretical model, Section 3 presents the empirical methods and results, and Section 4 concludes.

2 Theoretical Framework

We first develop a theoretical framework for modeling the effect of market concentration on fund managers’ effort, fund fees, fund performance, AFMI size, and potential benefits.

2.1 Setting

For brevity and parsimonious notation, we assume that variables and functions are real, continuous, and twice differentiable. Within a one-period market, there are two types of agents: fund managers of \( M \), \( M > 1 \) funds and \( N \), \( N \geq 1 \), investors. Acting competitively, each manager, conditional on fund size and market concentration level, sets a proportional management fee and chooses an effort level to maximize the fund expected net alpha to attract investments.\(^{15}\) In Case I, risk-neutral investors allocate investments among the \( M \)

\(^{15}\) In our framework, the competition among managers is Bertrand competition, where the number of competitors is \( M \), and the “prices” offered by managers in competition are fund net alphas net of management fees.
actively managed funds and a passive benchmark index fund to maximize their portfolios’ expected returns. In Case II, mean-variance risk-averse investors allocate their investments to maximize their portfolios’ Sharpe ratios.

Following PS, $r_F$, a vector of $M$ funds’ returns in excess of the riskless rate that investors receive, follows the regression model

$$r_F = \alpha + \beta r_p + u,$$

where $r_F$ is an $M \times 1$ vector with elements $r_{F,i}$, $i = 1, \ldots, M$.

The benchmark-adjusted returns on the $M$ funds that investors receive is

$$r = \hat{a} + u.$$

The variables $\beta$, $r$, $\alpha$, and $u$ are $M \times 1$ vectors. $\alpha$ is the vector of fund net alphas received by investors, and $\beta$ is the vector of fund betas. Funds hold fully diversified portfolios and, thus, have unit betas; therefore, $\beta$ is a unit vector. The scalar $r_p$ is the excess return on the passive benchmark portfolio, with mean $\mu_p$ and variance $\sigma^2_p$, $u$ is the residual vector, with elements that follow

$$u_i = x + \epsilon_i, \quad i = 1, \ldots, M,$$

where $\epsilon_i$’s are mean zero and variance $\sigma^2_\epsilon$ idiosyncratic risks, and are uncorrelated with each other, with $x$, and with $r_p$. The common factor $x$ has mean zero, variance of $\sigma^2_x$, and is uncorrelated with $r_p$. The values of $\mu_p$, $\sigma^2_p$, $\sigma^2_\epsilon$, and $\sigma^2_x$ are constants that are common knowledge of both investors and managers.

Each element in $\alpha$ has the following structure:

$$\alpha_i = a - b \frac{S_i}{W} + A(\epsilon_i; H) - f_i,$$

where $S$ is the aggregate size of the active management industry and is equal to the sum of all the funds’ sizes (i.e., $S = \sum_{i=1}^{M} s_i$); $W$ is equal to $S$ plus the amount invested in the passive benchmark; $a$ and $b$ are positive, unknown scalar parameters; $a$ is the expected return on an initial small fraction of wealth invested in active management, net of any costs; and $b$ is the absolute magnitude of the decreasing returns to scale at industry level. The first and second conditional moments of $a$ and $b$ are
where $D$ is investors’ information set. As we do not focus on $\sigma_{ab}$’s effects on the equilibrium, we assume that $\sigma_{ab} = 0$. In other words, conditional on current information, we assume that how $\hat{a}$ deviates from $a$ is unrelated to how $\hat{b}$ deviates from $b$. Our reasoning is that $a$ and $b$ are parameters requiring different estimation methods and $\sigma_{ab}$ tends to be small in comparison to $\sigma_a^2$ and $\sigma_b^2$; thus, the assumption of $\sigma_{ab} = 0$ is reasonable. Finally, with $f_i$ being a proportional management fee charged by manager $i$, manager $i$’s fund expected net alpha is

$$E(\alpha_i | D) = \hat{a} - \frac{\hat{b} S}{W} + \hat{A}(e_i; H) - f_i.$$  

(7)

Our model follows and builds on that of PS. In this partial equilibrium, the passive benchmark portfolio’s returns are exogenously given and are unaffected by interactions between investors and managers. Managers’ outperformance of the passive benchmark portfolio (i.e., net alphas), may come at the expense of “other investors,” who may be noise traders, liquidity seekers, misinformed, or irrational.\(^{17}\)

We allow manager $i$ to spend a non-negative amount of proportional effort $e_i$ (i.e., $e_i \in [0, \infty)$) to increase gross alpha (i.e., the alpha before subtracting the fee) by $A(e_i; H)$ under market concentration level $H$. Industrial organization theory suggests that market concentration not only depends on the number of incumbents, but also on threats of entry activity-limiting regulation and the competitiveness of related industries.\(^{18}\) We assume that

\(^{16}\) Investors observe the passive benchmark and the AFMI funds’ returns. The difference between these returns comes from three components: net alphas, the common risk factor, and idiosyncratic risks. As the distributions of the common risk and idiosyncratic risk are common knowledge, investors know the likelihood function of the net alphas. Given prior beliefs of net alphas, they form posteriors and update their beliefs. In our one-period model, there is no dynamic Bayesian updating, but we suggest that investors reached a fixed-point equilibrium. Further, because investors observe $f_i, H, S$ and $W$, they can also infer $A(e_i, H)$. Here, where equilibrium optimal effort levels of all managers are same, the estimate $\hat{A}(e_i, H)$ could be subsumed in $\hat{a}$; and in an equilibrium where managers’ optimal effort levels differ, the estimates $\hat{A}(e_i, H)$, could be subsumed in $f_i$. For simplicity and brevity, we depress the notation of $\hat{A}(e_i, H)$ in favor of $A(e_i, H)$ and follow PS formulation, Equations (5) and (6).

\(^{17}\) Please see the detailed discussion in PS, pp. 748–750.

\(^{18}\) Please see the discussion by Claessens and Laeven (2003).
$H$ is an exogenous constant because it depends mainly on some exogenous factors mentioned above.\textsuperscript{19} It belongs to $[0,1)$. If $H = 0$, there are infinitely many small funds in the market, and the market is fully competitive. If $H = 1$, the market is monopolistic. If the fund managers are competing with each other, $H$ belongs to $[0,1)$, and this is $H$’s domain in our framework. $A(e_i; H)$ is the same across funds and has the following functional characteristics:

- non-negative, i.e., $A(e_i; H) > 0$, $\forall e_i > 0, H$, $A(0; H) = 0$, $\forall H$, 
- increasing and concave in effort, i.e., $A_{e_i}(e_i; H) \equiv \partial A(e_i; H) / \partial e_i > 0$, $\forall e_i, H$, and $A_{e_i,e_i}(e_i; H) \equiv \partial^2 A(e_i; H) / \partial e_i^2 < 0$, $\forall e_i, H$, 
- increasing in market concentration, i.e., $A_H(e_i; H) \equiv \partial A(e_i; H) / \partial H > 0$, $\forall e_i > 0, H$, 
- positive cross-partial derivatives with respect to effort and market concentration, i.e., $A_{e_i,H}(e_i; H) = A_{H,e_i}(e_i; H) \equiv \partial^2 A(e_i; H) / \partial H \partial e_i > 0$, $\forall e_i, H$.

The economic sense of the structure of $A(e_i; H)$ is as follows. A particular positive level of effort has a positive impact on gross alpha production. If managers spend no effort, the increment in gross alpha due to their effort is zero (i.e., $A(e_i; H) \geq 0$, $\forall e_i, H$ and $A(0; H) \geq 0$, $\forall H$). Under a particular market concentration level $H$, an increase in effort $e_i$ may increase gross alpha, but the marginal increment is decreasing (i.e., $A_{e_i}(e_i; H) > 0$ and $A_{e_i,e_i}(e_i; H) < 0$, $\forall e_i, H$). The more concentrated the AFMI is, the relatively more investment opportunities there are, and the more marginally efficient is the use of industry resources.\textsuperscript{20} Thus, managers can generate a higher increment in gross alpha for a given effort level $e_i$ (i.e., $A_H(e_i; H) > 0$ ), and the marginal impact of $e_i$ on gross alpha is also larger (i.e., $A_{e_i,H}(e_i; H) = A_{H,e_i}(e_i; H) > 0$, $\forall e_i, H$). These two assumptions of the partial derivatives are consistent with Hoberg, Kumar and Prabhala (2015), who found that a higher competition level limits managers’ skills to create gross alpha persistently.

**Managers’ Cost**

We assume that funds’ fixed costs are zero because fixed costs to develop funds, such

\textsuperscript{19} In Section 13, we examine endogenous concentration.
\textsuperscript{20} In a more concentrated market, if a fund manager controls most of the industry resources and develops advanced strategies to produce gross alphas, other funds can mimic this fund’s strategy and produce higher gross alphas given a particular effort level, so this assumption is still valid when a dominant fund in the market controls the majority of the resources.
as registration fees and equipment expenditure, are usually trivial compared to variable costs related to employees’ salaries and managers’ compensation. We assume that average cost functions, \( C_i(e_i,s_i;H) \), contain three independent positive components: \( c_{0,i} \), the average cost for fund \( i \) to operate in the market before receiving investment and before manager \( i \) spends effort; \( c_{1,i} \), the average cost related to fund size; and \( c_{2,i}(e_i;H) \), the average cost of managerial effort under a particular market concentration.\(^{21}\) That is,

\[
C_i(e_i,s_i;H) = c_{0,i} + c_{1,i}s_i + c_{2,i}(e_i;H). \tag{8}
\]

To simplify our model, we let \( c_{0,i} \), and the function \( c_{2,i}(e_i;H) \) be the same across funds (we, thus, drop the subscript \( i \) from now on), but let \( c_{1,i} \) be different across funds. We discuss the effects of similarities and differences in these parameters across funds later. The function \( c_2(e_i;H) \) has the following characteristics:

- non-negative, i.e., \( c_2(e_i;H) > 0, \forall e_i > 0, H \) and \( c_2(0;H) = 0, \forall H \),
- increasing and convex in effort, i.e., \( c_{2,e_i}(e_i;H) = \partial^2 c_2(e_i;H) / \partial e_i^2 > 0, \forall e_i, H \),
- increasing in market concentration, i.e., \( c_{2,H}(e_i;H) = \partial^2 c_2(e_i;H) / \partial H > 0, \forall e_i > 0, H \),
- positive cross-partial derivatives with respect to effort and market concentration, i.e.,

\[
c_{2,e_i,H}(e_i;H) = c_{2,H,e_i}(e_i;H) = \partial^2 c_2(e_i;H) / \partial e_i \partial H > 0, \forall e_i, H. \tag{9}
\]

The average cost function implies that if fund size \( s_i \) increases, manager \( i \) ‘s average cost increases because the larger trades are associated with larger price impacts and higher execution costs and because of other factors that create diseconomies of scale in operation. \( c_{1,i} \) is the average cost sensitivity to fund size \( s_i \). Also, the cost of the fund is related to the incentive scheme that offers increasing and convex bonuses for employees’ performance. Thus, the average cost function is increasing and convex in effort \( e_i \) (\( c_{2,e_i}(e_i;H) > 0 \) and \( c_{2,e_i,H}(e_i;H) > 0, \forall e_i, H \)). In addition, in markets that are more concentrated, it is more costly to incentivize managerial efforts because market compensation levels for effort are higher.

\(^{21}\) To simplify our model, we assume there is no interaction between effort and size in the average cost function because it is unlikely that fund size affects managers’ per dollar effort. We also assume that there is no interaction between concentration and size in the average cost function because it is unlikely that concentration affects managers’ average cost sensitivities to fund sizes. Nevertheless, even if these interacting effects exist, they tend to be small in comparison to effects of other terms in the average cost function.
and/or because managers are less industrious. Thus, if markets are more concentrated, average costs due to effort are higher \( (c_{2H}(e_i; H) > 0, \forall e_i, H) \), and marginal costs due to effort are higher as well \( (c_{2e,H}(e_i; H) = c_{2H,e_i}(e_i; H) > 0, \forall e_i, H) \).

The total cost function of manager \( i \) is Equation (8) times fund size \( s_i \), so manager \( i \)'s total cost function is convex in \( s_i \). Therefore, our fund cost model is consistent with that of Berk and Green (2004), who assumed decreasing returns to scale at fund level. Empirically, PST also reported evidence consistent with fund-level decreasing returns to scale.

We define AFMI’s direct benefit function of manager \( i \)'s net alpha production as

\[
B(e_i; H) = A(e_i; H) - c_2(e_i; H).
\]

\( B(e_i; H) \) captures the direct benefit from effort exerted in active fund management, in terms of increase in gross alpha production minus the effort cost. We note that AFMI’s active search for net alphas might have indirect effects that we do not model here. It might drive security prices toward their true values; it might induce firms to improve governance and performance, and reduce agency costs. It might induce transfer of wealth from less productive firms/investors to more productive ones. We note that we should interpret benefits generally, allowing them to be positive or negative.

Whether the AFMI’s direct and or indirect benefits are non-zero or zero, here, as in the literature, gross alphas are zero-sum, because we measure gross alphas ex-post in the AFMI’s equilibrium. (See for example PS, pp. 748-750, including footnote 6, and references therein.)

### 2.2 Fund Managers’ Problem

Manager \( i \)'s economic profit is

\[
s_i\left(f_i - C'(e_i, s_i; H)\right),
\]

and for the fund \( i \) to survive,

\[
f_i - C'(e_i, s_i; H) \geq 0.
\]

Manager \( i \)'s problem can be written as

\[
\text{Max } s_i \left(f_i - C'(e_i, s_i; H)\right)
\]

subject to

\[
e_i \geq 0,
\]
\( f_i \geq 0.22 \)

Risk-neutral investors would invest only in funds that generate the highest expected net alphas. For mean-variance risk-averse investors, we assume that the benefit of diversification across funds approaches zero because funds are well diversified and because the uncertainty of the parameters \( x, a \) and \( b \) are likely to be much larger than the uncertainty of \( \epsilon_i \). Also, real-world investors tend not to invest in many funds and do not diversify across funds. Thus, it is likely that not only are total diversification benefits zero, but also likely that marginal diversification benefits are trivial. Thus, our risk-averse investors invest only in funds that generate the highest expected net alphas and managers have to compete over net alphas. Manager \( i \)'s problem can be transformed as,

\[
\max_{\epsilon \sim f} \mathbb{E}(\alpha_i | D) \quad \text{subject to} \quad f_i - C^i(\epsilon_i, s_i; H) \geq 0, \quad \epsilon_i \geq 0 \quad \text{and} \quad f_i \geq 0.
\]

Proof. See the Appendix.

The proof intuition is as follows. Under competition, funds that offer higher expected net alphas draw (all) investments. Thus, in equilibrium funds offer similar expected net alphas. The possibility that other managers increase fund profits by improving expected net alphas, and their fund sizes, induces managers to maximize expected net alphas in order to “survive.” We note that this aspect of the equilibrium is similar to that in PS, but in addition to their result, we show that it holds also in the case of finite number of managers, under Bertrand competition.

The following propositions provide results of fund managers’ equilibrium optimal effort levels and fees.

**PROPOSITION 1.** For manager \( i, i = 1, 2, ..., M \), if initial effort inputs generate non-positive AFMI’s direct benefits of net alpha production (i.e., \( B^i_b(0; H) \leq 0, \forall H \)), equilibrium optimal proportional effort levels \( e_i^* \) are zero (i.e., \( e_i^* = 0 \)) and the optimal proportional fee \( f_i^* \) equals the average cost of operating funds \( c_0 + c_{1,s_i} \) (i.e., \( f_i^* = c_{0,i} + c_{1,s_i} \)).

**COROLLARY to PROPOSITION 1.** Under the case of Proposition 1, the equilibrium is similar to that in PS. The managerial effort is not modeled, and managers optimally choose

---

22 For simplicity and brevity, we omit the condition in Equation (11) from the problem statement as it is implied by the optimization and, thus, is not binding.
not to charge fees above opportunity costs.

PROPOSITION 2. For manager \(i, i=1,2,...,M\), if initial effort inputs generate positive AFMI’s direct benefits of net alpha production (i.e., \(B_{ij}(0;H) > 0, \forall H\)), equilibrium optimal effort-fee combinations \((e^*_i, f^*_i)\) satisfy the following.

1) \(f^*_i - C^*(e^*_i, s_i; H) = 0\) (optimal fees are set to be equal to costs).

2) \(A_{ij}(e^*_i; H) - c_{2ij}(e^*_i; H) = B_{ij}(e^*_i; H) = 0\) (the impact of marginal efforts on gross alpha is set to be equal to the marginal average costs of effort, thus marginal AFMI’s direct benefits of net alpha production under the optimal effort level are zero).

3) \(e^*_i(H) \geq 0(<0)\) iff \(A_{ij}(e^*_i; H) - c_{2ij}(e^*_i; H) \geq 0(<0)\), where \(e^*_i'(H) \equiv \frac{de^*_i}{dH}\) (where concentration is higher, equilibrium optimal efforts are higher (lower) if and only if higher concentration induces a larger (smaller) marginal effort impact on gross alphas than on costs).

4) The signs of \(e^*_i'(H) \equiv \frac{de^*_i}{dH}\) depend on the signs of \(\frac{d(S/W)}{dH}\) and \(e^*_i'(H)\) (whether higher concentrations induce higher equilibrium optimal fees depends on whether they induce an increase in equilibrium industry sizes and whether they induce an increase in equilibrium optimal efforts).

5) \(B(e^*_i; H) \geq 0(<0)\) iff \(A_{ij}(e^*_i; H) - c_{2ij}(e^*_i; H) \geq 0(<0)\), where \(B'(e^*_i; H) \equiv \frac{dB(e^*_i; H)}{dH}\) (where concentrations are higher, equilibrium AFMI’s direct benefits of net alpha production are higher (lower) if and only if higher concentrations induce a larger (smaller) impact on gross alphas than on costs).

6) \(E(\alpha_i | D) = E(\alpha_j | D), \forall i, j\) (in equilibrium, managers offer market competitive net alphas).

7) \(E(r_{F,i} | D) / \sqrt{\text{Var}(r_{F,i} | D)} = E(r_{F,j} | D) / \sqrt{\text{Var}(r_{F,j} | D)}, \forall i, j\) (in equilibrium, managers offer market competitive Sharpe ratios).

Proofs of Proposition 1 and 2, and corollary. See the Mathematical Appendix.

The proof intuition is as follows. While competing for investments, managers maximize fund expected net alphas by choosing optimal efforts and fees, earning zero economic profits (break-even fees) in equilibrium. If higher concentration levels induce a higher (lower) marginal effort impact on gross alphas than a marginal effort impact on costs,
managers optimally choose higher (lower) effort levels in producing fund net alphas. Also, concentration affects managers’ costs by increasing levels of cost due to effort and by influencing levels of optimal (costly) efforts. If higher concentrations induce higher equilibrium optimal efforts, managers’ costs are driven higher, resulting in higher break-even fees. Otherwise, where concentrations are higher, increases in cost levels due to effort are cancelled out by decreases in optimal effort levels. In this case, negative relations between break-even fees, which are equal to costs, and concentration indicate negative relations between equilibrium optimal effort levels and concentration. In addition, higher concentrations have two effects on the AFMI’s direct benefits of net alpha production: directly increasing the levels of gross alphas and costs due to effort levels, and changing equilibrium optimal effort levels, consequently changing gross alphas and costs. In equilibrium, the latter effect is zero because the marginal effort impact on gross alphas is equal to the marginal effort impact on costs and the effect of higher concentration through effort on gross alphas is cancelled out by its effects through effort on costs. Therefore, if higher concentrations induce a higher direct impact on gross alphas than on costs, AFMI’s direct benefits of net alpha production are higher.

Also, as there are no diversification benefits across funds, managers who provide higher expected net alphas dominate, attracting investments. Consequently, their fund costs increase, inducing higher (break-even) fees and lowering expected net alphas. Thus, in equilibrium, allocation of investments, or fund sizes, set expected net alphas to be equal across funds. If fund managers cannot produce the AFMI highest expected net alpha, even for an infinitesimal fund size, their funds go out of the market. In addition, as funds have the same expected net alphas, they have the same expected returns. As the source of fund returns’ variance is the same across funds, the fund return variance is the same across funds. Therefore, managers offer the same competitive Sharpe ratio.

The following proposition identifies the relation of different managers’ equilibrium optimal effort levels, fees, benefits of effort, and AFMI share.

**PROPOSITION 3.** Under the same \( c_0 \) and the same functional form of \( c_i(e_i; H) \) but different \( c_{ij}’s \) across funds,

1) \( e_i^* = e_j^* \) and \( f_i^* = f_j^* \), \( \forall i, j \) (equilibrium efforts and fees are the same across funds).

2) Therefore, \( B(e_i^*; H) = B(e_j^*; H) \), \( \forall i, j \) (equilibrium AFMI’s direct benefits of net alpha production are the same across funds).
3) Fund sizes relate as \( s_i / s_j = c_{i,j} / c_{j,i} \), \( \forall i, j \).

4) AFMI shares, \( s_i / S \)'s are

\[
\frac{s_i}{S} = \left( \sum_{j=1}^{M} \left( c_{i,j}^{-1} \right) \right)^{-1}, \forall i.
\]

**Proof of Proposition 3.** See the Mathematical Appendix.

The third point of Proposition 3 shows that managers’ different costs of producing gross alphas, induce different fund sizes in equilibrium. The fourth point of Proposition 3 implies that funds’ market shares are deterministic functions of \( c_{i,i} \)'s and are, thus, unaffected by the AFMI weight in total wealth, \( S / W \). In other words, how investors weight the funds inside the AFMI is unaffected by how investors weight the AFMI as a whole relative to the passive benchmark. This property facilitates later results.

Proposition 3 is driven by the fact that \( c_0 \) and the functional form of \( c_2(e_i; H) \) are the same across funds but \( c_{1,i} \)'s are different across funds. In contrast, Proposition 1 and 2 are valid even without this assumption. In fact, from Proposition 2, we can see that if \( c_{0,i}, c_{1,i} \) and the functional form of \( c_2(e_i; H) \) are different across funds, managers end up with different levels of effort, different fees, and different fund sizes in equilibrium. If all fund managers have the same \( c_0, c_1, \) and functional form of \( c_2(e_i; H) \), they end up with the same equilibrium levels of effort, fees, and fund sizes.

We define the equilibrium optimal expected net alphas of an initial marginal investment in the AFMI as \( X(e^*_i, H) \). Quantitatively,

\[
X(e^*_i, H) \triangleq \hat{a} + A(e^*_i; H) - \left[ c_0 + c_2(e^*_i; H) \right].
\]  

(14)

For the AFMI to exist, we must have positive net alphas for initial infinitesimal investments into the AFMI:

\[
X(e^*_i, H) > 0.
\]  

(15)

If Inequality (15) is violated, investors receive negative or zero net alpha from the AFMI and invest all their wealth in the passive benchmark. Also, to offer meaningful results, we assume that initial marginal allocations of effort generate positive AFMI’s direct benefits, that is,

\[
B_{e_i}(0, H) > 0, \forall i, \forall H,
\]  

(16)
such that the optimal effort \( e_i^* \) is positive, finite, and attainable, i.e.,
\[ B_i(e_i^*, H) = 0, \quad e_i^* < K, \quad \forall i, \forall H \]
for some constant \( K \). We focus on the case under Proposition 2 in the following analyses.

2.3 Investors’ Problem

Let \( \delta_j \) denote the \( M \times 1 \) vector of weights that investor \( j \) places on the \( M \) funds, with elements \( \delta_{j,i}, \quad i = 1, \ldots, M \). Thus, investor \( j \)’s excess return is
\[
r_j = \delta_j^T r + (1 - \delta_j^T \mathbf{1}_M) r_p,
\]
where \( \mathbf{1}_M \) is an \( M \times 1 \) vector with elements equal to 1. Assuming all funds have beta loadings on the benchmark equal to 1 (i.e., \( \beta = \mathbf{1}_M \)), based on (1) and (17), we have
\[
r_j = r_p + \delta_j^T (\alpha + \mathbf{u}).
\]
Further, we have
\[
E(r_j | D) = \mu_p + \delta_j^T \mathbb{E}(\alpha | D) = \mu_p + \delta_j^T \left[ a - \hat{\delta} S W + A(e_i^*; H) - f^* \right] \mathbf{1}_M, \quad \forall j.
\]
Equation (19) is valid because the fund expected net alphas are the same across funds in equilibrium as implied by Proposition 2. Also,
\[
\text{Var}(r_j | D) = \sigma_p^2 + \left[ \sigma^2 + \sigma^2 + \sigma^2 \left( \frac{S}{W} \right)^2 \right] + \sigma^2 \delta_j^T \delta_j, \quad \forall j.
\]
We characterize equilibria in the cases where 1) there are infinitely many small risk-neutral investors, 2) there is a single large risk-neutral investor, 3) there are infinitely many small mean-variance risk-averse investors, and 4) there is a single large mean-variance risk-averse investor. Here, \( \text{infinitely many small investors} \) means \( N \to \infty \) investors have finite wealth, and their choices cannot affect fund sizes. \( \text{A single large investor} \) means \( N = 1 \), a single investor who controls all market wealth, and his or her choices determine fund sizes. In the following analyses, we focus on cases 3 and 4, whereas cases 1 and 2 are discussed in the Appendix.

Mean-Variance Risk-Averse (RA) Investors

If there are infinitely many (i.e., \( N \to \infty \)) small mean-variance risk-averse investors, none of them can affect fund sizes. Also, investors’ investment in the AFMI dilutes funds’ expected returns due to decreasing returns to scale in funds. In addition, mean-variance risk-
averse investors face risk-return tradeoffs in marginal allocations. Investor \( j \)'s objective is to maximize the portfolio's Sharpe ratio by choosing portfolio weights, \( \delta_j, j = 1, \ldots, M \). This investor's problem is

\[
\max_{\delta_j} \frac{\mathbb{E}(r_j | D)}{\sqrt{\text{Var}(r_j | D)}} = \max_{\delta_j} \left\{ \frac{\mu_p + \delta_j^T (\hat{a} - \hat{b} \frac{S}{W} + A(e_i^*; H) - f_i^*) \mathbf{1}_M}{\sqrt{\sigma_p^2 + \sigma_a^2 + \sigma_b^2 \left( \frac{S}{W} \right)^2} \left( \delta_j^T \mathbf{1}_M \right)^2 + \sigma_\varepsilon^2 (\delta_j^T \delta_1)} \right\},
\]

subject to

\[
\delta_j^T \mathbf{1}_M \leq 1,
\]

\[
\delta_{j,i} \geq 0, \quad \forall i,
\]

\[
f_i^* - C_i(e_i^*, s_i; H) = 0, \quad \forall i,
\]

\[
A_i(e_i^*; H) - c_{2a} (e_i^*; H) = B_i(e_i^*; H) = 0, \quad \forall i.
\]

Condition (22) is a form of wealth constraint, saying that investors cannot borrow from the passive benchmark to invest in the AFMI. Condition (23) says that there is no short sale of funds. Conditions (24) and (25) reflect managers' equilibrium optimal choices. Also, as we assume that there are no marginal diversification benefits across funds, we set the term \( \sigma_\varepsilon^2 (\delta_j^T \delta_1) \) to zero when solving the optimization problem (21). Because the equilibrium is symmetric, we have

\[
\delta_j^T \mathbf{1}_M = S / W.
\]

Where there is a single (i.e., \( N = 1 \)) large investor, he or she determines \( s_j, j = 1, \ldots, M \), thus, \( S / W \), to maximize the portfolio’s Sharpe ratio. The investor not only faces a tradeoff between allocating additional dollars to funds, taking advantage of fund net alphas, and diluting returns on wealth already in funds, but also faces a risk-return tradeoff. The problem is

\[
\max_{\delta_j} \frac{\mathbb{E}(r_j | D)}{\sqrt{\text{Var}(r_j | D)}} = \max_{\delta_j} \left\{ \frac{\mu_p + \delta_j^T (\hat{a} - \hat{b} \delta_1 \mathbf{1}_M + A(e_i^*; H) - f_i^*) \mathbf{1}_M}{\sqrt{\sigma_p^2 + \sigma_a^2 + \sigma_b^2 (\delta_1^T \mathbf{1}_M)^2 + \sigma_\varepsilon^2 (\delta_1^T \delta_1)}} \right\},
\]

subject to Conditions (22) – (25). We also set the term \( \sigma_\varepsilon^2 (\delta_1^T \delta_1) \) to zero and also have

\[
\delta_1^T \mathbf{1}_M = S / W.
\]
numerical solution.

The next proposition describes the equilibrium in the $N \to \infty$ case.

**PROPOSITION RA1, Unique Nash Equilibrium.**

Where $N \to \infty$, 

1) there exists a unique Nash equilibrium, $\{e^*, f^*, \delta^*\}$, where 

   - $e^*$ is an $M \times 1$ vector with managers’ optimal effort allocations, $e_i^*$, 
   - $f^*$ is an $M \times 1$ vector with managers’ optimal fee allocations, $f_i^*$, and 
   - $\delta^*$ is an $M \times N$ matrix with vectors of investors’ optimal wealth weights allocations to funds, $\delta_j^*$; 

2) in this equilibrium, managers produce the same expected net alpha, which drives their economic profits to zero, by charging only break-even fees; and investors allocate the same wealth proportions to each of the funds.

**Proof of Proposition RA1.** See the Mathematical Appendix.

The following proposition identifies equilibrium properties.

**PROPOSITION RA2, Equilibrium by Optimal Allocations.**

For $i = 1, 2, ..., M$, we have $E(\alpha_i \mid D)\{e^*, f^*, \delta^*\} > 0$; and where $N \to \infty$, the equilibrium optimal $S/W$ is either 1 or a real positive solution of the (constrained embedded) first-order condition of the investors’ problem (a cubic equation) substituting $\delta_j^{*T}1_M = S/W$, 

$$-\gamma \sigma_b^2 \left( \frac{S}{W} \right)^3 - \left[ \gamma \sigma_x^2 + \gamma \sigma_x^2 + \hat{b} + \left( \sum_{i=1}^{M} c_{i, t} \right)^{-1} W \right] \frac{S}{W} + X(e^*, H) = 0,$$

where $\gamma \equiv \mu_p / \sigma_p^2$.

Where $N = 1$, numerical solutions are required.

The intuition of Proposition RA2 is as follows. Whether $N \to \infty$ or $N = 1$, investors allocate investments to funds based on their risk-return tradeoffs. Investing too much wealth in the AFMI increases portfolio risk, so they choose to limit those investments, leaving $E(\alpha_i \mid D)\{e^*, f^*, \delta^*\} > 0$. Where $N \to \infty$, the properties of the cubic equation guarantee at least one real positive root. The solution of $S/W$ is the largest or the smallest real positive root of the cubic equation or 1, depending on which maximizes (21).
COROLLARY to PROPOSITION RA2. Where $N \to \infty$, for large enough $W$, such that $S/W < 1$, we have

$$
\frac{d(S/W)}{dX(e^*_i;H)} = \frac{1}{\gamma 3\sigma_s^2 \left( \frac{S}{W} \right)^2 + \sigma_u^2 + \sigma_s^2 + \hat{b} + \left( \sum_{i=1}^M c_{i,i} \right)^{-1} W} > 0, \text{ and}
$$

$$
\frac{d(S/W)}{d \left[ \hat{b} + \left( \sum_{i=1}^M c_{i,i} \right)^{-1} W \right]} = \frac{- (S/W)}{- \gamma 3\sigma_s^2 \left( \frac{S}{W} \right)^2 + \sigma_u^2 + \sigma_s^2 + \hat{b} + \left( \sum_{i=1}^M c_{i,i} \right)^{-1} W} < 0.
$$

That is, higher initial marginal fund expected net alphas induce a larger equilibrium AFMI size relative to total wealth, whereas a stronger decreasing returns to scale effect in the AFMI induces a smaller equilibrium AFMI size relative to total wealth.

The intuition of this corollary is as follows. Where $S/W < 1$, an increase (decrease) in $X(e^*_i;H)$ shifts up (down) the cubic function in Proposition RA2, inducing a larger (smaller) $S/W$ as the maximizer of investors’ objective function. The economic sense is that a higher level of equilibrium optimal expected net alpha of an initial marginal investment, $X(e^*_i;H)$, attracts more investments to the AFMI. Also, we can see that $\hat{b}$ is the expected decreasing returns to scale at the industry level, based on current information, whereas $\left( \sum_{i=1}^M c_{i,i} \right)^{-1} W$ may be regarded as the equilibrium decreasing returns to scale factor at the fund level because it is calculated by all the fund average cost sensitivities to size, $c_{i,i}$. Thus, the factor $\hat{b} + \left( \sum_{i=1}^M c_{i,i} \right)^{-1} W$ may be regarded as the combined decreasing returns to scale factor.

Investors invest less in funds if the effect of decreasing returns to scale is stronger in the AFMI. The following proposition offers the comparative statics.

PROPOSITION RA3 AFMI Size Sensitivity to Concentration.

1) Where $N \to \infty$ and $S/W < 1$, we have

$$
a. \quad \frac{d(S/W)}{dH} = \frac{d(S/W)}{dX(e^*_i;H)} \left[ A_H(e^*_i;H) - c_{2H}(e^*_i;H) \right]
$$

$$
d(S/W)/dH \geq 0(<0) \text{ iff } A_H(e^*_i;H) - c_{2H}(e^*_i;H) \geq 0(<0) \text{ (where concentration is higher, equilibrium industry size is larger (smaller) if and only if the higher concentration induces a larger (smaller) impact on gross alphas than on costs).}
$$
b. \[
\frac{d^2 (S/W)}{dH^2} = \frac{d (S/W)}{dX(e^*_i; H)} \frac{d^2 (e^*_i; H)}{dH^2} - 6g^2 S W \left[ A_H(e^*_i; H) - c_{2H}(e^*_i; H) \right] \left[ \frac{d (S/W)}{dX(e^*_i; H)} \right]^2.
\]

If \(d^2 B(e^*_i; H) / dH^2 \leq 0\), then \(d^2 (S/W) / dH^2 \leq 0\) (the fact that \(B(e^*_i; H)\) is concave in \(H\) indicates that \(S/W\) is concave in \(H\)), and if \(d^2 (S/W) / dH^2 \geq 0\), then \(d^2 B(e^*_i; H) / dH^2 \geq 0\) (the fact that \(S/W\) is convex in \(H\) indicates that \(B(e^*_i; H)\) is convex in \(H\)).

2) Where \(N \to \infty\) and \(S/W = 1\), \(S/W\) is unrelated to market concentration.

3) Where \(N = 1\) and \(S/W < 1\), numerical solutions are required to analyze the signs of \(d (S/W) / dH\) and \(d^2 (S/W) / dH^2\).

4) Where \(N = 1\) and \(S/W = 1\), \(S/W\) is unrelated to market concentration.

The intuition is as follows. Where \(N \to \infty\), a higher \(H\) affects industry size \(S/W\) through the equilibrium optimal expected net alpha of an initial marginal investment, \(X(e^*_i; H)\). If a higher \(H\) induces a larger (smaller) impact on gross alphas than on costs, then it creates a larger (smaller) \(X(e^*_i; H)\) and, consequently, attracts more (less) investments in the AFMI—if investors have additional wealth to allocate to funds (i.e., \(S/W < 1\)). From Proposition 2, we see that \(B'(e^*_i; H) = A_H(e^*_i; H) - c_{2H}(e^*_i; H)\); thus, in this case, a higher \(H\) induces a larger \(S/W\) if and only if it induces a higher \(B(e^*_i; H)\). Regarding the second-order derivative \(d^2 (S/W) / dH^2\) where \(N \to \infty\) and \(S/W < 1\), if investors are risk-neutral (see Proposition RN3 in the Appendix), \(d^2 (S/W) / dH^2\) is positively proportional to \(d^2 B(e^*_i; H) / dH^2\). That is, as \(H\) changes, the change of marginal \(S/W\) depends on the change of marginal \(B(e^*_i; H)\) because investors share all AFMI’s direct benefits of net alpha production in a market with competing managers. However, mean-variance risk-averse investors face, in addition, a risk-return tradeoff. Holding other parameters the same, if investors’ marginal portfolio risks are higher, they optimally invest less in funds, so higher \(H\) induces smaller marginal \(S/W\) in equilibrium. Thus, \(d^2 (S/W) / dH^2\) in the risk-averse case is reduced by an adjustment term for risk (the second component in the expression of \(d^2 (S/W) / dH^2\) in the first point of Proposition RA3). We can see that if \(B(e^*_i; H)\) is
concave in \( H \), \( S / W \) must be concave in \( H \); if \( S / W \) is convex in \( H \), \( B(e^*_i; H) \) must be convex in \( H \). Where \( N = 1 \), the situation is more complex because the single investor internalizes the whole market and, further, faces an additional tradeoff between allocating additional wealth to funds to increase returns and diluting returns on wealth already in funds. If investors have no additional wealth to allocate to funds (i.e., \( S / W = 1 \)), the market is at a corner solution and \( H \) has no effect on \( S / W \).

**PROPOSITION RA4, Net Alpha Sensitivity to Concentration.**

1) Where \( N \to \infty \) and \( S / W < 1 \), we have

\[
dE(\alpha_i | D) \bigg|_{e^*, r^*, \delta^*} = \left[ A_H(e^*_i; H) - c_{2H}(e^*_i; H) \right] \left[ 1 - \left( \hat{b} + \left( \sum_{i=1}^{M} c_{i,i}^{-1} \right)^{-1} \right) W \right] d\left( \frac{S}{W} \right) dX(e^*_i; H). 
\]

\[
dE(\alpha_i | D) / dH \bigg|_{e^*, r^*, \delta^*} \geq 0 < 0 \text{ iff } A_H(e^*_i; H) - c_{2H}(e^*_i; H) \geq 0 < 0 
\]

(Where concentration is higher, the equilibrium optimal expected net alphas are larger (smaller) if and only if this higher concentration induces a larger (smaller) impact on gross alphas than on costs).

\[
d^2 E(\alpha_i | D) / dH^2 \bigg|_{e^*, r^*, \delta^*} = d^2 B(e^*_i; H) / dH^2 \left\{ 1 - \left[ \hat{b} + \left( \sum_{i=1}^{M} c_{i,i}^{-1} \right)^{-1} \right] W \right\} + 6 \sigma^2 \gamma S / W 
\]

If \( d^2 E(\alpha_i | D) / dH^2 \bigg|_{e^*, r^*, \delta^*} \leq 0 \), then \( d^2 B(e^*_i; H) / dH^2 \leq 0 \), (the fact that \( E(\alpha_i | D) \bigg|_{e^*, r^*, \delta^*} \) is concave in \( H \) indicates that \( B(e^*_i; H) \) is concave in \( H \)). If \( d^2 B(e^*_i; H) / dH^2 > 0 \), then \( d^2 E(\alpha_i | D) / dH^2 \bigg|_{e^*, r^*, \delta^*} > 0 \) (the fact that \( B(e^*_i; H) \) is convex in \( H \) indicates that \( E(\alpha_i | D) \bigg|_{e^*, r^*, \delta^*} \) is convex in \( H \)).

2) Where \( N \to \infty \) and \( S / W = 1 \), we have

\[
dE(\alpha_i | D) / dH \bigg|_{e^*, r^*, \delta^*} = A_H(e^*_i; H) - c_{2H}(e^*_i; H). 
\]

\[
dE(\alpha_i | D) / dH \bigg|_{e^*, r^*, \delta^*} \geq 0 < 0 \text{ iff } A_H(e^*_i; H) - c_{2H}(e^*_i; H) \geq 0 < 0 
\]

(Where concentration is higher, the equilibrium optimal expected net alphas are larger (smaller) if and only if this higher concentration induces a larger (smaller) impact on gross alphas than on costs).
b. \( d^2E(\alpha_i \mid D) / dH^2 \mid_{e^*,r^*,s^*} = d^2B(e^*_i; H) / dH^2 \).

d^2E(\alpha_i \mid D) / dH^2 \mid_{e^*,r^*,s^*} \geq 0(0) \text{ iff } d^2B(e^*_i; H) / dH^2 \geq 0(0) \) (\( E(\alpha_i \mid D) \mid_{e^*,r^*,s^*} \) is convex (concave) in \( H \) if and only if \( B(e^*_i; H) \) is convex (concave) in \( H \)).

3) Where \( N = 1 \) and \( S/W < 1 \), numerical solutions are required to analyze the signs of 
\( dE(\alpha_i \mid D) / dH \mid_{e^*,r^*,s^*} \) and 
\( d^2E(\alpha_i \mid D) / dH^2 \mid_{e^*,r^*,s^*} \).

4) Where \( N = 1 \) and \( S/W = 1 \), we have

a. \( dE(\alpha_i \mid D) / dH \mid_{e^*,r^*,s^*} = A_H(e^*_i; H) - c_{2H}(e^*_i; H). \)

\( dE(\alpha_i \mid D) / dH \mid_{e^*,r^*,s^*} \geq 0(0) \text{ iff } A_H(e^*_i; H) - c_{2H}(e^*_i; H) \geq 0(0) \) (where concentration is higher, the equilibrium optimal expected net alphas are larger (smaller) if and only if this higher concentration induces a larger (smaller) impact on gross alphas than on costs).

b. \( d^2E(\alpha_i \mid D) / dH^2 \mid_{e^*,r^*,s^*} = d^2B(e^*_i; H) / dH^2 \).

\( d^2E(\alpha_i \mid D) / dH^2 \mid_{e^*,r^*,s^*} \geq 0(0) \text{ iff } d^2B(e^*_i; H) / dH^2 \geq 0(0) \) (\( E(\alpha_i \mid D) \mid_{e^*,r^*,s^*} \) is convex (concave) in \( H \) if and only if \( B(e^*_i; H) \) is convex (concave) in \( H \)).

The intuition of Proposition RA4 is as follows. Where \( N \rightarrow \infty \) and \( S/W \), a higher \( H \) influences \( E(\alpha_i \mid D) \mid_{e^*,r^*,s^*} \) at two stages. At the first stage, it changes managers’ ability to produce expected net alphas, which is represented by the first component of 
\( dE(\alpha_i \mid D) / dH \mid_{e^*,r^*,s^*} \). At the second stage, investors react to the changes in fund expected net alphas by adjusting the investment level to the funds, consequently affecting 
\( E(\alpha_i \mid D) \mid_{e^*,r^*,s^*} \) under a decreasing returns to scale framework. This effect is represented by the second component of 
\( dE(\alpha_i \mid D) / dH \mid_{e^*,r^*,s^*} \). If investors are risk-neutral, they adjust their investment level merely based on the changes in fund expected net alphas, driving 
\( E(\alpha_i \mid D) \mid_{e^*,r^*,s^*} \) to zero, so the second component of 
\( dE(\alpha_i \mid D) / dH \mid_{e^*,r^*,s^*} \) is zero (see Proposition RN4 in Appendix). However, if investors are risk-averse, their risk-return tradeoff makes their reaction to changes in fund expected net alphas less intense. That is, they subdue their additional investments to funds when inferring higher fund expected net alphas.
and limit their reduction in investments to funds when observing lower fund expected net alphas, inducing a positive value in the second component of $dE(\alpha_i|D)/dH|_{(\epsilon', \delta', \delta')}^*$ (i.e., $1 - \left[ \hat{b} + \left( \sum_{i=1}^{M} c_{1,i}^{-1} \right)^{-1} W \right] d(S/W)/dX(e_i^*;H) > 0$). Therefore, whether a higher $H$ increases $E(\alpha_i|D)|_{(\epsilon', \delta', \delta')}$ depends on whether it has a larger impact on gross alphas than on costs (i.e., the sign of $dE(\alpha_i|D)/dH|_{(\epsilon', \delta', \delta')}^*$ depends only on the sign of $A_{H}(e_i^*;H) - c_{2H}(e_i^*;H))$. As in equilibrium, $B'(e_i^*;H) = A_{H}(e_i^*;H) - c_{2H}(e_i^*;H)$, so whether a higher $H$ increases $E(\alpha_i|D)|_{(\epsilon', \delta', \delta')}$ depends on whether it increases $B(e_i^*;H)$. Also, as $H$ changes, the change of marginal $E(\alpha_i|D)|_{(\epsilon', \delta', \delta')}$ (i.e., $d^2E(\alpha_i|D)/dH^2|_{(\epsilon', \delta', \delta')}$) is positively proportional to the change of marginal $B(e_i^*;H)$, i.e., $d^2B(e_i^*;H)/dH^2$ plus a positive adjustment term that captures the effects of risk. This adjustment term is positive because, holding all other parameters the same, if investors’ marginal portfolio risks of investing in funds are higher, investors optimally invest less in funds. In doing so, they exert a smaller impact on net alphas; thus, a higher $H$ induces a higher marginal $E(\alpha_i|D)|_{(\epsilon', \delta', \delta')}$.

We can see that if $d^2B(e_i^*;H)/dH^2$ is positive, $d^2E(\alpha_i|D)/dH^2|_{(\epsilon', \delta', \delta')}$ must be positive, whereas if $d^2E(\alpha_i|D)/dH^2|_{(\epsilon', \delta', \delta')}$ is negative, $d^2B(e_i^*;H)/dH^2$ must be negative. Where $S/W = 1$, whether $N \to \infty$ or $N = 1$, $dE(\alpha_i|D)/dH|_{(\epsilon', \delta', \delta')}$ is equal to $A_{H}(e_i^*;H) - c_{2H}(e_i^*;H)$, and $d^2E(\alpha_i|D)/dH^2|_{(\epsilon', \delta', \delta')}$ is equal to $d^2B(e_i^*;H)/dH^2$. This is because investors have no additional wealth to allocate to funds, so they exert no impact on marginal $E(\alpha_i|D)|_{(\epsilon', \delta', \delta')}$

making the marginal equilibrium optimal expected net alphas depend only on the effect of $H$ on managers’ ability to produce net alphas. Where $N = 1$ and $S/W < 1$, the single investor faces an additional tradeoff between allocating additional dollars to funds to take advantage of fund net alphas and diluting returns on wealth already in funds. In this case, numerical solutions are required to analyze the signs of $dE(\alpha_i|D)/dH|_{(\epsilon', \delta', \delta')}$ and $d^2E(\alpha_i|D)/dH^2|_{(\epsilon', \delta', \delta')}$.
PROPOSITION RA5, Relation between Net Alpha and Market Share.

Whether \( N \to \infty \) or \( N=1 \), an increase (decrease) in \( c_{1,i} \), while \( c_{1,j}, \forall j \neq i \) are unchanged, induces a decrease (increase) in \( s_i / S \), and an increase (decrease) in \( s_j / S, \forall j \neq i \). Also,

1. Where \( N \to \infty \) and \( S / W < 1 \), or \( S / W = 1 \), an increase (decrease) in \( c_{1,i} \), while \( c_{1,j}, \forall j \neq i \) are unchanged, induces a decrease (increase) in \( E(\alpha_i | D)\) \( |_{(e^t,r^t,s^t)} \); thus, \( E(\alpha_i | D)\) \( |_{(e^t,r^t,s^t)} \) and \( s_i / S \) are positively related—internality effect; it induces a decrease (increase) in \( E(\alpha_j | D)\) \( |_{(e^t,r^t,s^t)} \), \( \forall j \neq i \); thus, \( E(\alpha_j | D)\) \( |_{(e^t,r^t,s^t)} \) and \( s_j / S \) are negatively related \( \forall j \neq i \)—externality effect.

2. Where \( N=1 \) and \( S / W < 1 \), numerical solutions are required to analyze the relation between \( E(\alpha_i | D)\) \( |_{(e^t,r^t,s^t)} \) and \( s_i / S \).

3. Where \( N=1 \) and \( S / W = 1 \), an increase (decrease) in \( c_{1,i} \), while \( c_{1,j}, \forall j \neq i \) are unchanged, induces a decrease (increase) in \( E(\alpha_j | D)\) \( |_{(e^t,r^t,s^t)} \); thus, \( E(\alpha_j | D)\) \( |_{(e^t,r^t,s^t)} \) and \( s_i / S \) are positively related—internality effect; it induces a decrease (increase) in \( E(\alpha_j | D)\) \( |_{(e^t,r^t,s^t)} \), \( \forall j \neq i \), thus \( E(\alpha_j | D)\) \( |_{(e^t,r^t,s^t)} \) and \( s_j / S \) are negatively related \( \forall j \neq i \)—externality effect.

The intuition of Proposition RA5 is as follows. Based on Proposition 3, we can see that any change in \( c_{1,i} \), keeping \( c_{1,j}, \forall j \neq i \) unchanged, results in a change in \( s_i / S \) in the opposite direction and a change in \( s_j / S, \forall j \neq i \) in the same direction. Also, a higher \( c_{1,i} \), affects \( E(\alpha_i | D)\) \( |_{(e^t,r^t,s^t)} \) at the two stages. At the first stage, it decreases manager \( i \)'s average cost and, thus, induces higher fund expected net alphas. As manager \( i \) offers a higher fund expected net alpha, investments shift into fund \( i \) from other funds, making all those funds’ fund expected net alphas higher due to decreasing returns to scale at fund level. At the second stage, an increase in fund expected net alphas attracts investments in the AFMI, which in turn drives down fund expected net alphas due to decreasing returns to scale at industry level. Where \( N \to \infty \) and \( S / W < 1 \), as investors’ portfolio risks increase (decrease) when they invest more (less) in the funds. They subdue investments to funds when observing an increase in fund expected net alphas and limit investment reduction when observing a decrease in fund.
expected net alphas. Thus investors’ risk-aversion mitigates the countered effect at the second stage and makes the first stage’s effect dominant. Where $S/W = 1$, whether $N = 1$ or $N \to \infty$, investors have no additional wealth to allocate to funds, so their investments have no impact on marginal equilibrium optimal expected net alphas, causing the first stage’s effect to dominate. In all these cases, we find $E(\alpha_i | D)_{|e^*, \epsilon^*, \delta}$’s are driven down by an increase in $c_{1,i}$, keeping $c_{1,j}, \forall j \neq i$ unchanged; consequently, we have a positive relation between $E(\alpha_i | D)_{|e^*, \epsilon^*, \delta}$ and $s_i / S$ (internality effect) and a negative relation between $E(\alpha_j | D)_{|e^*, \epsilon^*, \delta}$ and $s_j / S, \forall j \neq i$ (externality effect). Where $N = 1$ and $S/W < 1$, the single investor faces an additional tradeoff between allocating additional dollars to funds to take advantage of fund expected net alphas and diluting returns on wealth already in funds. This situation is more complex, and we rely on a numerical solution to solve it.

*Proof of Proposition RA2, RA3, RA4, RA5 and the corresponding corollary.* See the Mathematical Appendix.

### 2.4 Numerical Example

We provide a numerical analysis of the AFMI under our framework and set the parameter values as follows: $W = 100$, $M = 100$, $\mu_p = 0.05$, $\sigma_p = 0.1$, $\sigma_x = 0.05$, $\hat{a} = 0.15$, $\hat{b} = 0.3$, $\sigma_a = 0.4$, $\sigma_b = 0.4$, and $\sigma_{ab} = 0$.

To simplify the case, we assume the parameters in the average cost functions are the same across funds (thus we can drop the indicator $i$); so, in equilibrium, funds have same levels of effort, fees, and sizes. We assume the functional form of $A(e_i; H)$, for numerical analysis, is $A(e_i; H) = (H + A_0)\ln(1 + e_i)$, where $A_0$ is a positive parameter. Also, the functional form of $c_2(e_i; H)$ is $c_2(e_i; H) = c_2^1(H + c_2^2)e_i^2$. The parameters of $A(\bullet; H)$ and $C'(\bullet, \bullet; H)$ are set as follows: $c_0 = 0.005$, $c_1 = 0.1$, $c_1^1 = 2.5$, $c_1^2 = 0.01$, $A_0 = 0.5$.

We choose 100 points evenly spread on $[0, 0.999]$ to be the value of the market concentration level $H$ and study how equilibrium values change with $H$.

Figure 1 illustrates the numerical results for the case of infinitely many mean-variance risk-averse investors. We can see that $e_i^*$ decreases with market concentration because our numerical calibration makes $H$’s impact on marginal effort impact on costs larger than
marginal effort impact on gross alphas, across the domain of $H$. Also, $f_i^\pi$ decreases with $H$ when $H$ is small and increase with $H$ when $H$ becomes large because managers’ costs decrease with $H$ first and then slightly increase with $H$. We also have $S/W$ increase with $X(e_i^*, H)$, as we expect in our model. Moreover, $E(\alpha_i D)_{\{\delta, \theta, \delta\}}$ is always positive, consistent with our model. Also $S/W$, $B(e_i^*; H)$ and $E(\alpha_i D)_{\{\delta, \theta, \delta\}}$ first decrease with $H$ and then slightly increase with $H$ because the difference of $H$’s impact on gross alphas and its impact on costs first decreases with $H$ and then slightly increases with $H$. In addition, $S/W$, $B(e_i^*; H)$, and $E(\alpha_i D)_{\{\delta, \theta, \delta\}}$ are convex at the same time, in our calibration.

Figure 2 demonstrates the numerical results where there is a single mean-variance risk-averse investor in the market. The results are similar except that $S/W$ is much smaller and $E(\alpha_i D)_{\{\delta, \theta, \delta\}}$ is much larger. The reason is that the single investor internalizes the AFMI, limiting the investments in the funds and maximizing his or her portfolio Sharpe ratio.
Figure 1. Infinitely Many Small Mean-Variance Risk-Averse Investors—Comparative Statistics with Respect to Market Concentration

This figure presents the numerical results for the case of infinitely many small mean-variance risk-averse investors. The two subplots on the top illustrate the equilibrium optimal management effort and management fees at each market concentration level. The two subplots in the middle report $B(e_i^*, H)$ at each market concentration level and the equilibrium $S/W$ ratio at each $X(e_i^*, H)$ level. The two subplots at the bottom show the equilibrium $S/W$ ratio and the equilibrium fund expected net alphas at each market concentration level.
Figure 2. A Single Large Mean-Variance Risk-Averse Investor—Comparative Statistics with Respect to Market Concentration

This figure presents the numerical results for the case of a single large mean-variance risk-averse investor. The two subplots on the top illustrate the equilibrium optimal management effort and management fees at each market concentration level. The two subplots in the middle report $B(e^*_i, H)$ at each market concentration level and the equilibrium $S/W$ ratio at each $X(e^*_i, H)$ level. The two subplots at the bottom show the equilibrium $S/W$ ratio and the equilibrium fund expected net alphas at each market concentration level.
2.5 Agency Benefits Due to Market Concentration

A higher fund market concentration implies that fund managers may earn higher agency benefits and charge investors higher fees. Modeling that, we can decompose the management fee into two parts:

\[ f_i = fa(H) + fe_i, \quad \forall i, \]  

(28)

where \( fa(H) \) represents agency benefits, in terms of percentage fees, that managers (can) charge under a particular level of industry concentration \( H \), and \( fe_i \) is the endogenous component chosen by manager \( i \). We assume that \( fa(H) \) is the same for all managers, and

\[ fa(H) \geq 0, \quad \forall H, \]  

(29)

\[ fa'(H) \geq dfa(H) / dH > 0, \quad \forall H, \]  

(30)

\[ fa(l) < \infty, \quad \text{and} \]  

(31)

\[ fa(0) = 0. \]  

(32)

The rationale for these assumptions is as follows. Agency benefits are non-negative; the higher the market concentration is, the higher agency benefits are. Agency benefits are highest under monopolistic markets and are bounded from above; the agency costs on investors are zero if the AFMI is perfectly competitive with infinitely many small fund managers.

Competing Managers

In a market with managers competing for investments, \( fe_i \) can be positive or negative. If it is positive, manager \( i \) charges an additional fee on top of agency benefits; if it is negative, manager \( i \) subsidizes investors in order to increase investments in the fund. Let managers’ optimal fee be

\[ f_i^* = fa(H) + fe_i^*, \quad \forall i. \]  

(33)

Our previous analysis demonstrated that in equilibrium, fund managers charge break-even fees in order to compete for investments; so regardless of agency cost levels \( fa(H) \), managers choose \( fe_i^* \)’s such that \( f_i^* \)’s are break-even fees. As managers charge break-even fees in equilibrium, \( B(e_i^*; H) \) is transferred to investors in its entirety. We call the dollar amount of agency costs \( \Psi fa \); thus,

\[ \Psi fa = fa(H) \frac{S}{W} W. \]  

(34)
As $fa(H)$ increases with $H$, we know that if $S/W$ increases with $H$, then $\Psi fa$ increases with $H$. If $S/W$ decreases with $H$, then whether $\Psi fa$ increases with $H$ depends on the rate that $fa(H)$ increases with $H$ relative to the rate that $S/W$ decreases with $H$. The curvature of $\Psi fa$ in $H$ is inconclusive.

We call the dollar amount of AFMI’s direct benefits from managers’ optimal effort $\Psi B$; thus,

$$\Psi B = B(e^*_i; H) \frac{S}{W}.$$ (35)

Then, if $A_H(e^*_i; H) - c_{2H}(e^*_i; H)$ is positive (negative), $B(e^*_i; H)$ and $S/W$ both increase (decrease) with $H$, inducing $\Psi B$ to increase (decrease) with $H$. It is possible that $fa(H) > B(e^*_i; H)$ and $\Psi fa > \Psi B$. That is, under particular levels of market concentration, managers’ agency costs surpass AFMI’s direct benefit from managers’ effort. Still, because of managers’ competition for investments in their funds, they would set $fe^*_i$ to be negative to subsidize investors’ investments such that managers are earning break-even fees. Investors, however, only care about and only observe $f^*_i$; they do not observe $f^*_i$’s components and cannot base their decisions on either $fa(H)$ or $fe^*_i$.

**Colluding Managers**

Managers, may use market power to collude in charging fees higher than those that endogenously arise in the previous non-collusive equilibrium. We assume that their (market) power to charge higher fees is increasing with industry’s concentration level and that managers agree on a single collusive fee, $f(H)$. Thus, we now have

$$f_i = fa(H) = f(H), \ \forall i.$$ (36)

We consider the case where $N \to \infty$. Conditional on the industry collusive fee rate, managers maximize their corresponding funds’ profits by exerting optimal effort levels, incorporating investors’ optimal reactions. We can concisely write the total industry profit function, $\Pi$, as a sum of the industry funds profit functions:

$$\Pi^*(e^*_1, e^*_2, \ldots, e^*_M; H) \triangleq \text{Max}_{\{e_i^\prime\}_{i=1}} \Pi(e_1, e_2, \ldots, e_M; H) = \text{Max}_{\{e_i^\prime\}_{i=1}} \sum_{i=1}^M s_i \left( f(H) - C'(e_i, s_i; H) \right),$$ (37)

subject to

23 Under our functional assumptions, it is irrational for managers to collude in lowering fees.

24 Please note that Equations (29) to (32) still hold.
\[ f(H) - C'(e_i, s_i; H) \geq 0, \quad \forall i, H. \] (38)

Optimal effort levels \( e_i^* \) are given by the first-order-condition

\[
\sum_{i=1}^{M} \frac{A_{e_i}(e_i^*; H) \left[ f(H) - C'(e_i^*, s_i; H) \right] - \frac{S}{W} \left[ c_{i,1} \left( \sum_{i=1}^{M} c_{1,i} \right)^{-1} W A_{e_i}(e_i^*; H) + c_{2,1}(e_i^*; H) \right]}{\gamma \left[ 3 \sigma_b^2 \left( \frac{S}{W} \right)^2 + \sigma_a^2 + \sigma_s^2 \right] + \hat{b}} = 0, (39)
\]

where \( S/W \) is given by

\[
-\gamma \sigma_b^2 \left( \frac{S}{W} \right)^3 - [\gamma \sigma_a^2 + \gamma \sigma_s^2 + \hat{b}] \frac{S}{W} + A(e_i^*, H) - f(H) = 0, \quad (40)
\]

if \( S/W < 1 \), and \( S/W = 1 \) otherwise. The second-order-condition is satisfied, so \( e_i^* \) is a maximum point. Also, in equilibrium, where \( S/W < 1 \), we have

\[
d\left( \frac{S}{W} \right) = \frac{A_{e_i}(e_i^*; H) e_i^*(H) + A_{e_h}(e_i^*; H) - f'(H)}{\gamma \left[ 3 \sigma_b^2 \left( \frac{S}{W} \right)^2 + \sigma_a^2 + \sigma_s^2 \right] + \hat{b}}, \quad (41)
\]

where \( f'(H) = df(H)/dH \), and

\[
\frac{dE(\alpha_{i|D})}{dH} \bigg|_{e_i^*, e_i^*} = \frac{\gamma \left[ 3 \sigma_b^2 \left( \frac{S}{W} \right)^2 + \sigma_a^2 + \sigma_s^2 \right] \left[ A_{e_i}(e_i^*; H) e_i^*(H) + A_{e_h}(e_i^*; H) - f'(H) \right]}{\gamma \left[ 3 \sigma_b^2 \left( \frac{S}{W} \right)^2 + \sigma_a^2 + \sigma_s^2 \right] + \hat{b}}. \quad (42)
\]

Therefore, both \( d\left( \frac{S}{W} \right)/dH \geq 0(<0) \) and \( dE(\alpha_{i|D})/dH \bigg|_{e_i^*, e_i^*} \geq 0(<0) \) if and only if \( A_{e_i}(e_i^*; H) e_i^*(H) + A_{e_h}(e_i^*; H) - f'(H) \geq 0(<0) \). The economic sense is that if in equilibrium, a higher concentration increases fund gross alphas more than the exogenous fees, managers optimally choose efforts to produce higher fund expected net alphas to attract investments that increase industry profits. Where \( S/W = 1 \), managers choose the minimum \( e_i^* \) to make \( S/W = 1 \). The reason is that if managers can optimally induce investors to invest all their wealth in funds, they choose the minimum effort to do so. In this case, both \( S/W \) and \( E(\alpha_{i|D}) \bigg|_{e_i^*, e_i^*} \) are unaffected by \( H \), so

\[
\frac{d\left( \frac{S}{W} \right)}{dH} = 0, \quad (43)
\]
\[
\frac{dE(\alpha_i, D)}{dH} \bigg|_{(\epsilon^{*}, \epsilon, s^{*})} = 0. 
\] (44)

The results of \( d^2 (S/W) / dH^2 \) and \( d^2 E(\alpha_i, D) / dH^2 \bigg|_{(\epsilon^{*}, \epsilon, s^{*})} \) require numerical analysis.

To choose the optimal collusion fee, a “collusion planner” would write a mapping from \( H \) to \( f(H) \), where each point in the range is the one for which industry profits, \( \Pi^{*} \), are the highest.

2.6 Discussion of Covariance between \( \hat{a} \) and \( \hat{b} \)

We assume that \( \sigma_{ab} = 0 \), but we note that the value of \( \sigma_{ab} \) affects the equilibrium results because it affects portfolio risks. If \( \sigma_{ab} \) (in absolute value) is large relative to other risk sources, such as \( \sigma_{a}^2 \), \( \sigma_{b}^2 \), and \( \sigma_{c}^2 \), changes in investors’ wealth allocations to funds, would induce changes in their portfolio risks, affecting in turn their optimal demands. This would make our theoretical results in propositions RA3, RA4, RA5 and results in Section 2.5 more complex. We believe that consequences of such an analysis would not be directly material to the issues that we explore here and would obfuscate the analysis. We, thus, assume that precisions of estimates of \( a \) and \( b \), conditional on current information, are not closely related, making \( \sigma_{ab} \to 0 \).

3 Endogenous Market Concentration Level and Empirical Analysis

Our model allows for an endogenous market concentration level. If we define \( H \) as the Herfindahl-Hirschman index (HHI), which is the sum of market shares squared, then for an \( M \) firms’ market \( H \in \left[\frac{1}{M}, 1\right) \). 25 Using funds’ equilibrium market share, as identified in Proposition 3, we can write the equilibrium market concentration \( H^{*} \) as

\[
H^{*} \triangleq \sum_{i=1}^{M} \left( \frac{S_i}{S} \right)^2 = \sum_{i=1}^{M} \left( c_{1,i} \sum_{j=1}^{M} \left( c_{1,j}^{-1} \right) \right)^2. 
\] (45)

We can see that \( H^{*} \) is determined by \( c_{1,i} \)'s. Specifically, depending on the size of \( c_{1,i} \) relative to that \( c_{1,j}, \forall j \neq i \), an increase in \( c_{1,i} \), holding \( c_{1,j}, \forall j \neq i \) constant, increases or decreases \( H^{*} \).

25 In an \( M \) AFMI, for example, the Herfindahl-Hirschman index could have values between the highest concentration, 1, where one of the funds captures practically all the market share, and the lowest concentration, \( 1/M \), where market shares are evenly divided.
In the case where there are infinitely many risk-averse investors, an increase in $c_{i,a}$ affects the equilibrium fund expected net alphas in two ways: 1) its direct impact leads to lower equilibrium fund expected net alphas (Proposition RA5), and 2) depending on fund $i$’s size relative to rivals, it increases or decreases $H^*$, which consequently increases (decreases) equilibrium fund expected net alphas if and only if $A_H(e_i^*;H^*) - c_{2H}(e_i^*;H^*) \geq (<)0$ (Proposition RA4). Similarly, an increase in $c_{i,a}$ affects the equilibrium AFMI size in two ways: 1) its direct impact leads to an (inverse direction) AFMI size change, and 2) it increases or decreases $H^*$, which consequently increases (decreases) the equilibrium AFMI size if and only if $A_H(e_i^*;H^*) - c_{2H}(e_i^*;H^*) \geq (<)0$. Thus, in the endogenous market concentration case, the relation between the market concentration and the equilibrium fund expected net alphas and AFMI size is more complex.

We now proceed with an empirical analysis of the benefits and costs of changing market concentration levels of the AFMI using the version of our model with endogenous concentration. In this sense, this version of our model befits available data of empirical market concentration levels, such as the HHI. Popular empirical market concentration measures, such as HHI, are functions of rivals’ relative sizes. We expect that market characteristics, such as regulation, transaction costs, tax rates, and barriers to entry, affect funds’ cost sensitivity to size (i.e., $c_{i,a}$’s). As a result, they affect relative fund sizes and, thus, the level of empirical market concentration measures. We use empirical techniques to control potential endogeneity of market concentration measures.

Whether fund net alphas and AFMI size move in the same direction with market concentration and whether both are concave (convex) in it become empirical questions. Further, in cases where active fund management creates value, if fund net alphas and AFMI size increase (decrease) with market concentration, our model predicts positive (negative) benefits of marginal managerial effort, for plausible parameter values. We note that both signs of the benefits of changing concentration levels are plausible alternatives to a null

---

26 We believe that our cost function, Equation (8), is a concise one that essential effects within our model. To assure that all our functional form restrictions of the non-specialized model (exogenous concentration levels), which we deem basic and simple, hold in the specialized one (endogenous concentration levels); however, we need to impose additional, technical, “second order,” parameter restrictions. For brevity and simplicity, we do not impose these restrictions. We call the parameter values that make the specialized model abide by these restrictions plausible. We, later, confirm that the said technical restrictions are not empirically binding. That is, imposing these restrictions would not have changed our empirical results. In other words, the empirically estimated parameters fall within the plausible parameters range.
hypothesis of no benefit of active fund manager effort. We now provide evidence regarding
the market concentration-net alpha and market concentration-AFMI size relations for the U.S.
active equity mutual fund industry.

3.1 Methodology

Our goal is to analyze how the AFMI size and fund net alphas change with market
concentration of an AFMI. Next, we describe how we measure concentration, fund net alpha,
and our econometric strategy to estimate the effect of changing concentration on net alpha
and AFMI size, controlling for endogeneity and omitted-variable bias-related issues.

Measures of AFMI Concentration

Following the literature, we measure competitiveness of an AFMI using three indices
(see, for example, Berger and Hannan (1989), Geroski (1990), Berger (1995), Goldberg and
Rai (1996), Nickell (1996), Berger, Bonime, Covitz and Hancock (1999), Cremers, Nair, and
Peyer (2008), and Giroud and Mueller (2011)):
1) the HHI:

\[ H_i = \sum_{i} m_i \times S_{i,t}^2; \]  \hspace{1cm} (46)

2) the normalized HHI (NHHI):

\[ NHHI_i = \frac{m_i \times H_i - 1}{m_i - 1}; \]  \hspace{1cm} (47)

3) the sum of the first five largest funds’ market shares (5-Fund-Index):

\[ 5_{\text{-Fund-Index}} = \sum_{i=1}^{5} MS_{i,t}, \]  \hspace{1cm} (48)

where \( MS_{i,t} \) is the market share of fund \( i \) at time \( t \), measured by the fund’s assets under
management at time \( t \) over the total assets under management in its market at time \( t \), and \( m_t \)
is the number of funds at time \( t \). As the HHI is related to the number of funds, for a
robustness check, we also use two measures not related to the number of funds to measure
market concentration, the NHHI used by Cremers, Nair and Peyer (2008) and the 5-Fund-
Index, one of the common measures of market concentration.

Choice of Benchmarks and Net Alpha Estimation

Berk and Binsbergen (2015) argue that to measure the value added by a fund, its
performance should be compared to the next-best investment opportunity available to
investors. In our model, we assume that a single passive benchmark exists and is common
knowledge to investors and managers. We make these assumptions because they keep our theoretical analysis parsimonious and relaxing them does not alter the key insights from our model. However, in our empirical section, we allow for multiple benchmarks and match each active equity mutual fund to a set of tradable index funds that reasonably replicate passive alternatives available to an average mutual fund investor. Specifically, we assume the following return-generating process:

\[ R_{i,t} = \alpha_{i,t} + b_{i,t}^1 F_{t}^1 + b_{i,t}^2 F_{t}^2 + \ldots + b_{i,t}^n F_{t}^n, \]  

where the indices \( i \) and \( t \) represent the fund and time indices, whereas \( n \) indicates the number of tradable index funds in the market. \( R_{i,t} \) is the return net of management fee of a fund, and \( F_{t}^1 \) through \( F_{t}^n \) are the returns net of management fees of tradable index funds in different asset classes. We use net alpha (instead of gross alpha) in our empirical analyses because we are interested in aspects of market competition and societal welfare predicted by our model, where net alpha is the relevant quantity. We treat the index funds \( F_{t}^1 \) through \( F_{t}^n \) as a basis fund set that may be used to replicate the returns on any passive benchmarks used by mutual fund investors.

To perform our analysis, we first need to calculate fund net alphas \((\alpha_{i,t})\). For each active fund in our sample, we calculate a set of weights on our basis fund set that sum to one and minimize the tracking error between the active fund return and a corresponding passive benchmark portfolio return (Sharpe (1992)). We note that our empirical design of identifying passive benchmarks, using matching tradable index funds, fits our theoretical structure, which assumes the appropriate passive benchmarks for each fund.

We perform this analysis on a rolling basis, using returns from months \((t - 60)\) to \((t - 1)\) to avoid look-ahead bias. That is, we identify coefficients \( b_{i,t}^1 \) to \( b_{i,t}^n \) to minimize the variance of the residual. These coefficients are constrained to be between zero and one (we do not allow short selling), and their sum is constrained to be one. These coefficients identify the portfolio weights in our basis index fund set that provides the estimated minimum “tracking error” passive benchmark of a fund.

Next, to calculate a fund’s net alphas in month \( t \), we subtract the returns on the identified passive portfolio (the style benchmark) for month \( t \) from the active equity fund’s returns in month \( t \) and that of the style benchmark in month \( t \). This provides us with fund net alphas in each month for each fund.

To ensure the robustness of our results, we also use an alternative method to fund net
alphas. Specifically, it is possible that traded index funds do not capture unobservable risk factors that drive excess returns. Errors in our set of passive benchmarks or our matching strategy may result in net alphas that measure exposure to such unobservable risk factors instead of fund manager performance. Using the method developed by Connor and Korajczyk (1988), we estimate unobserved common factors in our estimated fund net alphas using the principal components of our estimated fund net alphas series. We use these estimated principal components to control for unobserved common factors in the fund net alphas. Specifically, we regress each fund’s fund net alphas on the first two principal components without a constant term. We refer to the residuals of these regressions as PC-adjusted fund net alphas and use them as the dependent variable for robustness checks.

**Controlling for Endogeneity and Omitted-Variable Bias**

Estimating the effect of market concentration on performance is a challenge because market concentration is determined endogenously. PST explain why a simple regression is likely to deliver biased estimates and introduce a recursive demeaning (RD) estimator to avoid the biases. In analyzing the relation between fund net alphas and market concentration, we use the RD estimation procedure of PST. We estimate the effects of size ($\beta_1$) and market concentration ($\beta_2$ and $\beta_3$) on fund net alphas using the following panel regression:

$$\tilde{\alpha}_{i,t} = \beta_1 \overline{MS}_{i,t-1} + \beta_2 \overline{H}_{i,t-1} + \beta_3 \overline{H}_{i,t-1}^2 + \overline{\epsilon}_{i,t},$$

(50)

The bar above the variables denotes forward-demeaned variables, defined below:

$$\overline{\alpha}_{i,t} = \alpha_{i,t} - \frac{1}{T_i - t + 1} \sum_{s=t}^{T_i} \alpha_{i,s},$$

(51)

$$\overline{MS}_{i,t} = MS_{i,t} - \frac{1}{T_i - t + 1} \sum_{s=t}^{T_i} MS_{i,s},$$

(52)

$$\overline{H}_{i,t} = H_{i,t} - \frac{1}{T_i - t + 1} \sum_{s=t}^{T_i} H_{i,s},$$

(53)

$$\overline{H}_{i,t}^2 = H_{i,t}^2 - \frac{1}{T_i - t + 1} \sum_{s=t}^{T_i} H_{i,s}^2,$$

(54)

where $T_i$ is the number of time-series observations of fund $i$. We run robustness checks by replacing $H_i$ (HHI) with the NHHI and with the 5-Fund-Index.

The RD method in Equation (50) can control for the fund fixed effect. We include market share as a control, not only because the equilibrium market share provides information on a fund’s cost sensitivity to fund size (Proposition 3), but also because current
empirical studies show a linear relation between changes in market share and fund performance (Spiegel and Zhang (2013)) and use it as a firm-level market power measure (e.g., Berger, Bonime, Covitz and Hancock (1999) and Nickell (1996)). There may be potential endogeneity (reverse causality) between AFMI shares and fund net alphas because when fund net alphas are higher, corresponding asset values increase and funds attract investments, both leading to a higher market share. This endogeneity issue may bias our results. We address this endogeneity issue using an instrumental variable method. In the first stage, we regress $\overline{MS}_{i,t-1}$ (recursively forward-demeaned market share) on $\overline{MS}_{i,t-1}$ (recursively backward-demeaned market share) without a constant term. In the second stage, we use the fitted value from the first stage to run Equation (50),27 where

$$\overline{MS}_{i,t} = MS_{i,t} - \frac{1}{t-1} \sum_{s=1}^{t-1} MS_{i,s},$$

(55)

To be a valid instrument of $\overline{MS}_{i,t-1}$, $\overline{MS}_{i,t-1}$ must satisfy the relevance and exclusion conditions. The relevance condition is likely to hold because both $\overline{MS}_{i,t-1}$ and $\overline{MS}_{i,t-1}$ are derived from $MS_{i,t-1}$ and are, thus likely to be closely related. The exclusion condition is also likely to hold because the backward-looking information in $\overline{MS}_{i,t-1}$ is unlikely to be helpful in predicting the forward-looking net alpha information in $\overline{\epsilon}_{i,t}$, where $\overline{\epsilon}_{i,t}$ is the residual in the RD method.

On the other hand, there is no reason to believe that fund net alphas, as individual fund performance, are endogenous with the market concentration ratios as industry-level measures. In particular, there is no reason to believe that innovations in market concentration are correlated with the error term in the regression. Thus, we directly use the recursive forward-demeaned market concentration ratios in the model.

In analyzing the relation between the AFMI size and market concentration, we use the vector auto-regression (VAR) method. Although theoretically we assume market concentration is exogenous, empirically this industry-level variable may be endogenous with another industry-level variable, the industry size. The VAR method can address this potential endogeneity issue, and our model is

---

27 We correct the second-stage standard error estimates of $\beta_1$ by incorporating the estimation errors from the first-stage regression.
where $IS_t$ is AFMI size at time $t$ and $e_{a,t}$, $e_{b,t}$, and $e_{c,t}$ are the residuals.

### 3.2 Data

We obtain our data from Morningstar Direct. Our sample contains 1,374 actively managed U.S. equity mutual funds from January 1979 to December 2014. The Data Appendix supplements the data description below.

We use keywords in Morningstar to identify U.S. mutual funds (both open-end and closed-end) and exclude index funds, enhanced index funds, funds of funds, and in-house funds of funds. Also, we require funds to be classified as Equity in the Global Broad Category Group, and we further exclude international funds, real estate funds, and sector funds. Next, we use the Fund ID provided by Morningstar to aggregate fund share class-level information to fund-level information. Since we use a 5-year rolling window to estimate fund net alphas, we require each of our active equity mutual funds to have at least 10 years’ return observations. Using these filters, we obtain our sample of 1,374 actively managed U.S. equity mutual funds.

The index funds used in the style-matching model are also from Morningstar. We require index funds to have no missing observations in our sample period so that the style-matching model is consistent and stable. The factors used in the style-matching model include index funds with the Morningstar Institutional Categories of Small Core, Large Core and S&P 500 Tracking, and the CRSP Fama-French risk-free rate. All the fund returns are net of administrative and management fees and other costs taken out of fund assets.

Table 1 reports the summary statistics. It shows that monthly fund returns are positive on average but vary a great deal, from smaller than -14% to more than 13%, with a standard deviation of more than 5%. Monthly fund net alphas are positive on average, also with a wide variation. We also report summary statistics of the fit of our passive benchmark-matching method using $R$-squared, which is measured as

$$Rsqr_{t,i} = 1 - \frac{Var(\alpha_{i,t})}{Var(R_{t,i})},$$

where $Var(\alpha_{i,t})$ is the variance of the residuals of the regression, and $Var(R_{t,i})$ is the
variance of $R_{i,t}$. The average R-squared in our style-matching model are quite high at 0.86, and the variation around the mean is relatively small.

To analyze robustness, we redo our analyses using fund net alphas adjusted by the first two principal components. The values of the monthly relative industry size (total funds’ net assets divided by stock market capitalization) and the monthly fund sizes in December 2014 dollars (funds’ net assets divided by stock market capitalization in the same month, multiplied by the stock market capitalization in December 2014) are similar to the sample in PST.

The number of active equity mutual funds in our sample increases over time. The market concentration measures, such as the HHI, NHHI, and 5-Fund Index, with fluctuations, tend to decrease over time. Also, all three market concentration measures do not seem to perform with skewness.
Table 1. Statistical Summary

Our sample period is from January 1979 to December 2014, and monthly data is used. Panel A reports the summary statistics for fund-level data, and Panel B reports those for industry-level data. Fund Net Return and Fund Net Alpha are in percentages, and both are net of administrative and management fees and other costs taken out of fund assets. The Style-Matching Model R-sqr, AFMI Share, HHI, NHHI, and 5-Fund-Index are in decimals. Fund Size is measured in $100 millions and is equal to the fund’s total net assets under management, divided by the stock market capitalization in the same month, and multiplied by the stock market capitalization in December 2014. Industry Size is the sum of funds’ net assets under management divided by the stock market capitalization in the same month. Number of Funds is in units.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs.</th>
<th>Mean</th>
<th>Std.</th>
<th>1st</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
<th>99th</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Fund-Level Data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fund Net Return (%)</td>
<td>321456</td>
<td>0.8736</td>
<td>5.1508</td>
<td>-14.4922</td>
<td>-1.7976</td>
<td>1.2998</td>
<td>3.8907</td>
<td>13.0053</td>
</tr>
<tr>
<td>Fund Net Alpha (%)</td>
<td>246553</td>
<td>0.0349</td>
<td>1.9499</td>
<td>-5.4465</td>
<td>-0.8570</td>
<td>0.0215</td>
<td>0.9156</td>
<td>5.5982</td>
</tr>
<tr>
<td>Style-Matching Model R-sqr (decimal)</td>
<td>246557</td>
<td>0.8607</td>
<td>0.1175</td>
<td>0.4223</td>
<td>0.8178</td>
<td>0.8953</td>
<td>0.9408</td>
<td>0.9894</td>
</tr>
<tr>
<td>Fund Size (in 100 Million of 2014 Dec Dollars)</td>
<td>314083</td>
<td>28.7796</td>
<td>95.3306</td>
<td>0.0399</td>
<td>1.3833</td>
<td>5.5718</td>
<td>20.1835</td>
<td>416.9203</td>
</tr>
<tr>
<td>Fund Market Share (decimal)</td>
<td>314083</td>
<td>0.0012</td>
<td>0.0041</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0002</td>
<td>0.0007</td>
<td>0.0185</td>
</tr>
<tr>
<td><strong>Panel B: Industry-Level Data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industry Size (decimal)</td>
<td>432</td>
<td>0.0982</td>
<td>0.0591</td>
<td>0.0200</td>
<td>0.0389</td>
<td>0.1035</td>
<td>0.1638</td>
<td>0.1801</td>
</tr>
<tr>
<td>Number of Funds (No.)</td>
<td>432</td>
<td>850.2</td>
<td>659.5</td>
<td>86.0</td>
<td>249.0</td>
<td>677.5</td>
<td>1468.5</td>
<td>2126.0</td>
</tr>
<tr>
<td>HHI (decimal)</td>
<td>432</td>
<td>0.0191</td>
<td>0.0230</td>
<td>0.0061</td>
<td>0.0101</td>
<td>0.0157</td>
<td>0.0243</td>
<td>0.0382</td>
</tr>
<tr>
<td>NHHI (decimal)</td>
<td>432</td>
<td>0.0157</td>
<td>0.0139</td>
<td>0.0057</td>
<td>0.0094</td>
<td>0.0141</td>
<td>0.0201</td>
<td>0.0269</td>
</tr>
<tr>
<td>5-Fund-Index (decimal)</td>
<td>432</td>
<td>0.2166</td>
<td>0.0765</td>
<td>0.1240</td>
<td>0.1640</td>
<td>0.1986</td>
<td>0.2650</td>
<td>0.3438</td>
</tr>
</tbody>
</table>

Because our sample differs from PST, we check for any alarming systematic differences by evaluating the returns to scale relation in our sample. Table 2 reports the estimated relation of fund net alpha and fund size and fund industry size, using PST’s RD model. We find that fund net alpha is significantly negatively associated with lagged industry size and is insignificantly negatively associated with lagged fund size. In an unreported robustness check, we replace lagged fund size by lagged log of fund size and find consistent results. Thus, we find evidence of decreasing returns to scale at the industry level. These findings are consistent with PST’s.
### Table 2. Sample Check for Decreasing Returns to Scale Assumption

This table reports the results of RD panel regression model, following PST. Fund Net Alpha is the dependent variable. Fund Size is measured in $100 millions, and is equal to the fund’s total net assets under management, divided by the stock market capitalization in the same month, multiplied by the stock market capitalization in December 2014. Industry Size is the sum of AFMI’s net assets under management divided by the stock market capitalization in the same month. The unit of coefficients is percentage. Standard errors are clustered by fund and presented in parentheses. The symbols ***, **, and * represent the 1%, 5%, and 10% significant level in a two-tail t-test, respectively.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Fund Net Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Lagged Fund Size</td>
<td>-0.0005</td>
</tr>
<tr>
<td></td>
<td>(0.0010)</td>
</tr>
<tr>
<td>Lagged Industry Size</td>
<td>-1.0211***</td>
</tr>
<tr>
<td></td>
<td>(0.1302)</td>
</tr>
<tr>
<td>Observations</td>
<td>239,537</td>
</tr>
<tr>
<td>R-Sqr</td>
<td>0.0000</td>
</tr>
<tr>
<td>Adjusted R-Sqr</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
3.3 Empirical Results

In this section, we implement our theoretical model to empirically evaluate effects of market concentration levels on the U.S. mutual fund industry.

Figure 3 shows the HHI value from January 1984 to December 2014. We can see that before 1990, the HHI value was relatively high, fluctuating from 0.02 to 0.03. After that, it continued decreasing; and in the current years, it has reached 0.006, which is around a quarter of the values before 1990. This figure shows that the concentration of the U.S. active equity mutual fund market decreased substantially. Alternative market concentration measures, such as NHHI and 5-Fund Index, show similar trends.

Figure 3. HHI Value from January 1984 to December 2014

The HHI value is in decimals. The gray bars represent the recession periods.

We first evaluate the relation between fund net alphas and market concentration. The results using the RD method are shown in Table 3. Panel A reports the results using fund net alpha as the dependent variable. In the first two columns, we find that the coefficient of the first-order term of lagged HHI is significantly positive, whereas the coefficient of the second-order term is significantly negative. This result is robust to including lagged market share and lagged industry size as controls. This suggests that the effect of concentration is distinct from the effect of decreasing returns to scale at the fund and industry level. To control for the possibility of unaccounted common factors in the estimated net alphas, we also use principal
component (PC)-adjusted fund net alphas as the dependent variable (Panel B) and find similar results.

The main result of this table is that fund net alphas, on average, are increasing concave in fund market concentration. Our theoretical results, then, indicate that for plausible parameter values, higher levels of market concentration induce increases in gross alpha production opportunities that are higher than those in managers’ effort costs.

Table 3. Fund Net Alpha and Market Concentration Relation

This table reports the results of our RD panel regression model. Panel A reports the results using the Fund Net Alpha as the dependent variable, whereas Panel B reports the results using the PC-Adjusted Fund Net Alpha (adjusted by the first two principal components of fund net alphas) as the dependent variable. Market Share is equal to a fund’s net assets under management divided by the sum of all funds’ net assets under management in the same month. Industry Size is the sum of AFMI’s net assets under management divided by the stock market capitalization in the same month. HHI is calculated as the sum of squares of each fund’s market share, and HHI^2 is the square of HHI. The unit of coefficients is percentage. Standard errors are clustered by fund and presented in parentheses. The symbols ***, **, and * represent the 1%, 5%, and 10% significant level in a two-tail t-test, respectively.

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Panel A</th>
<th>Panel B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fund Net Alpha</td>
<td>PC-Adjusted Fund Net Alpha</td>
</tr>
<tr>
<td>Lagged HHI</td>
<td>6.5277*** 40.0796*** 39.1271** 35.3591*** 34.8497**</td>
<td>2.5033*** 11.8816*** 9.4626**</td>
</tr>
<tr>
<td></td>
<td>(1.0362) (4.8085) (17.7560) (4.6913) (17.4796)</td>
<td>(0.8485) (4.3336) (4.1665)</td>
</tr>
<tr>
<td>Lagged HHI^2</td>
<td>-1.110.4260*** -1.081.2070* -1.402.0231*** -1.394.8592** -310.3817** -459.8112***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-156.2080) (-589.6182) (-184.4615) (-700.8128) (-142.7130) (-161.0837)</td>
<td></td>
</tr>
<tr>
<td>Lagged Market Share</td>
<td>-12.0701 (-23.6434)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-15.1453 (24.8670)</td>
<td></td>
</tr>
<tr>
<td>Lagged Industry Size</td>
<td>-1.8946*** -1.9440</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.3977) (1.4847)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>245,178 245,178 239,537 245,178 239,537</td>
<td>245,179 245,179 245,179</td>
</tr>
<tr>
<td>R-Sqr</td>
<td>0.0002 0.0004 0.0004 0.0006 0.0006</td>
<td>0.0000 0.0001 0.0002</td>
</tr>
<tr>
<td>Adjusted R-Sqr</td>
<td>0.0002 0.0004 0.0004 0.0006 0.0006</td>
<td>0.0000 0.0001 0.0002</td>
</tr>
</tbody>
</table>
We use a VAR to evaluate the relation between industry size and market concentration. The results are shown in Table 4. The first column of each model specification shows how AFMI size is associated with HHI. The result of interest in this table is that AFMI size is significantly positively associated with lagged HHI (model specification 1) and is significantly negatively associated with the second order of lagged HHI (model specification 2). If we further include a time trend or year dummies into the model, we find consistent results (model specifications 3 and 4). Thus, AFMI size is increasing concave in market concentration. The positive relation between industry size and market concentration indicates that, consistent with our previous tests, higher market concentration levels, on average, for plausible parameter values, increase gross alphas more than they increase managers’ effort costs. Also, we find concavity of both industry size and fund net alphas in concentration levels, again consistent with our model’s theoretical prediction. As an aside, we also note that in the second column, AFMI size has little effect on HHI: the magnitude of the coefficient of the lagged AFMI size is almost zero.

Table 4. AFMI Size and Market Concentration Relation

This table reports the results of our VAR model. Industry Size is the sum of funds’ net assets under management divided by the stock market capitalization in the same month. HHI is calculated as the sum of the squares of each fund’s market share, and HHI^2 is the square of HHI. Time Trend is set to be one for January 1984 and to increase by one each month. Robust standard errors are used and presented in parentheses. The symbols ***, **, and * represent the 1%, 5%, and 10% significant level in a two-tail t-test, respectively.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Industry Size</td>
<td>HHI</td>
<td>Industry Size</td>
<td>HHI</td>
</tr>
<tr>
<td>Lagged Industry Size</td>
<td>1.0033*** (-0.0014)</td>
<td>-0.0001*** (0.0000)</td>
<td>1.0032*** (0.0035)</td>
<td>-0.0000*** (0.0000)</td>
</tr>
<tr>
<td>Lagged HHI</td>
<td>28.7108*** (3.7047)</td>
<td>0.525*** (0.0234)</td>
<td>49.1392*** (27.8090)</td>
<td>0.9419*** (0.0084)</td>
</tr>
<tr>
<td>Lagged HHI^2</td>
<td>-136.0583* (76.1298)</td>
<td>0.929*** (0.0085)</td>
<td>-204.4277*** (84.6050)</td>
<td>0.9177*** (0.0094)</td>
</tr>
<tr>
<td>Time Trend</td>
<td>-0.4798*** (0.1997)</td>
<td>0.0145*** (0.0013)</td>
<td>-0.9431 (0.0024)</td>
<td>0.0010 (0.0007)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.4798* (0.0035)</td>
<td>0.0145*** (0.0000)</td>
<td>-0.9431 (0.0024)</td>
<td>0.0010 (0.0007)</td>
</tr>
<tr>
<td>Year Dummies</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>431</td>
<td>431</td>
<td>431</td>
<td>431</td>
</tr>
<tr>
<td>R-Sq</td>
<td>0.999</td>
<td>0.663</td>
<td>0.999</td>
<td>0.406</td>
</tr>
</tbody>
</table>
Robustness

In addition to reported tables, we have examined the robustness of our main empirical results in various cases. We analyze the sensitivity of our results to various measures of market concentration such as the NHHI and 5-Fund Index. We find consistent results. We also analyze the sensitivity of the results in Table 2 and Table 3 by using fund fixed-effect regressions instead of the RD method. Most of the results are consistent, except when regressing the PC-adjusted fund net alpha on market concentration measures; we find that the significance of market concentration measure is reduced. Furthermore, we analyze whether our results are driven by small funds. We redo our main analyses using observations after restricting our sample to funds with a net asset value above $15 million in any month of our sample period. Again, we find consistent results.

To test whether our main results are stable across sub-samples, we redo our analyses in Table 3 for three sub-periods. We find a significantly positive relation between fund net alphas and lagged HHI in all three sub-periods.

4 Conclusion

We develop a theoretical model to analyze an AFMI equilibrium where we investigate performance, size, and managers’ costly (optimal) effort under a continuum of exogenous market concentration levels. We use Pastor and Stambaugh’s (2012) framework, where net alpha production is of decreasing returns to scale at the industry level, and we similarly model the decreasing returns to scale effect at the fund level. Higher market concentration levels imply better utilization of industry resources and the existence of more unexplored investment opportunities, making managers’ efforts more productive. At the same time, however, higher market concentration levels allow managers to require higher compensation for effort, making effort costs higher.

Our model’s comparative statistics characterize the association between fund net alphas and a continuum of exogenous market concentration levels, and that between AFMI size and market concentration. In particular, we consider the case of infinitely many mean-variance risk-averse investors whose portfolio risks increase with investments in funds. The funds’ expected net alphas increase with market concentration if and only if higher concentration induces a larger impact on gross alpha production than on the costs of effort (i.e., higher concentration induces higher AFMI’s direct benefits of net alpha production). Observing an increase in fund expected net alphas, due to higher market concentration, mean-variance risk-averse investors increase their mutual fund holdings but reach optimum
investment levels at higher expected net alphas than before. Thus, the equilibrium fund expected net alphas become positively associated with market concentration. In addition, the concavity of fund expected net alphas in market concentration indicates that the AFMI’s direct benefits of net alpha production are concave in market concentration. This further induces concavity of AFMI size in market concentration.

We also study the consequences of increased agency costs under higher market concentration levels in a market where managers can collude to pursue agency benefits. In addition, we specialize our model to allow for endogenous market concentration levels, which befits empirical market concentration measures and enables us to study the model empirically.

We use Morningstar’s U.S. active equity mutual fund data. First, we find that on average, fund net alphas are negatively associated with fund size and AFMI size, confirming decreasing returns to scale at both fund and industry levels. More importantly, we also find that, on average, both fund net alphas and AFMI size are increasing concave with market concentration.

Our findings have policy implications for the U.S. AFMI. Under the current, empirically identified, tradeoff between changes in managerial productivity and in effort costs due to changes in the AFMI concentration level, increases in concentration levels are likely to increase fund net alphas, AFMI size, and AFMI’s direct benefits of net alpha production. These implications support the efficiency of the prevailing real-world AFMI structure of competing fund families, which is more concentrated than that of competing individual funds. Future research could extend our analysis to international fund markets, pension funds, and hedge funds.
References


Appendix

Proof of Equivalence of Managers’ Problems (12) and (13)

Due to competition, if funds offer higher (lower) expected net alphas, investments shift into (out of) it. Thus, in equilibrium all funds offer the same expected net alphas. We show in Proposition 2 and 3 that the equilibrium expected net alpha is the highest one that each manager can achieve at zero profit, and that equilibrium fund sizes are determined by manages costs (which can be viewed as a reflection of their skills).

Suppose that the funds market expected net alpha is $\bar{\alpha}$, where $\bar{\alpha}$ is below the highest level of fund expected net alpha that managers can produce (implying positive profits). While producing expected net alpha of $\bar{\alpha}$, manager $i$ maximizes profits by choosing optimal effort $e_i^*$ that maximizes fund expected net alpha (i.e., the conditions in Proposition 2 hold), and then charges a fee $f_i$ such that her fund expected net alpha becomes $\bar{\alpha}$. Then, the managerial fee becomes

$$f_i = \hat{a} - \hat{b} \frac{S}{W} + A(e_i^*; H) - \bar{\alpha}.$$  \hfill (A1)

Define the profit rate of manager $i$, $\text{pro}_i$, as $\text{pro}_i \triangleq f_i - C^i(e_i^*, s_i; H)$. Then, from the last definition and Equation (A1), we have

$$\bar{\alpha} = \hat{a} - \hat{b} \frac{S}{W} + A(e_i^*; H) - \text{pro}_i - c_{0,i} - c_{1,i}s_i - c_{2,i}(e_i^*; H).$$  \hfill (A2)

As all managers produce the same level of expected net alphas, Equation (A2) implies an equilibrium condition,

$$\text{pro}_i + c_{1,i}s_i = \text{pro}_j + c_{1,j}s_j, \quad \forall i, j.$$  \hfill (A3)

Next, we consider manager $i$’s profit function

$$s_i[f_i - c_{0,i} - c_{1,i}s_i - c_{2,i}(e_i^*; H)],$$  \hfill (A4)

and by the first-order condition, the optimal fund size given manager $i$’s profit level is

$$s_{i_{\text{opt}}} = \frac{f_i - c_{0,i} - c_{2,i}(e_i^*; H)}{2c_{1,i}} = \frac{\text{pro}_i}{2} + \frac{s_i}{2}. \hfill (A5)$$

The latter equality is useful in presenting the optimal size relative to current size. Note that if manager $i$ maximizes her fund’s expected net alpha, her profit rate $\text{pro}_i = 0$, and the condition in Equation (A5), for $s_{i_{\text{opt}}}$, does not exist. For a particular manager $j, j \neq i$, it is possible that $\text{pro}_j$ so high that $s_j < s_{j_{\text{opt}}}$. In other words, although manager $i$ does not
observe other managers’ cost functions and profit rates, she knows that it is possible that some other manager(s) might have incentive to lower down the profit rate to attract investments, increase their fund size and increase their fund’s profits.

Following this argument, we analyze a simple game between manager \( i \) and other managers, grouped as an entity “\(-i\)”. If manager \( i \) improves her fund expected net alpha infinitesimally, then, other managers receive no investments and get zero profit, but her profits will be changed by an infinitesimal amount \( \varepsilon_i \). If, on the other hand, manager \(-i\) (any of the other managers) increases her fund’s expected net alpha infinitesimally, then manager \( i \) receives no investment and earns zero profit, but manager \(-i\)’s profit changes by \( \varepsilon_{-i} \). If all managers produce the same level of fund expected net alphas, then they can make profit. Note that \( \varepsilon_i (\varepsilon_{-i}) \) can be positive or negative, depending on whether manager \( i \)’s \((-i)\)’s fund size is below optimal level or above optimal level. Assume manager \(-i\)’s strategy is to improve the fund expected net alpha with probability \( p \) and maintain \( \bar{\alpha} \) with probability \( 1 - p \). This does not mean that manager \(-i\) randomly chooses her action. Instead, it means that manager \( i \) knows that it is possible that some other manager(s) improve fund expected net alpha to attract investments in order to improve fund profits, and this probability, \( p \), is nontrivial. Suppose that manager \( i \)’s strategy is to improve her fund expected net alpha with probability \( \theta \) and maintain \( \bar{\alpha} \) with probability \( 1 - \theta \). The payoffs of the game are illustrated in the following table, with the row (column) representing manager \( i \)’s \((-i)\)’s action, and with manager \( i \)’s \((-i)\)’s payoffs are in the first (second) figures in the brackets.

<table>
<thead>
<tr>
<th></th>
<th>Maintain ( \bar{\alpha} )</th>
<th>Improve Infinitesimally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maintain ( \bar{\alpha} )</td>
<td>( 1 - \theta )</td>
<td>((0, pro_{-i}s_{-i} + \varepsilon_{-i}))</td>
</tr>
<tr>
<td>Improve Infinitesimally</td>
<td>( \theta )</td>
<td>((0, pro_i s_i + \varepsilon_i, 0))</td>
</tr>
</tbody>
</table>

The expected payoff of manager \( i \) is

\[
\pi_i = (1-p)\left((1-\theta)pro_is_i + \theta(pro_{-i}s_{-i} + \varepsilon_i)\right) + p\theta(pro_is_i + \varepsilon_i) \quad (A6)
\]

The first-order condition is

\[
\frac{d\pi_i}{d\theta} = \varepsilon_i + p \times pro_is_i \quad (A7)
\]

With \( \varepsilon_i \to 0 \), \( d\pi_i/d\theta > 0 \). Thus, manager \( i \) chooses \( \theta = 1 \) to maximize \( \pi_i \). Note that if \( \bar{\alpha} \) is the feasibly maximum fund expected net alpha, managers’ profit rates are zero, and \( \varepsilon_i \) and \( \varepsilon_{-i} \) are negative. Then, in this case, the unique Nash equilibrium is (Maintain \( \bar{\alpha} \), Maintain \( \bar{\alpha} \)).
Therefore, each manager will improve his or her fund expected net alpha as long as it is below the maximum fund expected net alpha, thus, managers’ problems of maximizing profits is equivalent to their maximizing funds’ expected net alphas.

Theories for the Risk-Neutral Case

Where there are infinitely many small risk-neutral investors (i.e., $N \rightarrow \infty$), each investor $j$’s choice of $\delta_j$ has no effect on funds’ sizes, thus no effect on $S / W$; but each of them imposes a negative externality on other investors: as investors increase their investments in funds when they observe positive expected net alphas, the net alpha received by each investor diminishes because of the decreasing returns to scale at both the industry level and fund level. In other words, investors dilute each other’s returns, driving the expected net alphas on all active funds toward zero. A small investor $j$’s problem is

$$\max_{\delta_j} \mathbb{E}(r_j | D) = \max_{\delta_j} \left\{ \mu_p + \delta_j^T \left[ \bar{a} - \hat{b} \frac{S}{W} + A(e_i^*; H) - f_i^* \right] t_M \right\}$$  \hspace{1cm} (A8)

subject to

$$\delta_j^T t_M \leq 1, \quad (A9)$$

$$\delta_{j,i} \geq 0, \forall i, \quad (A10)$$

$$f_i^* - c_i(e_i^*; s_i; H) = 0, \forall i, \quad (A11)$$

$$A_i(e_i^*; H) - c_2i(e_i^*; H) = B_i(e_i^*; H) = 0, \forall i. \quad (A12)$$

These constraints are the same for their counterparts, investors who are mean-variance risk-averse. We assume a symmetric equilibrium such that each investor makes the same equilibrium optimal investment allocation (i.e., $\delta_j^*$ is the same for all $j$), so we have

$$\delta_j^T t_M = S / W. \quad (A13)$$

Where $N=1$, the single large investor internalizes the whole market and can determine $s_i, i=1,\ldots,M$ and $S / W$ to maximize his or her expected portfolio return. In this case, the investor faces a tradeoff between allocating additional dollars to funds to take advantage of fund net alphas and diluting returns on wealth already in funds. Notice that this situation includes the case where there are $N$ investors and they collude to act as a single entity. The single investor’s problem is

$$\max_{\delta_i} \mathbb{E}(r_i | D) = \max_{\delta_i} \left\{ \mu_p + \delta_i^T \left[ \bar{a} - \hat{b} \delta_i^T t_M + A(e_i^*; H) - f_i^* \right] t_M \right\}, \quad (A14)$$

subject to the conditions from (A9) to (A12). Similarly, we have
\[ \delta_i^T t_M = S/W . \] (A15)

The next proposition defines the equilibrium in the risk-neutral case.

PROPOSITION RN1, Unique Nash Equilibrium.
Whether \( N \to \infty \) or \( N = 1 \),
1) There exists a unique Nash equilibrium, \( \{ e^*, f^*, \delta^* \} \), where
   \( e^* \) is a \( M \times 1 \) vector that aggregates individual managers’ optimal effort allocations, \( e_i^* \);
   \( f^* \) is a \( M \times 1 \) vector that aggregates individual managers’ optimal fee allocations, \( f_i^* \);
   and
   \( \delta^* \) is a \( M \times N \) matrix that aggregates the vector of individual investors’ optimal wealth allocations to funds, \( \delta_{ij}^* \).
2) In this equilibrium, managers produce the same expected net alpha that drives their economic profits to zero, by charging only break-even fees, and investors allocate the same wealth proportions to each of the funds.

Proof of Proposition RN1. See the Mathematical Appendix.

The following proposition offers some equilibrium results.

PROPOSITION RN2, Equilibrium by Optimal Allocations.
1) \( N \to \infty \),
\[ \delta_i^T t_M = S/W = \min \frac{X(e_i^*, H)}{\hat{b} + \left( \sum_{i=1}^M c_{i,i}^{-1} \right)^{-1} W}, \forall j. \]
Where \( S/W < 1 \), we have \( E(\alpha_i | D)_{[e^*, f^*, \delta^*]} = 0 \);
where \( S/W = 1 \), we have \( E(\alpha_i | D)_{[e^*, f^*, \delta^*]} = X(e_i^*, H) - \left[ \hat{b} + \left( \sum_{i=1}^M c_{i,i}^{-1} \right)^{-1} W \right] \geq 0 \).
2) \( N = 1 \),
\[ \delta_i^T t_M = S/W = \min \frac{X(e_i^*, H) / 2}{\hat{b} + \left( \sum_{i=1}^M c_{i,i}^{-1} \right)^{-1} W}. \]
Where \( S/W < 1 \), we have \( E(\alpha_i | D)_{[e^*, f^*, \delta^*]} = X(e_i^*, H) / 2 > 0 \);
Where $S/W = 1$, we have $E(\alpha_i \mid D)|_{e',t',\delta'} = X(e', H) - \left[ \hat{b} + \left( \sum_{i=1}^{M} c_{i,i}^{-1} \right)^{-1} \right] W > 0$.

Where $N \to \infty$, small investors invest in the AFMI as long as they infer positive fund expected net alphas. If they have additional wealth to allocate (i.e., $S/W < 1$), they drive equilibrium fund expected net alphas to zero. If they have no additional wealth to allocate (i.e., $S/W = 1$), the equilibrium fund expected net alphas are not driven to 0. At this time, the equilibrium fund expected net alphas are higher if $X(e', H)$ and the equilibrium optimal expected net alpha of an initial marginal investment in the AFMI is higher, whereas the equilibrium fund expected net alphas are lower if the decreasing returns to scale effect in the AFMI (represented by the factor $\hat{b} + \left( \sum_{i=1}^{M} c_{i,i}^{-1} \right)^{-1} W$), is stronger. Where $N = 1$, the single investor internalizes the whole market; and to maximize the expected portfolio returns, he or she never allocates investments to drive the equilibrium fund expected net alphas to zero. When the investor has additional wealth to allocate, he or she allocates wealth to funds such that the effect of decreasing returns to scale in the market is eliminated. If the investor has no additional wealth to allocate, the equilibrium fund expected net alphas are similar to those where there are infinitely many risk-neutral investors.

Also, we can see that if $N = 1$, the industry size is half as large as the counterpart in the setting with infinitely many small risk-neutral investors. The reason is that the single large investor can internalize the market, in the sense that his or her own investment determines $S/W$. When observing positive fund expected net alphas, the investor chooses the optimal level of investment based on the tradeoff of increasing investment in funds (i.e., a larger amount investment captures higher expected returns from the AFMI, but the expected net alpha of each unit of investment reduces due to decreasing returns to scale).

COROLLARY 1 to PROPOSITION RN2. Where $N = 1$, the equilibrium fund expected net alphas $E(\alpha_i \mid D)|_{e',t',\delta'}$ where $S/W = 1$ are higher than those where $S/W < 1$. That is, the equilibrium fund expected net alphas are higher when the single investor has no additional investments to allocate than those when he or she has additional wealth.

The intuition is as follows. Suppose there is a threshold $\bar{W}$ such that the equilibrium $S/W|_{W=\bar{W}} = 1$ and $S/W|_{W=\bar{W}}$ also achieves the internal solution. Any additional wealth above $\bar{W}$ is optimally invested in the passive benchmark, making $S/W < 1$ and not affecting the
fund expected net alphas. All wealth is optimally invested in funds if the wealth level is below $\bar{W}$, and at that time $S/W = 1$. Due to the decreasing returns to scale feature in our model, $E(\alpha_i | D)_{r,s}^{\epsilon_i, r,s}$ decreases with wealth invested in funds until $W = \bar{W}$ and become unaffected by wealth after $W > \bar{W}$. Thus, the equilibrium fund expected net alphas $E(\alpha_i | D)_{r,s}^{\epsilon_i, r,s}$ where $S/W = 1$ are higher than those where $S/W < 1$.

COROLLARY 2 to PROPOSITION RN2. For (large enough) $W$ such that $S/W < 1$, whether $N \to \infty$ or $N = 1$, we have $d(S/W)/dX(e_i^*;H) > 0$ and $d(S/W)/d\left[\hat{b} + \left(\sum_{i=1}^{M} c_{i,i}^{-1}\right)^{-1} W\right] < 0$. That is, higher initial marginal fund expected net alphas, induce a larger equilibrium AFMI size relative to total wealth, whereas stronger decreasing returns to scale effect in the AFMI induces a smaller equilibrium AFMI size relative to total wealth.

PROPOSITION RN3, AFMI Size Sensitivity to Concentration.

Whether $N \to \infty$ or $N = 1$,

1) where $S/W < 1$, we have

a. $\frac{d(S/W)}{dH} = \frac{d(S/W)}{dX(e_i^*;H)} \left[A_{iH}(e_i^*;H) - c_{2H}(e_i^*;H)\right]$.

$d(S/W)/dH \geq 0(<0)$ iff $A_{iH}(e_i^*;H) - c_{2H}(e_i^*;H) \geq 0(<0)$ (where concentration is higher, equilibrium industry size is larger (smaller) if and only if higher concentration induces a larger (smaller) impact on gross alphas than on costs.).

b. $\frac{d^2(S/W)}{dH^2} = \frac{d(S/W)}{dX(e_i^*;H)} \frac{d^2B(e_i^*;H)}{dH^2}$.

$d^2(S/W)/dH^2 \geq 0(<0)$ iff $d^2B(e_i^*;H)/dH^2 \geq 0(<0)$ ($S/W$ is concave (convex) in $H$ if and only if $B(e_i^*;H)$ is concave (convex) in $H$).

2) Where $S/W = 1$, $S/W$ is unrelated to market concentration.

The intuition is as follows. Whether $N \to \infty$ or $N = 1$, if an increase in $H$ induces a higher (lower) impact on net alphas than on costs, the initial marginal fund expected net alpha, $X(e_i^*;H)$, is higher (lower). Consequently, this higher (lower) initial marginal fund expected net alpha attracts (discourages) investments if investors have additional wealth to allocate to
funds (i.e., \( S/W < 1 \)). In this case, the change of the rate at which \( H \) drives up \( S/W \) is positively proportional to the change of the rate at which \( H \) drives up \( B(e^*_i; H) \). If investors have no additional wealth to allocate to funds (i.e., \( S/W = 1 \)), the market is at a corner solution and \( H \) has no effect on \( S/W \).

**PROPOSITION RN4, Net alpha Sensitivity to Concentration.**

1) Where \( N \to \infty \) and \( S/W < 1 \), we have

a. \( dE(\alpha_i \mid D)/dH \mid_{e^*, r^*, s^*} = 0 \),

b. \( d^2E(\alpha_i \mid D)/dh^2 \mid_{e^*, r^*, s^*} = 0 \).

(Equilibrium fund expected net alphas are unrelated to \( H \)).

2) Where \( N \to \infty \) and \( S/W = 1 \), we have

a. \( dE(\alpha_i \mid D)/dH \mid_{e^*, r^*, s^*} = A_H(e^*_i; H) - c_{2H}(e^*_i; H) \).

\[ dE(\alpha_i \mid D)/dH \mid_{e^*, r^*, s^*} \geq 0(<0) \quad \text{iff} \quad A_H(e^*_i; H) - c_{2H}(e^*_i; H) \geq 0(<0) \] (where concentration is higher, equilibrium optimal expected net alphas are larger (smaller) if and only if higher concentration induces a larger (smaller) impact on gross alphas than on costs).

b. \( d^2E(\alpha_i \mid D)/dh^2 \mid_{e^*, r^*, s^*} = d^2B(e^*_i; H)/dh^2 \)

\[ d^2E(\alpha_i \mid D)/dh^2 \mid_{e^*, r^*, s^*} \geq 0(<0) \quad \text{iff} \quad d^2B(e^*_i; H)/dh^2 \geq 0(<0) \] (\( E(\alpha_i \mid D) \mid_{e^*, r^*, s^*} \) is convex (concave) in \( H \) if and only if \( B(e^*_i; H) \) is convex (concave) in \( H \)).

3) Where \( N = 1 \) and \( S/W < 1 \), we have

a. \( dE(\alpha_i \mid D)/dH \mid_{e^*, r^*, s^*} = 0.5 \left[ A_H(e^*_i; H) - c_{2H}(e^*_i; H) \right] \)

\[ dE(\alpha_i \mid D)/dH \mid_{e^*, r^*, s^*} \geq 0(<0) \quad \text{iff} \quad A_H(e^*_i; H) - c_{2H}(e^*_i; H) \geq 0(<0) \] (where concentration is higher, equilibrium optimal expected net alphas are larger (smaller) if and only if higher concentration induces a larger (smaller) impact on gross alphas than on costs).

b. \( d^2E(\alpha_i \mid D)/dh^2 \mid_{e^*, r^*, s^*} = 0.5d^2B(e^*_i; H)/dh^2 \)

\[ d^2E(\alpha_i \mid D)/dh^2 \mid_{e^*, r^*, s^*} \geq 0(<0) \quad \text{iff} \quad d^2B(e^*_i; H)/dh^2 \geq 0(<0) \] (\( E(\alpha_i \mid D) \mid_{e^*, r^*, s^*} \) is convex (concave) in \( H \) if and only if \( B(e^*_i; H) \) is convex (concave) in \( H \)).
4) Where $N=1$ and $S/W=1$, we have

a. \[
dE(\alpha_i \mid D) / dH \mid_{(e^*, t^*, s^*)} = A_H(e^*_i; H) - c_{2H}(e^*_i; H)
\]

\[
dE(\alpha_i \mid D) / dH \mid_{(e^*, t^*, s^*)} \geq 0(\neq 0) \iff A_H(e^*_i; H) - c_{2H}(e^*_i; H) \geq 0(\neq 0)
\]

(where concentration is higher, equilibrium optimal expected net alphas are larger (smaller) if and only if higher concentration induces a larger (smaller) impact on gross alphas than on costs).

b. \[
d^2E(\alpha_i \mid D) / dH^2 \mid_{(e^*, t^*, s^*)} = d^2B(e^*_i; H) / dH^2
\]

\[
d^2E(\alpha_i \mid D) / dH^2 \mid_{(e^*, t^*, s^*)} \geq 0(\neq 0) \iff d^2B(e^*_i; H) / dH^2 \geq 0(\neq 0) \quad (E(\alpha_i \mid D) \mid_{(e^*, t^*, s^*)} \text{ is convex (concave) in } H \text{ if and only if } B(e^*_i; H) \text{ is convex (concave) in } H).
\]

The intuition of Proposition RN4 is as follows. An increase in $H$ affects $E(\alpha_i \mid D) \mid_{(e^*, t^*, s^*)}$ at two stages. At the first stage, an increase in $H$ changes managers’ ability to produce net alphas. At the second stage, investors react to the changes in fund expected net alphas by changing the investment level, consequently affecting $E(\alpha_i \mid D) \mid_{(e^*, t^*, s^*)}$ under a decreasing returns to scale framework. Where $N \rightarrow \infty$ and $S/W < 1$, investors adjust their investments to drive $E(\alpha_i \mid D) \mid_{(e^*, t^*, s^*)}$ to zero, so $E(\alpha_i \mid D) \mid_{(e^*, t^*, s^*)}$ is unrelated to $H$. Where $N \rightarrow \infty$ and $S/W = 1$, investors have no additional wealth to allocate to funds even though funds’ expected net alphas are changed by a higher $H$. Thus, whether $E(\alpha_i \mid D) \mid_{(e^*, t^*, s^*)}$ increases depends on whether managers are able to produce higher net alphas under this higher concentration level (i.e., whether $A_H(e^*_i; H) - c_{2H}(e^*_i; H) > 0$). Also, in this case, if $H$ is higher, the change of marginal $E(\alpha_i \mid D) \mid_{(e^*, t^*, s^*)}$ depends on the change of marginal $B(e^*_i; H)$. Where $N=1$, if the single investor observes an increase in fund expected net alphas due to a higher $H$, he or she would not allocate wealth to completely offset the increase in fund expected net alphas due to a tradeoff, if the single investor has additional wealth to allocate (i.e., $S/W < 1$). If the investor has no additional wealth to allocate (i.e., $S/W = 1$), his or her choice does not affect the marginal $E(\alpha_i \mid D) \mid_{(e^*, t^*, s^*)}$ due to a higher $H$. In either situation, with a higher $H$, whether
E(\alpha_i \mid D)_{[e',r',s']}$ increases depends on whether managers can produce higher net alphas, and the change of marginal $E(\alpha_i \mid D)_{[e',r',s']}$ depends on the change of marginal $B(e^*_i; H)$.

**PROPOSITION RN5, Relation between Net Alpha and Market Share.**

Whether $N \to \infty$ or $N = 1$, an increase (decrease) in $c_{i,j}$, while $c_{i,j}, \forall j \neq i$ are unchanged, induces a decrease (increase) in $s_i \mid S$ and an increase (decrease) in $s_j \mid S, \forall j \neq i$. Also,

1) Where $S/W < 1$, it induces no change in $E(\alpha_i \mid D)_{[e',r',s']}$ and $E(\alpha_j \mid D)_{[e',r',s']}, \forall j \neq i$. Thus, $E(\alpha_i \mid D)_{[e',r',s']}$ is unrelated to $s_i \mid S, \forall i$.

2) Where $S/W = 1$, it induces a decrease (increase) in $E(\alpha_i \mid D)_{[e',r',s']}$, thus $E(\alpha_i \mid D)_{[e',r',s']} \mid s_i \mid S$ are positively related—an internality effect; it induces a decrease (increase) in $E(\alpha_j \mid D)_{[e',r',s']}, \forall j \neq i$, thus $E(\alpha_i \mid D)_{[e',r',s']}$ and $s_j \mid S$ are negatively related $\forall j \neq i$—an externality effect.

The intuition of Proposition RN5 is as follows. First, based on Proposition 3, we can see that any change in $c_{i,j}$, keeping $c_{i,j}, \forall j \neq i$ unchanged, results in a change in $s_i \mid S$ in the opposite direction and a change in $s_j \mid S, \forall j \neq i$ in the same direction. Also, an increase in $c_{i,j}$, affects $E(\alpha_i \mid D)_{[e',r',s']}$ at two stages. At the first stage, it increases manager $i$’s average cost sensitivity to size and induces a decrease in fund expected net alphas. As manager $i$ offers a lower fund expected net alpha, investments shift from fund $i$ to other funds, making all other funds’ fund expected net alphas lower due to decreasing returns to scale at the fund level. At the second stage, a decrease in fund expected net alphas discourages investments in the AFMI, which in turn drives up fund expected net alphas due to the effect of decreasing returns to scale. Where $S/W < 1$, infinitely many small risk-neutral investors drive fund expected net alphas to zero, whereas a single large risk-neutral investor strategically allocates wealth to funds. In either case, the two stages’ effects of changes in $c_{i,j}$ on $E(\alpha_i \mid D)_{[e',r',s']}$’s are completely cancelled out; thus, for each fund $i$, $E(\alpha_i \mid D)_{[e',r',s']}$ is unrelated to $c_{i,j}$ and, consequently, unrelated to $s_i \mid S$. Where $S/W = 1$, investors have no additional wealth to allocate, so the second stage’s effect does not exist.
and we find $E(\alpha_i | D)|_{(e^*, r^*, s^*)}$’s are driven down by an increase in $c_{i,j}$, keeping $c_{i,j}$, $\forall j \neq i$ unchanged. Consequently, we have a positive relation between $E(\alpha_i | D)|_{(e^*, r^*, s^*)}$ and $s_i / S$, and a negative relation between $E(\alpha_j | D)|_{(e^*, r^*, s^*)}$ and $s_j / S$, $\forall j \neq i$. In other words, in this case, any change in manager $i$’s average cost sensitivity to size induces an internality effect, a positive relation between its market share and equilibrium fund expected net alphas, and induces an externality effect, a negative relation between other funds’ market shares and their equilibrium fund expected net alphas.

Proof of Proposition RN2, RN3, RN4, RN5 and the corresponding corollaries. See the Mathematical Appendix.

Figure 4 and Figure 5 illustrate the numerical results for the case of risk-neutral investors using the same numerical parameters as the risk-averse case. The results of $e_i^*$, $f_i^*$, and $B(e_i^*; H)$ are the same as those in the risk-averse case, whether there are infinitely many small risk-neutral investors or there is a single large investor. Where there are infinitely many small risk-neutral investors, when $H$ is small, we have $S / W = 1$, and $E(\alpha_i | D)|_{(e^*, r^*, s^*)}$ is positive and is decreasing with $H$. When $H$ becomes larger, $S / W$ starts to decrease and slightly increase after it achieves the minimum point. In this case, $S / W$ increases with $B(e_i^*; H)$, and we have $E(\alpha_i | D)|_{(e^*, r^*, s^*)}$ equal to zero. This is because in our calibration, the difference of $H$’s impact on gross alphas and on costs decreases with $H$ when $H$ is small and then slightly increases with $H$ when $H$ is large. In addition, the curvature of $S / W$ and $E(\alpha_i | D)|_{(e^*, r^*, s^*)}$ in $H$ are convex.

Where there is a single large investor in the market, the results are similar except that the levels of $S / W$ are much smaller than those where there are infinitely many risk-neutral investors, and $E(\alpha_i | D)|_{(e^*, r^*, s^*)}$’s are positive across the levels of $H$. This is because the single investor can internalize the AFMI, limiting the investments in the funds and maximizing portfolio expected returns.
Figure 4. Infinitely Many Small Risk-Neutral Investors—Comparative Statistics with Respect to Market Concentration

This figure presents the numerical results for the case of infinitely many small risk-neutral investors. The two subplots on the top illustrate the equilibrium optimal management effort and management fees at each market concentration level. The two subplots in the middle report $B(e_i^*, H)$ at each market concentration level and the equilibrium $S/W$ ratio at each $X(e_i^*, H)$ level. The two subplots at the bottom show the equilibrium $S/W$ ratio and the equilibrium fund expected net alphas at each market concentration level.
Figure 5. A Single Large Risk-Neutral Investor—Comparative Statistics with Respect to Market Concentration

This figure presents the numerical results for the case of a single large risk-neutral investor. The two subplots on the top illustrate the equilibrium optimal management effort and management fees at each market concentration level. The two subplots in the middle report $B(e^*_i, H)$ at each market concentration level and the equilibrium $S/W$ ratio at each $X(e^*_i, H)$ level. The two subplots at the bottom show the equilibrium $S/W$ ratio and the equilibrium fund expected net alphas at each market concentration level.
**Agency Benefits Due to Market Concentration**

We use the same settings as those in Section 2.5. We consider only the case where there are infinitely many risk-neutral investors. Managers are colluding to choose $e_i$ to maximize their objective function (37), subject to the constraint (38). The first-order-condition gives

$$
\sum_{j=1}^{M} \left[ A_i(e_i^*;H)f(H) - C_i(e_i^*, s_i; H) \right] - \left[ A_i(e_i^*;H) - f(H) \right] \left[ W\left(\sum_{t=1}^{M} c_{i,t}^{-1}\right)^{-1} A_i(e_i^*;H) - c_{2i}(e_i^*;H) \right] = 0.
$$

(A16)

The second-order condition with respect to $e_i$ is satisfied, so $e_i^*$ given by (A16) is a maximum point. Also, in equilibrium, where $S/W < 1$, we have

$$
S = \frac{\hat{a} + A(e_i^*;H) - f(H)}{\hat{b}},
$$

(A17)

$$
\frac{d(S/W)}{dH} = \frac{A_{ii}(e_i^*;H)e_i^*(H) + A_{ii}(e_i^*;H) - f'(H)}{\hat{b}},
$$

(A18)

$$
\frac{dE(\alpha_i | D)}{dH}_{|e^*,s^*} = 0.
$$

(A19)

Thus, $d(S/W) / dH \geq 0$ if and only if $A_{ii}(e_i^*;H)e_i^*(H) + A_{ii}(e_i^*;H) - f'(H) \geq 0$. In other words, if and only if higher concentration induces higher net alphas than agency benefits, investors are willing to invest more in funds, eventually driving equilibrium optimal expected net alphas to zero. Where $S/W = 1$, managers choose $e_i^*$ such that

$$
S = \frac{\hat{a} + A(e_i^*;H) - f(H)}{\hat{b}} = 1,
$$

(A20)

so

$$
\frac{d(S/W)}{dH} = 0,
$$

(A21)

and (A19) still holds because if managers can optimally induce investors to invest all their wealth in funds, they choose the minimum effort to do so. In this case, both $S/W$ and $E(\alpha_i | D)_{|e^*,s^*}$ are unaffected by $H$. 

59