ENDOGENOUS DYNAMIC CONCENTRATION OF THE ACTIVE FUND MANAGEMENT INDUSTRY

David Feldman[a] and Jingrui Xu[b,c]

Latest revision, October 13, 2021

Abstract. We introduce continuous-time rational models of concentration in the active fund management industry, where managers with heterogenous dynamic unobservable abilities compete for investments of risk-neutral or risk-averse investors. Fund flows increase with managers' inferred abilities, inducing endogenous dynamic distribution of fund sizes, thus, dynamic industry concentration [measured by the Herfindahl-Hirschman Index (HHI)]. In equilibrium, increases in managers' inferred abilities that are sufficiently larger (smaller) than others', increase (decrease) the HHI. Changes in inferred abilities affect the HHI dynamics and curvature and, consequently, funds' performances and sizes. These combined effects explain corresponding stylized facts and offer guidance for empirical studies.

JEL Codes: G11, G14, G23, J24, L11

Keywords: Active fund management, Market concentration, Fund performance, Alpha, Dynamic unobservable manager abilities, Learning, Linear filtering

---

[a] Banking and Finance, UNSW Business School, UNSW Sydney, NSW 2052, Australia. Phone: +61-2-93855748. Email: d.feldman@unsw.edu.au

[b] Wang Yanan Institute for Studies in Economics (WISE) and Department of Finance, School of Economics, Xiamen University, China. Email: jingrui.xu@xmu.edu.cn

[c] We thank Linda Pesante for helpful discussion.

1 Introduction

Active fund management industry (AFMI) investors seek excess returns over passive indices allocating, based on preference, wealth to AFMI managers, who incur costs and charge fees. [e.g., Berk and Green (2004), Berk (2005)]. This implies that in AFMI it is the buyers (investors) who determine the quantity of production (i.e., investments). The producers (fund managers), with fixed prices (management fees), stimulate the quantity of production by offering higher returns to improve profits. These features make AFMI different from classical production industries, in which the producers determine the product price and quantity of production, and the buyers decide the amount to buy. As AFMI manages a huge amount of wealth,[1] the study of AFMI structure, in particular AFMI concentration, and its dynamics, offers special economic insights.

Current literature has shown that AFMI concentration has impact on AFMI size and performance [Feldman, Saxena, and Xu (2020, 2021)]. This implies that the dynamics of AFMI concentration exerts significant effect on AFMI over time. However, there is little study on the dynamics of AFMI concentration, and the goal of this paper is to fill the gap.

We develop a continuous-time framework to model multiple heterogeneous active funds and a passive benchmark portfolio. In our model, fund managers' abilities to create abnormal returns are dynamic and unobservable to both investors and managers. Both learn these abilities by observing fund returns.[2] Managers set constant management fees and, over time, choose the size of wealth they actively manage to determine fund net alpha[3] while maximizing fund profits. Risk-neutral investors supply capital with infinite elasticity to funds that have positive excess expected returns compared to the benchmark return. There are decreasing returns to scale in AFMI such that funds' total costs are increasing and convex in

---

[1] According to the Investment Company Institute (ICI), the total net assets of worldwide regulated open-end funds (including mutual funds, exchange-traded funds, and institutional funds), were $63.1 trillion in 2020. See the 2021 Investment Company Fact Book at ICI website, https://www.ici.org/system/files/2021-05/2021_factbook.pdf, accessed on October 12th, 2021.

[2] The active funds' observable gross alphas follow Itô processes, in which the drift terms depend on the dynamic unobservable manager ability levels. These ability levels also follow Itô processes. Their diffusions are (locally, imperfectly) correlated with those of funds' gross alpha processes.

[3] Berk and Green (2004) shows that the case in which the fund manager actively manages the whole fund and chooses his/her management fee at each time is equivalent to the case in which the fund manager chooses the amount of the fund to actively manage at each time under a fixed management fee. As the latter case is more realistic, we focus on it to conduct our analyses.

the size of assets under active management. All these settings are similar to those in the current models [e.g., Berk and Green (2004), Dangl, Wu, and Zechner (2008), Brown and Wu (2013, 2016), and Feldman and Xu (2021)].

In the unique Nash equilibrium in which each market participant makes an optimal decision, a fund's size is increasing and convex in the expectation of its manager's ability to create abnormal returns (hereafter, inferred ability). Also, a fund's profit is unaffected by the fixed fee but is increasing and convex with its manager's inferred ability.[4] This implies that in equilibrium, better managers manage larger funds and receive larger rewards. Further, many common measures of the AFMI's industrial organization are less informative. For example, as fund costs are transferred to investors as deductions in fund returns, a fund's profit margin (the difference of revenue and costs, divided by the revenue) and Lerner Index (the difference of fee and marginal cost, divided by fee) are equal to one, and profit on each dollar under management is the fixed percentage management fee. As profit margin and Lerner Index are indicators of funds' profitability and market power, respectively, the above results imply that there are no dynamics in funds' profitability or market power. This makes AFMI's concentration dynamics more relevant in studying the AFMI's industrial organization dynamics.

We use the Herfindahl-Hirschman index (HHI) to measure AFMI concentration, which is the sum of funds' market shares squared,[5] for several reasons. First, the HHI reflects the combined influence of both unequal fund sizes and the concentration of activity in a few large funds, so it has advantage over other concentration measures, such as a concentration ratio, which only sums up the market shares of a few largest funds and ignores the information of other funds. Second, some regulatory agencies use the HHI to measure concentration.[6] Third,

---

[4] The intuition is that, with a fixed fee, to maximize fund profit, a fund manager tries to attract as much investment as possible by offering positive expected net alpha to investors. Under decreasing returns to scale, the manager's inferred ability determines the expected net alphas that he/she can produce and then determines the equilibrium fund size. A manager with higher inferred ability puts a larger amount of the fund under active management to offer higher expected net alpha, and investors respond to this higher inferred ability more intensively when investing in this fund.

[5] A higher (lower) HHI implies a more (less) concentrated AFMI. The highest value of HHI is one, which implies a monopolistic AFMI. The lowest value of HHI is the inverse of the number of funds, which implies identical funds in the AFMI.

[6] For example, the U.S. Census calculates industry concentration as HHI, used by regulatory agencies such as the Federal Trade Commission and Department of Justice [e.g., Ali, Klasa, and Yeung (2009) and Azar, Schmalz, and Tecu (2018)].

the HHI is a common measure of concentration in current theoretical and empirical studies.[7] Fourth, new concentration measures are calculated based on the HHI. For example, the normalized Herfindahl-Hirschman index adjusts the effects of the number of rivals,[8] and the modified Herfindahl-Hirschman index captures the concentrations of producers and of shareholders' ownership.[9]

As managers' inferred abilities determine equilibrium fund sizes, fund market shares depend on managers' inferred abilities relative to those of other managers. Then, the heterogeneity in managers' inferred abilities, or the relative inferred abilities, determines the equilibrium AFMI HHI (hereafter HHI). Consequently, the dynamics of the HHI depends on the change in managers' relative inferred abilities. We find that if a fund's inferred ability is sufficiently large (small) relative to other funds, then an increase in the manager's inferred ability, holding other managers' inferred abilities unchanged, has a positive (negative) impact on the change in the HHI. This positive (negative) impact is stronger if this manager's inferred ability, fund size factor, which is the inverse of the product of management fee and the decreasing returns to scale coefficient, and/or sensitivity of gross alpha to ability are larger.

The intuition is that if a manager's inferred ability is sufficiently large, then the fund's equilibrium size is sufficiently large such that the fund dominates in the market. A higher inferred ability attracts more investment to this fund, making it larger and the AFMI more concentrated at this fund. Further, the manager's higher inferred ability, a higher fund size factor, and a higher sensitivity of gross alpha to ability all induce a larger fund size, making this fund even larger and the AFMI more concentrated. On the other hand, if this manager's inferred ability is sufficiently small, then the fund's equilibrium size is sufficiently small relative to other funds' sizes. A higher inferred ability attracts more investment to this fund, making its size closer to other funds and making the AFMI less concentrated. Also, if this inferred ability, fund size factor, and/or sensitivity of gross alpha to ability are larger, then the

---

[7] See, for example, theoretical models, such as Bustamante and Donangelo (2017) and Corhay, Kung, and Schmid (2020), that study firm concentration; and Feldman, Saxena, and Xu (2020, 2021) that study AFMI concentration; and empirical models, such as Cornaggia, Mao, Tian, and Wolfe (2015) that study labor concentration and industry concentration; Spiegel and Tookes (2013) and Gu (2016) that study product market concentration;and Giannetti and Saidi (2019) that study credit concentration.

[8] See, for example, Cremers, Nair, and Peyer (2008).

[9] See, for example, O'Brien and Salop (2000), Azar, Schmalz, and Tecu (2018), and Koch, Panayides, and Thomas (2021).

fund's size becomes larger and closer to other funds, making the AFMI less concentrated.

Regarding second-order effects, the HHI is concave in a manager's inferred ability, if this manager's inferred ability is sufficiently large or sufficiently small relative to that of other managers. Then, over the next infinitesimal period, this concavity has a negative impact on the change in the HHI. The HHI is convex in this manager's inferred ability if all managers' inferred abilities are sufficiently close. Then, over the next infinitesimal period, this convexity has a positive impact on the change in the HHI. If the HHI is concave (convex) in this manager's inferred ability, then both a higher fund size factor and a higher sensitivity of gross alpha to ability make the effect of this concavity (convexity) on the change in the HHI stronger.

The intuition is that if a manager's inferred ability is sufficiently large (small), then this fund's market share is sufficiently large (small), making the AFMI concentrated at this fund (at other funds). Although the higher (lower) the manager's inferred ability makes the AFMI more concentrated at this fund (at other funds), it becomes more and more difficult to increase the HHI in this way. On the other hand, if all managers' inferred abilities are sufficiently close, then their fund sizes are sufficiently close. In that case, both a larger and a smaller inferred ability of a manager can make this fund's size deviate from other funds' sizes, making the AFMI more concentrated. It is easier to make this fund's size deviate from other funds' sizes and to increase the HHI if the absolute change in this manager's inferred ability is larger. Further, both a higher fund size factor and a higher sensitivity of gross alpha to ability make this fund's equilibrium size larger, thus more relevant, in the AFMI. So, if the HHI is concave (convex) in a manager's inferred ability, this fund's larger size intensifies the effect of this concavity (convexity) on the change in the HHI.

Moreover, in a special case in which managers' unobservable abilities are constant, over time the estimation precision of inferred abilities monotonically improves, and the sensitivities of inferred abilities to innovation shocks decrease monotonically. As time goes to infinity, AFMI reaches a steady state in which investors know managers' abilities and managers' inferred abilities stay unchanged. Consequently, investments in funds stay unchanged, making fund market shares and the HHI constant.

We also consider the case in which investors are mean-variance risk-averse and maximize portfolio instantaneous Sharpe ratios by allocating wealth to funds and the passive benchmark. We find that investors' risk considerations decrease equilibrium fund sizes. However, the way to compare fund sizes relative to those of others does not depend on investors' risk considerations, so the dynamics of the HHI relates to managers' relative inferred abilities in a way similar to that in the risk-neutral case.

Further, our framework is compatible with the effect of fund entrances and exits on the HHI, if we allow the total number of funds to change over time and set fund survival inferred ability levels to zero, that is, funds exit (enter) the market if their inferred abilities decrease (increase) to zero. These survival inferred ability levels can be regarded as those endogenously chosen by profit-maximizing managers. This is because funds with positive inferred abilities earn positive equilibrium profits and optimally choose to stay in the market to earn the profits, whereas without short selling of assets, funds with negative inferred abilities optimally choose to put zero assets under active management to avoid losses and exit AFMI. Also, equilibrium fund sizes become zero when managers' inferred abilities are zero. Then, funds exit (enter) the market when their equilibrium fund sizes continuously decrease from positive to negative (increase from negative to positive) without causing jumps in fund sizes or jumps in the HHI. Consequently, our theoretical results on the dynamics of the HHI, which is an Itô process, still apply when funds exit and enter the market. In other words, funds' entrances and exits do not affect the dynamics of the HHI immediately, but they change the set of funds in AFMI and affect the dynamics of the HHI after that.

To model the effect of AFMI concentration on AFMI size and net alphas, we assume that a higher AFMI concentration increases both the productivity and cost of active management, and first set AFMI concentration to be exogenous.[10] Then, in equilibrium higher AFMI concentration induces a larger (smaller) AFMI size and higher (lower) expected net alphas if and only if higher AFMI concentration exerts a larger (smaller) impact on the productivity of active management than on the cost of active management. Consequently, AFMI size and expected net alphas move in the same direction: either both increase with AFMI concentration or both decrease with it. Furthermore, AFMI size and expected net alphas are either both convex in AFMI concentration or both concave in AFMI concentration. These

[10] This assumption is consistent with those of Feldman, Saxena, and Xu (2020, 2021). See discussions in Section 4.2.

theoretical predictions are consistent with those of Proposition RA3 in Feldman, Saxena, and Xu (2020). If we endogenize AFMI concentration and calculate it as the HHI in this framework, then a positive shock in a fund's return, which induces a positive change in the manager's inferred ability, has direct instantaneous positive effects on the fund's net alpha and on AFMI size. It also has indirect instantaneous effects on the fund's net alpha and AFMI size through the HHI. That is, the change in this manager's inferred ability increases or decreases the HHI, depending on parameter values; consequently, the change in the HHI increases (decreases) both the fund's net alpha and AFMI size, if the change in the HHI induces larger (smaller) productivities of active management than the costs of active management. These results on the dynamics in AFMI over time are new compared to those of Feldman, Saxena, and Xu (2020), which studies one-period fixed-point equilibria.

Regarding the empirical AFMI concentration levels, we find that in the U.S. active equity mutual fund market, HHIs at fund level and at fund family level have fluctuated over the last thirty-one years, consistent with a framework with dynamic unobservable manager abilities and inconsistent with a framework with constant such abilities. This finding is consistent with those of Feldman and Xu (2021). Also, we find that the HHI at fund level (at fund family level) are more correlated with the large funds' (fund families') market shares than the number of funds (fund families) in the market, showing that the relative inferred abilities that induce market shares are more relevant than the number of competitors in analyzing the HHI. Therefore, it is important to study heterogeneous managers where the HHI captures managers' relative inferred abilities, instead of homogeneous managers, where the HHI is simply the inverse of the number of competitors. In addition, although the number of funds and the number of fund families increase over time and their time series have a correlation coefficient close to 1, HHIs at fund level and at fund family level evolve with quite different time patterns and have a correlation coefficient of only 0.17. This implies that AFMI concentrated at fund level is not necessarily concentrated at fund family level, and vice versa. Using only the HHI at fund level or the HHI at fund family level might not represent AFMI concentration well.

We also find that from 1995 to the early 2000s, the number of funds and fund families keeps increasing, and the HHIs at both fund level and fund family level decrease. Wahal and

Wand (2011) shows that in this period, incumbents in the mutual fund market that have a high overlap in their portfolio holdings with those of new entrants experience lower fund flows and lower alphas. Kosowski, Timmermann, Wermers, and White (2006) and Fama and French (2010) find a decrease in fund manager performance in similar periods. Our model can explain these stylized empirical findings coherently. As new funds and fund families hold portfolios similar to those of the incumbents, as Wahal and Wand (2011) discover, funds' inferred abilities become more similar. Similar funds' inferred abilities result in similar equilibrium fund sizes, so the HHI decreases. The decrease in the HHI might decrease the productivity of funds' active management more than the cost of active management, because when more managers hold similar active portfolios and trade on similar assets, it is much more difficult for them to find opportunities for abnormal returns. Consequently, the decrease in the HHI induces lower fund sizes and fund net alphas.

We contribute to the literature in the following ways. To our best knowledge, we develop the first model of dynamic equilibrium AFMI concentration, under a framework of multiple heterogeneous managers with dynamic unobservable abilities. We show that our theoretical results are valid whether investors are risk neutral or mean-variance risk averse, and whether there are fund entrances or exits. Our model can explain stylized findings regarding AFMI concentration, size, and performance. In addition, our theoretical results provide guidance for future research on empirical AFMI concentration. For example, we highlight potential discrepancies of AFMI concentration dynamics at fund and family levels.

The rest of this paper is organized as follows. Section 2 introduces our model. Section 3 provides simulation results of our equilibria. Section 4 discusses an extension of our model where AFMI HHI affects fund net alpha productions. Section 5 illustrates the empirical HHI, explains some stylized findings of AFMI concentration, size, and performance, and provides guidance for future research. Section 6 concludes.

2 The Model

We introduce a rational equilibrium framework to study the dynamics of AFMI concentration. We first demonstrate how investors and managers infer managers' dynamic unobservable abilities and form equilibrium fund sizes by solving, respectively, investors' portfolio optimization problems and managers' profit-maximizing problems. Our baseline model assumes risk-neutral investors; then we introduce a model with mean-variance risk-averse investors. We also discuss how funds' entrances and exits affect our analyses of the dynamics of AFMI concentration.

Some of our model settings are similar to those of Berk and Green (2004), Brown and Wu (2016), and Feldman and Xu (2021).[11] Also, in our model, investors can invest in multiple independent heterogeneous active funds, each with one manager, and in a passive benchmark portfolio. This multiple-fund setting is similar to the one in Brown and Wu (2016). Within a continuous-time framework, we study the active fund managers and investors over a finite time interval, at times t, $t \in [0, T]$, where T , $T > 0$ is a constant.

2.1 Observable Fund Returns and Unobservable Manager Abilities: Filtering

Let $\boldsymbol{\xi}_{\mathbf{t}}$, $0 \leq t \leq T$ be an $n \times 1$ vector of active funds' gross share prices, i.e., share price before fund costs and fees, where the ith element is $\xi_{i,t}$, $i = 1, \ldots, n$, and $n \geq 2$. Then, $\mathbf{I^{-1}}(\boldsymbol{\xi}_{\mathbf{t}})\mathbf{d}\boldsymbol{\xi}_{\mathbf{t}}$ is the $n \times 1$ vector of the instantaneous fund gross rates of return, where $\mathbf{I}(\boldsymbol{\xi}_{\mathbf{t}})$ is an $n \times n$ diagonal matrix with $\xi_{i,t}$ as the ith diagonal element.[12] For simplification, we assume that active funds have beta loads of one on the passive benchmark portfolio. To focus on the active funds' return, similar to Feldman and Xu (2021), we normalize the passive benchmark portfolio's return to zero so that the vector of instantaneous fund gross returns in excess of the passive benchmark is also $\mathbf{I^{-1}}(\boldsymbol{\xi}_{\mathbf{t}})\mathbf{d}\boldsymbol{\xi}_{\mathbf{t}}$. Hereafter, we call the active funds' instantaneous gross alphas $\mathbf{I^{-1}}(\boldsymbol{\xi}_{\mathbf{t}})\mathbf{d}\boldsymbol{\xi}_{\mathbf{t}}$, briefly, gross alphas.

The active funds' gross alphas depend on the $n \times 1$ vector of active fund managers' instantaneous abilities, $\boldsymbol{\theta}_{\mathbf{t}}$, $0 \leq t \leq T$, to beat the benchmark, where the ith element is $\theta_{i,t}$, $i = 1, \ldots, n$. We call them briefly, abilities. The abilities are unobservable to both fund managers and investors. Fund managers and investors learn about $\boldsymbol{\theta}_{\mathbf{t}}$ by observing the history of fund gross alphas $\mathbf{I^{-1}}(\boldsymbol{\xi}_{\mathbf{t}})\mathbf{d}\boldsymbol{\xi}_{\mathbf{t}}$, $0 \leq s \leq t$ (or equivalently by observing gross fund share prices $\boldsymbol{\xi}_{\mathbf{s}}$, $0 \leq s \leq t$). We assume a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration

---

[11] Similar to Berk and Green (2004), Brown and Wu (2016), and Feldman and Xu (2021), managers and investors are symmetrically informed; the model is in partial equilibrium; managers' actions do not affect the passive benchmark returns; and we do not model sources of managers' abilities to outperform the passive benchmarks portfolios.
[12] The $n \times 1$ vector $\mathbf{d}\boldsymbol{\xi}_{\mathbf{t}}$ has its ith element as $d\xi_{i,t}$, which is the differential of $\xi_{i,t}$, $i = 1, \ldots, n$. Hereafter, a vector with $\mathbf{d}$ on the left has a similar definition.

$\{\mathcal{F}_t\}_{0\le t\le T}$, a right-continuous, nondecreasing family of sub-σ-algebras. An $n \times 1$ vectors of independent Wiener processes, $\mathbf{W_{1,t}}$ and $\mathbf{W_{2,t}}$, $0 \le t \le T$, are adapted to this filtration, where their ith elements are $W_{1i,t}$ and $W_{2i,t}$, $i = 1, \dots, n$, respectively.[13] The unobservable $\boldsymbol{\theta}_\mathbf{t}$ and the observable $\boldsymbol{\xi}_\mathbf{t}$ evolve as follows:

$$\mathbf{d}\boldsymbol{\theta}_\mathbf{t} = (\mathbf{a_0} + \mathbf{a_1}\boldsymbol{\theta}_\mathbf{t})dt + \mathbf{b_1}\mathbf{dW_{1,t}} + \mathbf{b_2}\mathbf{dW_{2,t}} \tag{1}$$

$$\mathbf{I^{-1}}(\boldsymbol{\xi}_\mathbf{t})\mathbf{d}\boldsymbol{\xi}_\mathbf{t} = \mathbf{A}\boldsymbol{\theta}_\mathbf{t}dt + \mathbf{B}\mathbf{dW_{2,t}} \tag{2}$$

with initial conditions $\boldsymbol{\theta}_\mathbf{0}$ and $\boldsymbol{\xi}_\mathbf{0}$, respectively. The $n \times n$ constant diagonal matrices $\mathbf{a_0}$, $\mathbf{a_1}$, $\mathbf{b_1}$, $\mathbf{b_2}$, $\mathbf{A}$, and $\mathbf{B}$ have diagonal elements $a_{i,0}$, $a_{i,1}$, $b_{i,1}$, $b_{i,2}$, A_i, and B_i, $i = 1, \dots, n$, respectively. To make economic sense, we assume that $A_i > 0$ (otherwise an ability becomes a "disability"). For simplicity and without loss of generality, we assume $B_i > 0$. While abilities are unobservable to managers and investors, the evolution processes ("laws of motion") and all parameter values are common knowledge.

This setting implies the following. First, the abilities, $\boldsymbol{\theta}_\mathbf{t}$, to beat the benchmark follow dynamic processes. Second, the fund gross alphas, $\mathbf{I^{-1}}(\boldsymbol{\xi}_\mathbf{t})\mathbf{d}\boldsymbol{\xi}_\mathbf{t}$, depend on the managers' abilities and on random shocks. As $A_i > 0$, $i = 1, \dots, n$, a manager with positive (negative) ability tends to create positive (negative) fund gross alpha, and the larger A_i is, the higher is the sensitivity of gross alpha to ability. Also, B_i, $i = 1, \dots, n$ is the diffusion coefficient of fund gross alpha, which positively corresponds to its volatility. Third, as $\mathbf{a_0}$, $\mathbf{a_1}$, $\mathbf{b_1}$, $\mathbf{b_2}$, $\mathbf{A}$, and $\mathbf{B}$ are diagonal matrices, over time a manager's ability and gross alpha are independent of other managers'.[14,15] Fourth, where $b_{i,2} > 0$ ($b_{i,2} < 0$), that is, b_{i2} is strictly positive (negative), the shock $W_{2i,t}$ affects manager i's ability and fund gross alpha, which, consequently, are instantaneously positively (negatively) correlated, as $b_{i,2}B_i > 0$ ($b_{i,2}B_i < 0$). Where $b_{i,2} = 0$, and $b_{i,1} > 0$, manager i's ability and gross alpha are affected by independent shocks, thus are instantaneously uncorrelated. A larger $b_{i,2}$ relative to $b_{i,1}$

---

[13] For any i and j, $dW_{1i,t}dW_{2j,t} = 0$; and for any $i \neq j$, $dW_{1i,t}dW_{1j,t} = 0$ and $dW_{2i,t}dW_{2j,t} = 0$.

[14] For any $i \neq j$, $d\theta_{i,t}d\theta_{j,t} = 0$, $d\theta_{i,t}(d\xi_{j,t}/\xi_{j,t}) = 0$, and $(d\xi_{i,t}/\xi_{i,t})(d\xi_{j,t}/\xi_{j,t}) = 0$.

[15] Current literature shows that in some fund families, as funds are managed by the same team of managers, their abilities and alphas are correlated such that we can learn the ability of a fund from another fund's performance [e.g., Brown and Wu (2016) and Choi, Kahraman, and Mukherjee (2016)]. In this sense, we can think of a "fund" in our model as a fund family in the real world such that the ability and alpha of a fund family are independent of other fund families'. Under a similar framework, we can analyze the AFMI concentration based on the market shares of fund families. The insights of this model of fund family concentration are similar to those of our model. To simplify our discussion, we call each institution as a "fund" in this paper.

implies a higher instantaneous correlation between manager i's gross alpha and ability.

To facilitate our analysis, we define the following terms:

- $\mathcal{F}_t^{\boldsymbol{\xi}} \triangleq$: the σ-algebras generated by $\{\boldsymbol{\xi}_\mathbf{s}, 0 \leq s \leq t\}$, with $\left\{\mathcal{F}_t^{\boldsymbol{\xi}}\right\}_{0\leq t\leq T}$ as the corresponding filtration over $0 \leq t \leq T$;
- $\mathbf{m_t} \triangleq$: the $n \times 1$ vector of mean of $\boldsymbol{\theta}_\mathbf{t}$ conditional on the observations $\boldsymbol{\xi}_\mathbf{s}$, $0 \leq s \leq t$, i.e., $\mathbf{m_t} \triangleq \mathrm{E}\left(\boldsymbol{\theta}_\mathbf{t}|\mathcal{F}_t^{\boldsymbol{\xi}}\right)$;
- $\boldsymbol{\gamma}_\mathbf{t} \triangleq$: the $n \times n$ covariance matrix of $\boldsymbol{\theta}_\mathbf{t}$ conditional on the observations $\boldsymbol{\xi}_\mathbf{s}$, $0 \leq s \leq t$, i.e., $\boldsymbol{\gamma}_\mathbf{t} \triangleq \mathrm{E}\left[(\boldsymbol{\theta}_\mathbf{t} - \mathbf{m_t})(\boldsymbol{\theta}_\mathbf{t} - \mathbf{m_t})'|\mathcal{F}_t^{\boldsymbol{\xi}}\right]$.

We assume that the conditional distribution of $\boldsymbol{\theta}_\mathbf{0}$, given $\boldsymbol{\xi}_\mathbf{0}$ (the prior distribution), is Gaussian, $N(\mathbf{m_0}, \boldsymbol{\gamma}_\mathbf{0})$, where $\boldsymbol{\gamma}_\mathbf{0}$ is a $n \times n$ diagonal matrix, and elements of $\boldsymbol{\xi}_\mathbf{0}$, $\mathbf{m_0}$, and $\boldsymbol{\gamma}_\mathbf{0}$ have finite values.

Managers and investors update their estimates of $\boldsymbol{\theta}_\mathbf{t}$ using their observations of $\boldsymbol{\xi}_\mathbf{t}$ in a Bayesian fashion. As $\mathbf{m_t}$ is the expected abilities inferred from observable fund returns, hereafter we call $\mathbf{m_t}$ as inferred abilities for short. This type of model is presented in Liptser and Shiryaev (2001a, Ch. 8; 2001b, Ch. 12).[16] The techniques are called optimal linear filtering and are used in numerous previous studies.[17] The following proposition describes how managers and investors form and update their estimates of managers' abilities $\boldsymbol{\theta}_\mathbf{t}$.

Proposition 1.

a. Let $\mathcal{F}_t^{\boldsymbol{\xi}_\mathbf{0},\overline{\mathbf{W}}}$, $0 \leq t \leq T$, be the σ-algebras generated by $\{\boldsymbol{\xi}_\mathbf{0}, \overline{\mathbf{W}}_\mathbf{s}, 0 \leq s \leq t\}$. Then,

$$\overline{\mathbf{W}}_\mathbf{t} = \int_0^t \mathbf{B^{-1}}[\mathbf{I^{-1}}(\boldsymbol{\xi}_\mathbf{t})\mathbf{d}\boldsymbol{\xi}_\mathbf{t} - \mathbf{A}\mathbf{m_s}ds] \tag{3}$$

is a vector of independent Wiener process with respect to the filtration $\left\{\mathcal{F}_t^{\boldsymbol{\xi}}\right\}_{0\leq t\leq T}$, with the ith element as $\overline{W}_{i,t}$ and with $\overline{\mathbf{W}}_\mathbf{0}$ being a zero $n \times 1$ vector. The σ-algebras $\mathcal{F}_t^{\boldsymbol{\xi}}$ and $\mathcal{F}_t^{\boldsymbol{\xi}_\mathbf{0},\overline{\mathbf{W}}}$ are equivalent.

[16] The models presented by Liptser and Shiryaev (2001a,b) allow all model parameters to be stochastic, functions of the stochastic gross alpha. For simplicity, we introduce a linear framework with constant parameters.
[17] See, for example, Dothan and Feldman (1986), Feldman (1989, 2007), Berk and Stanton (2007), Dangl, Wu, and Zechner (2008), Brown and Wu (2013, 2016), and Feldman and Xu (2021).

b. $\bar{\mathbf{W}}_{\mathbf{t}}$ innovates the inferred abilities $\mathbf{m_t}$. The variables $\mathbf{m_t}$, $\boldsymbol{\xi}_{\mathbf{t}}$, and $\boldsymbol{\gamma}_{\mathbf{t}}$ are the unique, continuous, $\mathcal{F}_t^{\boldsymbol{\xi}}$-measurable solutions of the system of equations

$$\mathbf{dm_t} = (\mathbf{a_0} + \mathbf{a_1 m_t})dt + \boldsymbol{\sigma}_{\mathbf{m}}(\boldsymbol{\gamma}_{\mathbf{t}})\mathbf{d}\bar{\mathbf{W}}_{\mathbf{t}}, \quad (4)$$

$$\mathbf{I}^{-1}(\boldsymbol{\xi}_{\mathbf{t}})\mathbf{d}\boldsymbol{\xi}_{\mathbf{t}} = \mathbf{Am_t}dt + \mathbf{B}\mathbf{d}\bar{\mathbf{W}}_{\mathbf{t}}, \quad (5)$$

$$\mathbf{d}\boldsymbol{\gamma}_{\mathbf{t}} = [\mathbf{b_1 b_1} + \mathbf{b_2 b_2} + 2\mathbf{a_1}\boldsymbol{\gamma}_{\mathbf{t}} - \boldsymbol{\sigma}_{\mathbf{m}}(\boldsymbol{\gamma}_{\mathbf{t}})\sigma'_m(\boldsymbol{\gamma}_{\mathbf{t}})]dt, \quad (6)$$

where

$$\boldsymbol{\sigma}_{\mathbf{m}}(\boldsymbol{\gamma}_{\mathbf{t}}) \triangleq (\mathbf{b_2 B} + \mathbf{A}\boldsymbol{\gamma}_{\mathbf{t}})'\mathbf{B}^{-1}, \quad (7)$$

with initial conditions $\boldsymbol{\xi}_{\mathbf{0}}$, $\mathbf{m_0}$, and $\boldsymbol{\gamma}_{\mathbf{0}}$.

c. The random process $(\boldsymbol{\theta}_{\mathbf{t}}, \boldsymbol{\xi}_{\mathbf{t}})$, $0 \le t \le T$ is conditionally Gaussian given $\mathcal{F}_t^{\boldsymbol{\xi}}$.

Proof. Theorem 8.1 of Liptser and Shiryaev (2001a) and Theorem 12.5 of Liptser and Shiryaev (2001b) jointly provide the proof of Proposition 1a. Theorem 12.5 of Liptser and Shiryaev (2001b) provides the proof of Proposition 1b. Theorem 11.1 of Liptser and Shiryaev (2001b) provides the proof of Proposition 1c. □

The technical requirements to prove the theorems are regular conditions over the period $0 \le t \le T$, such as boundedness of parameter values, integrality of variables, and finite moments of variables.[18] The intuition of these requirements is that, over a finite time period, almost surely manager abilities, fund gross alphas, and their variations should be finite so that the learning processes are well defined. These requirements are satisfied, due to our finite parameter values, finite initial values, and the finite horizon within which we study our model. In the real world, abilities that keep improving or deteriorating over a short period, or abilities that revert to a finite mean over a long period, would satisfy the technical requirements and follow our learning processes.

The Wiener process $\bar{\mathbf{W}}_{\mathbf{t}}$ represents the innovation shocks to estimates of manager unobservable abilities. By Proposition 1a, the process $\boldsymbol{\xi}_{\mathbf{t}}$ and the innovation process $\bar{\mathbf{W}}_{\mathbf{t}}$ with $\boldsymbol{\xi}_{\mathbf{0}}$ generate equivalent information.

Proposition 1b implies that investors make their optimal decisions in two steps. First, they observe the history of the funds' share prices $\boldsymbol{\xi}_{\mathbf{t}}$ and restructure the state space to consist

[18] See the requirements of the corresponding theorems in Liptser and Shiryaev (2001a, 2001b).

of only observable processes while maintaining informational equivalence,[19] and generate a posterior distribution of the fund manager abilities, i.e., the first moment of inferred abilities $\mathbf{m_t}$, and so on. Second, they use their posterior estimate, $\mathbf{m_t}$, to predict the fund gross alphas in the forthcoming future, as shown by Equation (5). They use this prediction in solving their problems. Notice that in these optimization processes, the unobservable manager abilities $\boldsymbol{\theta_t}$ is replaced by its observable conditional mean, $\mathbf{m_t}$, updated by a new Wiener process $\bar{\mathbf{W}}_\mathbf{t}$, and that $\mathbf{m_t}$ is continuously updated as a function of the dynamic conditional covariance matrix $\boldsymbol{\gamma_t}$. Hence, investors' problems become Markovian, which makes the problems tractable (allowing a state vector solution).[20]

Taking a closer look at $\mathbf{d}\boldsymbol{\gamma_t}$, we find that as $\boldsymbol{\gamma_0}$ and the parameter matrices in Equations (6) and (7) are diagonal, $\boldsymbol{\gamma_t}$ and $\boldsymbol{\sigma_m}(\boldsymbol{\gamma_t})$ are diagonal. Then, we can define the ith diagonal element of $\boldsymbol{\gamma_t}$ as $\gamma_{i,t}$, $i = 1, \ldots, n$, which is the variance of $\theta_{i,t}$ conditional on the observations of share prices, representing the imprecision of the estimate $m_{i,t}$. We have

$$d\gamma_{i,t} = \left[b_{i,1}^2 + b_{i,2}^2 + 2a_{i,1}\gamma_{i,t} - \sigma_{i,m}^2\left(\gamma_{i,t}\right)\right]dt, \tag{8}$$

where $\sigma_{i,m}\left(\gamma_{i,t}\right)$, $i = 1, \ldots, n$, is the ith diagonal element of $\boldsymbol{\sigma_m}(\boldsymbol{\gamma_t})$ that

$$\sigma_{i,m}\left(\gamma_{i,t}\right) \triangleq \left(b_{i,2}B_i + A_i\gamma_{i,t}\right)/B_i. \tag{9}$$

As $\boldsymbol{\gamma_t}$ and $\boldsymbol{\sigma_m}(\boldsymbol{\gamma_t})$ are diagonal, by Equation (4), $m_{i,t}$ is unaffected by $\bar{W}_{j,t}$ or $\gamma_{j,t}$ for any $i \neq j$. Thus, a manager's inferred ability and its precision are independent of those of other managers, which simplifies our analyses in the following sections.[21]

Depending on parameter values, $d\gamma_{i,t}$ can be positive, negative, or zero; that is, the precision of the future estimates of manager i's ability level can increase, decrease, or be unchanged for the next small time period. In particular, where $b_{i,2}B_i$, the instantaneous

[19] See Feldman (1992).
[20] The elliptical nature our conditionally Gaussian structure allows closure of the filter after two conditional moments. Otherwise, all the conditional higher moments would be part of the filter, and the choice of which higher moments to ignore would be a function of the desired precision.
[21] If the parameter matrices in Equations (6) and (7) and the initial values are not diagonal, then a manager's inferred ability could depend on innovation shocks to other funds, and the precision of the inferred ability could depend on the correlations of this manager's ability and gross alpha with other managers'. Consequently, a fund's equilibrium size, could depend on other fund managers' inferred abilities. This complicates our discussions and does not affect our main insights, so we do not introduce this complexity.

covariance between $d\xi_{i,t}/\xi_{i,t}$ and $\theta_{i,t}$,[22] and/or A_i, the sensitivity of the drift of $d\xi_{i,t}/\xi_{i,t}$ to $\theta_{i,t}$, are sufficiently large (small), $d\gamma_{i,t}$ would be negative (positive). In other words, if manager i's gross alpha and ability are more (less) correlated instantaneously and/or the change in the gross alpha is more (less) sensitive to the ability, then the precision of investors' posterior estimates of manager i's ability increases (decreases). Also, where B_i^2, the instantaneous variance of $d\xi_{i,t}/\xi_{i,t}$,[23] is sufficiently small (large), $d\gamma_{i,t}$ also is negative (positive). In other words, if manager i's gross alpha process is less (more) volatile, then the precision of investors posterior estimates of manager i's ability increases (decreases).

To make economic sense, we assume a nonnegative $b_{i,2}$, $i = 1, \dots, n$, which induces a positive correlation $(b_{i,2}B_i + A_i\gamma_{i,t})$ between inferred ability and performance shocks (because B_i and A_i are positive).[24] Then, the sensitivity of inferred ability to innovation shocks in fund gross alpha, $\sigma_{i,m}(\gamma_{i,t})$ is always positive. In other words, under this setting, for each fund, a positive (negative) shock in fund gross alpha induces an increase (a decrease) in the manager's inferred ability. Then, depending on parameter values, the dynamics of $d\gamma_{i,t}$, induces a $\gamma_{i,t}$ that monotonically increases, decreases, or stays unchanged over time. Consequently, $\sigma_{i,m}(\gamma_{i,t})$, monotonically increases, decreases, or stays unchanged, respectively, over time.

By Proposition 1c, the conditional distribution of $\boldsymbol{\theta}_\mathbf{t}$ is Gaussian. Then, conditional distribution of $\boldsymbol{\theta}_\mathbf{t}$ is determined by the first two moments, $\mathbf{m}_\mathbf{t}$ and $\boldsymbol{\gamma}_\mathbf{t}$. As the parameters $\mathbf{a_1}$, $\mathbf{b_1}$, $\mathbf{b_2}$, $\mathbf{A}$, and $\mathbf{B}$ are constant matrices and $\boldsymbol{\gamma}_\mathbf{0}$ is given, $\boldsymbol{\gamma}_\mathbf{t}$ is deterministic, as shown in Proposition 1b. Consequently, $\boldsymbol{\sigma}_\mathbf{m}(\boldsymbol{\gamma}_\mathbf{t})$, is also deterministic but dynamic. However, $\mathbf{m}_\mathbf{t}$ is stochastic and its future values are unknown. Therefore, investors know the precision of their future estimates of manager abilities in advance, although they do not know the future estimates of these abilities. The fact that the random process $(\boldsymbol{\theta}_\mathbf{t}, \boldsymbol{\xi}_\mathbf{t})$, $0 \le t \le T$ is conditionally

[22] This is because $\mathrm{Cov}(d\theta_{i,t}, d\xi_{i,t}/\xi_{i,t}) = b_{\mathrm{i},2}B_i dt$.
[23] This is because $\mathrm{Cov}(d\xi_{i,t}/\xi_{i,t}, d\xi_{i,t}/\xi_{i,t}) = B_i^2 dt$.
[24] This is because a negative $b_{i,2}$ induces a negative instantaneous/idiosyncratic correlation, which can give rise to negative total correlation. If $\gamma_{i,t}$ weighs the positive systematic source of correlation, A_i, insufficiently high; then the negative instantaneous/idiosyncratic source of correlation $(b_{i,2}B_i)$ dominates. Thus, under these special parameter values, which we do not allow here, the dynamics $\gamma_{i,t}$ may induce correlation between inferred ability and performance shocks, which changes sign over time, resulting in a transient nonmonotonic relation between performance shocks and inferred ability even under the linear structure that we analyze in this section. For detailed analysis of this nonmonotonicity, see Feldman (1989, Proposition 4).

Gaussian, given $\mathcal{F}_t^{\boldsymbol{\xi}}$, facilitates the generation of the posterior estimate of gross alphas in closed form.

2.2 Investors' Optimizations and Fund Managers' Optimizations

Using the above filter to re-represent the state space $\{\boldsymbol{\theta}_t, \boldsymbol{\xi}_t\}$ in terms of observable variables $\{\boldsymbol{\xi}_t, \mathbf{m}_t, \boldsymbol{\gamma}_t\}$, we can solve investors' and fund managers' optimization problems.

We assume that there are infinitely many small risk-neutral investors in the market and that each investor's investment decision does not affect the funds' returns and sizes, although all investors together do affect them. An investor's portfolio return depends on three components: fund gross alphas, management fees, and fund costs. We assume that each fund manager chooses the amount of the fund to actively manage at each time t under fixed management fees f_i, $i = 1, \dots, n$.

Similar to Berk and Green (2004), Feldman, Saxena, and Xu (2020, 2021), and Feldman and Xu (2021) we assume decreasing returns to scale at the fund level. For fund i, $i = 1, \dots, n$, at time t, fund costs variable $C_i\left(q_{i,t}^a\right)$ is a function of the fund amount that is under active management $q_{i,t}^a$,

$$C_i\left(q_{i,t}^a\right) = c_i {q_{i,t}^a}^2. \tag{10}$$

Of $q_{i,t}$, the total asset managed by fund i, i.e., fund i's size, the amount $q_{i,t} - q_{i,t}^a$ ($q_{i,t} - q_{i,t}^a \geq 0$) is invested in the passive benchmark, earning the passive benchmark portfolio return and inducing no fund costs. The amount $q_{i,t}^a$ generates fund gross alphas.

At time t, let the price of fund i's assets under management net of fund costs and fees be $S_{i,t}$, $0 \leq t \leq T$. Then, the active fund's net return is $dS_{i,t}/S_{i,t}$. As we normalize the passive benchmark portfolio's return to zero, the active fund's net return in excess of the passive benchmark is $dS_{i,t}/S_{i,t} - 0 = dS_{i,t}/S_{i,t}$. Hereafter, we call $dS_{i,t}/S_{i,t}$ fund i's instantaneous net alpha, or briefly net alpha. Based on the above discussion, we have,

$$\frac{dS_{i,t}}{S_{i,t}} = \frac{q_{i,t}^a}{q_{i,t}} \frac{d\xi_{i,t}}{\xi_{i,t}} - \frac{C_i\left(q_{i,t}^a\right)}{q_{i,t}} dt - f_i dt. \tag{11}$$

Similar to Berk and Green (2004) and Feldman and Xu (2021), we assume that risk-neutral investors supply capital with infinite elasticity to funds that have positive expected fund net alphas. With sufficient capital, investors' fund allocations drive the conditional expectation of fund net alphas to zero at each time t. Thus, we have the following condition in equilibrium:

$$\mathrm{E}\left[\frac{dS_{i,t}}{S_{i,t}}\middle|\mathcal{F}_t^{\xi}\right] = 0,\ \forall t, i = 1, \dots, n. \tag{12}$$

Taking conditional expectation on Equation (11) and setting it to zero, we have

$$\frac{q_{i,t}^a}{q_{i,t}} A_i m_{i,t} - \frac{c_i {q_{i,t}^a}^2}{q_{i,t}} - f_i = 0. \tag{13}$$

Rearranging,

$$f_i q_{i,t} = A_i m_{i,t} q_{i,t}^a - c_i {q_{i,t}^a}^2. \tag{14}$$

As any fund costs are deducted from investment returns before the returns are transferred to investors [as shown by the fund net alpha Equation (11)], the term $f_i q_{i,t}$ is fund manager i's profit. Fund manager i wants to maximize fund profit $f_i q_{i,t}$ by choosing $q_{i,t}^a$. Then, manager i's problem is

$$\max_{q_{i,t}^a} f_i q_{i,t} = \max_{q_{i,t}^a} A_i m_{i,t} q_{i,t}^a - c_i {q_{i,t}^a}^2 \tag{15}$$

subject to the constraint

$$0 \le q_{i,t}^a \le q_{i,t},\ \forall\ i = 1, \dots, n. \tag{16}$$

As in Berk and Green (2004) and Feldman and Xu (2021), we define the lowest level of inferred ability that makes a fund survive as $\underline{m}_{i,t}$, $i = 1, \dots, n$. If $m_{i,t} < \underline{m}_{i,t}$, fund i receives no investments from investors and exits the market. Hereafter, we briefly call $\underline{m}_{i,t}$, $i = 1, \dots, n$ the survival levels. Here we assume $\underline{m}_{i,t} \ge 0$.[25] The optimal amount under active management and the optimal total assets under management, ${q_{i,t}^a}^*$ and $q_{i,t}^*$, are not trivial where $m_{i,t} \ge \underline{m}_{i,t}$; otherwise, they are both zero.

Solving investors' and managers' problems, we can obtain the equilibrium optimal solutions for funds surviving in the market[26]:

---

[25] The reason is that given updated information, for fund i, the expected instantaneous gross alpha accumulated in dt is $\mathrm{E}(d\xi_{i,t}/\xi_{i,t}|\mathcal{F}_t^{\xi}) = A_i m_{i,t} dt$, with $A_i > 0$. If $m_{i,t} < 0$, the expected instantaneous gross alpha is negative. With positive fund costs and fees, the expected instantaneous net alpha earned by investors in dt would be substantially smaller than zero, so they would switch their investments to the passive benchmark portfolio. Thus, we do not allow $m_{i,t} < 0$ for a surviving fund.

[26] Similar to Berk and Green (2004) and Feldman and Xu (2021), we assume that managers choose f_i such that the constraint $0 \le {q_{i,t}^a}^* \le q_{i,t}^*$, is satisfied for $i = 1, \dots, n$, so this constraint does not affect the optimization. See

$$q_{i,t}^{a\ *} = \frac{A_i m_{i,t}}{2c_i} \tag{17}$$

$$q_{i,t}^{*} = \frac{\left(A_i m_{i,t}\right)^2}{4c_i f_i}. \tag{18}$$

To simplify the notations, we define fund i's size factor as X_i, such that

$$X_i \triangleq \frac{1}{4c_i f_i}. \tag{19}$$

The higher the decreasing returns to scale coefficient c_i and the higher the management fee f_i are, the lower is fund i's size factor and, then, the lower is the equilibrium fund size $q_{i,t}^{*}$. Then,

$$q_{i,t}^{*} = X_i\left(A_i m_{i,t}\right)^2. \tag{20}$$

Proof. See the Appendix. □

2.3 Equilibrium Market Power and Market Structure

We demonstrate now that AFMI dynamic concentration is the key measure to study the AFMI's industrial organization, while other common measures are less informative.

As investors receive net alphas from funds, any fund costs are transferred to investors as reductions in fund net alphas so that fund managers bear no costs in operation. Then, in equilibrium, for $i = 1, \dots, n$, manager i's profit is the revenue $f_i q_{i,t}^{*}$, and the profit rate on each dollar under management is f_i, a constant. A manager's profit margin, i.e., the difference between revenue and costs, divided by the revenue, is always one $\left[\left(f_i q_{i,t}^{*} - 0\right)/f_i q_{i,t}^{*} = 1\right]$. These results imply that there are no dynamics in a manager's profit rate or profit margin. Also, if we calculate a manager's profit markup, i.e., revenue divided by costs, we find that the profit markup $\left[= f_i q_{i,t}^{*}/0\right]$ is positive infinity. This does not imply that a manager has infinite profitability. Notice again that it is the investors who determine the quantity of production (fund sizes), and investors choose the quantity to capture any positive expected net alpha. As a manager's profit rate is fixed at its constant management fee, he or she needs to attract investments by maximizing the expected fund net alpha as much as possible; as the manager's ability to create the fund net alpha is limited, the equilibrium profit is limited.

---

the proof of the solutions of the optimization problems in the Appendix.

A fund's market power can be measured by its Lerner Index, which is the difference between fee and marginal cost, divided by fee. From the above discussion, we can see that a fund's Lerner Index is always one $[= (f_i - 0)/f_i]$, so in equilibrium, a fund's market power is not dynamic either. In contrast, the market structure of AFMI is dynamic, as funds' relative sizes change over time. Thus, to understand the dynamics of AFMI industrial organization, we focus on the dynamics of its market structure, in particular, the dynamics of AFMI concentration.

2.4 Equilibrium AFMI Concentration

We use the Herfindahl-Hirschman Index (HHI) to measure AFMI concentration for the reasons discussed in our Introduction section. Let $\mathbf{q_t^*}$ be the $n \times 1$ vector of the equilibrium fund sizes, with the ith element as $q_{i,t}^*$. We have, using Equations (20) and (19),

$$\mathbf{q_t^*} = \mathbf{A^2 I^2 (m_t) X}, \tag{21}$$

where $\mathbf{I(m_t)}$ is a $n \times n$ diagonal matrix with the ith element as the ith element of $\mathbf{m_t}$, and $\mathbf{X}$ is a $n \times 1$ vector with the ith element as X_i. Then, the $n \times 1$ vector of the equilibrium fund market shares, $\mathbf{w_t^*}$, is

$$\mathbf{w_t^*} = \frac{\mathbf{q_t^*}}{\mathbf{q_t^{*\prime} 1}}, \tag{22}$$

where $\mathbf{1}$ is an $n \times 1$ vector of ones. By definition, the equilibrium AFMI HHI (henceforth we briefly call it HHI) is

$$HHI_t^* \triangleq \mathbf{w_t^{*\prime} w_t^*} = \frac{\mathbf{q_t^{*\prime} q_t^*}}{(\mathbf{q_t^{*\prime} 1})^2}. \tag{23}$$

Substituting Equation (20) into Equation (23), we have

$$HHI_t^* = \frac{\mathbf{X' A^4 I^4 (m_t) X}}{[\mathbf{X' A^2 I^2 (m_t) 1}]^2} = \frac{\sum_{i=1}^n X_i^2 \left(A_i m_{i,t}\right)^4}{\left[\sum_{i=1}^n X_i \left(A_i m_{i,t}\right)^2\right]^2}. \tag{24}$$

From (24), we can see that HHI's dynamics is determined by inferred abilities' dynamics, i.e., $HHI_t^* = HHI_t^*(\mathbf{m_t})$. Then, by Itô's Lemma, we have

$$dHHI_t^* = \frac{\partial HHI_t^*}{\partial \mathbf{m_t}'} \mathbf{dm_t} + \frac{1}{2} \mathbf{dm_t'} \frac{\partial^2 HHI_t^*}{\partial \mathbf{m_t}' \partial \mathbf{m_t}} \mathbf{dm_t}. \tag{25}$$

Substituting Equation (4) into (25) and due to the independence of $\overline{W}_{i,t}$, $i = 1, \dots, n$, we have

$$dHHI_t^* = \frac{\partial HHI_t^*}{\partial \mathbf{m_t}'}(\mathbf{a_0} + \mathbf{a_1}\mathbf{m_t})dt + \frac{\partial HHI_t^*}{\partial \mathbf{m_t}'}\boldsymbol{\sigma}_\mathbf{m}(\boldsymbol{\gamma}_\mathbf{t})\mathbf{d\overline{W}_t} + \frac{1}{2}\mathbf{trace}\left[\boldsymbol{\sigma}_m'(\boldsymbol{\gamma}_\mathbf{t})\frac{\partial^2 HHI_t^*}{\partial \mathbf{m_t}'\partial \mathbf{m_t}}\boldsymbol{\sigma}_\mathbf{m}(\boldsymbol{\gamma}_\mathbf{t})\right]dt. \tag{26}$$

To facilitate our discussion, we rewrite Equation (26) in scalar form:

$$dHHI_t^* = \sum_{i=1}^{n}\left[\frac{\partial HHI_t^*}{\partial m_{i,t}}(a_{i,0} + a_{i,1}m_{i,t})dt + \frac{\partial HHI_t^*}{\partial m_{i,t}}\sigma_{i,m}(\gamma_{i,t})d\overline{W}_{i,t} + \frac{1}{2}\frac{\partial^2 HHI_t^*}{\partial {m_{i,t}}^2}\sigma_{i,m}^2 dt\right], \tag{27}$$

where

$$\frac{\partial HHI_t^*}{\partial m_{i,t}} = 4X_iA_i^2 m_{i,t} \times \frac{q_i\sum_{j=1}^n q_j - \sum_{j=1}^n q_j^2}{\left(\sum_{j=1}^n q_j\right)^3} \tag{28}$$

and

$$\frac{\partial^2 HHI_t^*}{\partial {m_{i,t}}^2} = 4X_iA_i^2 \times \left[\frac{3q_i\sum_{j=1}^n q_j + \frac{6q_i\left(\sum_{j=1}^n q_j^2\right)}{\left(\sum_{j=1}^n q_j\right)} - 8q_i^2 - \sum_{j=1}^n q_j^2}{\left(\sum_{j=1}^n q_j\right)^3}\right]. \tag{29}$$

From Equation (28), we can see that if fund i's inferred ability is sufficiently large (small) relative to other funds' such that its size is sufficiently large (small) relative to other funds' sizes, then $\frac{\partial HHI_t^*}{\partial m_{i,t}}$ is positive (negative).[27] Then, as shown in Equation (27), holding other managers' inferred abilities unchanged, an increase in manager i's inferred ability, due to a sufficiently large drift term in inferred ability, $a_{i,0} + a_{i,1}m_{i,t}$ or a sufficiently large innovation shock in performance, $d\overline{W}_{i,t}$, has a positive (negative) impact on the change in the HHI, $dHHI_t^*$. Also, higher $m_{i,t}$, X_i, and A_i all make this positive (negative) impact stronger.

The intuition is that, if manager i's inferred ability is sufficiently large relative to other managers' inferred abilities, then fund i's size is sufficiently large relative to other funds' sizes, dominating in the market. A higher inferred ability attracts more investment to fund i, making

[27] If q_i is sufficiently (small) large relative to q_j, $\forall j \neq i$, then $q_i\sum_{j=1}^n q_j = \sum_{j=1}^n q_iq_j \geq \sum_{j=1}^n q_j^2$ ($q_i\sum_{j=1}^n q_j = \sum_{j=1}^n q_iq_j \leq \sum_{j=1}^n q_j^2$). These two inequalities are due to the nonnegativity of fund sizes.

it larger and the AFMI more concentrated at fund i. Further, if manager i's inferred ability, fund size factor, and sensitivity of gross alpha to ability, are larger, then fund i's size becomes even larger, making the AFMI more concentrated. On the other hand, if manager i's inferred ability is sufficiently small relative to other managers' inferred abilities, then fund i's size is sufficiently small relative to other funds' sizes. A higher inferred ability attracts more investment to fund i, making its size closer to other funds' and then making the AFMI less concentrated. In this case, if manager i's inferred ability, fund size factor, and sensitivity of gross alpha to ability, are larger, then fund i's size becomes larger and closer to other funds', making the AFMI less concentrated.

Regarding the second-order effect shown in Equation (29), if manager i's inferred ability is sufficiently large (small) relative to other managers' such that fund i's size is sufficiently large (small) relative to other funds' sizes, then $\frac{\partial^2 HHI_t^*}{\partial {m_{i,t}}^2}$ is negative and HHI_t^* is concave in $m_{i,t}$.[28] Then, over the next infinitesimal period, this concavity has a negative impact on $dHHI_t^*$. If all managers' inferred abilities are sufficiently close to each other's such that funds' sizes are sufficiently close, making HHI_t^* close to its minimum value $1/n$, then $\frac{\partial^2 HHI_t^*}{\partial {m_{i,t}}^2}$ is positive and HHI_t^* is convex in $m_{i,t}$.[29] Then, over the next infinitesimal period, this convexity has a positive impact on $dHHI_t^*$. Also, if HHI_t^* is concave (convex) in $m_{i,t}$, then a higher X_i and a higher A_i both make this concavity (convexity) stronger.

The intuition is that, if fund i's market share is sufficiently large (small) due to manager i's sufficiently large (small) inferred ability, then the AFMI is concentrated at fund i (at other funds). Although a higher (lower) inferred ability of manager i can make the AFMI more concentrated at fund i (at other funds), it becomes more and more difficult to increase the concentration in this way. On the other hand, if all managers' inferred abilities are close to each

---

[28] If q_i is sufficiently small relative to q_j, $\forall j \neq i$, then the term $-\sum_{j=1}^n q_j^2$ dominates in the expression $3q_i \sum_{j=1}^n q_j + \frac{6q_i\left(\sum_{j=1}^n q_j^2\right)}{\left(\sum_{j=1}^n q_j\right)} - 8q_i^2 - \sum_{j=1}^n q_j^2$, making it negative. If q_i is sufficiently large relative to q_j, $\forall j \neq i$, then $3q_i \sum_{j=1}^n q_j + \frac{6q_i\left(\sum_{j=1}^n q_j^2\right)}{\left(\sum_{j=1}^n q_j\right)} < 9q_i^2$ and $-8q_i^2 - \sum_{j=1}^n q_j^2 < -9q_i^2$, making $3q_i \sum_{j=1}^n q_j + \frac{6q_i\left(\sum_{j=1}^n q_j^2\right)}{\left(\sum_{j=1}^n q_j\right)} - 8q_i^2 - \sum_{j=1}^n q_j^2 < 9q_i^2 - 9q_i^2 = 0$.

[29] If all funds' sizes are sufficiently close, then the expression $3q_i \sum_{j=1}^n q_j + \frac{6q_i\left(\sum_{j=1}^n q_j^2\right)}{\left(\sum_{j=1}^n q_j\right)} - 8q_i^2 - \sum_{j=1}^n q_j^2 \approx (2n-2)q_i^2 > 0$ as $n \geq 2$.

other's such that funds' sizes are close, then a larger and a smaller inferred ability of manager i can both make fund i's size deviate from other funds' sizes, making the AFMI more concentrated. It is easier to make fund i's size deviate from other funds' sizes and to increase the HHI if the absolute change in manager i's inferred ability is larger in this case. Further, a higher fund size factor and a higher sensitivity of gross alpha to ability both make fund i's size larger and, thus, more relevant in AFMI. Then, holding the relative abilities thus the relative sizes unchanged, if the HHI is concave (convex) in manager i's inferred ability, the larger size of fund i intensifies the concavity (convexity) effect.

The following proposition summarizes the above results.

Proposition RN. If a manager's inferred ability is higher than the survival level, then we have the following results of this manager's inferred ability and the HHI.

a. If this manager's inferred ability is sufficiently large (small) relative to other managers' such that the fund's size is sufficiently large (small) relative to other funds', then an increase in this manager's inferred ability, due to a sufficiently large drift term in inferred ability or a sufficiently large innovation shock in performance, has a positive (negative) impact on the change of the HHI. This positive (negative) impact is stronger if this manager's inferred ability, fund size factor, and sensitivity of gross alpha to ability are larger.

b. If this fund's size is sufficiently large (small) relative to other funds' due to the manager's sufficiently large (small) inferred ability, then the HHI is concave in this manager's inferred ability. Over the next infinitesimal period, this concavity has a negative impact on the change in the HHI. If all managers' inferred abilities are sufficiently close to each other's such that funds' sizes are sufficiently close, then the HHI is convex in this manager's inferred ability. Then, over the next infinitesimal period, this convexity has a positive impact on the change in the HHI. If the HHI is concave (convex) in this manager's inferred ability, then both a higher fund size factor and a higher sensitivity of gross alpha to ability make the concavity (convexity) effect stronger. □

2.5 Constant Abilities and the HHI

Consider a special case in which managers' abilities are unobservable constants, such that $\mathbf{a_0}$, $\mathbf{a_1}$, $\mathbf{b_1}$, and $\mathbf{b_2}$ are $n \times n$ zero matrices. We have

$$\mathbf{dm_t} = \boldsymbol{\sigma}_\mathbf{m}(\boldsymbol{\gamma}_\mathbf{t})\mathbf{d\overline{W}_t}, \tag{30}$$

$$\boldsymbol{\sigma}_\mathbf{m}(\boldsymbol{\gamma}_\mathbf{t}) = (\mathbf{A}\boldsymbol{\gamma}_\mathbf{t})'\mathbf{B^{-1}}, \tag{31}$$

$$\boldsymbol{\gamma}_\mathbf{t} = [\mathbf{I} + \boldsymbol{\gamma}_\mathbf{0}\mathbf{A}\mathbf{B}^{-2}\mathbf{A}t]^{-1}\boldsymbol{\gamma}_\mathbf{0}, \tag{32}$$

where $\mathbf{I}$ is an $n \times n$ identity matrix. Theorem 12.8 of Liptser and Shiryaev (2001b) provides the proof of the above results. These results show that for fund i, $i = 1, \dots, n$, $\gamma_{i,t} = \frac{\gamma_{i,0}B_i^2}{B_i^2 + A_i^2\gamma_{i,0}t}$ decreases over time monotonically, so the sensitivity of inferred ability to innovation shocks, $\sigma_{i,m}(\gamma_{i,t}) \triangleq (A_i\gamma_{i,t})/B_i$, also decreases over time monotonically. As $t \to \infty$, $\gamma_{i,t}$ converges to zero. Consequently, $\sigma_{i,m}(\gamma_{i,t})$ converges to zero, thus $dm_{i,t} = \sigma_{i,m}(\gamma_{i,t})d\overline{W}_{i,t}$ becomes zero. Then, by Equation (25), $dHHI_t^*$ is zero. This is the static state of this constant-ability framework.

The intuition is that as managers' abilities are unobservable constants, over time the estimation precisions improve monotonically such that the inferred abilities are less and less sensitive to the future observations of fund gross alphas. As time goes to infinity, eventually, people know the managers' abilities and do not change their estimates. In this case, investors will not change their investments flows to funds anymore (i.e., fund sizes stay unchanged), making the HHI stay unchanged.

The following proposition summarizes the above results.

Proposition CA. If managers' unobservable abilities are constant and their inferred abilities are higher than the survival levels, then, over time, the estimation precision of inferred abilities monotonically improves and the sensitivity of inferred abilities to innovation shocks decreases monotonically. As time goes to infinity, AFFMI reaches a static state where the managers' inferred abilities do not change, inducing equilibrium fund sizes to stay unchanged and the HHI unchanged. □

2.6 Mean-Variance Risk-Averse Investors and the HHI

To study the effect of investors' risk aversion on the HHI, we assume that investors are mean-variance risk averse who maximize their portfolios' instantaneous Sharpe ratios. These investors' optimal portfolios are growth optimal and are the same as those of investors with Bernoulli logarithmic preferences, who maximize expected utility.[30] This setting is also similar to the one in Pastor and Stambaugh (2012), Feldman, Saxena, and Xu (2020, 2021), and Feldman and Xu (2021).

As risk-averse investors trade off risk and return, we need to redefine our model. First, we cannot normalize the passive benchmark portfolio return to be zero, as the level of this return is relevant.[31] Here, we define the share price of the passive benchmark portfolio at time t, $0 \leq t \leq T$, as η_t. This, in turn, includes redefinitions of net and gross alphas. We assume that the passive benchmark portfolio return $d\eta_t/\eta_t$ follows

$$\frac{d\eta_t}{\eta_t} = \mu_p dt + \sigma_p dW_{p,t}, \tag{33}$$

where μ_p and σ_p are positive known constants and $W_{p,t}$ is a Wiener Process.

Second, for $i = 1, \ldots, n$, we still define $d\xi_{i,t}/\xi_{i,t}$, as the fund gross alphas, which follow the process defined in Equations (1) and (2), and define $dS_{i,t}/S_{i,t}$ as the fund net alphas. As the active funds have beta loading of one on the passive benchmark portfolio, the fund gross return is $d\xi_{i,t}/\xi_{i,t} + d\eta_t/\eta_t$ and the fund net return is $dS_{i,t}/S_{i,t} + d\eta_t/\eta_t$. Also, we assume that the risk source of the benchmark return, $W_{p,t}$, is independent of that of gross alphas, so

$$dW_{p,t} d\overline{W}_{i,t} = 0, \ \forall t, i = 1, \ldots, n. \tag{34}$$

Third, to simplify our discussion, we normalize the risk-free rate to zero.[32] All other settings are the same as before.

An investor invests in n active funds and the passive benchmark to maximize the portfolio's instantaneous Sharpe ratio:

---

[30] See the discussions of mean-variance risk-averse investors in Feldman and Xu (2021).

[31] As risk-averse investors' preferences are defined over their whole portfolios, they do not form their decision based on a marginal analysis of the active funds' risk alone. [See, for example, Equation (43), below, which collapses if the passive benchmark return is normalized to zero.]

[32] Alternatively, we can regard $d\eta_t/\eta_t$ as the passive benchmark portfolio return in excess of the risk-free rate.

$$\max_{w_t} \frac{\mathrm{E}\left[\frac{dp_t}{p_t}\middle|\mathcal{F}_t^{\xi}\right]}{\sqrt{\mathrm{Var}\left[\frac{dp_t}{p_t}\middle|\mathcal{F}_t^{\xi}\right]}} \tag{35}$$

subject to

$$\mathbf{v_t}'\mathbf{1} = 1 \tag{36}$$

$$0 \leq v_{i,t} \leq 1,\ \forall\, i = 1, \dots, n+1, \tag{37}$$

where $\mathbf{v_t}$ is the $(n+1) \times 1$ portfolio weight vector, with the ith element $v_{i,t}$ as the weight allocated to the ith fund $i = 1, \dots, n$, and the last element $v_{n+1,t}$ as the weight allocated to the passive benchmark portfolio. Condition (37) is to prevent short selling of active funds or the passive benchmark portfolio. Also, p_t is the portfolio's value, and dp_t/p_t is the investor's instantaneous portfolio return. We define $\mathbf{R_t}$ as the $(n+1) \times 1$ return vector of these $n+1$ assets, with the ith element $i = 1, \dots, n$

$$\begin{aligned} R_{i,t} &= \frac{dS_{i,t}}{S_{i,t}} + \frac{d\eta_t}{\eta_t} \\ &= \left(\frac{q_{i,t}^a}{q_{i,t}} A_i m_{i,t} - \frac{c_i {q_{i,t}^a}^2}{q_{i,t}} - f_i + \mu_p\right) dt + \frac{q_{i,t}^a}{q_{i,t}} B_i d\overline{W}_{i,t} + \sigma_p dW_{p,t} \end{aligned} \tag{38}$$

and

$$R_{n+1,t} = \frac{d\eta_t}{\eta_t}. \tag{39}$$

Then, the investor's portfolio return is

$$\frac{dp_t}{p_t} = \mathbf{v_t}'\mathbf{R_t}. \tag{40}$$

Solving the investor's problem, we have the optimal weight allocations $\mathbf{v_t^*}$. As investors face the same risk-return tradeoff and have the same objective function, they all make the same optimal decision of $\mathbf{v_t^*}$. We define the part of the total wealth of all investors that is allocated to financial assets (i.e., allocated to the active fund and the passive benchmark portfolio) as V, $V \in (0, +\infty)$, $0 \leq t \leq T$. To simplify our analyses and focus on how managers' heterogeneity affects the dynamics of the HHI, we assume that V is constant and exogenous to both investors and managers.[33] Then, the amount of wealth allocated to fund i,

[33] In reality, this wealth not only depends on the returns from financial assets, but also depends on production activities, research and development expenditures, consumptions, taxes, and many other aspects of the economy that we do not model here. Also, it can change over time and its dynamics can affect the dynamics of the HHI. To simplify our model, we do not introduce these complexities of this wealth.

i.e., fund i's size, is $q_{i,t}^* = v_{i,t}^* V,\ i = 1, \ldots, n$.

As in the risk-neutral case, we can write the fund manager's profit as a function of $q_{i,t}^a$, i.e., $g_i(q_{i,t}^a)$, where g_i is some (smooth, increasing, concave) function. Then, manager i's problem is

$$\max_{q_{i,t}^a} f_i q_{i,t} = \max_{q_{i,t}^a} g_i(q_{i,t}^a) \tag{41}$$

subject to

$$0 \le q_{i,t}^a \le q_{i,t},\ \forall\, i = 1, \ldots, n. \tag{42}$$

By solving the investors' and managers' problems,[34] we obtain the equilibrium fund size:

$$q_{i,t}^* = \frac{\left(A_i m_{i,t}\right)^2 V \sigma_p^2}{4 f_i \left(B_i^2 \mu_p + c_i V \sigma_p^2\right)}. \tag{43}$$

We define the size factor of fund i where investors are mean-variance risk-averse, as

$$X_i^{RA} \triangleq \frac{V \sigma_p^2}{4 f_i \left(B_i^2 \mu_p + c_i V \sigma_p^2\right)} = \frac{1}{4 f_i c_i + \frac{4 f_i B_i^2 \mu_p}{V \sigma_p^2}}. \tag{44}$$

Similar to the results of X_i, a higher decreasing returns to scale coefficient c_i and a higher management fee f_i, both decrease the size factor X_i^{RA}. Additionally, higher B_i^2 and μ_p both decrease X_i^{RA}, and higher V and σ_p^2 both increase X_i^{RA}. The intuition is that, holding other parameters unchanged, mean-variance risk-averse investors invest more (less) in active fund i and less (more) in the passive benchmark, if the risk of the passive benchmark's return σ_p^2 (the risk of fund i's gross alpha B_i^2) is higher. Also, investors invest more in active funds if they have more wealth V to invest, and invest more in the passive benchmark and less in active funds, if the passive benchmark's mean return μ_p is higher. Further, we can see that, holding other parameters unchanged, X_i^{RA} is smaller than X_i; i.e., comparing to AFMI with risk-neutral investors, AFMI with mean-variance risk-averse investors has smaller equilibrium fund sizes. That is because investors' risk considerations reduce their investment to risky active funds. Using this new definition of fund i's size factor, we have

---

[34] We assume that managers choose f_i such that the constraint $0 \le {q_{i,t}^a}^* \le q_{i,t}^*$, is satisfied for $i = 1, \ldots, n$, so this constraint does not affect the managers' optimization processes. Also, we assume that μ_p is sufficiently large or σ_p^2 is sufficiently small so that $0 \le v_{i,t}^* \le 1$, is satisfied for $i = 1, \ldots, n + 1$, so this constraint does not affect the investors' optimization processes. See the proof in the appendix.

$$q_{i,t}^{*}=\left(A_i m_{i,t}\right)^2 X_i^{RA}. \tag{45}$$

Then, substituting $q_{i,t}^{*}$ into the formula of HHI_t^{*}, we have

$$HHI_t^{*}=\frac{\mathbf{X}^{\mathbf{RA}\prime}\mathbf{A}^4\mathbf{I}^4(\mathbf{m_t})\mathbf{X}^{\mathbf{RA}}}{[\mathbf{X}^{\mathbf{RA}\prime}\mathbf{A}^2\mathbf{I}^2(\mathbf{m_t})\mathbf{1}]^2}=\frac{\sum_{i=1}^{n} {X_i^{RA}}^2\left(A_i m_{i,t}\right)^4}{\left[\sum_{i=1}^{n} X_i^{RA}\left(A_i m_{i,t}\right)^2\right]^2} \tag{46}$$

where $\mathbf{X}^{\mathbf{RA}}$ is a $n\times 1$ vector with the ith element as X_i^{RA}. We can see that the form of HHI_t^{*} in (46) is the same as the one in (24) of the risk-neutral case. The only difference is that here we use $\mathbf{X}^{\mathbf{RA}}$ instead of $\mathbf{X}$ as the size factors. Thus, the relation of the dynamics of HHI_t^{*} and managers' inferred abilities in Proposition RN still holds. The intuition is that investors' risk considerations decrease the equilibrium fund sizes, but HHI_t^{*} depends on relative fund sizes, and the way to compare funds' sizes does not depend on investors' risk considerations. Thus, the dynamics of HHI_t^{*} are related to managers' relative inferred abilities, similar to the risk-neutral case.

The following proposition summarizes the results in this section.

Proposition RA. The results of Proposition RN hold if investors are mean-variance risk-averse, except that, all else equal, equilibrium fund sizes are smaller, and funds' size factors depend on more parameters. In particular, the size factors not only decrease with the decreasing returns to scale coefficients and management fees, but also increase with the total wealth and the passive benchmark portfolio's risks and decrease with variances of active funds' gross alphas and the passive benchmark portfolio's mean return.

Proof. See the Appendix. □

2.7 Funds' Entrances and Exits and the HHI

Besides the dynamics of fund managers' relative abilities, a fund's entrance and exit could affect the dynamics of AFMI concentration. Although we do not analyze funds' entrances and exits explicitly, we show in this section that our framework is compatible with the effects of them, if we allow the total number of funds to change over time, i.e., $n=n_t$, and require funds to exit the market if their managers' inferred abilities reduce to zero, i.e., the survival ability level $\underline{m}_{i,t}=0,\ i=1,\dots,n_t$. Notice that in equilibrium, funds with positive (zero) inferred abilities earn positive (zero) profits, as implied by the equilibrium fund sizes in

Equation (18) in the risk-neutral case and those in Equation (43) in the mean-variance risk-averse case. When $\underline{m}_{i,t} = 0,\ i = 1, \dots, n_t$, managers with positive inferred abilities optimally stay in the market to earn positive profits. On the other hand, as managers cannot short sell investors' wealth, managers with negative inferred abilities optimally choose to put zero assets under active management to avoid losses, thus exit the market. Therefore, the setting of $\underline{m}_{i,t} = 0,\ i = 1, \dots, n_t$ is consistent with profit-maximizing managers, and these survival ability levels can be regarded as those endogenously chosen by fund managers.

To see how our framework is compatible with the effects of funds' entrances and exits, notice that equilibrium fund sizes $q_{i,t}^*$, are functions of managers' inferred abilities $m_{i,t}$, [Equations (18) in the risk-neutral case and (43) in the mean-variance risk-averse case]. As the value of $m_{i,t}$ changes continuously, the value of.$q_{i,t}^*$ also changes continuously. When $m_{i,t}$ decreases to zero, $q_{i,t}^*$ (and fund i's market share) decreases to zero, such that when the fund exits the market, the exit does not cause a jump in HHI_t^*. On the other hand, a potential entrant can be regarded as a fund with negative inferred ability. When its inferred ability $m_{i,t}$ increases to zero, it enters the market with an equilibrium fund size $q_{i,t}^*$ equal to zero. After that, if $m_{i,t}$ increases, then $q_{i,t}^*$ increases. As the changes in $m_{i,t}$ and $q_{i,t}^*$ are continuous, the entrance does not cause a jump in HHI_t^* either. Then, in these two cases, $dHHI_t^*$ can still be expressed by Equation (25), and the results from Section 2.3 to Section 2.5 are still valid. In other words, funds' entrances and exits do not affect $dHHI_t^*$ immediately, but change the set of funds in AFMI and affect $dHHI_t^*$ after that.

However, if $\underline{m}_{i,t} > 0,\ i = 1, \dots, n_t$, a fund's exit or entrance creates a jump in HHI_t^*, and we need to incorporate this jump effect when analyzing $dHHI_t^*$. The reason is that when fund i exits the market with $m_{i,t}$ decreasing to $\underline{m}_{i,t}$, its equilibrium fund size $q_{i,t}^*$ jumps from a value larger than (but not close to) zero, to zero, creating a jump in HHI_t^*. On the other hand, when fund i enters the market with $m_{i,t}$ increasing to $\underline{m}_{i,t}$, its equilibrium fund size $q_{i,t}^*$ jumps from zero to a value larger than (but not close to) zero, creating a jump in HHI_t^*, too. In these two cases, $dHHI_t^*$ cannot be expressed by Equation (25) because the jump effects should be added.

In reality, we observe that investors keep withdrawing investments from badly performing funds, so when a fund with a history of bad performance eventually exits AFMI, its size is negligible compared to AFMI's size. Also, when a new fund enters market, it starts with a size that is trivial compared to AFMI's size, and if it performs well later, it grows. When these exits and entrances happen in the real world, we do not observe jumps in AFMI concentration levels. Therefore, our model can sufficiently explain the dynamics of AFMI concentration when funds exit and enter.

3 Simulation of AFMI Concentration

In our following numerical analyses, we consider a two-fund AFMI, i.e., $n = 2$, and assume that investors are risk neutral. The numerical analyses with mean-variance risk-averse investors are similar, and we omit them for brevity.

We first illustrate how the HHI changes with different values of relative inferred manager abilities, fund size factors, and sensitivity of gross alphas to abilities. We set $m_{2,t} = 1$, $A_2 = 1$, and $X_2 = 100$. We set the range of $m_{1,t}$ as $[0, 4]$. As $m_{2,t} = 1$, the value of $m_{1,t}$ can be regarded as manager 1's inferred ability relative to manager 2's. We simulate the values of the HHI for three cases,

- Case One: $A_1 = A_2 = 1$ and $X_1 = X_2 = 100$;
- Case Two: $A_1 = A_2 = 1$ and $X_1 = 2X_2 = 200$;
- Case Three: $A_1 = 2A_2 = 2$ and $X_1 = X_2 = 100$.

Figure 1 illustrates the results. In Case One, the two funds have the same size factor and sensitivity of gross alpha to ability. Where $m_{1,t}$ is smaller (larger) than one, fund 1's equilibrium size is smaller (larger) than fund 2's, and the AFMI is concentrated at fund 2 (fund 1). Then, a higher $m_{1,t}$ increases fund 1's size and makes the AFMI less (more) concentrated. The lowest level of HHI_t^* is 0.5, achieved where $m_{1,t} = 1$, i.e., the two managers have the same inferred ability thus the same equilibrium size. The highest HHI_t^* is 1, achieved where $m_{1,t} = 0$ and $m_{1,t} \to \infty$, i.e., either manager 2 or manager 1 has infinite relative ability such that AFMI becomes monopolistic. Moreover, in the figure, we can see that where $m_{1,t}$ is close to zero (close to four), HHI_t^* is concave in $m_{1,t}$, as it is more difficult to increase HHI_t^* by further decreasing (increasing) $m_{1,t}$. Also, where $m_{1,t}$ is close to one, HHI_t^* is convex in $m_{1,t}$, as it is easier to increase HHI_t^* if $m_{1,t}$ has a larger deviation from one and makes fund 1's size deviate farther from fund 2's.

In Case Two, fund 1 has a larger size factor but the same sensitivity of gross alpha to ability. Comparing Case Two with Case One, we can see that the graph of Case Two shrinks to the left. In particular, where HHI_t^* decreases (increases) with $m_{1,t}$, at the same $m_{1,t}$ level, HHI_t^* has a lower (higher) value because the larger size factor enhances the negative (positive) impact of a higher $m_{1,t}$ on HHI_t^*. Also, in Case Two, where HHI_t^* is concave (convex) in $m_{1,t}$, HHI_t^* is more sensitive with $m_{1,t}$ because the larger size factor also intensifies the concavity (convexity).

In Case Three, fund 1 has a larger sensitivity of gross alpha to ability but the same size factor. Because a higher sensitivity of gross alpha to ability has a stronger effect on equilibrium fund size than the size factor [by Equation (20), A_i has a power of two whereas X_i has a power of one], the graph of Case Three shrinks more to the left and has larger concavity and convexity in the corresponding intervals, compared with Case Two.

Next, we simulate these two funds' inferred abilities, $m_{1,t}$ and $m_{2,t}$, and then HHI_t^*. We discretize our continuous-time processes into discrete-time processes, setting $dt = \Delta t$ to be one month and $d\overline{W}_{1,t} = \Delta\overline{W}_{1,t}$ and $d\overline{W}_{2,t} = \Delta\overline{W}_{2,t}$, to follow a normal distribution of mean zero and variance Δt. We set some of the two funds' parameter values the same and set them similar to those of Feldman and Xu (2021): for $i = 1, 2$, $f_i = 0.094\%$, $B_i = 5.24\%$, $m_{i,0} = 1.12\%$, and $\gamma_{i,0} = 0.0006$. Additionally, we set $c_i = 0.0002$ and $A_i = 1$, $i = 1, 2$. We conduct the simulation for two frameworks, one with dynamic abilities and the other with constant abilities. In particular, the parameters specific to these two frameworks are set as follows.

- Dynamic Abilities: for $i = 1, 2$, $a_{0,i} = 0.01$, $a_{1,i} = -0.02$, $b_{1,i} = 0.02$, and $b_{2,i} = 0.01$.
- Constant Abilities: for $i = 1, 2$, $a_{0,i} = 0$, $a_{1,i} = 0$, $b_{1,i} = 0$, and $b_{2,i} = 0$.

We simulate $\Delta\overline{W}_{1,t}$ and $\Delta\overline{W}_{2,t}$ as two independent series of increments of Brownian motions and use the same set of simulated $\Delta\overline{W}_{1,t}$ and $\Delta\overline{W}_{2,t}$ values for both cases.

We simulate the results for 400 months. Figure 2 plots the simulation results. In both frameworks, we can see that, when $m_{1,t}$ is farther away from (closer to) $m_{2,t}$, HHI_t^* becomes larger (smaller). Also, with constant abilities, the two managers' inferred abilities change little after 300 months. This is because the estimation precisions are very large after 300 months, so the inferred abilities are not sensitive to innovation shocks. Consequently, equilibrium fund sizes change little after 300 months, making HHI_t^* stable after 300 months in the interval from 0.90 to 0.92. On the other hand, with dynamic abilities, the two managers' inferred abilities fluctuate greatly over time, even after 300 months, because the estimation precisions are low, so the inferred abilities are still sensitive to innovation shocks. Consequently, equilibrium fund sizes fluctuate greatly after 300 months, making HHI_t^* volatile after 300 months in the interval from 0.50 to 0.75.

4 AFMI Concentration and Fund Net Alpha Production

If we allow AFMI concentration to affect fund net alpha production as Feldman, Saxena, and Xu (2020) (hereafter, FSX) does, then our model generates results of how equilibrium AFMI size and net alphas change with AFMI concentration, thus providing the key insights of the FSX model. In addition to FSX's fixed-point equilibrium results, we also generate results of the dynamics of AFMI.

4.1 Comparison of Our Model and the FSX Model

Notice that some of our models' settings are consistent with those of FSX. In particular, FSX assumes that when investors are risk-neutral, as long as they have sufficient wealth, they invest until expected fund net alphas drops to zero due to decreasing returns to scale, and when investors are mean-variance risk-averse, they maximize their portfolio instantaneous Sharpe ratios. In our model, investors' problems are similar. We also assume that when investors are risk-neutral, they invest until expected fund net alphas (conditional on current information) drops to zero, and when investors are mean-variance risk-averse, they maximize their portfolio Sharpe ratios instantaneously. Also, in FSX, fund managers maximize expected fund net alphas in order to survive, and under the settings of FSX, maximizing managers' expected fund net alpha is equivalent to maximizing managers' profits. In our model, our fund managers' problem is to maximize fund profits, similar to those in FSX.

However, as FSX is a one-period model, it provides results of AFMI concentration only at the fixed-point equilibrium. Our continuous-time model generates results regarding the dynamics of AFMI concentration and offers new insights to this area.

4.2 Exogenous AFMI Concentration, Equilibrium AFMI Size, and Performance

We can degenerate our model to provide the key insights of the FSX model where AFMI concentration is exogenous. We set the following parameters the same across funds to simplify our discussions: $B_i = B,\ c_i = c,\ f_i = f,\ i = 1, \dots, n$. Also, we assume that investors are mean-variance risk-averse. In addition, similar to the baseline model of FSX, we assume a continuum of AFMI concentration levels, denoted by H, which is exogenous and affects production and cost of managers' activities.

FSX assumes that managers spend costly efforts in producing fund returns. A higher AFMI concentration increases the productiveness of their efforts as it implies more unexplored investment opportunities and simultaneously affects the costs of managerial efforts as it affects managers' salaries. Then, in the FSX model, the tradeoff of a higher AFMI concentration affects the equilibrium AFMI size and expected net alphas.[35] Our model does not explicitly define managerial efforts but assumes that managers choose the amount of assets to be under active management, $q_{i,t}^a$. Then, we can regard $q_{i,t}^a$ as a variable representing managerial effort. To model how AFMI concentration affects the effectiveness of a higher $q_{i,t}^a$ (notice that a higher $q_{i,t}^a$ induces higher returns and higher costs), we assume that funds' decreasing returns to scale coefficient is a function of AFMI concentration, i.e., $c = c(H)$. This is because a higher H implies more unexplored investment opportunities that mitigate the decreasing returns to scale. In other words, if a unit of $q_{i,t}^a$ can produce more returns, then the "net costs" of a unit of $q_{i,t}^a$ is lower, implying a lower level of decreasing returns to scale $c(H)$. Also, if H is higher, managers' salaries on per unit of $q_{i,t}^a$ can be higher, so the decreasing returns to scale level $c(H)$ is also higher. Then, whether $c(H)$ increases with H depends on this tradeoff.

Based on these settings and performing a similar analysis as shown before, we find the equilibrium AFMI size and (conditional) expected net alphas:

$$IS_t^* \triangleq \sum_{i=1}^{n} q_{i,t}^* = \frac{V\sigma_p^2}{4f\left[B^2\mu_p + c(H)V\sigma_p^2\right]} \sum_{i=1}^{n} \left(A_i m_{i,t}\right)^2 \qquad (47)$$

[35] See the discussions of direct benefits and Proposition RA3 in FSX Section 2.

$$\mathrm{E}\left[\frac{dS_{i,t}}{S_{i,t}}\middle|\mathcal{F}_t^{\xi}\right]=\frac{fB^2\mu_p}{B^2\mu_p+c(H)V\sigma_p^2}dt. \tag{48}$$

With simple mathematics, we can show that both IS_t^* and $\mathrm{E}\left[\frac{dS_{i,t}}{S_{i,t}}\middle|\mathcal{F}_t^{\xi}\right]$ increase (decrease) with H if and only if $c'(H)\triangleq\frac{dc(H)}{dH}<0$ $(c'(H)>0)$. Also, IS_t^* is convex (concave) in H if and only if $\mathrm{E}\left[\frac{dS_{i,t}}{S_{i,t}}\middle|\mathcal{F}_t^{\xi}\right]$ is convex (concave) in H.

The intuition is that, if a higher AFMI concentration decreases the decreasing returns to scale level by inducing more productivity than costs of active management, it induces a higher expected net alpha for each fund. The higher expected net alpha attracts investors to invest more in active funds, increasing the fund sizes and then the AFMI size. When risk-averse investors observe higher expected fund net alphas, they invest more in active funds, but as investing more in active funds increases their portfolios risk, they do not invest "too much more" in active funds. Thus, although larger fund sizes imply larger costs due to decreasing returns to scale, this effect is mitigated by investors' risk aversion, such that the effect of decreasing returns to scale does not overwhelm the increment in expected net alphas in equilibrium. Then, in equilibrium, AFMI size and expected net alphas both increase with AFMI concentration. Also, AFMI concentration affects the rates of changes of equilibrium AFMI size and expected net alphas in a similar way, so equilibrium AFMI size and expected net alphas are both convex or both concave in AFMI concentration.

The theoretical results above are the same as those of Proposition RA3 in FSX, so our model provides the key insights of FSX. The following proposition summarizes these results.

Proposition FSX. In equilibrium, higher concentration induces larger (smaller) equilibrium AFMI size and higher (lower) equilibrium expected net alphas if and only if higher concentration induces a larger (smaller) impact on productivity of active management than on the costs of active management. Then, AFMI size and expected net alphas either both increase with AFMI concentration, or both decrease with AFMI concentration. In addition, AFMI size and expected net alphas are either both convex in AFMI concentration or both concave in AFMI concentration.

Proof. See the Appendix. □

4.3 Endogenous AFMI Concentration, Equilibrium AFMI Size, and Performance

Next, similar to Section 2.4 of FSX, we endogenize the AFMI concentration and calculate it as HHI, the sum of funds' market shares squared. We replace H in the above analysis by HHI. By Equation (23), the equilibrium AFMI HHI is derived as

$$\begin{aligned} HHI_t^* &= \frac{\mathbf{X^{RA\prime}}(HHI_t^*)\mathbf{A^4 I^4(m_t)X^{RA}}(HHI_t^*)}{[\mathbf{X^{RA\prime}}(HHI_t^*)\mathbf{A^2 I^2(m_t)1}]^2} \\ &= \frac{\sum_{i=1}^n X_i^{RA^2}(HHI_t^*)\left(A_i m_{i,t}\right)^4}{\left[\sum_{i=1}^n X_i^{RA}(HHI_t^*)\left(A_i m_{i,t}\right)^2\right]^2} = \frac{\sum_{i=1}^n \left(A_i m_{i,t}\right)^4}{\left[\sum_{i=1}^n \left(A_i m_{i,t}\right)^2\right]^2} \end{aligned} \tag{49}$$

where

$$X_i^{RA}(HHI_t^*) = \frac{V\sigma_p^2}{4f\left[B^2\mu_p + c(HHI_t^*)V\sigma_p^2\right]}. \tag{50}$$

As we assume $B_i = B,\ c_i = c(HHI_t^*),\ f_i = f,\ i = 1, \dots, n$, the size factors $X_i^{RA}(HHI_t^*)$ are the same across funds, so they cancel out with each other in the numerator and denominator of Equation (49). Therefore, the dynamics of HHI_t^* depends on the changes in managers' relative inferred abilities. This result is similar to those in our Propositions RN and RA.

Substituting HHI_t^* into equilibrium AFMI size and fund net alphas, we have

$$IS_t^* = \frac{V\sigma_p^2}{4f\left[B^2\mu_p + c(HHI_t^*)V\sigma_p^2\right]}\sum_{i=1}^n \left(A_i m_{i,t}\right)^2 \tag{51}$$

$$\frac{dS_{i,t}}{S_{i,t}} = \frac{fB^2\mu_p}{B^2\mu_p + c(HHI_t^*)V\sigma_p^2}dt + \frac{2fB}{A_i m_{i,t}}d\overline{W}_{i,t}. \tag{52}$$

Then, we can see that a positive shock in a fund's return, $d\overline{W}_{i,t}$, which induces a positive change in this manager's inferred ability, $dm_{i,t}$, has direct positive effects on the fund's net alpha and the AFMI size. This is because a positive shock in a fund's return implies a larger fund net alpha to investors, inducing more investments to this fund and, consequently, a larger AFMI size. It also has indirect effects on the fund's net alpha and the AFMI size through HHI_t^*. The change in this manager's inferred ability increases or decreases HHI_t^*, depending on parameter values. Then the change in HHI_t^* increases (decreases) both the fund's net alpha and the AFMI size, if the change in HHI_t^* induces larger (smaller) productivities of active management than costs of active management, i.e., $c(HHI_t^*)$ is smaller (larger). These results of the dynamics of the AFMI performance and size are new compared to those of FSX, which studies one-period fixed-point equilibria.

If we want to study how the dynamics of HHI_t^* affect the dynamics of other AFMI variables in equilibrium, we can extend our model in a way similar to the one shown above to study the corresponding relations or effects. We leave these extensions for future research.

5 Empirical AFMI Concentration and Related Empirical Literature

In this section, we illustrate the empirical U.S. AFMI concentration and show that some of its patterns can be explained by our theories. We also show that some stylized findings in the empirical AFMI literature can be explained by our model and offer some directions of future research in AFMI concentration.

5.1 Data

We collect our active fund data from the survivor-bias-free mutual fund database of the Center for Research in Security Prices (CRSP). Our sample period is from January 1990 to December 2020, and monthly data is used. We exclude index funds, variable annuity funds, and exchange-traded funds (ETFs) then choose U.S. domestic equity-only mutual funds by using the Lipper fund classification.[36] We use the MFLINKS database to aggregate fund share class-level information to fund-level information. In particular, we calculate funds' total net assets under management by summing up its share classes' net assets under management. Fund family is identified by the management company code.[37] We calculate fund families' net assets under management by summing up member funds' net assets under management. Then, we calculate the following: funds' (fund families') market shares based on their net assets under management, the HHI at fund level (fund family level), and the sum of market shares of the biggest five funds (fund families), i.e., the 5FI at fund level (fund family level). The equity

[36] We use funds in the following Lipper classes: Large-Cap Core, Large-Cap Growth, Large-Cap Value, Mid-Cap Core, Mid-Cap Growth, Mid-Cap Value, Small-Cap Core, Small-Cap Growth, Small-Cap Value, Multi-Cap Core, Multi-Cap Growth, and Multi-Cap Value. If a fund has a missing Lipper class in some months, we use its Lipper class in the previous months; if there is no information on a Lipper class in the previous months, we use its Lipper class in the later months.

[37] If a fund has a missing management company code in some months, we use the fund's management company code in the previous months; if there is no information of management company code in the previous months, we use the fund's management company code in the later months.

fund filter and data processing are similar to Brown and Wu (2016), Feldman, Saxena, and Xu (2020), and Feldman and Xu (2021).

5.2 The U.S. AFMI Concentration

Table 1 reports the summary statistics. We can see that there is a big variation in funds' sizes (the net asset under management) with a standard deviation equal to 4.2 billion dollars. The fund families' sizes have an even bigger variation, with a standard deviation equal to 72.4 billion dollars. A few fund families only have one equity fund, and others have multiple family member funds. The number of funds and number of fund families tend to be symmetric, whereas the HHIs at fund level and at fund family level tend to skew to the right.

Figure 3 plots the HHIs and the 5FIs at the fund and fund family levels, and the number of funds and fund families during our sample period. The two recessions, from March 2001 to November 2001, and from December 2007 to June 2009, are marked in gray. First, notice that the HHIs at fund level and at fund family level both fluctuate a lot, and do not converge to a particular level. This finding is consistent with the framework with dynamic manager abilities but not consistent with the one with constant manager abilities. Therefore, the finding here is consistent with those of Feldman and Xu (2021).[38]

At both the fund level and fund family level, the 5FI tends to move closely with the HHI. If we calculate the correlation, we find that at the fund level (fund family level), the HHI and the 5FI have a correlation coefficient of 0.94 (0.89), significant at 1% significance level. In other words, a higher 5FI tends to increase the HHI. This is consistent with our theoretical prediction in Propositions RN and RA, that the increase in the biggest funds' inferred manager abilities (implied by their larger fund sizes) has a positive impact on the dynamics of the HHI. The impact from these biggest funds is larger than those from smaller funds because of these funds' large inferred manager abilities, size factors, and/or sensitivities of gross alpha to ability (implied by their large fund sizes).

At the fund level, the HHI tends to move in the opposite direction with the number of funds, and these two variables' correlation coefficient is -0.85, significant at 1% significance

[38] Feldman and Xu (2021) shows that the fund flows' sensitivities to fund performance are nonmonotonic over time, which is consistent with a nonlinear filtering framework of dynamic unobservable managing abilities and inconsistent with a framework of constant unobservable managing abilities.

level. This implies that more competitors tend to make the market less concentrated. On the other hand, at the fund family level, the HHI is insignificantly associated with the number of fund families. Also, as we can see above, the HHI is more correlated with the 5FI than with the number of competitors. This is consistent with our theoretical framework that the managers' relative inferred abilities are more relevant than the number of rivals in analyzing the HHI. Therefore, it is important to study heterogeneous managers where the HHI captures managers' relative inferred abilities, instead of homogeneous managers where the HHI is simply the inverse of the number of competitors.

After the financial crisis from December 2007 to June 2009, the HHIs at the fund level and fund family level keep gradually decreasing for a few years. Our model shows that these decreasing patterns in HHIs are due to the fact that funds' and fund families' inferred abilities become more similar with each other. Then, it is interesting to check which policies or economic conditions make investors expect that manager abilities are more similar than before. Also, in the current few years, the HHIs at fund level and at fund family level both increase. Our model shows that the increasing patterns are due to the inferred abilities of big (small) funds and big (small) fund families increase (decrease). It is also interesting to analyze the factors that change these inferred abilities.

5.3 Related Empirical Literature

Wahal and Wand (2011) shows that from late 1990s to 2005, incumbents in the mutual fund market that have a high overlap in their portfolio holdings with those of new entrants experience lower fund flows and lower alphas. Similarly, Kosowski, Timmermann, Wermers, and White (2006) shows that outperforming managers become scarce after 1990 and speculates that this might be due to the competition among the large number of new funds that reduces the gains from trading. Fama and French (2010) also reports a decline in the persistence of alphas after 1992 and speculates that the cause is either diseconomies of scale or the entry of hordes of mediocre funds that make it difficult to uncover truly informed managers.

Our model can explain these phenomena coherently. From Figure 3, we can see that the number of funds and fund families keep increasing from 1995 to early 2000s; and in the same period, the HHIs at both fund level and fund family level both decrease. The new entrants of funds and fund families in this period hold portfolios similar to the incumbents, as Wahal and Wand (2011) concludes, inducing similarity in funds' inferred abilities. As similarity in funds' inferred abilities leads to similarity in equilibrium fund sizes, the HHI decreases. The decrease in the HHI might decrease the productivity of funds' active management more than the costs of active management because when more funds hold similar active portfolios and trade on similar assets, it is much more difficult to find investment opportunities to create abnormal returns. Thus, the decrease in the HHI induces larger decreasing returns to scale level during that period. The higher decreasing returns to scale level consequently induces lower fund sizes and fund net alphas, as we discuss previously, consistent with the findings of these empirical studies.

Also, current literature finds that family members can compete or cooperate with each other [e.g., Evans, Prado, and Zambrana (2020) and Eisele, Nefedova, Parise, and Peijnenburg (2020)]. Figure 3 shows that although the number of funds and the number of fund families move closely with each other (the correlation coefficient of their time series is close to 1), the HHI at fund level and the HHI at fund family level do not move closely with each other. Their correlation coefficient is only 0.17 and significant at 1% significance level. This implies that AFMI concentrated at fund level is not necessarily concentrated at fund family level and vice versa. Using only the HHI at fund level or the HHI at fund family level might not represent AFMI concentration well. To more precisely measure AFMI concentration, as implied by these current studies, we might need to regard family members that tend to cooperate (compete) with each other as one competitor (separate competitors).

Other literature shows that mutual funds compete in different dimensions, such as by trading assets in specific industries and style markets (defined by e.g., stock's total capitalization and book-to-market-ratio), by selling fund shares in specific retail market segments (e.g., direct-sold and broker-sold), and by offering unique products [e.g., Kacperczyk, Sialm, and Zheng (2005), Guercio and Reuter (2014), Hoberg, Kumar, and Prabhala (2018), and Kostovetsky and Warner (2020)]. Thus, we might need to use new methods to define competitors when analyzing the market structure of sub-sectors in AFMI.

6 Conclusion

We introduce continuous-time rational models of dynamics of AFMI equilibrium HHI, in which unobservable fund manager abilities are dynamic. In equilibrium, managers with higher inferred abilities receive larger fund sizes, so managers' relative inferred abilities determine the HHI. Our model predicts that if a manager's inferred ability is sufficiently larger (smaller) than others', then an increase in this manager's inferred ability has a positive (negative) impact on the change of the equilibrium HHI. If this fund has a larger inferred ability, fund size factor, and sensitivity of gross alpha to ability, then the positive (negative) impact is stronger.

If a manager has sufficiently large (small) inferred ability relative to others', then the HHI is concave in this manager's inferred ability, and the concavity has a negative impact on the change of the equilibrium HHI. Also, if all funds' inferred abilities are sufficiently close to each other's, then the equilibrium HHI is convex in this manager's inferred ability, and this convexity has a positive impact on the change of the equilibrium HHI.

We also show a special case in which unobservable fund manager abilities are constant. In this case, as time goes to infinity, managers' inferred ability converges to their true ability and does not change, making both equilibrium fund sizes and equilibrium HHI stay unchanged. All our results hold whether investors are risk-neutral or mean-variance risk-averse and whether there are fund entrances or exits.

Our framework can explain phenomena of the dynamics of AFMI. In particular, the fluctuation of the empirical HHI over time is consistent with our theoretical results, in which manager abilities are dynamic and unobservable, but is inconsistent with a model with constant unobservable manager abilities. Also, the fact that the empirical HHIs are mainly driven by the largest funds and fund families is also consistent with our theoretical prediction. Furthermore, the fact that the HHI is more correlated with the large competitors' market shares than the number of competitors shows the importance of modeling heterogeneous managers where the HHI captures managers' relative inferred abilities, instead of homogeneous managers where the HHI is simply the inverse of the number of competitors. The stylized findings that new entrants who have portfolio holdings similar to incumbents' decrease funds' performance and fund flows can also be explained by our model.

In addition, our paper shows that future research on the dynamics of AFMI concentration can focus on factors that affect fund managers' relative inferred abilities, new measures of AFMI concentration, and how the dynamics of AFMI concentration affects other aspects of AFMI. Although our paper studies the dynamics of AFMI concentration, our framework can be extended to study the dynamics of concentration in other industries in which incomplete information exists: producers' performance depends on dynamic states that are unobservable to customers and producers.

Appendix

This section provides the proofs of the results in the corresponding sections.

Proof of Results in Section 2.2

In the managers' problems shown in Equation (15), to maximize $A_i m_{i,t} q_{i,t}^a - c_i {q_{i,t}^a}^2$, we apply the first-order condition with respect to $q_{i,t}^a$, and find the optimal value ${q_{i,t}^a}^*$ as

$$ {q_{i,t}^a}^* = \frac{A_i m_{i,t}}{2c_i}. \tag{A1} $$

The second-order condition $-2c_i < 0$ shows that ${q_{i,t}^a}^*$ induces a maximum. Substituting Equation (A1) into Equation (14) and rearranging, we find the fund i' optimal fund sizes as

$$ q_{i,t}^* = \frac{\left(A_i m_{i,t}\right)^2}{4c_i f_i}. \tag{A2} $$

Here we assume that manager i, $i = 1, \dots, n$, sets f_i sufficiently low such that the constraint $0 \le {q_{i,t}^a}^* \le q_{i,t}^*$ is automatically satisfied and we do not incorporate this constraint in the optimization.

Q.E.D.

Proof of Results in Section 2.6

First, we define the following:

- mean return vector of the $n+1$ assets, $\boldsymbol{\mu}_\mathbf{t}$, which is an $(n+1)\times 1$ vector, with $\mu_{i,t} = \left(\frac{q_{i,t}^a}{q_{i,t}} A_i m_{i,t} - \frac{c_i {q_{i,t}^a}^2}{q_{i,t}} - f_i + \mu_p\right) dt$, $i = 1, \dots, n$, and $\mu_{n+1,t} = \mu_p dt$;
- covariance matrix of the $n+1$ assets, $\mathbf{Q}_\mathbf{t}$, which is an $(n+1)\times(n+1)$ positive definite symmetric matrix, with diagonal elements $Q_{ii,t} = \left[\left(\frac{q_{i,t}^a}{q_{i,t}}\right)^2 B_i^2 + \sigma_p^2\right] dt$, $i = 1, \dots, n$, and $Q_{ii,t} = \sigma_p^2$, $i = n+1$, and off-diagonal elements $Q_{ij,t} = \sigma_p^2 dt$, $\forall i \ne j$.

Then, we have

$$ \mathrm{E}\left[\frac{dp_t}{p_t}\middle| \mathcal{F}_t^{\boldsymbol{\xi}}\right] = \mathbf{v_t}'\boldsymbol{\mu}_\mathbf{t} \tag{A3} $$

$$ \mathrm{Var}\left[\frac{dp_t}{p_t}\middle| \mathcal{F}_t^{\boldsymbol{\xi}}\right] = \mathbf{v_t}'\mathbf{Q_t}\mathbf{v_t}. \tag{A4} $$

Next, we write down the Lagrange function

$$F_t(\mathbf{v_t}, \lambda_t) = \frac{\mathbf{v_t}'\boldsymbol{\mu}_\mathbf{t}}{\sqrt{\mathbf{v_t}'\mathbf{Q_t}\mathbf{v_t}}} + \lambda_t(1 - \mathbf{v_t}'\mathbf{1}). \tag{A5}$$

We later will argue that the condition $0 \le v_{i,t} \le 1,\ \forall t, i = 1, \dots, n+1$ is automatically satisfied in our model, so it does not affect our optimization process and is not incorporated in Equation (A5). First-order conditions generate

$$\nabla_{\mathbf{v_t}} F_t(\mathbf{v}_\mathbf{t}^*, \lambda_t^*) = \frac{(\mathbf{v}_\mathbf{t}^{*\prime}\mathbf{Q_t}\mathbf{v}_\mathbf{t}^*)^{\frac{1}{2}}\boldsymbol{\mu}_\mathbf{t} - (\mathbf{v}_\mathbf{t}^{*\prime}\mathbf{Q_t}\mathbf{v}_\mathbf{t}^*)^{-\frac{1}{2}}\mathbf{Q_t}\mathbf{v}_\mathbf{t}^*\mathbf{v}_\mathbf{t}^{*\prime}\boldsymbol{\mu}_\mathbf{t}}{\mathbf{v}_\mathbf{t}^{*\prime}\mathbf{Q_t}\mathbf{v}_\mathbf{t}^*} - \lambda_t^*\mathbf{1} = \mathbf{0} \tag{A6}$$

$$\nabla_{\lambda_t} F_t(\mathbf{v}_\mathbf{t}^*, \lambda_t^*) = 1 - \mathbf{v}_\mathbf{t}^{*\prime}\mathbf{1} = \mathbf{0}. \tag{A7}$$

Multiplying both sides of Equation (A6) by $\mathbf{v}_\mathbf{t}^{*\prime}$ on the left, we have

$$\frac{(\mathbf{v}_\mathbf{t}^{*\prime}\mathbf{Q_t}\mathbf{v}_\mathbf{t}^*)^{\frac{1}{2}}\mathbf{v}_\mathbf{t}^{*\prime}\boldsymbol{\mu}_\mathbf{t} - (\mathbf{v}_\mathbf{t}^{*\prime}\mathbf{Q_t}\mathbf{v}_\mathbf{t}^*)^{-\frac{1}{2}}\mathbf{v}_\mathbf{t}^{*\prime}\mathbf{Q_t}\mathbf{v}_\mathbf{t}^*\mathbf{v}_\mathbf{t}^{*\prime}\boldsymbol{\mu}_\mathbf{t}}{\mathbf{v}_\mathbf{t}^{*\prime}\mathbf{Q_t}\mathbf{v}_\mathbf{t}^*} = \lambda_t^* = 0. \tag{A8}$$

Then,

$$(\mathbf{v}_\mathbf{t}^{*\prime}\mathbf{Q_t}\mathbf{v}_\mathbf{t}^*)^{\frac{1}{2}}\boldsymbol{\mu}_\mathbf{t} - (\mathbf{v}_\mathbf{t}^{*\prime}\mathbf{Q_t}\mathbf{v}_\mathbf{t}^*)^{-\frac{1}{2}}\mathbf{Q_t}\mathbf{v}_\mathbf{t}^*\mathbf{v}_\mathbf{t}^{*\prime}\boldsymbol{\mu}_\mathbf{t} = \mathbf{0}. \tag{A9}$$

The second-order condition is satisfied and omitted here for brevity. Then, $\mathbf{v}_\mathbf{t}^*$ is a maximizer. Next we solve $\mathbf{v}_\mathbf{t}^*$ explicitly. Define $\mu_v^* dt \triangleq \mathbf{v}_\mathbf{t}^{*\prime}\boldsymbol{\mu}_\mathbf{t}$ and $\sigma_v^{2*} dt \triangleq \mathbf{v}_\mathbf{t}^{*\prime}\mathbf{Q_t}\mathbf{v}_\mathbf{t}^*$, which are the portfolio mean return and variance of return at the optimal weight allocations in dt, respectively. Rearranging Equation (A9), we have

$$\mathbf{Q_t}\mathbf{v}_\mathbf{t}^* = \boldsymbol{\mu}_\mathbf{t}\frac{\sigma_v^{2*}}{\mu_v^*}. \tag{A10}$$

Then, the ith element of $\mathbf{Q_t}\mathbf{v}_\mathbf{t}^*$ is $\left[v_{i,t}^*\left(\frac{q_{i,t}^a}{q_{i,t}}\right)^2 B_i^2 + \sigma_p^2\right] dt,\ i = 1, \dots, n$. The last element of $\mathbf{Q_t}\mathbf{v}_\mathbf{t}^*$ is $\sigma_p^2 dt$. Also, the ith element of $\boldsymbol{\mu}_\mathbf{t}\frac{\sigma_v^{2*}}{\mu_v^*}$ is $\frac{\sigma_v^{2*}}{\mu_v^*}\left(\frac{q_{i,t}^a}{q_{i,t}}A_i m_{i,t} - \frac{c_i {q_{i,t}^a}^2}{q_{i,t}} - f_i + \mu_p\right) dt,\ i = 1, \dots, n$. The last element of $\boldsymbol{\mu}_\mathbf{t}\frac{\sigma_v^{2*}}{\mu_v^*}$ is $\frac{\sigma_v^{2*}\mu_p}{\mu_v^*} dt$. We have the following relation:

$$\frac{v_{i,t}^*\left(\frac{q_{i,t}^a}{q_{i,t}}\right)^2 B_i^2 + \sigma_p^2}{\sigma_p^2} = \frac{\frac{\sigma_v^{2*}}{\mu_v^*}\left(\frac{q_{i,t}^a}{q_{i,t}}A_i m_{i,t} - \frac{c_i {q_{i,t}^a}^2}{q_{i,t}} - f_i + \mu_p\right)}{\frac{\sigma_v^{2*}\mu_p}{\mu_v^*}} \tag{A11}$$

for $i = 1, \dots, n$. Rearranging the expression above, we have

$$v_{i,t}^* = \frac{\left(\frac{q_{i,t}^a}{q_{i,t}} A_i m_{i,t} - \frac{c_i {q_{i,t}^a}^2}{q_{i,t}} - f_i\right)\sigma_p^2}{\left(\frac{q_{i,t}^a}{q_{i,t}}\right)^2 B_i^2 \mu_p} \tag{A12}$$

for $i = 1, \dots, n$.

Then, funds' sizes can be expressed as, for $i = 1, \dots, n$:

$$q_{i,t} = V v_{i,t}^* = V \frac{\left(\frac{q_{i,t}^a}{q_{i,t}} A_i m_{i,t} - \frac{c_i {q_{i,t}^a}^2}{q_{i,t}} - f_i\right)\sigma_p^2}{\left(\frac{q_{i,t}^a}{q_{i,t}}\right)^2 B_i^2 \mu_p}. \tag{A13}$$

Substitute the expression above into Equation (41), and rearrange to get

$$f_i q_{i,t} = -\frac{{q_{i,t}^a}^2 B_i^2 \mu_p}{V \sigma_p^2} - c_i {q_{i,t}^a}^2 + q_{i,t}^a A_i m_{i,t}. \tag{A14}$$

Manager i's problem is to maximize $f_i q_{i,t}$ by choosing $q_{i,t}^a$. Applying the first-order condition on the right-hand side of Equation (A14), we have

$$q_{i,t}^{a*} = \frac{A_i m_{i,t} V \sigma_p^2}{2\left(B_i^2 \mu_p + c_i V \sigma_p^2\right)}. \tag{A15}$$

The second-order condition is $-\frac{2B_i^2 \mu_p}{V\sigma_p^2} - 2c_i < 0$, showing that $q_{i,t}^{a*}$ is a maximizer. Then substituting $q_{i,t}^{a*}$ back to Equation (A13), we have

$$q_{i,t}^* = \frac{\left(A_i m_{i,t}\right)^2 V \sigma_p^2}{4 f_i \left(B_i^2 \mu_p + c_i V \sigma_p^2\right)}. \tag{A16}$$

We can see that

$$\frac{q_{i,t}^{a*}}{q_{i,t}^*} = \frac{2 f_i}{A_i m_{i,t}}. \tag{A17}$$

We assume that manager i sets f_i sufficiently low such that the condition $0 \leq q_{i,t}^{a*} \leq q_{i,t}^*$ is automatically satisfied and we do not incorporate this constraint in the optimization problem in Equation (41). Also, by Equations (A13) and (A16), we have, for $i = 1, \dots, n$,

$$v_{i,t}^* = \frac{q_{i,t}^*}{V} = \frac{\left(A_i m_{i,t}\right)^2 \sigma_p^2}{4 f_i \left(B_i^2 \mu_p + c_i V \sigma_p^2\right)}. \tag{A18}$$

As $m_{i,t} \geq \underline{m_{i,t}} \geq 0$ and all other parameters on the right-hand side of Equation (A18) are positive, $v_{i,t}^*$, $i = 1, \dots, n$ is nonnegative, i.e., investors do not short sell active funds. That is, as long as funds provide positive expected net alphas, investors do not short sell them. Also, summing up Equation (A18) for $i = 1, \dots, n$, we have

$$\sum_{i=1}^{n} v_{i,t}^* = \sum_{i=1}^{n} \frac{\left(A_i m_{i,t}\right)^2}{4f_i\left(\frac{B_i^2 \mu_p}{\sigma_p^2} + c_i V\right)}. \tag{A19}$$

With a sufficiently large μ_p or a sufficiently small σ_p^2, we have $\sum_{i=1}^{n} v_{i,t}^* \leq 1$. As $v_{i,t}^*$, $i = 1, \dots, n$ is nonnegative and $\sum_{i=1}^{n} v_{i,t}^* \leq 1$, we have $v_{i,t}^* \leq 1$, for $i = 1, \dots, n$. With all these conditions, we also have $0 \leq v_{n+1,t}^* \leq 1$; i.e., investors invest part of their wealth into the passive benchmark. The intuition is that as long as the passive benchmark portfolio provides sufficiently high expected return or sufficiently low risk, investors do not short sell it. These results are realistic because in reality, we observe investors invest part of their wealth in active funds and another in passive benchmark portfolios. Then, the condition $0 \leq v_{i,t} \leq 1$, $\forall, i = 1, \dots, n+1$ is automatically satisfied and we do not incorporate this constraint in solving the investors' optimization problems.

Q.E.D.

Proof of Results in Section 4

Summing up both sides of Equation (A16) over i to get AFMI equilibrium size IS_t^*, with $B_i = B$, $c_i = c(H)$, $f_i = f$, for $i = 1, \dots, n$ we have

$$IS_t^* \triangleq \sum_{i=1}^{n} q_{i,t}^* = \frac{V\sigma_p^2}{4f\left(B^2\mu_p + c(H)V\sigma_p^2\right)} \sum_{i=1}^{n} \left(A_i m_{i,t}\right)^2. \tag{A20}$$

By Equation (5), the ith element of $\mathbf{I^{-1}(\xi_t) d\xi_t}$ is

$$\frac{d\xi_{i,t}}{\xi_{i,t}} = A_i m_{i,t} dt + B_i d\overline{W}_{i,t}. \tag{A21}$$

Substituting Equations (A15), (A16), and (A21) into Equation (11), with $B_i = B$, $c_i = c(H)$, $f_i = f$, for $i = 1, \dots, n$, to get the equilibrium fund net alpha, we have

$$\frac{dS_{i,t}}{S_{i,t}} = \frac{fB^2\mu_p}{B^2\mu_p + c(H)V\sigma_p^2} dt + \frac{2fB}{A_i m_{i,t}} d\overline{W}_{i,t}. \tag{A22}$$

Then, the equilibrium expected fund net alpha conditional on the current information is

$$\mathrm{E}\left[\frac{dS_{i,t}}{S_{i,t}}\middle|\mathcal{F}_t^{\xi}\right]=\frac{fB^2\mu_p}{B^2\mu_p+c(H)V\sigma_p^2}dt. \tag{A23}$$

By differentiating Equations (A20) and (A23) with respect to H, we have

$$\frac{\partial IS_t^*}{\partial H}=\frac{-V^2\sigma_p^4c'(H)}{4f\left(B^2\mu_p+c(H)V\sigma_p^2\right)^2}\sum_{i=1}^{n}\left(A_i m_{i,t}\right)^2 \tag{A24}$$

$$\frac{\partial \mathrm{E}\left[\frac{dS_{i,t}}{S_{i,t}}\middle|\mathcal{F}_t^{\xi}\right]}{\partial H}=\frac{-fB^2\mu_pV\sigma_p^2c'(H)}{\left(B^2\mu_p+c(H)V\sigma_p^2\right)^2}dt, \tag{A25}$$

where $c'(H)\triangleq\frac{dc(H)}{dH}$. Thus, given that all other parameters are positive, both $\frac{\partial IS_t^*}{\partial H}$ and $\frac{\partial \mathrm{E}\left[\frac{dS_{i,t}}{S_{i,t}}\middle|\mathcal{F}_t^{\xi}\right]}{\partial H}$ are positive (negative) if and only if $c'(H)<0$ $(c'(H)>0)$. In other words, both IS_t^* and $\mathrm{E}\left[\frac{dS_{i,t}}{S_{i,t}}\middle|\mathcal{F}_t^{\xi}\right]$ increase (decrease) with H if and only if $c'(H)<0$ $(c'(H)>0)$.

We differentiate Equations (A24) and (A25) with respect to H, and we have

$$\frac{\partial^2 IS_t^*}{\partial H^2}=16f^2V^2\sigma_p^4\times \frac{-c''(H)\left(B^2\mu_p+c(H)V\sigma_p^2\right)+2V\sigma_p^2(c'(H))^2}{\left[4f\left(B^2\mu_p+c(H)V\sigma_p^2\right)\right]^3}\sum_{i=1}^{n}\left(A_i m_{i,t}\right)^2 \tag{A26}$$

$$\frac{\partial^2 \mathrm{E}\left[\frac{dS_{i,t}}{S_{i,t}}\middle|\mathcal{F}_t^{\xi}\right]}{\partial H^2}=fB^2\mu_pV\sigma_p^2\times \frac{-c''(H)\left(B^2\mu_p+c(H)V\sigma_p^2\right)+2V\sigma_p^2(c'(H))^2}{\left(B^2\mu_p+c(H)V\sigma_p^2\right)^3}dt. \tag{A27}$$

Thus, given that all other parameters are positive, both $\frac{\partial^2 IS_t^*}{\partial H^2}$ and $\frac{\partial^2 \mathrm{E}\left[\frac{dS_{i,t}}{S_{i,t}}\middle|\mathcal{F}_t^{\xi}\right]}{\partial H^2}$ are positive (negative) if and only if $-c''(H)\left(B^2\mu_p+c(H)V\sigma_p^2\right)+2V\sigma_p^2(c'(H))^2>0$ $(-c''(H)\left(B^2\mu_p+c(H)V\sigma_p^2\right)+2V\sigma_p^2(c'(H))^2<0)$. Consequently, IS_t^* is convex (concave) in H if and only if $\mathrm{E}\left[\frac{dS_{i,t}}{S_{i,t}}\middle|\mathcal{F}_t^{\xi}\right]$ is convex (concave) in H.

Q.E.D.

References

Ali, A., Klasa, S., Yeung, E., 2009. The limitations of industry concentration measures constructed with compustat data: implications for finance research. Review of Financial Studies 22, 3839–3871.

Azar, J., Schmalz, M.C., Tecu, I., 2018. Anticompetitive effects of common ownership. Journal of Finance 73, 1513–1565.

Berk, J. B., 2005. Five myths of active portfolio management. The Journal of Portfolio Management 31, 27–31.

Berk, J. B., Green, R. C., 2004. Mutual fund flows and performance in rational markets. Journal of Political Economy 112, 1269–1295.

Brown, D., Wu, Y., 2013. Mutual fund families and performance evaluation. Unpublished working paper.

Brown, D., Wu, Y., 2016. Mutual fund flows and cross-fund learning within families. Journal of Finance 71, 383–424.

Bustamante M.C., Donangelo, A., 2017. Product market competition and industry returns. Review of Financial Studies 30, 4216–4266.

Choi, D, Kahraman, B., Mukherjee, A., 2016. Learning about mutual fund managers. Journal of Finance 71, 383–424.

Corhay, A., Kung, H., Schmid, L., 2020. Competition, markups, and predictable returns. Review of Financial Studies 33, 5906–5939.

Cornaggia, J., Mao, Y., Tian, X., Wolfe, B., 2015. Does banking competition affect innovation? Journal of Financial Economics 115, 189–209.

Cremers, M., Nair, V. B., Peyer, U., 2008. Takeover defenses and competition: The role of stakeholders. Journal of Empirical Legal Studies 5, 791–818.

Dangl, T., Wu, Y., Zechner, J., 2008. Market discipline and internal governance in the mutual fund industry. Review of Financial Studies 21, 2307–2343.

Dothan, M. U., Feldman, D., 1986. Equilibrium interest rates and multiperiod bonds in a partially observable economy. The Journal of Finance 41, 369–382.

Eisele, A., Nefedova, T., Parise, G., Peijnenburg, K., 2020. Trading out of sight: An analysis of cross-trading in mutual fund families. Journal of Financial Economics 135, 359–378.

Evans, R.B., Prado, M.P., Zambrana, R., 2020. Competition and cooperation in mutual fund families. Journal of Financial Economics 136, 168–188.

Fama, E., French, K., 2010. Luck versus skill in the cross-section of mutual fund returns. Journal of Finance 65, 1915–1947.

Feldman, D., 1989. The term structure of interest rates in a partially observable economy. Journal of Finance 44, 789–812.

Feldman, D., 1992. Logarithmic preferences, myopic decisions, and incomplete information. Journal of Financial and Quantitative Analysis 27, 619–629.

Feldman, D., Xu, J., 2021. Fund flows, performance, and exit under dynamic unobservable managing ability. Unpublished working paper.

Feldman, D., Saxena, K., Xu, J., 2020. Is the active fund management industry concentrated enough? Journal of Financial Economics 136, 23–43.

Feldman, D., Saxena, K., Xu, J., 2021. One global village? Competition in the international active fund management industry. Unpublished working paper.

Giannetti, M., Saidi, F., 2019. Shock propagation and banking structure. Review of Financial Studies 32, 2499–2540.

Gu, L., 2016. Product market competition, R&D investment, and stock returns. Journal of Financial Economics 119, 441–455.

Guercio, D. D., Reuter, J., 2014. Mutual fund performance and the incentive to generate alpha. Journal of Finance 69, 1673–1704.

Hoberg, G., Kumar, N., Prabhala, N., 2018. Mutual fund competition, managerial skill, and alpha persistence. Review of Financial Studies 31, 1896–1929.

Kacperczyk, M., Sialm, C., Zheng, L., 2005. On the industry concentration of actively managed equity mutual funds. Journal of Finance 60, 1983–2011.

Koch, A., Panayides, M., Thomas, S., 2021. Common ownership and competition in product markets. Journal of Financial Economics 139, 109–137.

Kosowski, R., Timmermann, A., Wermers, R., White, H., 2006. Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. Journal of Finance 61, 2551–2595.

Kostovetsky, L., Warner, J. B., 2020. Measuring innovation and product differentiation: evidence from mutual funds. Journal of Finance 75, 779–823.

Liptser, R. S., Shiryaev, A. N., 2001a. *Statistics of Random Processes I*. NY: Springer-Verlag.

Liptser, R. S., Shiryaev, A. N., 2001b. *Statistics of Random Processes II*. NY: Springer-Verlag.

O'Brien, D. P., Salop S. C., 2000. Competitive effects of partial ownership: Financial interest and corporate control. Antitrust Law Journal 67, 559–614.

Pastor, L., Stambaugh, R. F., 2012. On the size of the active management industry. Journal of Political Economy 120, 740–781.

Spiegel, M., Tookes, H., 2013. Dynamic competition, valuation, and merger activity. Journal of Finance 68, 125–172.

Wahal, S., Wang, Y., 2011. Competition among mutual funds. Journal of Financial Economics 99, 40–59.

Figure 1. AFMI Equilibrium HHI and Relative Inferred Abilities

Figure 1 illustrates the results of an AFMI with two funds, fund 1 and fund 2. The vertical axis is the equilibrium AFMI Herfindahl-Hirschman Index, HHI_t^*, and the horizontal axis is manager 1's inferred ability, $m_{1,t}$. Manager 2's inferred ability $m_{2,t}$ is set to be one, so that $m_{1,t}$ can be regarded as manager 1's inferred ability relative to manager 2's. In Case One, the two managers have the same size factor, $X_1 = X_2 = 100$, and the same sensitivity of gross alpha to ability, $A_1 = A_2 = 1$. In Case Two, $X_1 = 2X_2 = 200$ and $A_1 = A_2 = 1$, whereas in Case Three, $X_1 = X_2 = 100$ and $A_1 = 2A_2 = 2$. The solid curve, dashed curve, and dotted dashed curve illustrate the results of Case One, Case Two, and Case Three, respectively.

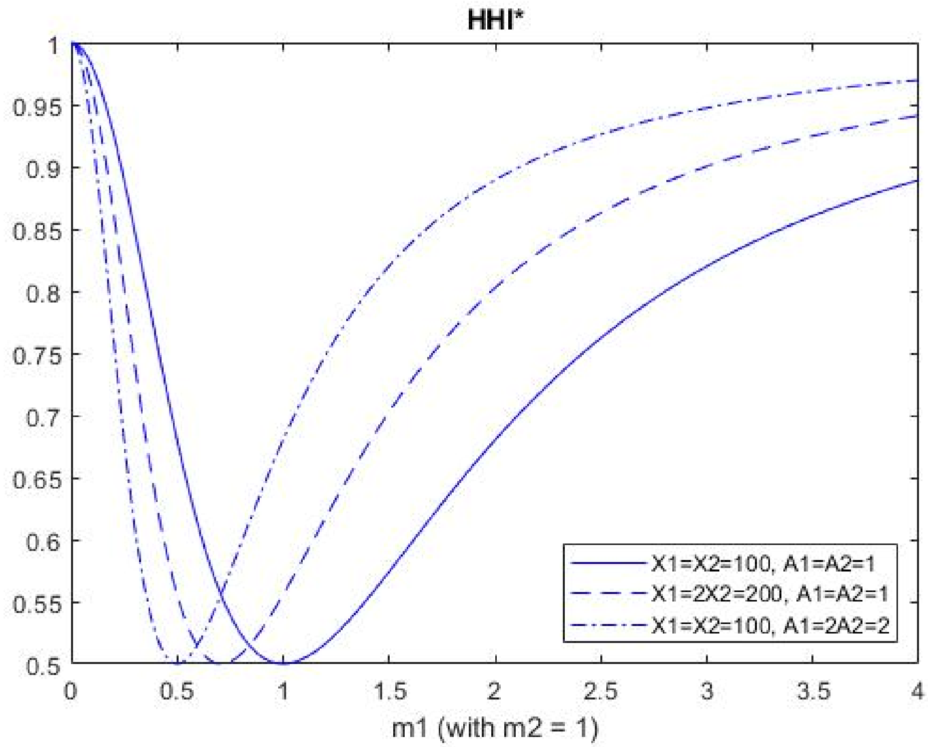

Figure 2. AFMI Equilibrium HHI and Inferred Abilities with Dynamic Abilities and Constant Abilities

Figure 2 illustrates the results of an AFMI with two funds, fund 1 and fund 2, with dynamic abilities in the two upper subplots and with constant abilities in the two lower subplots, respectively. For each case, on the left-hand side, we illustrate the simulated inferred abilities, $m_{1,t}$ and $m_{2,t}$, in blue lines and red stars, respectively. On the right-hand side, we illustrate the equilibrium AFMI Herfindahl-Hirschman Index, HHI_t^*. We plot these simulation results from Month 0 to Month 400.

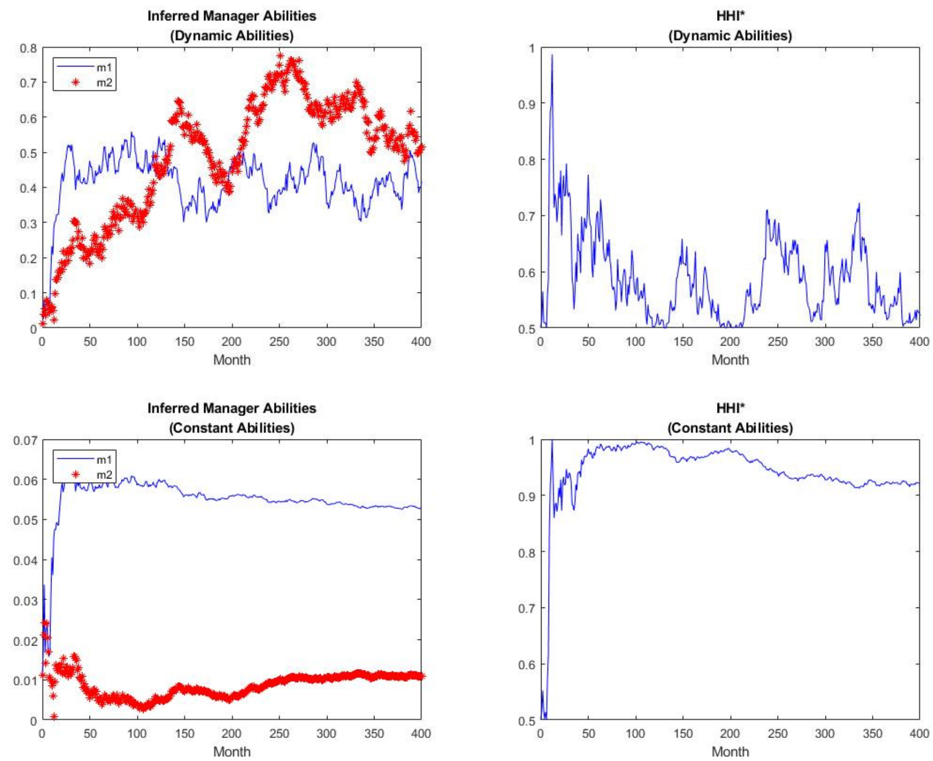

Figure 3 U.S. AFMI Concentration Dynamics

Figure 3 plots the monthly values of variables from January 1990 to December 2020 using the U.S. active equity mutual fund data from the Center for Research in Security Prices (CRSP). The two graphs on the top plot the HHIs calculated at fund level and fund family level, respectively. The two graphs in the middle plot the 5FIs calculated at fund level and fund family level, respectively. The two graphs at the bottom plot the number of funds and the number of fund families in the market, respectively. The HHI is the Herfindahl-Hirschman Index, calculated as the sum of market shares squared. Funds' (fund families') market shares are calculated based on their net assets under management. The 5FI is the sum of the largest five market shares. The number of funds (the number of fund families) is counted as the number of the U.S. active equity mutual funds (mutual fund families) that have observations of net assets under management. The gray areas represent the two recessions, from March 2001 to November 2001, and from December 2007 to June 2009, respectively.

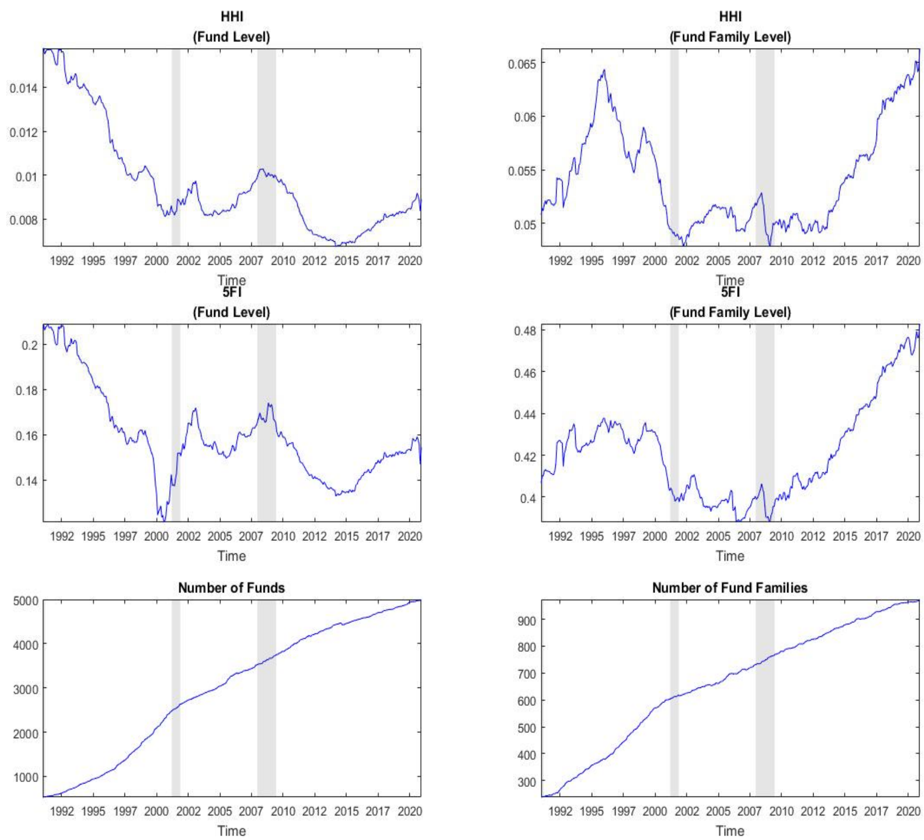

Table 1 Summary Statistics

Table 1 reports the summary statistics of our monthly observations from January 1990 to December 2020. Fund Net Asset is the fund's net assets under management measured in million dollars. Fund Family Net Asset is the sum of funds' net assets under management in the fund family, measured in million dollars. Fund Family Size is the number of funds in the fund family, and it is a number. Industry Size is the sum of all funds' net assets under management, measured in million dollars. HHI at Fund Level (HHI at Fund Family Level) is the Herfindahl-Hirschman Index, calculated as the sum of market shares squared of funds (fund families), and it is in decimal. Funds' (fund families') market shares are calculated based on their net assets under management. 5FI at Fund Level (5FI at Fund Family Level) is the sum of the market shares of the largest five funds (fund families), and it is in decimal. The number of funds (the number of fund families) is counted as the number of the U.S. active equity mutual funds (mutual fund families) that have observations of net assets under management.

				Percentile				
Variable	Observation	Mean	Standard Deviation	1st	25th	50th	75th	99th
Fund Net Asset (in 1 Million Dollar)	1079671	752.29	4169.00	0.10	6.90	50.70	301.10	11567.51
Fund Family Net Asset (in 1 Million Dollar)	1079671	22512.52	72388.38	0.40	211.80	2742.70	16429.76	448008.20
Fund Family Size (Number of Funds)	1079671	24.48	26.87	1	5	16	35	114
Industry Size (in 1 Million Dollar)	372	2183399	1220398.0	27564.8	1297342.0	2284910.0	3128818.0	4322742.0
HHI at Fund Level (Decimal)	372	0.0118	0.0140	0.0068	0.0082	0.0090	0.0103	0.1053
5FI at Fund Level (Decimal)	372	0.1698	0.0715	0.1245	0.1475	0.1562	0.1668	0.6346
Number of Funds in the Industry (Number)	372	2902.3410	1501.5510	53	1422	3110	4305	4977
HHI at Fund Family Level (Decimal)	372	0.0635	0.0632	0.0481	0.0502	0.0523	0.0576	0.4886
5FI at Fund Family Level (Decimal)	372	0.4310	0.0714	0.3886	0.4023	0.4208	0.4334	0.8900
Number of Fund Families in the Industry (Number)	372	643.0188	237.3335	34	452	676	838	967