ENDOGENOUS DYNAMIC CONCENTRATION OF THE ACTIVE FUND MANAGEMENT INDUSTRY, HETEROGENEOUS MANAGER ABILITIES, AND STOCK MARKET VOLATILITY

David Feldman and Jingrui Xu

CONTRIBUTION

- Merging two strands of literature

- Introducing advances

FIRST STRAND OF LITERATURE

- Size and performance of active fund management industry (AFMI) and their sensitivities to market concentration.

- Representatives:
  - Pastor and Stambaugh (2012, *JPE*)
  - Feldman, Saxena and Xu (2020, *JFE*)

SECOND STRAND OF LITERATURE

- Flow-performance relations of a single AFMI fund's sensitivity to inferred unobservable manager's skills.

- Representatives:
  - Berk and Green (2004, *JPE*)
  - Feldman and Xu (2022)

ADVANCES WRT THE FIRST STRAND

- First strand – *exogenous* AFMI concentration
  here – *endogenous*

- AFMI concentration is determined in equilibrium:
  - Investors learn managers skills by observing managers' performance
  - Investors determine funds' sizes by adjusting their investment flows according to their inferred managers' (unobservable dynamic) skills

The resulting set of fund sizes determines the market concentration

ADVANCES WRT TO THE SECOND STRAND

- Berk and Green (2004) assumes constant managers ability and risk neutrality

- Constant unobservable abilities ⇒
  - (quick) convergence to true abilities' values
  - flow performance sensitivities that quickly and monotonically decreasing to zero
  - no exit of "older" funds

- all 3 above points are *refuted* by the real-world and by the literature

- Feldman and Xu (2022) allow nonlinear (unobservable) dynamic abilities and solve two models for risk-neutral and for risk-averse investors, explaining
  - ever dynamic managers' abilities (as observed in the real-world)
  - ever nonmonotonic flow-performance sensitivities (as observed in the real-world)
  - exit of "older" funds (as observed in the real-world)
  - all 3 points above hold, both, for risk-neutral investors and risk-averse ones

ADVANCES WRT TO BOTH STRANDS

- Effect of performance shocks on AFMI concentration and its dynamics, and on flow-performance sensitivities

- Effects of performance variations on AFMI concentration and its dynamics, and on flow-performance sensitivities

- Effects of stock market volatility on managers' inferred abilities, on (the endogeneous) AFMI concentration and its dynamics, and on flow-performance sensitivities
(Requires nonlinear filtering)

EVIDENCE

- Dynamic nonmonotonic flow-performance sensitivities and exit of old funds [e.g., Feldman and Xu (2022)]

- Fund manager performance does not persist [e.g., Carhart (1997), Berk and Tonks (2007), Wang (2014)]

- Fund family activities vary across time and affect managers' ability to outperform benchmarks [e.g., Gaspar, Massa, and Matos (2006), Evans (2010), Brown and Wu (2016)]

- Managers' replacements and change of attention allocation affect manager abilities [e.g., Dangl, Wu, and Zechner (2008), Kacperczyk, Nieuwerburgh, and Veldkamp (2016)]

- Dynamic macroeconomic factors affect fund managers' abilities [e.g., Ferreira, Keswani, Miguel, and Ramos (2012, 2013), Feldman, Saxena, and Xu (2020, 2022)]

THIS PAPER

- Continuous-time rational models of (endogeneous) concentration in the active fund management industry

- Managers, facing decreasing returns to scale investment opportunities, with heterogenous dynamic unobservable abilities maximize profits by competing for investments of investors

- Infinitely many small risk-neutral or risk-averse investors

- Investors allocate wealth between active funds and a passive benchmark portfolio

- Investors drive returns down (if risk-averse) or all the way to zero (if risk-neutral)

- Funds' sizes incentivize managers

- Investors and managers are symmetrically informed, know all "law of motions" and parameter values, but do not observe abilities

- 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; we also determine second order effects

- Impact of changes in managers' inferred abilities increase in "fund size factors" (inverse of the product of decreasing returns to scale coefficient and fees) and sensitivity of gross alpha to ability

- 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. Including a comparison to the fixed-point equilibrium results in Feldman, Saxena and Xu (2022), which both confirms their results and offers new questions for future research, generated by the dynamic analysis

- We use three performance measures: Amihud-Goyenko $(1 - R^2)$, and standard deviations of net and gross alphas

- Empirical results
  - Follow the theoretical predictions above
  - US HHI fluctuates over last few decades and does not converge (consistent with our framework
  - From 1990 to early 2000s number of funds keeps increasing while the HHI keeps decreasing (Consistent with new funds that hold portfolios similar to those of incumbents, fund manager inferred abilities converge, inducing similar fund sizes, and lower HHI)

CONTRIBUTION

- First model of equilibrium dynamic AFMI concentration under heterogeneous managers with dynamic unobservable abilities

- Show how AFMI's concentration evolves with different factors, for risk neutral and for risk averse investors, under funds' entrants and exits.

- Empirical evidence on how funds relative performances, performance variations, and stock market volatility drive AFMI HHI dynamics, supporting our theory.

- Show that our nonlinear framework of manger abilities explains and predict AFMI HHI dynamics better than linear frameworks

FILTERING LITERATURE (FINANCIAL ECONOMICS)

Application of filtering techniques in asset pricing, e.g.,

- incomplete information equilibria with unobservable state variables [e.g., Dothan and Feldman (1986), Detemple (1986), Feldman (1989, 1992, 2002, 2003, 2007)],
- fund management [Berk and Green (2004), Berk and Stanton (2007), Dangl, Wu, and Zechner (2008), Brown and Wu (2013, 2016)], Feldman and Xu (2022)

A Model of AFMI Concentration

We introduce a rational equilibrium model to study the dynamics of AFMI concentration. In our model, investors can invest in multiple independent heterogeneous active funds, each with one manager, and in a passive benchmark portfolio.[1] Within a continuous-time framework, we study the active fund managers and investors over a time interval, at times t, $t \in [0, T]$, where T ,$T > 0$ is a constant, allowed to be sufficiently large (i.e., $T \rightarrow \infty$) when we study the steady state in some special cases. Our baseline model uses a linear framework as shown in Section **Error! Reference source not found.** to study the dynamics of AFMI concentration. Then, we extend our framework to a nonlinear one as shown in Section **Error!**

[1] This multiple-fund setting is similar to the one in Brown and Wu (2016).

Reference source not found. to study how the dynamics of stock market volatility affects that of AFMI concentration. Other settings of our model are similar to those in the current literature.[2]

Observable Fund Returns and Unobservable Manager Abilities: Filtering

There are n, $n \geq 2$, active funds in the market, which create returns for investors by investing their wealth in the stock market. 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 $\xi_{i,t}$, $i = 1, \dots, n$. Then, $\mathbf{I}^{-\mathbf{1}}(\boldsymbol{\xi}_{\mathbf{t}})\mathbf{d}\boldsymbol{\xi}_{\mathbf{t}}$ is the $n \times 1$ vector of the instantaneous fund gross rates of return,

[2] Similar to Berk and Green (2004), Brown and Wu (2016), and Feldman and Xu (2022), 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.

where $\mathbf{I}(\boldsymbol{\xi}_\mathbf{t})$ is an $n \times n$ diagonal matrix with $\xi_{i,t}$ as the ith diagonal element.[3] For simplification, we assume that active funds have beta loads of one on the passive benchmark portfolio. To focus on the active funds' returns, similar to Feldman and Xu (2022), 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 $\mathbf{I}^{-1}(\boldsymbol{\xi}_\mathbf{t})\mathbf{d}\boldsymbol{\xi}_\mathbf{t}$ the funds' instantaneous gross alphas, or briefly, gross alphas.

Fund gross alphas depend on the $n \times 1$ vector of 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

[3] 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.

them, briefly, abilities. These 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 $\boldsymbol{\xi}_{\mathbf{s}}$, $0 \leq s \leq t$). We assume a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $\{\mathcal{F}_t\}_{0 \leq t \leq T}$. The $n \times 1$ vectors of independent Wiener processes, $\mathbf{W}_{\mathbf{1,t}}$ and $\mathbf{W}_{\mathbf{2,t}}$, $0 \leq t \leq T$, are adapted to this filtration, where their ith elements are $W_{1i,t}$ and $W_{2i,t}$, $i = 1, \dots, n$, respectively.[4] The unobservable $\boldsymbol{\theta}_{\mathbf{t}}$ and the observable $\boldsymbol{\xi}_{\mathbf{t}}$ evolve as follows:

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

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

[4] 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 $dW_{2i,t}dW_{2j,t} = 0$.

with initial conditions $\boldsymbol{\theta}_{\mathbf{0}}$ and $\boldsymbol{\xi}_{\mathbf{0}}$, respectively. The $n \times 1$ constant vector $\mathbf{a}_{\mathbf{0}}$ has its i th element $a_{i,0}$, $i = 1, \ldots, n$, whereas the $n \times n$ constant diagonal matrices $\mathbf{a}_{\mathbf{1}}$, $\mathbf{b}_{\mathbf{1}}$, $\mathbf{b}_{\mathbf{2}}$, $\mathbf{A}$, and $\mathbf{B}$ have their ith diagonal elements $a_{i,1}$, $b_{i,1}$, $b_{i,2}$, A_i, and B_i, $i = 1, \ldots, n$, respectively. We assume that $A_i > 0$, $i = 1, \ldots, n$ and, without loss of generality, we assume $B_i > 0$, $i = 1, \ldots, n$. 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, $\boldsymbol{\theta}_{\mathbf{t}}$, follow dynamic processes. Second, the fund gross alphas, $\mathbf{I}^{-\mathbf{1}}(\boldsymbol{\xi}_{\mathbf{t}})\mathbf{d}\boldsymbol{\xi}_{\mathbf{t}}$, depend on the managers' abilities and on random shocks. As $A_i > 0$, $i = 1, \ldots, n$, 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 i's gross alpha, which positively corresponds to the variation of fund i's gross alpha.[5] Third, as $\mathbf{a_1}$, $\mathbf{b_1}$, $\mathbf{b_2}$, $\mathbf{A}$, and $\mathbf{B}$ are diagonal matrices, over time a manager's ability and gross alpha are independent of those of other managers.[6,7] Fourth, where $b_{i,2} > 0$ ($b_{i,2} < 0$), the shock $W_{2i,t}$ affects manager i's ability and fund gross alpha, which, consequently, are instantaneously positively (negatively)

---

[5] Notice that for fund i, $i = 1, \dots, n$, the parameter B_i determines the instantaneous variance of $d\xi_{i,t}/\xi_{i,t}$ at time t, as $\mathrm{Var}(d\xi_{i,t}/\xi_{i,t}|\mathcal{F}_t) = B_i^2 dt$, and determines the instantaneous quadratic variation of $d\xi_{i,t}/\xi_{i,t}$ at time t, as $(d\xi_{i,t}/\xi_{i,t})^2 = B_i^2 dt$. Thus, we can regard B_i as a parameter indicating fund i's performance variation.

[6] 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 $(d\xi_{i,t}/\xi_{i,t})(d\xi_{j,t}/\xi_{j,t}) = 0$.

[7] 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 about 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 those 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.

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 $b_{i,1}$ 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., $\boldsymbol{\gamma}_{\mathbf{t}} \triangleq \mathrm{E}\left[(\boldsymbol{\theta}_{\mathbf{t}} - \mathbf{m}_{\mathbf{t}})(\boldsymbol{\theta}_{\mathbf{t}} - \mathbf{m}_{\mathbf{t}})' | \mathcal{F}_t^{\boldsymbol{\xi}}\right]$.

As $\mathbf{m}_{\mathbf{t}}$ is the expected abilities inferred from observable fund returns, hereafter, we briefly call $\mathbf{m}_{\mathbf{t}}$ as inferred abilities. We assume that the conditional distribution of $\boldsymbol{\theta}_{\mathbf{0}}$ given $\boldsymbol{\xi}_{\mathbf{0}}$ (the prior distribution) is Gaussian, $N(\mathbf{m}_{\mathbf{0}}, \boldsymbol{\gamma}_{\mathbf{0}})$, where $\boldsymbol{\gamma}_{\mathbf{0}}$ is a $n \times n$ diagonal matrix, and elements of $\boldsymbol{\xi}_{\mathbf{0}}$, $\mathbf{m}_{\mathbf{0}}$, 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.[8] The techniques are

[8] This type of model is solved in Liptser and Shiryaev (2001a, Ch. 8; 2001b, Ch. 12). More general models with settings similar to those presented by Liptser and Shiryaev (2001a,b) allow model parameters to be functions of the stochastic gross alphas.

called optimal filtering and are used in numerous previous studies.[9] In our case, let $\mathcal{F}_t^{\boldsymbol{\xi}_0,\bar{\mathbf{W}}}$, $0 \leq t \leq T$ be the σ-algebras generated by $\{\boldsymbol{\xi}_0, \ \bar{\mathbf{W}}_s, 0 \leq s \leq t\}$. Then,

$$\bar{\mathbf{W}}_t = \int_0^t \mathbf{B}^{-1}[\mathbf{I}^{-1}(\boldsymbol{\xi}_t)d\boldsymbol{\xi}_t - \mathbf{A}\mathbf{m}_s ds] \tag{3}$$

is an $n \times 1$ 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 $\bar{W}_{i,t}$ and with its initial value $\bar{\mathbf{W}}_0$ being a zero $n \times 1$ vector. The σ-algebras $\mathcal{F}_t^{\boldsymbol{\xi}}$ and $\mathcal{F}_t^{\boldsymbol{\xi}_0,\bar{\mathbf{W}}}$ are equivalent. $\bar{\mathbf{W}}_t$ innovates the inferred abilities $\mathbf{m}_t$. The variables $\mathbf{m}_t$, $\boldsymbol{\xi}_t$, and $\boldsymbol{\gamma}_t$ are the

[9] 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 (2022).

unique, continuous, $\mathcal{F}_t^{\xi}$-measurable solutions of the system of equations

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

$$\mathbf{I}^{-1}(\boldsymbol{\xi}_\mathbf{t})\mathbf{d}\boldsymbol{\xi}_\mathbf{t} = \mathbf{Am_t}dt + \mathbf{Bd\overline{W}_t}, \tag{5}$$

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

where

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

with initial conditions $\boldsymbol{\xi_0}$, $\mathbf{m_0}$, and $\boldsymbol{\gamma_0}$. The random process $(\boldsymbol{\theta_t}, \boldsymbol{\xi_t})$, $0 \leq t \leq T$ is conditionally Gaussian given $\mathcal{F}_t^{\boldsymbol{\xi}}$.[10,11] 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 fund share prices, representing the imprecision of the estimate $m_{i,t}$. We have

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

---

[10] The technical requirements to prove the theorems are regular conditions over the period $0 \leq t \leq T$, such as boundedness of parameter values, integrality of variables, and finite moments of variables. See the requirements of the corresponding theorems in Liptser and Shiryaev (2001a, 2001b). 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.

[11] Notice that the processes $(\boldsymbol{\xi_t}, \overline{\mathbf{W}}_\mathbf{t})$ or, equivalently, $(\boldsymbol{\xi_t}, \mathbf{m_t}, \boldsymbol{\gamma_t})$ provide the same information as $(\boldsymbol{\theta_t}, \boldsymbol{\xi_t})$ over $0 \leq t \leq T$, where $\boldsymbol{\theta_t}$ is unobservable. Hence, investors' original non-Markovian problem can be stated as an equivalent Markovian one, which allows a state vector solution.

where $\sigma_{i,m}(\gamma_{i,t})$, $i = 1, \dots, n$, is the ith diagonal element of $\boldsymbol{\sigma}_{\mathbf{m}}(\boldsymbol{\gamma}_{\mathbf{t}})$ that

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

As $\boldsymbol{\gamma}_{\mathbf{t}}$ and $\boldsymbol{\sigma}_{\mathbf{m}}(\boldsymbol{\gamma}_{\mathbf{t}})$ are diagonal, by Equation (4), $m_{i,t}$ is unaffected by $\overline{W}_{j,t}$ 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.[12]

To make economic sense, we assume a nonnegative $b_{i,2}$, $i = 1, \dots, n$, which induces a positive $\sigma_{i,m}(\gamma_{i,t})$ as shown in Equation **Error! Reference source not found.** (because B_i

[12] If the parameter matrices in Equations (6) and (7) and/or 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 those of other managers. Consequently, a fund's equilibrium size, shown in the next sections, 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.

and A_i are positive).[13] 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. Also, 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.[14]

The above results imply that investors make their optimal decisions in two steps. First, they observe the history of the

[13] 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).

[14] In this linear structure, $\gamma_{i,t}$ has a steady state such that $d\gamma_{i,t} = 0$. $\gamma_{i,t}$ converges to this steady state monotonically. Consequently, $\sigma_{i,m}(\gamma_{i,t})$ also has a steady state to which it converges monotonically.

funds' share prices, $\boldsymbol{\xi}_{\mathbf{t}}$, restructure the state space to consist of only observable processes while maintaining informational equivalence,[15] and generate a posterior distribution of the fund manager abilities. In this way, they convert the problem from a non-Markovian one to an equivalent tractable Markovian one.[16,17] Second, they use their posterior estimate, $\mathbf{m}_{\mathbf{t}}$, to predict the fund gross alphas in the forthcoming future, as shown by Equation (5). They use this prediction in solving their optimization problems.

[15] See Feldman (1992).

[16] Notice that in these optimization processes, the unobservable manager abilities $\boldsymbol{\theta}_{\mathbf{t}}$ is replaced by its observable conditional mean, $\mathbf{m}_{\mathbf{t}}$, updated by a new Wiener process $\bar{\mathbf{W}}_{\mathbf{t}}$, and that $\mathbf{m}_{\mathbf{t}}$ is continuously updated as a function of the dynamic conditional covariance matrix $\boldsymbol{\gamma}_{\mathbf{t}}$. Hence, investors' problems become Markovian, which makes the problems tractable (allowing a state vector solution).

[17] 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.

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 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 these variables. An investor's portfolio return depends on three components: fund gross alphas, management fees, and fund costs. Similar to Berk and Green (2004), Feldman, Saxena, and Xu (2020, 2021), Feldman and Xu (2022), and other related models, we assume the following. Each fund manager chooses the amount of the fund to actively manage at each time t under fixed management fees f_i, $i = 1, \ldots, n$. There are decreasing returns to scale at the fund level. For fund i, $i = 1, \ldots, n$, at time t, fund costs variable $C_i\left(q_{i,t}^a\right)$ is an increasing and convex function of the fund amount that is under active management $q_{i,t}^a$, such that

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

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 asset under management net of fund costs and fees be $S_{i,t}$, $0 \le t \le T$. 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,

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

Similar to Berk and Green (2004) and Feldman and Xu (2022), we assume that risk-neutral investors supply capital with infinite elasticity to funds that have positive expected fund net alphas, driving 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 manager i's profit. Manager i wants to maximize 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

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

As in Berk and Green (2004) and Feldman and Xu (2022), we define $\underline{m}_{i,t}$, $i = 1, \ldots, n$, such that if $m_{i,t} < \underline{m}_{i,t}$, fund i receives no investments from investors and exits the market. Hereafter, we briefly call $\underline{m}_{i,t}$, $i = 1, \ldots, n$ the survival levels.

Here we assume $\underline{m}_{i,t} \geq 0$.[18] 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 $m_{i,t} \geq \underline{m}_{i,t}$; otherwise, they are both zero.

Solving investors' and managers' problems, we obtain the equilibrium optimal solutions for funds surviving in the market[19]

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

[18] The reason is that given updated information, for fund i, the expected instantaneous gross alpha accumulated in dt is $\mathrm{E}\left(d\xi_{i,t}/\xi_{i,t}|\mathcal{F}_t^{\xi}\right) = A_i m_{i,t} dt$, 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.

[19] Similar to Berk and Green (2004) and Feldman and Xu (2022), we assume that managers choose f_i such that the constraint $0 \leq {q_{i,t}^{a}}^{*} \leq q_{i,t}^{*}$, is satisfied for $i = 1, \ldots, n$, so this constraint does not affect the optimization. See the proof of the solutions of the optimization problems in the Internet Appendix.

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

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 parameter 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 Internet Appendix. □

Equilibrium Market Power and Market Structure

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

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 $\left[\left(f_i q^*_{i,t} - 0\right)/f_i q^*_{i,t} = 1\right]$. 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 the 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 as much as possible by maximizing the expected fund net alpha; as the manager's ability to create the fund net alpha is limited, the equilibrium profit is limited.

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 $[= (f_i - 0)/f_i]$.

The above results show that in this framework and those with similar settings commonly used by the literature, there are no dynamics in the common measures of a manager's profitability and market power. Simply calculating these measures does not offer much insight to the dynamics of AFMI. 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 need to focus on the dynamics of its market structure, in particular, the dynamics of AFMI concentration.

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 $q_{i,t}^*$. Based on Equation (20), we have,

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

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

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

Substituting Equations **Error! Reference source not found.** into Equation (23), we have the following result.

Proposition RN1. HHI and Relative Inferred Abilities

In equilibrium, HHI relates to managers' inferred abilities as follows

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

and we can denote $HHI_t^* \triangleq HHI_t^*(\mathbf{m_t})$. □

Proposition RN1 shows that funds' size factors, sensitivities of gross alphas to abilities, and managers' relative inferred abilities together determine HHI. If managers are homogeneous such that these factors are the same for all managers, then funds' sizes are the same and HHI_t^* is constant at its minimum value $1/n$. If managers are heterogeneous such that these parameters are different for different managers, then HHI_t^* can take any value between $1/n$ and its maximum value 1, where AFMI is monopolistic. To offer more insights to the market equilibrium, we focus on the case of heterogeneous managers in this paper. As $\mathbf{m_t}$ is the only variable in Equation (24), HHI_t^* can be regarded as a function driven by $\mathbf{m_t}$.

Notice that Feldman, Saxena, and Xu (2020) (hereafter, FSX), in a one-period model, also derive the endogenous HHI, which is a function of the constant decreasing returns to scale parameters in the fixed-point equilibrium.[20] Our continuous-time model not only derives this result because the constant decreasing returns to scale parameters are captured by the fund size factors in our model, but also suggests that investors' expectations of managers' (relative) abilities are relevant in determining fund sizes, thus HHI. As these expectations are dynamic over time, HHI is also dynamic over time; factors affecting the dynamics of these expectations also affect that of HHI. Thus, our model offers new and important insights into HHI over the FSX model. The following proposition shows

[20] See the Equation (33) in Section 2.4 of FSX.

how the changes of investors' inferences of manager abilities influence the dynamics of HHI.

Proposition RN2. Dynamics of HHI and Changes in Relative Inferred Abilities

HHI evolves as follows

$$dHHI_t^* = \frac{\partial HHI_t^*}{\partial \mathbf{m_t}'}\mathbf{dm_t} + \frac{1}{2}\mathbf{dm}_\mathbf{t}'\frac{\partial^2 HHI_t^*}{\partial \mathbf{m_t}'\partial \mathbf{m_t}}\mathbf{dm_t}$$
$$= \frac{\partial HHI_t^*}{\partial \mathbf{m_t}'}\boldsymbol{\sigma}_\mathbf{m}(\boldsymbol{\gamma}_\mathbf{t})\mathbf{d\overline{W}_t} + \frac{\partial HHI_t^*}{\partial \mathbf{m_t}'}(\mathbf{a_0} + \mathbf{a_1 m_t})dt \quad (25)$$
$$+ \frac{1}{2}\mathbf{trace}\left[\boldsymbol{\sigma}_m'(\boldsymbol{\gamma}_\mathbf{t})\frac{\partial^2 HHI_t^*}{\partial \mathbf{m_t}'\partial \mathbf{m_t}}\boldsymbol{\sigma}_\mathbf{m}(\boldsymbol{\gamma}_\mathbf{t})\right]dt.$$

To facilitate our discussion, we rewrite $dHHI_t^*$ in scalar form:

$$
\begin{aligned}
dHHI_t^* &= \sum_{i=1}^{n}\left[\frac{\partial HHI_t^*}{\partial m_{i,t}} dm_{i,t} + \frac{1}{2}\frac{\partial^2 HHI_t^*}{\partial {m_{i,t}}^2}\left(dm_{i,t}\right)^2\right] \\
&= \sum_{i=1}^{n}\left[\frac{\partial HHI_t^*}{\partial m_{i,t}} \sigma_{i,m}\left(\gamma_{i,t}\right) d\overline{W}_{i,t} + \frac{\partial HHI_t^*}{\partial m_{i,t}}\left(a_{i,0} + a_{i,1} m_{i,t}\right) dt + \frac{1}{2}\frac{\partial^2 HHI_t^*}{\partial {m_{i,t}}^2} \sigma_{i,m}^2 dt\right],
\end{aligned}
\tag{26}
$$

where

$$\frac{\partial HHI_t^*}{\partial m_{i,t}} = 4X_i A_i^2 m_{i,t} \times \frac{q_{i,t}^* \sum_{j=1}^n q_{j,t}^* - \sum_{j=1}^n q_{j,t}^{*2}}{\left(\sum_{j=1}^n q_{j,t}^*\right)^3}, \quad (27)$$

and

$$\frac{\partial^2 HHI_t^*}{\partial m_{i,t}{}^2} = 4X_i A_i^2 \times \left[\frac{3q_{i,t}^* \sum_{j=1}^n q_{j,t}^* + \frac{6q_{i,t}^*\left(\sum_{j=1}^n q_{j,t}^{*2}\right)}{\left(\sum_{j=1}^n q_{j,t}^*\right)} - 8q_{i,t}^{*2} - \sum_{j=1}^n q_{j,t}^{*2}}{\left(\sum_{j=1}^n q_{j,t}^*\right)^3}\right]. \quad (28)$$

Proof. Apply Itô's Lemma on $HHI_t^*(\mathbf{m_t})$ and substitute Equation (4) into the expression, using the property of independence of $\overline{W}_{i,t}$, $i = 1, \dots, n$. □

Proposition RN2 shows how HHI changes with inferred abilities over time. We summarize the key insights directly from Proposition RN2 in the following two corollaries, followed by explanations and intuitions.

Corollary RN2.1. Size of Inferred Ability and Impact on Dynamics of HHI

If $m_{i,t} > \underline{m}_{i,t}$, then we have the following.

a. If $m_{i,t}$ is sufficiently large (small) such that $q_{i,t}^* > \frac{\sum_{j=1}^{n} q_{j,t}^{*2}}{\sum_{j=1}^{n} q_{j,t}^*}$ ($q_{i,t}^* < \frac{\sum_{j=1}^{n} q_{j,t}^{*2}}{\sum_{j=1}^{n} q_{j,t}^*}$), then an increase in $m_{i,t}$ has a positive (negative) impact on $dHHI_t^*$.

b. If $m_{i,t}$ is sufficiently large or sufficiently small such that $3q_{i,t}^* \sum_{j=1}^{n} q_{j,t}^* + \frac{6q_{i,t}^*\left(\sum_{j=1}^{n} q_{j,t}^{*2}\right)}{\left(\sum_{j=1}^{n} q_{j,t}^*\right)} - 8q_{i,t}^{*2} - \sum_{j=1}^{n} q_{j,t}^{*2} < 0$, then HHI_t^* is concave in $m_{i,t}$. Over the next infinitesimal period dt, this concavity has a negative impact on $dHHI_t^*$. If all $m_{i,t}$ for $i = 1, \dots, n$ are sufficiently close to each other, making $q_{i,t}^*$ for $i = 1, \dots, n$ sufficiently close such that

$$3q_{i,t}^{*}\sum_{j=1}^{n}q_{j,t}^{*}+\frac{6q_{i,t}^{*}\left(\sum_{j=1}^{n}q_{j,t}^{*2}\right)}{\left(\sum_{j=1}^{n}q_{j,t}^{*}\right)}-8q_{i,t}^{*2}-\sum_{j=1}^{n}q_{j,t}^{*2}>0 \quad ,$$

then the HHI_t^* is convex in $m_{i,t}$. Then, over dt, this convexity has a positive impact on $dHHI_t^*$. □

To understand Corollary RN2.1a, we observe from Equation (27) that if fund i's inferred ability $m_{i,t}$ is sufficiently large (small) relative to those of other funds, such that $q_{i,t}^* > \frac{\sum_{j=1}^{n} q_{j,t}^{*2}}{\sum_{j=1}^{n} q_{j,t}^{*}}$ ($q_{i,t}^* < \frac{\sum_{j=1}^{n} q_{j,t}^{*2}}{\sum_{j=1}^{n} q_{j,t}^{*}}$), then $\frac{\partial HHI_t^*}{\partial m_{i,t}}$ is positive (negative). Then, as shown in Equation (26), 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 HHI, $dHHI_t^*$.

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, and fund i dominates in the market. A higher inferred ability attracts more investment to fund i, making it larger and making AFMI more concentrated at fund i. 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 those of other funds and then making AFMI less concentrated.

To understand Corollary RN2.1b, consider the second-order partial derivative shown in Equation (28). If manager i's inferred ability $m_{i,t}$ is sufficiently large (small) relative to those of other managers, such that $q_{i,t}^*$ is sufficiently large (small) relative to $q_{j,t}^*$'s for $j \neq i$, then $\frac{\partial^2 HHI_t^*}{\partial {m_{i,t}}^2} < 0$ and HHI_t^* is concave in $m_{i,t}$.[21] Then, over the next infinitesimal period dt, this concavity has a negative impact on $dHHI_t^*$. If all managers'

[21] If $q_{i,t}^*$ is sufficiently small relative to $q_{j,t}^*$'s for $j \neq i$, then the term $-\sum_{j=1}^n q_{j,t}^{*2}$ dominates in the expression $3q_{i,t}^* \sum_{j=1}^n q_{j,t}^* + \frac{6q_{i,t}^*\left(\sum_{j=1}^n q_{j,t}^{*2}\right)}{\left(\sum_{j=1}^n q_{j,t}^*\right)} - 8q_{i,t}^{*2} - \sum_{j=1}^n q_{j,t}^{*2}$, making this expression negative. If $q_{i,t}^*$ is sufficiently large relative to $q_{j,t}^*$'s for $j \neq i$, then $3q_{i,t}^* \sum_{j=1}^n q_{j,t}^* + \frac{6q_{i,t}^*\left(\sum_{j=1}^n q_{j,t}^{*2}\right)}{\left(\sum_{j=1}^n q_{j,t}^*\right)} < 9q_{i,t}^{*2}$ and $-8q_{i,t}^{*2} - \sum_{j=1}^n q_{j,t}^{*2} < -9q_{i,t}^{*2}$, making $3q_{i,t}^* \sum_{j=1}^n q_{j,t}^* + \frac{6q_{i,t}^*\left(\sum_{j=1}^n q_{j,t}^{*2}\right)}{\left(\sum_{j=1}^n q_{j,t}^*\right)} - 8q_{i,t}^{*2} - \sum_{j=1}^n q_{j,t}^{*2} < 9q_{i,t}^{*2} - 9q_{i,t}^{*2} = 0$.

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 $\frac{\partial^2 HHI_t^*}{\partial {m_{i,t}}^2} > 0$ and HHI_t^* is convex in $m_{i,t}$.[22] Then, over the next infinitesimal period dt, this convexity has a positive impact on $dHHI_t^*$.

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 AFMI is concentrated at fund i (at other funds). Although a higher (lower) inferred ability of manager i can make AFMI more concentrated at fund i (at other funds), it becomes more and more difficult to increase the concentration

---

[22] If all funds' sizes are sufficiently close, then the expression is $3q_{i,t}^* \sum_{j=1}^n q_{j,t}^* + \frac{6q_{i,t}^*\left(\sum_{j=1}^n q_{j,t}^{*2}\right)}{\left(\sum_{j=1}^n q_{j,t}^*\right)} - 8q_{i,t}^{*2} - \sum_{j=1}^n q_{j,t}^{*2} \approx (2n-2)q_{i,t}^{*2} > 0$ as $n \geq 2$.

in this way. On the other hand, if all managers' inferred abilities are close, such that funds' sizes are close, then a larger and a smaller inferred ability of manager i both can make fund i's size deviate from other funds' sizes, making AFMI more concentrated. It is easier to make fund i's size deviate from other funds' sizes and to increase HHI if the absolute change in manager i's inferred ability is larger in this case.

For illustration, we simulate HHI over different levels of inferred abilities in the Internet Appendix.

Corollary RN2.2. Interaction Effect of Performance Shock and Performance Variation

If $m_{i,t} > \underline{m}_{i,t}$, then we have the following result. If $m_{i,t}$ is sufficiently large (small) such that $q_{i,t}^* > \frac{\sum_{j=1}^{n} q_{j,t}^{*2}}{\sum_{j=1}^{n} q_{j,t}^*}$ ($q_{i,t}^* < \frac{\sum_{j=1}^{n} q_{j,t}^{*2}}{\sum_{j=1}^{n} q_{j,t}^*}$), then a positive $d\overline{W}_{i,t}$ exerts a positive (negative) impact on $dHHI_t^*$, and a higher B_i mitigates this positive (negative) impact. □

Corollary RN2.2 shows that the interaction effect of $d\overline{W}_{i,t}$ and B_i on $dHHI_t^*$ is negative (positive) if $m_{i,t}$ is sufficiently large (small) relative to $m_{j,t}$'s for $j \neq i$. This is because $\sigma_{i,m}(\gamma_{i,t}) > 0$, and a higher B_i decreases $\sigma_{i,m}(\gamma_{i,t})$, as shown in Equation **Error! Reference source not found.**. Also, a higher B_i does not affect $\frac{\partial HHI_t^*}{\partial m_{i,t}}$, as implied by Equation (27). Thus, a higher B_i decreases the absolute value of $\frac{\partial HHI_t^*}{\partial m_{i,t}}\sigma_{i,m}(\gamma_{i,t})$, which is the coefficient of $d\overline{W}_{i,t}$ in the expression of $dHHI_t^*$, as shown in Equation (26). If $m_{i,t}$ is sufficiently large (small) relative to $m_{j,t}$'s for $j \neq i$, then $\frac{\partial HHI_t^*}{\partial m_{i,t}}$ and thus $\frac{\partial HHI_t^*}{\partial m_{i,t}}\sigma_{i,m}(\gamma_{i,t})$ are positive (negative). Then, a smaller absolute value of $\frac{\partial HHI_t^*}{\partial m_{i,t}}\sigma_{i,m}(\gamma_{i,t})$ induced by a higher B_i makes $\frac{\partial HHI_t^*}{\partial m_{i,t}}\sigma_{i,m}(\gamma_{i,t})$ smaller (larger).

The intuition of the above result is as follows. A positive shock in manager i's performance induces higher manager i's inferred ability thus higher fund i's size. If this manager's inferred ability is sufficiently large (small) relative to those of other managers, a higher manager i's inferred ability increases (decreases) HHI, as mentioned in the earlier discussion. In this case, this positive performance shock increases (decreases) HHI. Moreover, if manager i's performance variation is higher, then investors allocate smaller weights on manager i's performance shocks when learning about her ability. Consequently, a positive shock in manager i's performance induces smaller impact on her inferred ability, and thus induces a positive (negative) impact with a smaller absolute value on HHI.

Equilibrium AFMI Concentration and Stock Market Volatility: Extension to a Nonlinear Framework

We analyze how stock market volatility affects manager abilities and then AFMI concentration by extending our linear framework shown in Equations (1) and (2) to a nonlinear one. Higher stock market volatility increases market stress and redemption risk. The consequential higher redemption from investors and the need of larger cash buffers to manage the higher redemption risk impede managers when implementing investment strategies to produce abnormal returns, making fund gross alphas less related to manager abilities and more related to luck.[23] Thus, we assume that sensitivities of gross alphas to manager abilities is a decreasing function of stock

[23] See, for example, the discussion of how market stress affects fund performance in Jin, Kacperczyk, Kahraman, and Suntheim (2022).

market volatility. Let λ_t be a variable that captures the impact of stock market volatility on the sensitivities of gross alphas to manager abilities, i.e., $A_i \triangleq A_i(\lambda_t)$ and $\frac{\partial A_i(\lambda_t)}{\partial \lambda_t} < 0$, $i = 1, \dots, n$, following

$$d\lambda_t = \mu_\lambda dt + \sigma_\lambda dz_t. \tag{29}$$

While, in general, μ_λ and σ_λ could be functions of λ_t and other market variables,[24] for brevity and simplicity, we assume here that μ_λ and σ_λ are constant, and that z_t is a Brownian motion adapted to $\{\mathcal{F}_t\}_{0 \le t \le T}$ and independent of $\mathbf{W}_{1,t}$ and $\mathbf{W}_{2,t}$. Using the analysis in Section **Error! Reference source not found.**, we derive the dynamics of HHI in the following proposition.[25]

[24] For example, λ_t could follow an autoregressive process.

[25] Learning about manager abilities is unaffected by λ_t because λ_t is unaffected by unobservable manager abilities, $\boldsymbol{\theta}_t$, and z_t is independent of $\mathbf{W}_{1,t}$ and $\mathbf{W}_{2,t}$. Consequently, λ_t is independent of $m_{i,t}$, $i = 1, \dots, n$.

Proposition RNV. Dynamics of HHI and Changes in Relative Inferred Abilities

HHI evolves as follows (in scalar form):

$$dHHI_t^* = dX_t + \left(\sum_{i=1}^{n} \frac{\partial HHI_t^*}{\partial A_i(\lambda_t)} \frac{\partial A_i(\lambda_t)}{\partial \lambda_t}\right) d\lambda_t + \frac{1}{2}\sum_{i=1}^{n} \left[\frac{\partial^2 HHI_t^*}{\partial A_i(\lambda_t)^2}\left(\frac{\partial A_i(\lambda_t)}{\partial \lambda_t}\right)^2 + \frac{\partial HHI_t^*}{\partial A_i(\lambda_t)} \frac{\partial^2 A_i(\lambda_t)}{\partial {\lambda_t}^2}\right] \sigma_\lambda^2 dt, \tag{30}$$

where dX_t equals the $dHHI_t^*$ in Equation (26) with A_i replaced by $A_i(\lambda_t)$,

$$\frac{\partial HHI_t^*}{\partial A_i(\lambda_t)} = 4X_i m_{i,t}^2 A_i(\lambda_t) \times \frac{q_{i,t}^* \sum_{j=1}^n q_{j,t}^* - \sum_{j=1}^n q_{j,t}^{*2}}{\left(\sum_{j=1}^n q_{j,t}^*\right)^3}, \tag{31}$$

and

$$\frac{\partial^2 HHI_t^*}{\partial A_i(\lambda_t)^2} = 4X_i m_{i,t}^2 \times \tag{32}$$

$$\left[\frac{3q_{i,t}^{*}\sum_{j=1}^{n}q_{j,t}^{*}+\frac{6q_{i,t}^{*}\left(\sum_{j=1}^{n}q_{j,t}^{*2}\right)}{\left(\sum_{j=1}^{n}q_{j,t}^{*}\right)}-8q_{i,t}^{*2}-\sum_{j=1}^{n}q_{j,t}^{*2}}{\left(\sum_{j=1}^{n}q_{j,t}^{*}\right)^{3}}\right].$$

Proof. Apply Itô's Lemma on $HHI_t^*(\mathbf{m_t},\lambda_t)$, using the property that λ_t is independent of $m_{i,t}$, $i = 1, \dots, n$.□

Proposition RNV shows how the dynamics of stock market volatility affects that of HHI. A higher stock market volatility, λ_t, decreases sensitivities of gross alphas to manager abilities, $A_i(\lambda_t)$, as $\frac{\partial A_i(\lambda_t)}{\partial \lambda_t} < 0$, $i = 1, \dots n$. Given the same inferred manager abilities, it decreases fund expected gross alphas; this consequently decreases all funds' equilibrium sizes. If fund i's inferred ability $m_{i,t}$ is sufficiently large (small) relative to those of other funds, such that $q^*_{i,t} > \frac{\sum_{j=1}^{n} q^{*2}_{j,t}}{\sum_{j=1}^{n} q^*_{j,t}}$ ($q^*_{i,t} < \frac{\sum_{j=1}^{n} q^{*2}_{j,t}}{\sum_{j=1}^{n} q^*_{j,t}}$), then the decrease in fund i's size exerts a negative (positive) impact on HHI, as shown in the earlier discussions. In this case, we have $\frac{\partial HHI^*_t}{\partial A_i(\lambda_t)} > 0$ ($\frac{\partial HHI^*_t}{\partial A_i(\lambda_t)} < 0$) and, consequently, $\frac{\partial HHI^*_t}{\partial A_i(\lambda_t)} \frac{\partial A_i(\lambda_t)}{\partial \lambda_t} < 0$ ($\frac{\partial HHI^*_t}{\partial A_i(\lambda_t)} \frac{\partial A_i(\lambda_t)}{\partial \lambda_t} > 0$). Then, whether HHI increases (decreases) with λ_t depends on whether the aggregate effect of λ_t , $\sum_{i=1}^{n} \frac{\partial HHI^*_t}{\partial A_i(\lambda_t)} \frac{\partial A_i(\lambda_t)}{\partial \lambda_t}$, is positive (negative). From Equation **Error! Reference source not found.**, we can see that if fund i is extremely large relative to other funds, due to its large X_i, $A_i(\lambda_t)$, and/or $m_{i,t}$, then $\frac{\partial HHI_t^*}{\partial A_i(\lambda_t)}$ is positive with a large absolute value, which would drive the value of $\sum_{i=1}^{n} \frac{\partial HHI_t^*}{\partial A_i(\lambda_t)} \frac{\partial A_i(\lambda_t)}{\partial \lambda_t}$ when the magnitude of $\frac{\partial A_i(\lambda_t)}{\partial \lambda_t}$ is similar to those of other funds.[26] As $\frac{\partial A_i(\lambda_t)}{\partial \lambda_t} < 0$, we would have a negative $\sum_{i=1}^{n} \frac{\partial HHI_t^*}{\partial A_i(\lambda_t)} \frac{\partial A_i(\lambda_t)}{\partial \lambda_t}$ when some extremely large funds exist in AFMI. In other words, when the distribution of funds' sizes is highly skewed to the right (which is the case in reality[27]), the effect of the decrease in extremely large funds'

[26] As the change in stock market volatility affects active equity funds in a similar way, it is likely that the magnitudes of $\frac{\partial A_i(\lambda_t)}{\partial \lambda_t}$, $i = 1, \ldots, n$ are close to each other.

[27] In our empirical analysis in Section **Error! Reference source not found.**, we also show that the distribution of funds' sizes is highly skewed to the right in our sample.

sizes due to an increase in stock market volatility dominates those of small funds, inducing a lower HHI.

A nonlinear frame allowing coefficients of processes of manager abilities and gross alphas to be functions of other observable economic factors can also model how the dynamics of these factors affect that of HHI. Linear frameworks of manager abilities and gross alphas that are used in the current literature[28] cannot directly incorporate the effects of economic factors on manager abilities and gross alphas and, consequently, cannot easily model these effects on the dynamics of HHI as we do here. We study only the effect of the stock market volatility on HHI in this section; the effects of other economic factors are left for future research.

[28] See, for example, Berk and Green (2004) and their followers.

Constant Manager Abilities and HHI

We illustrate a special case of HHI in which manager abilities are constant under the linear framework shown in Section **Error! Reference source not found.**. In this case, $\mathbf{a_0}$ is an $n \times 1$ zero vector and $\mathbf{a_1}$, $\mathbf{b_1}$, and $\mathbf{b_2}$ are $n \times n$ zero matrices, making $\mathbf{d\theta_t}$ a zero vector. We have

$$\mathbf{dm_t} = \boldsymbol{\sigma}_\mathbf{m}(\boldsymbol{\gamma}_\mathbf{t})\mathbf{d\overline{W}_t}, \tag{33}$$

$$\boldsymbol{\sigma}_\mathbf{m}(\boldsymbol{\gamma}_\mathbf{t}) \triangleq (\mathbf{A}\boldsymbol{\gamma}_\mathbf{t})'\mathbf{B}^{-1}, \tag{34}$$

$$\boldsymbol{\gamma}_\mathbf{t} = [\mathbf{I} + \boldsymbol{\gamma}_\mathbf{0}\mathbf{A}\mathbf{B}^{-2}\mathbf{A}t]^{-1}\boldsymbol{\gamma}_\mathbf{0}, \tag{35}$$

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, $i = 1, \ldots, n$, we have that the imprecision of the estimate $m_{i,t}$, $\gamma_{i,t} = \frac{\gamma_{i,0} B_i^2}{B_i^2 + A_i^2 \gamma_{i,0} t}$ decreases to zero over time monotonically, so the sensitivity of inferred ability to performance shocks, $\sigma_{i,m}(\gamma_{i,t}) \triangleq (A_i \gamma_{i,t})/B_i$, also decreases to zero monotonically. Thus, we have the following proposition.

Proposition CA. Constant Manager Abilities and Steady State of HHI

If $\boldsymbol{\theta}_\mathbf{t}$ is a constant vector and $m_{i,t} > \underline{m}_{i,t}$ for $i = 1, \dots, n$, then over time, $\gamma_{i,t}$ and $\sigma_{i,m}(\gamma_{i,t})$ decrease monotonically to zero. As $t \to \infty$, for $i = 1, \dots, n$, $dm_{i,t} = \sigma_{i,m}(\gamma_{i,t}) d\overline{W}_{i,t} \to 0$ and $m_{i,t}$, becomes constant, making HHI_t^* a constant. □

Proposition CA shows the steady state of this constant-ability framework. The intuition is that as managers' abilities are unobservable constants, estimation precisions improve monotonically over time, inducing inferred abilities to be increasingly less sensitive to funds' gross alpha realizations. As time goes to infinity, people learn managers' abilities, thus do not change their estimates. Then, investors stop changing their investments flows to funds (i.e., fund sizes stay unchanged), making HHI stay unchanged. As empirical HHI does not converge to a constant in the long term, as shown in Feldman, Saxena, and Xu (2020, 2021) and our following empirical section, theoretical models with this framework[29] lack the explanatory and predictive power of HHI dynamics.

[29] See, for example, Berk and Green (2004), Choi, Kahraman, and Mukherjee (2016), and Brown and Wu (2016).

For illustration, we simulate HHI in the cases of constant ability and of dynamic ability in the Internet Appendix.

Mean-Variance Risk-Averse Investors and HHI

To study the effect of investors' risk aversion on HHI, we use our linear framework shown in Section **Error! Reference source not found.** and 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 (2022).

As risk-averse investors trade off risk and return, we need to redefine our model. First, we cannot normalize the passive

[30] See the discussions of mean-variance risk-averse investors in Feldman and Xu (2022).

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, as η_t. We assume that the passive benchmark portfolio return $d\eta_t/\eta_t$ follows.

$$\frac{d\eta_t}{\eta_t} = \mu_p dt + \sigma_p dW_{p,t}, \tag{36}$$

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

[31] As mean-variance 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 (46), which collapses if the passive benchmark return is normalized to zero.]

portfolio, the fund gross return is $d\xi_{i,t}/\xi_{i,t} + d\eta_t/\eta_t$ and the fund net return is $dS_{i,t}/S_{i,t} + d\eta_t/\eta_t$. We assume that the risk source of the benchmark return, $W_{p,t}$, is independent of that of gross alphas, so

$$dW_{p,t}d\overline{W}_{i,t} = 0,\ \forall t, i = 1, \dots, n. \tag{37}$$

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:

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

$$\max_{w_t} \frac{\mathrm{E}\left[\left.\frac{dp_t}{p_t}\right| \mathcal{F}_t^{\xi}\right]}{\sqrt{\mathrm{Var}\left[\left.\frac{dp_t}{p_t}\right| \mathcal{F}_t^{\xi}\right]}} \tag{38}$$

subject to

$$\mathbf{v_t}'\mathbf{1} = 1, \tag{39}$$

$$0 \leq v_{i,t} \leq 1,\ \forall\, i = 1, \dots, n+1, \tag{40}$$

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 (40) 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 $(n+1) \times 1$ net return vector of these $n+1$ assets, with the ith element $i = 1, \dots, n$.

$$R_{i,t} = \frac{dS_{i,t}}{S_{i,t}} + \frac{d\eta_t}{\eta_t}$$

$$= \left(\frac{q_{i,t}^a}{q_{i,t}} A_i m_{i,t} - \frac{c_i {q_{i,t}^a}^2}{q_{i,t}} - f_i + \mu_p \right) dt \tag{41}$$

$$+ \frac{q_{i,t}^a}{q_{i,t}} B_i d\overline{W}_{i,t} + \sigma_p dW_{p,t}$$

and

$$R_{n+1,t} = \frac{d\eta_t}{\eta_t} = \mu_p dt + \sigma_p dW_{p,t}. \tag{42}$$

Then, the investor's portfolio net return is

$$\frac{dp_t}{p_t} = \mathbf{v_t}'\mathbf{R_t}. \tag{43}$$

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 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 HHI, we assume that V is constant and exogenous to both investors and managers.[33] Then, the amount of wealth allocated to fund i, i.e., fund i's size, is $q_{i,t}^{*} = v_{i,t}^{*} V, i = 1, \dots, n$.

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

$$\max_{q_{i,t}^{a}} f_i q_{i,t} = \max_{q_{i,t}^{a}} g_i\left(q_{i,t}^{a}\right) \tag{44}$$

subject to

$$0 \leq q_{i,t}^{a} \leq q_{i,t}, \ \forall\, i = 1, \dots, n. \tag{45}$$

[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 HHI. To simplify our model, we do not introduce these complexities of this wealth value.

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

$$q_{i,t}^{*} = \frac{\left(A_i m_{i,t}\right)^2}{4 f_i \left(\frac{B_i^2 \mu_p}{V \sigma_p^2} + c_i\right)}. \tag{46}$$

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

$$X_i^{RA} \triangleq \frac{1}{4 f_i \left(\frac{B_i^2 \mu_p}{V \sigma_p^2} + c_i\right)}. \tag{47}$$

[34] We assume that managers choose f_i such that the constraint $0 \leq q_{i,t}^{a\ *} \leq q_{i,t}^{*}$, 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 $0 \leq v_{i,t}^{*} \leq 1$, is satisfied for $i = 1, \ldots, n+1$, so this constraint does not affect the investors' optimization processes. See the proof in the Internet Appendix.

Similar to the results of X_i, a larger parameter of decreasing returns to scale, 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 fund i 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 fund i if they have more wealth V to invest, and switch from fund i to the passive benchmark if the 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. In other words, compared to AFMI with risk-neutral investors, AFMI with mean-variance risk-averse investors has smaller equilibrium fund sizes. This is because investors' risk considerations reduce their investment to risky active funds. Using this new definition of fund i's size factor, we have

$$q_{i,t}^{*} = \left(A_i m_{i,t}\right)^2 X_i^{RA}. \tag{48}$$

Proof. See the Internet Appendix. □

We substitute $q_{i,t}^{*}$ shown above into the formula of HHI_t^{*} and derive the following results,

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

where $\mathbf{X^{RA}}$ is an $n \times 1$ vector with the ith element as X_i^{RA}. We can see that the form of HHI_t^* in (49) is the same as the one in (24) in the case of risk-neutral investors. The only difference is that here we use $\mathbf{X^{RA}}$ instead of $\mathbf{X}$ as the size factors. Thus, the relation of the dynamics of HHI_t^* and managers' inferred abilities in Proposition RN1 still holds; consequently, the results of Proposition RN2 and Corollaries RN2.1 and RN2.2 hold. Also, if we allow A_i to be a decreasing function of stock market volatility, λ_t, as we do in Section **Error! Reference source not found.**, then the results of Proposition RNV 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 these sizes does not depend on investors' risk considerations. Thus, the dynamics of HHI_t^* relates to managers' relative inferred abilities in a way similar to that of the risk-neutral case.

The following proposition summarizes the results in this section.

Proposition RA. HHI and Mean-Variance Risk-Averse Investors

When investors are mean-variance risk averse, $q_{i,t}^*$, $i = 1, \ldots, n$, are smaller than those when investors are risk neutral, and funds' size factors X_i^{RA}, $i = 1, \ldots, n$, not only decrease with c_i and f_i, but also increase with V and σ_p^2 and decrease with B_i^2 and μ_p. Besides the size factors, the other results of Propositions RN1 and RN2, Corollaries RN2.1 and RN2.2, and Proposition RNV still hold. □

Fund Entrances and Exits and 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 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 (46) 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,[35] managers with negative inferred abilities optimally choose to put zero assets under active management

[35] Managers can short sell some stocks when constructing a portfolio to pursue alphas, but they cannot short the whole portfolio or short the "active management amount", as shown by the constraint $q_{i,t}^{a} \geq 0$ for any i.

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 again that equilibrium fund sizes, $q^*_{i,t}$, are functions of managers' inferred abilities, $m_{i,t}$. 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^{*}. Then, in these two cases, $dHHI_t^{*}$ can still be expressed by Equation (25), and the results from Section **Error! Reference source not found.** to Section **Error! Reference source not found.** are still valid. In other words, funds' entrances and exits do not affect $dHHI_t^{*}$ immediately, but they change the set of funds in AFMI and affect $dHHI_t^{*}$ after that.

However, if $\underline{m}_{i,t} > 0$ for any $i = 1, \dots, n_t$, then fund i'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 value, 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, also creating a jump in HHI_t^{*}. 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 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.

Empirical Study

Based on Corollaries RN2.1 and RN2.2, we have the following two predictions, respectively. For funds that are sufficiently large (small) relative to others,

a. increase in these funds' performances relative to those of other funds exerts positive (negative) impacts on HHI;
b. higher performance variations in these funds mitigate these positive (negative) impacts on HHI such that the interaction

effects of shocks in relative performance and performance variations are negative (positive) in these funds.

Also, based on Proposition RNV, we have the following prediction.

c. When the distribution of funds' sizes is highly skewed to the right, an increase in stock market volatility decreases HHI.

We test the above predictions empirically.

Methodology

We first develop the measures of fund performance and performance variation. We estimate fund performance using empirical asset pricing models in the current literature, such as the five-factor model developed by Fama and French (2015) (hereafter, FF5) and the four-factor model developed by Fama and French (1993) and Carhart (1997) (hereafter, FFC4). For each fund i, we estimate the following:

$$r_{i,t} = \sum_{j=1}^{M} \beta_{i,j} F_{j,t} + \varepsilon_{i,t}, \tag{50}$$

where $r_{i,t}$ is fund i's net return in excess of risk-free return, $F_{j,t}$ is the return of factor j, $\beta_{i,j}$ is the factor loading of fund i to factor j, M is the number of factors, and $\varepsilon_{i,t}$ is the residual. This model is estimated on a rolling-window basis.

Our first measure of fund performance variation is the $1 - R^2$ of the regression model calculated as $\frac{\sum_t (\hat{\varepsilon}_{i,t})^2}{\sum_t (r_{i,t} - \bar{r}_i)^2}$, where $\hat{\varepsilon}_{i,t}$ is the estimated residual and $\bar{r}_i$ is the average excess return of fund i over the rolling window period. Notice that $\hat{\varepsilon}_{i,t} = r_{i,t} - \sum_{j=1}^{M} \hat{\beta}_{i,j} F_{j,t}$, where $\hat{\beta}_{i,j}$ is the estimate of factor loading to factor j and $\hat{\varepsilon}_{i,t}$ can be regarded as the in-sample estimate of abnormal net return, or net alpha. Consequently, $1 - R^2$ can be regarded as the in-sample estimate of fund performance variation (normalized by total variation of the excess return). Amihud and Goyenko (2013) also find that the measure $1 - R^2$ in such regression models is highly related to fund performance. Similar to Amihud and Goyenko (2013), we use a 24-month rolling window to estimate the models for each fund i, and we denote the $1 - R^2$ estimated by the previous 24-month period (from $t - 1$ to $t - 24$) as $OMR2_{i,t-1}$.

We estimate the (out-of-sample) fund net alpha at time t as the $NetAlpha_{i,t} = r_{i,t} - \sum_{j=1}^{M} \hat{\beta}_{i,j} F_{j,t}$, where $\hat{\beta}_{i,j}$ is estimated using the observations in the previous 24 months. Our second measure of fund performance variation is the standard deviation of the net alphas in the previous 12 months, denoted as $NetAlpha\_Std_{i,t-1}$. For robustness, we also calculate the fund gross alpha as the fund net alpha plus the fund annual expense ratio divided by 12, and then calculate the standard deviation of this gross alpha in the previous 12 months as a measure of fund performance variation, denoted as $GrossAlpha\_Std_{i,t-1}$. These two measures of fund performance variation are the same as the performance volatility measures used by Huang, Wei, and Yan (2021).

We next choose the option-implied volatility index (VIX) as our measure of stock market volatility. VIX not only measures stock market volatility but also captures investors' expectation of such volatility, so current literature commonly uses VIX to measure market stress and panic.[36] Thus, we expect that at a higher VIX level, the stock market is more volatile and stressful, impeding fund managers to implement their investment strategies and consequently reducing the sensitivities of gross alphas to manager abilities.

Fund-Level Analysis: Effectiveness of Our Measures of Stock Market Volatility and Performance Variation

Feldman and Xu (2022) show that the equilibrium fund flow–net alpha sensitivity decreases with performance variation and

[36] See, for example, Jin, Kacperczyk, Kahraman, and Suntheim (2022).

increases with the sensitivity of gross alpha to manager ability.[37] If our measure of stock market volatility is effective, then an increase in this measure should decrease the flow–net alpha sensitivity because it decreases the sensitivity of gross alpha to manager ability; if our measures of performance variation are effective, then higher values in these measures should decrease the flow–net alpha sensitivity. We test the effectiveness of our measures using the following model:

[37] In our model, we can also easily show that the equilibrium fund flow–net alpha is $\frac{dq_{i,t}^*}{q_{i,t}^*} = \frac{A_i(\lambda_t)\sigma_{i,m}(\gamma_{i,t})}{f_i B_i}\left(\frac{dS_{i,t}}{S_{i,t}}\right) + \frac{A_i^2(\lambda_t)\sigma_{i,m}^2(\gamma_{i,t})}{4f_i^2 B_i^2}\left(\frac{dS_{i,t}}{S_{i,t}}\right)^2 + 2\left[\frac{a_{i,0}}{m_{i,t}} + a_{i,1}\right]dt$, $i = 1, \dots, n$, by applying Itô's Lemma on $q_{i,t}^*$ to calculate $dq_{i,t}^*$ and then divide it by $q_{i,t}^*$. Then, the flow–net alpha sensitivity decreases with B_i and increases with $A_i(\lambda_t)$.

$$
\begin{aligned}
&Flow_{i,t} \\
&= \delta_0 + \delta_1 NetAlpha_{i,t-1} \\
&+ \delta_2 NetAlpha_{i,t-1} \times VIX_{t-1} \\
&+ \delta_3 VIX_{t-1} \\
&+ \delta_4 NetAlpha_{i,t-1} \times Perf\_Var_{i,t-1} \\
&+ \delta_5 Perf\_Var_{i,t-1} + \delta Controls_{i,t-1} \\
&+ \phi_t + \upsilon_i + \varepsilon_{i,t},
\end{aligned} \tag{51}
$$

where $Flow_{i,t}$ is the fund percentage flow calculated as the difference between the monthly growth rate of the fund's total net asset under management (TNA) and the fund's monthly net return, VIX_t is the VIX value, and $Perf\_Var_{i,t}$ is a measure of fund performance variation, which is $OMR2_{i,t}$, $NetAlpha\_Std_{i,t}$, or $GrossAlpha\_Std_{i,t}$. We follow the literature [38] to choose control variables in the vector $Controls_{i,t-1}$, which include the lagged values of the natural logarithm of the fund size ($\ln Size_{i,t-1}$); the natural logarithm of fund age ($lnAge_{i,t-1}$); fund expense ratio ($Expense_{i,t-1}$); fund turnover ratio ($Turnover_{i,t-1}$); the weighted average flow of the fund class based on the Lipper fund classification; i.e., the style flow, ($StyleFlow_{i,t-1}$); fund flow ($Flow_{i,t-1}$); fund family net alpha ($FamAlpha_{i,t-1}$); and the natural logarithm of fund family size ($\ln FamSize_{i,t-1}$). Variables ϕ_t and υ_i represent year effects and fund effects, respectively. Detailed definitions and constructions of these variables are shown in the Data Appendix. When analyzing the flow–net

[38] See, for example, Brown and Wu (2016), Franzoni and Schmalz (2017), Harvey and Liu (2019), Jiang, Starks, and Sun (2021), Huang, Wei, and Yan (2021), and Feldman and Xu (2022).

alpha relations, we also include the interaction terms of $\ln Size_{i,t-1}$ and $lnAge_{i,t-1}$ with $NetAlpha_{i,t-1}$ because the current literature shows that the flow–net alpha sensitivity is affected by fund size [Brown and Wu (2016)] and fund age [Feldman and Xu (2022)]. To account for potential time-series and cross-sectional correlations in residuals, we cluster the standard error by year and by fund.

If our measure of stock market volatility is effective, we should find that δ_2 is significantly negative; if our three measures of fund performance variation are effective, we should find that δ_4 is significantly negative.

Market-Level Analysis: Dynamics of HHI and Changes in Stock Market Volatility, Fund Performances, and Performance Variations

We test our three theoretical predictions using our measures of stock market volatility and fund performance variation. We measure the changes in funds' performances relative to these changes in other funds by the changes in these funds' market shares. The reason is that a fund's equilibrium size is a positive function of the fund manager's inferred ability shown in our theoretical model, and then market share, which is a fund's size relative to the sum of all fund sizes, indicates a fund manager's inferred ability relative to other managers. Consequently, change in a fund's market share indicates change in relative inferred ability due to the change in the fund's performance relative to that of other funds.

Then, we test the following model.

$$\begin{aligned}
&dif_HHI_t \\
&= \delta_0 + \delta_1 dif_VIX_{t-1} \\
&+ \delta_2 dif_MarketShare^B_{t-1} \\
&+ \delta_3 dif_MarketShare^S_{t-1} \\
&+ \delta_4 dif_MarketShare^B_{t-1} \\
&\times Perf_Var^B_{t-1} \\
&+ \delta_5 dif_MarketShare^S_{t-1} \\
&\times Perf_Var^S_{t-1} + \delta_6 Perf_Var^B_{t-1} \\
&+ \delta_7 Perf_Var^S_{t-1} + \delta_8 NumGrowth_{t-1} \\
&+ \phi_t + \varepsilon_t,
\end{aligned} \tag{52}$$

where dif_HHI_t is the change in HHI from time $t-1$ to t and dif_VIX_{t-1} is the change in VIX from time $t-2$ to $t-1$. The superscripts B and S denote the large-fund group and small-fund group, respectively. We define the large-fund group as the largest five funds (based on fund TNA values) and the small-fund group as the funds with fund TNA values from the fifth percentile to the tenth percentile because these funds are likely to be sufficiently large and sufficiently small, respectively, relative to other funds.[39] We redefine the large-fund group and small-fund group in each month. The explanatory variable $dif\_MarketShare_{t-1}^{B}$ ($dif\_MarketShare_{t-1}^{S}$) is the change in market share of the

[39] Because the performances and sizes of funds with fund size values from the lowest five percentiles are very volatile and contain much noise, we choose the funds with fund size values from the fifth percentile to the tenth percentile to construct the small-fund group. We do robustness checks with different classifications of the large-fund group and small-fund group, as shown in the following discussion of the empirical study.

large-fund group (small-fund group) from time $t-2$ to $t-1$. Also, $Perf\_Var_{t-1}^{B}$ ($Perf\_Var_{t-1}^{S}$) is the weighted average of the measure of performance variation within the large-fund group (small-fund group) at time $t-1$, using funds' TNAs at this time as weights. We also include $NumGrowth_{t-1}$ as a control variable, which is the change in the number of funds in the market from time $t-2$ to $t-1$, divided by the number of funds at $t-2$. To account for potential serial correlation in residuals, we use Newey-West estimates of standard error with the maximum lag of 12 to be considered in the autocorrelation structure.

In the above model, without including explanatory variables $Perf\_Var_{t-1}^{B}$, $Perf\_Var_{t-1}^{S}$, $dif\_MarketShare_{t-1}^{B} \times$

$Perf\_Var_{t-1}^{B}$, and $dif\_MarketShare_{t-1}^{S} \times Perf\_Var_{t-1}^{S}$, we expect δ_1 to be negative when the distribution of funds' sizes is highly skewed to the right because, in this case, higher stock market volatility should induce negative impact on HHI; we expect δ_2 (δ_3) to be positive (negative) because shocks in the relative performance of the large-fund group (small-fund group), measured by the changes in the market share, should induce a positive (negative) impact on HHI. When including these four explanatory variables in this model, we expect δ_4 (δ_5) to be negative (positive) because performance variation of the large-fund group (small-fund group) should mitigate the positive (negative) impact of shocks in the relative performance of this group on HHI.

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 January 1990 to December 2020, and we use monthly data.[40] We exclude index funds, variable annuity funds, and exchange-traded funds (ETFs), and then choose U.S. domestic equity-only mutual funds by using the Lipper fund classification.[41] This equity fund filter is similar to many of the current empirical studies such as those of Amihud and Goyenko (2013), Brown and Wu

[40] Information on the Lipper fund classification and most of the information on the management company code to identify fund families begins in December of 1999. As we use a 24-month rolling window to estimate fund net alpha, and we need 12 months to estimate alpha standard deviation, our test period starts from January 1993.

[41] 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.

(2016), Choi, Kahraman, and Mukherjee (2016), Huang, Wei, and Yan (2021), and Feldman and Xu (2022).

We use the MFLINKS database to aggregate fund share class-level information to fund-level information. In particular, we calculate a fund's TNA by summing up its share classes' TNA and calculate fund size as fund TNA normalized to the December 2020 dollar value[42]. We calculate a fund-level variable's value as the weighted average of share class-level values using share classes' TNAs as weights. Fund family is identified by the management company code,[43] and we use

[42] We divide a fund's TNA by the total market capitalization of the U.S. equity market in that month, and then multiply it by the total market capitalization of the U.S. equity market in December 2020. The U.S. equity market information is offered by the CRSP US stock database, and we calculate the total market capitalization using only ordinary common shares, with the share type code in CRSP equal to 10 and 11.

[43] 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.

funds' TNAs as weights in calculating fund family performance.

To estimate the FFC4 model, we collect the risk-free rate and the corresponding factors from the Fama-French database in Wharton Research Data Services (WRDS). To estimate the FF5 model, we collect the factors from the Fama-French website.[44] We collect daily observations of VIX from WRDS and calculate the average value of VIX in each month to develop the monthly VIX values.

When conducting our market-level analysis on the dynamics of HHI, we include the observations of fund net alpha and $1 - R^2$ of the empirical asset pricing model in our sample only if observations of fund net returns are available and fund TNA is

[44] The website address is https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html, accessed on July 19, 2022.

positive in all of the previous 24 months (i.e., the estimation window). We include the observations of fund net alpha (gross alpha) standard deviation only if observations of fund net alphas (gross alphas) are available in all the previous 12 months. We also exclude fund observations if the fund's size (in the December 2020 dollar value) is below 15 million. In doing the fund-level analysis on the effects of stock market volatility and performance variation on the flow–net alpha sensitivity, we further require a fund to have at least 24 months' observations of all the variables in Equation **Error! Reference source not found.**.[45] We also winsorize all the fund-level variables at the 1% and 99% levels when doing this analysis.[46]

---

[45] We also require a fund family to have at least two funds so that the fund family-level variables are meaningful.
[46] We winsorize the variables except VIX_{t-1}, which is a market-level variable.

The above criteria and process are similar to those in the fund management literature, such as Amihud and Goyenko (2013). We have 3,158 funds in our sample for our market-level analysis and have 2,437 funds for the fund-level analysis. The Data Appendix details the constructions of all the variables.

Empirical Results

Table 1 reports the summary statistics of the variables for our fund-level analysis on the flow–net alpha sensitivity. It shows that distributions of fund flow and style flow are slightly skewed to the right, whereas that of fund size is highly skewed to the right with a large standard deviation, implying that some extremely large funds exist in the market. Also, on average, fund net returns are slightly positive, whereas fund net alphas are slightly negative whether estimated by FF5 or FFC4. On average, the values of $1 - R^2$ of FF5 and FFC4 are close to 0.08, implying that on average, around 8% of the total variation of fund net returns in excess of risk-free return is due to active management and cannot be explained by these models. The standard deviation of net alpha and that of gross alpha are very close to each other, as the fund expense ratio is very stable.[47] The VIX value is close to symmetric with a large variation.

Table 2 illustrates the results of the regression model in Equation **Error! Reference source not found.**. It shows that in all model specifications, the interaction term of fund net alpha and VIX is significantly negative, suggesting that a higher VIX level significantly decreases the flow–net alpha

[47] The differences in the values of these two variables' statistics exist in the sixth or seventh digit after the decimal.

sensitivity. In all these model specifications, a one-unit increase in VIX decreases the flow–net alpha sensitivity by around 0.002, holding other variables unchanged. Also, all the interaction terms of fund net alpha and performance variation measure are negative and highly significant, suggesting that higher performance variation significantly reduces the flow–net alpha sensitivity. Particularly, the first three columns report the results for which fund performance and performance variation are estimated by the FF5 model. We find that, holding other variables unchanged, if $OMR2_{i,t-1}$ increases by 0.01, the flow–net alpha sensitivity decreases by 0.0016 on average [model specification (1)]; if $NetAlpha\_Std_{i,t-1}$ or $GrossAlpha\_Std_{i,t-1}$ increases by 0.01, the flow–net alpha sensitivity decreases by 0.0004 on average [model specifications (2) and (3)].[48] The last three columns report the results for which fund performance and performance variation are estimated by the FFC4 model, and the results are highly consistently with those reported in the first three columns.

The above results imply that higher stock market volatility decreases the sensitivity of gross alpha to manager ability, so we observe that it decreases the flow–net alpha sensitivity. The finding that a higher VIX level decreases the flow–net alpha sensitivity is consistent with that in Jin, Kacperczyk, Kahraman, and Suntheim (2022).[49] Also, higher performance

[48] The results of model specifications (2) and (3), and model specifications (5) and (6) are very close because the standard deviation of net alpha and that of gross alpha are very close to each other. The difference exists in the sixth digit after the decimal in the coefficients and standard errors.

[49] Similarly, other studies also suggest that the flow–net alpha sensitivity decreases when the market is in extreme condition, more volatile, and accompanies with more economic uncertainty [Franzoni and Schmalz (2017), Harvey and Liu (2019), and Jiang, Starks, and Sun (2021)].

variation makes investors rely less on fund performance to infer manager abilities and react less intensively to fund performance. This finding is consistent with that in Huang, Wei, and Yan (2021). In short, in these tests, we find that our measures of stock market volatility and performance variation are effective, so they should also affect the dynamics of HHI, as stated in our empirical prediction.

Table 3 reports the summary statistics of the variables for our market-level analysis on the dynamics of HHI. It shows that on average, HHI is around 0.01 in the U.S. active equity mutual fund market, showing that this market is competitive. The large-fund group, which contains only five funds, on average occupies 17% of the market share, whereas the small-fund group, which contains around seventy funds on average over time, occupies only 0.07% of the market share on average. Also, the small-fund group tends to have a larger performance variation than the large-fund group, as implied by its larger mean values of $1 - R^2$, standard deviation of net alpha, and standard deviation of gross alpha. The change in VIX is small on average but variates a lot, implying that stock market volatility changes substantially over time.

To offer more insights before we report the test results, we plot HHI, the number of funds in the market, and market shares of the large-fund and small-fund groups in Figure 1.

First, we can see that HHI fluctuates a great deal over the last few decades and does not converge to a particular level. This finding is consistent with the framework with dynamic manager abilities but inconsistent with a linear framework with constant manager abilities, where HHI converges to a constant level. Therefore, the finding here is consistent with those of Feldman and Xu (2022).[50] Second, HHI moves more closely with the market share of the large-fund group than with the inverse of the number of funds. As the market share value indicates the relative inferred ability of this group, this finding is consistent with our theoretical framework that the managers' relative inferred abilities are more relevant than the number of funds when analyzing HHI. Therefore, it is important to study heterogeneous managers for whom HHI captures managers'

[50] Feldman and Xu (2022) shows that 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.

relative inferred abilities, instead of homogeneous managers because for them, HHI is simply the inverse of the number of funds.

Further, our theory can explain some of the results in this figure in a way that is compatible with the stylized facts shown in the literature. For example, Wahal and Wand (2011) show that from the 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. Kosowski, Timmermann, Wermers, and White (2006) show that outperforming managers become scarce after 1990 and speculates that this might be due to the competition among the large number of new funds, which reduces the gains from trading. Fama and French (2010) also report 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. In Figure 1, we observe that the number of funds keeps increasing from the early 1990s to the early 2000s, whereas HHI keeps decreasing in this period. If the new entrants in this period hold portfolios similar to those of the incumbents and/or outperformance become scarce in this period, then fund managers' inferred abilities become more similar. By our theoretical results, similarity in fund managers' inferred abilities leads to similarity in equilibrium fund sizes, so HHI decreases.

Table 4 reports the results of our market-level analysis on the dynamics of HHI, results of the regression model in Equation

Error! Reference source not found.. It shows that the coefficient of dif_VIX_{t-1} is significantly negative in all model specifications. In particular, results in column (1) (other columns) indicate that holding other variables unchanged, a one-unit increase in VIX decreases HHI in the next month by around 0.0002 (0.0001). This finding is consistent with our third prediction that when the distribution of funds' sizes is highly skewed to the right (as shown in Table 1), an increase in stock market volatility decreases HHI.

Also, in column (1) the coefficient of $dif\_MarketShare_{t-1}^{B}$ is significantly positive, implying that a positive shock in the large-fund group's market share induces an increase in HHI in the next month. In particular, holding other variables unchanged, if the large-fund group's market share increases by 0.01, then HHI in the next month would increase by around 0.012 . Also, the coefficient of $dif\_MarketShare_{t-1}^{S}$ is negative but is insignificant. The insignificance is probably due to the noise in the small funds' market shares. As change in market share indicates change in relative performance in AFMI in general, the results in column (1) are consistent with our first prediction that, for sufficiently large (small) funds, increase in their performances relative to those of other funds exerts positive (negative) impacts on HHI.

Columns (2) to (4) offer the results when fund performance and performance variation measures are estimated by the FF5 model. The coefficients of the interaction terms of

$dif\_MarketShare_{t-1}^{B}$ and the measures of the large-fund group's performance variation are significantly negative. In particular, holding other variables unchanged, if $OMR2_{i,t-1}^{B}$ increases by one basis point, then the impact of $dif\_MarketShare_{t-1}^{B}$ on dif_HHI_{t} decreases by around 0.0014; if $NetAlpha\_Std_{i,t-1}^{B}$ or $GrossAlpha\_Std_{i,t-1}^{B}$ increases by one basis point, then the impact of $dif\_MarketShare_{t-1}^{B}$ on dif_HHI_{t} decreases by around 0.015. Also, the coefficients of the interaction terms of $dif\_MarketShare_{t-1}^{S}$ and the measures of the small-fund's performance variation are positive and marginally significant. The results in columns (5) to (7) when measures of fund performance and performance variation are estimated by the

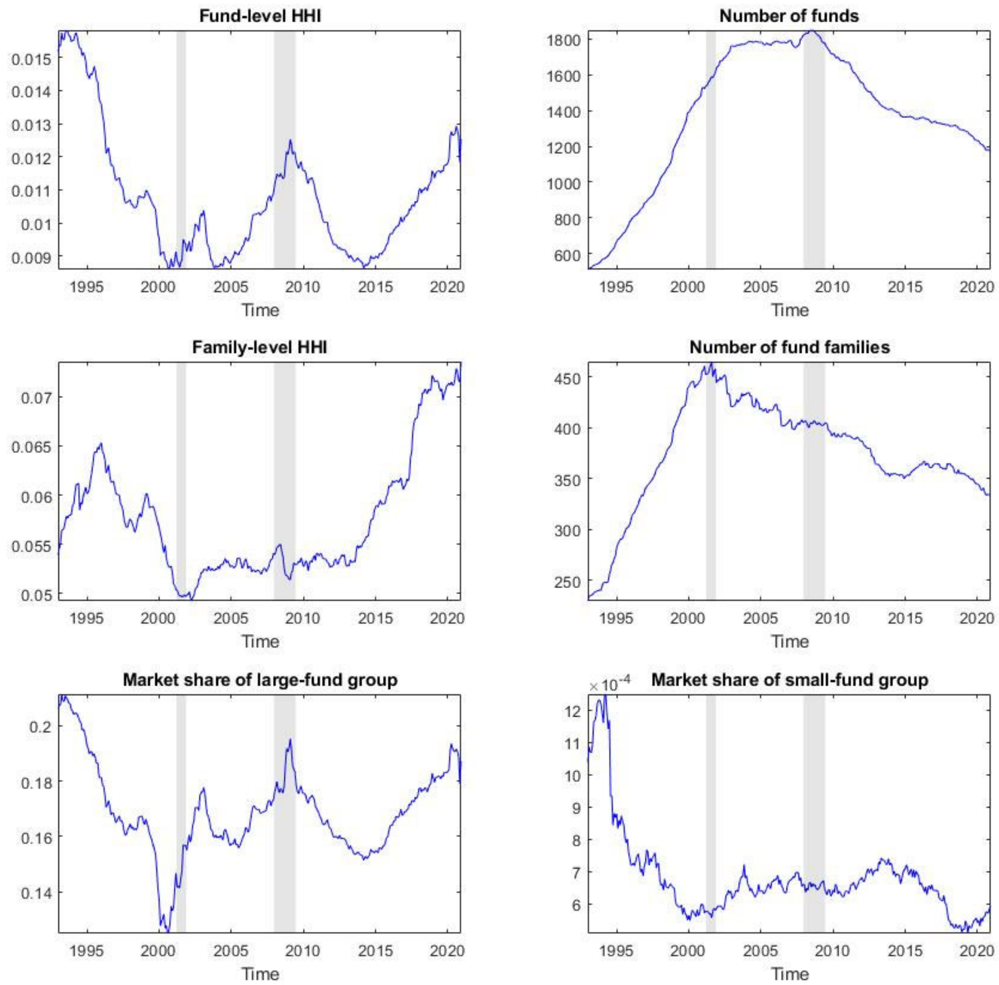
FFC4 model are consistent with those in columns (2) to (4). We also find that the coefficients of the interaction term of $dif\_MarketShare^{S}_{t-1}$ and $NetAlpha\_Std^{S}_{i,t-1}$, and that of $dif\_MarketShare^{S}_{t-1}$ and $GrossAlpha\_Std^{S}_{i,t-1}$ become more significant in these model specifications. In general, these results are consistent with our second prediction that higher performance variations in sufficiently large (small) funds mitigate the positive (negative) impacts of the increase in their relative performance on HHI.

We also do multiple robustness checks on our test results. We estimate the shocks in VIX as the out-of-sample residuals of an AR(1) model or an AR(2) model on VIX on a 24-month rolling-widow basis, and use these shocks to measure the (unexpected) changes in VIX instead of dif_VIX_t. We redefine the large-fund group as the largest ten funds. We also redefine the small-fund group as the funds with fund TNA values from the tenth percentile to the fifteenth percentile, or as those with fund TNA values from the fifth percentile to the fifteenth percentile. Furthermore, we use standard error clustered by year instead of Newey-West estimates of standard error. We redo the tests and find results that are highly consistent with those in Table 4. For brevity, we omit the results of these robustness checks here.

In summary, the above empirical results are consistent with our theoretical predictions.

Figure 1 U.S. AFMI Concentration Dynamics

Figure 1 plots the monthly values of variables from January 1993 to December 2020 using the U.S. active equity mutual fund data from the Center for Research in Security Prices (CRSP). The two graphs at the top plot the HHI and the number of funds in the market, respectively. The two graphs at the bottom plot the market shares of the large-fund group and small-fund group, respectively. HHI is the Herfindahl-Hirschman Index, calculated as the sum of funds' market shares squared. The number of funds is counted as the number of the U.S. active equity mutual funds that have observations satisfying our criteria. Funds' market shares are calculated based on their total net assets under management. The large-fund group contains the largest five funds in the market, whereas the small-fund group contains funds that have fund size values from the fifth percentile to the tenth percentile. These two groups are redefined each month. The gray areas represent the two recessions, from March 2001 to November 2001, and from December 2007 to June 2009, respectively.



Fund-level HHI
0.015
0.014
0.013
0.012
0.011
0.01
0.009
1995
2000
2005
2010
2015
2020
Time
Number of funds
1800
1600
1400
1200
1000
800
600
1995
2000
2005
2010
2015
2020
Time
Family-level HHI
0.07
0.065
0.06
0.055
0.05
1995
2000
2005
2010
2015
2020
Time
Number of fund families
450
400
350
300
250
1995
2000
2005
2010
2015
2020
Time
Market share of large-fund group
0.2
0.18
0.16
0.14
1995
2000
2005
2010
2015
2020
Time
$\times 10^{-4}$
Market share of small-fund group
12
11
10
9
8
7
6
1995
2000
2005
2010
2015
2020
Time

Table 1. Summary Statistics on Variables for Fund-Level Analysis

Table 1 reports the summary statistics on the variables for our fund-level analysis. Our sample period is from January 1990 to December 2020, and we use monthly data. FF5 is the five-factor model developed by Fama and French (2015), and FFC4 is the four-factor model developed by Fama and French (1993) and Carhart (1997). We estimate the models on a 24-month rolling-window basis, and over time, calculate the $1 - R^2$ and out-of-sample prediction of fund net alphas. The definitions and constructions of all the variables are reported in the Data Appendix.

Variable	Observation	Mean	Standard deviation	Percentile 25th	50th	75th
Fund characteristics						
Fund flow (decimal)	369589	0.0027	0.8675	-0.0152	-0.0050	0.0068
Fund net return (decimal)	369589	0.0077	0.0624	-0.0191	0.0118	0.0381
Fund TNA (in 1 billion December 2020 dollars)	369589	4.6323	16.1613	0.2767	0.9473	3.1540
Fund age (number of months)	369589	203.5	171.4	89.0	155.0	250.0
Fund expense (decimal)	369589	0.0117	0.0042	0.0093	0.0114	0.0139
Fund turn over ratio (decimal)	369589	0.7868	0.6987	0.3400	0.6167	1.0200
Style flow (decimal)	369589	-0.0012	0.0103	-0.0068	-0.0024	0.0035
Estimates from FF5						
Fund net alpha (decimal)	369589	-0.0009	0.0427	-0.0100	-0.0011	0.0076
1 - R^2 of the factor model (decimal)	369589	0.0769	0.0746	0.0312	0.0566	0.0977
Fund net alpha standard deviation (decimal)	369589	0.0170	0.0388	0.0093	0.0133	0.0195
Fund gross alpha standard deviation (decimal)	369589	0.0170	0.0388	0.0093	0.0133	0.0195
Estimates from FFC4						
Fund net alpha (decimal)	369589	-0.0010	0.0437	-0.0098	-0.0011	0.0074
1 - R^2 of the factor model (decimal)	369589	0.0813	0.0763	0.0339	0.0610	0.1037
Fund net alpha standard deviation (decimal)	369589	0.0166	0.0391	0.0092	0.0131	0.0191
Fund gross alpha standard deviation (decimal)	369589	0.0166	0.0391	0.0092	0.0131	0.0191
Fund family characteristics						
Fund family net alpha by FF5 (decimal)	68552	-0.0006	0.1020	-0.0089	-0.0012	0.0064
Fund family net alpha by FFC4 (decimal)	68552	-0.0006	0.1027	-0.0088	-0.0012	0.0062
Number of member funds in fund families (number)	68552	5.9023	6.0191	2.0000	4.0000	7.0000
Fund family TNA (in 1 billion December 2020 dollars)	68006	25.1030	90.1054	0.8643	3.8815	15.7357
Market characteristics						
VIX (decimal)	336	0.1954	0.0805	0.1361	0.1750	0.2345

Table 2. Flow–Net Alpha Sensitivity, Stock Market Volatility, and Performance Variation

Table 2 reports the results of the model in Equation **Error! Reference source not found.**. The dependent variable is the fund percentage flow, $Flow$, and it is in decimal. The independent variables are lagged by one month. The first three columns report the results of the model using the measures of fund performance and performance variation estimated by the FF5 model, and the last three columns report the results of the model using the measures estimated by the FFC4 model. The detailed definitions of the variables are in the Data Appendix. Standard errors that are clustered by fund and by year are presented in parentheses. The symbols ***, **, and * represent the 1%, 5%, and 10% significance levels, respectively, in a two-tail t-test.

	FF5			FFC4		
	(1)	(2)	(3)	(4)	(5)	(6)
NetAlpha	0.3075***	0.3143***	0.3143***	0.4089***	0.3935***	0.3935***
	(0.0731)	(0.0771)	(0.0771)	(0.0848)	(0.0817)	(0.0817)
*NetAlpha*VIX*	-0.1719**	-0.2077**	-0.2077**	-0.1704*	-0.2142**	-0.2142**
	(0.0814)	(0.0867)	(0.0867)	(0.0884)	(0.0815)	(0.0815)
VIX	-0.0035	-0.0042	-0.0042	-0.0031	-0.0040	-0.0040
	(0.0040)	(0.0039)	(0.0039)	(0.0039)	(0.0037)	(0.0037)
*NetAlpha*OMR2*	-0.1648***			-0.2035***		
	(0.0342)			(0.0337)		
OMR2	0.0040			0.0040		
	(0.0087)			(0.0082)		
*NetAlpha*NetAlpha_Std*		-0.0443***			-0.0543***	
		(0.0102)			(0.0109)	
NetAlpha_Std		0.0135			0.0132	
		(0.0135)			(0.0129)	
*NetAlpha*GrossAlpha_Std*			-0.0443***			-0.0543***
			(0.0102)			(0.0109)
GrossAlpha_Std			0.0135			0.0132
			(0.0135)			(0.0129)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Fund fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	369,589	369,589	369,589	369,589	369,589	369,589
R-squared	0.0451	0.0451	0.0451	0.0454	0.0453	0.0453
Adjusted R-squared	0.0387	0.0387	0.0387	0.0390	0.0389	0.0389

Table 3. Summary Statistics on Variables for Market-Level Analysis

Table 3 reports the summary statistics on the variables for our market-level analysis. Our sample period is from January 1990 to December 2020, and we use monthly data. HHI is the Herfindahl-Hirschman Index, calculated as the sum of market shares squared, and it is in decimal. VIX is the average of daily option-implied volatility index values in each month. The large-fund group contains the largest five funds (based on fund size values), and the small-fund group contains those with fund size values from the fifth percentile to the tenth percentile. FF5 is the five-factor model developed by Fama and French (2015), and FFC4 is the four-factor model developed by Fama and French (1993) and Carhart (1997). We estimate these models on a 24-month rolling-window basis, and over time, calculate the $1 - R^2$ and the out-of-sample prediction of fund net alphas. The definitions and constructions of all the variables are reported in the Data Appendix.

Variable	Observation	Mean	Standard deviation	Percentile 25th	50th	75th
Market characteristics						
Fund-level HHI (decimal)	336	0.0108	0.0019	0.0092	0.0104	0.0115
Change in fund-level HHI (decimal)	336	-0.0003	0.0047	-0.0001	0.0000	0.0001
Family-level HHI (decimal)	336	0.0575	0.0061	0.0528	0.0550	0.0608
Change in family-level HHI (decimal)	336	-0.0006	0.0120	-0.0002	0.0001	0.0003
VIX (decimal)	336	0.1954	0.0805	0.1361	0.1750	0.2345
Change in VIX (decimal)	336	0.0003	0.0433	-0.0177	-0.0027	0.0119
Market share of large-fund group (decimal)	336	0.1692	0.0163	0.1594	0.1674	0.1774
Change in market share of large-fund group (decimal)	336	-0.0012	0.0206	-0.0012	-0.0001	0.0010
Market share of small-fund group (decimal)	336	0.0007	0.0001	0.0006	0.0007	0.0007
Change in market share of small-fund group (decimal)	336	-3.50E-07	2.86E-05	-9.75E-06	4.59E-07	1.02E-05
Number of funds (number)	336	1379	374	1219	1405	1727
Growth rate of the number of funds (decimal)	336	0.0198	0.3164	-0.0028	0.0011	0.0072
Estimates from FF5						
Large-fund group's 1 - R2 of the factor model (decimal)	336	0.0744	0.0506	0.0313	0.0580	0.1155
Large-fund group's net alpha standard deviation (decimal)	336	0.0092	0.0031	0.0076	0.0082	0.0104
Large-fund group's gross alpha standard deviation (decimal)	336	0.0092	0.0031	0.0076	0.0082	0.0104
Small-fund group's 1 - R^2 of the factor model (decimal)	336	0.1111	0.0459	0.0759	0.1065	0.1405
Small-fund group's net alpha standard deviation (decimal)	336	0.0157	0.0058	0.0125	0.0144	0.0195
Small-fund group's gross alpha standard deviation (decimal)	336	0.0155	0.0058	0.0124	0.0142	0.0190
Estimates from FFC4						
Large-fund group's 1 - R2 of the factor model (decimal)	336	0.0818	0.0572	0.0357	0.0613	0.1284
Large-fund group's net alpha standard deviation (decimal)	336	0.0092	0.0029	0.0076	0.0087	0.0108
Large-fund group's gross alpha standard deviation (decimal)	336	0.0092	0.0029	0.0076	0.0087	0.0108
Small-fund group's 1 - R^2 of the factor model (decimal)	336	0.1165	0.0471	0.0812	0.1117	0.1421
Small-fund group's net alpha standard deviation (decimal)	336	0.0153	0.0057	0.0123	0.0142	0.0183
Small-fund group's gross alpha standard deviation (decimal)	336	0.0151	0.0057	0.0122	0.0141	0.0179

Table 4. Dynamics of HHI, Changes in Stock Market Volatility and Fund Performance, and Performance Variation

Table 4 reports the results of the model in Equation **Error! Reference source not found.**, where the dependent variable is the change in fund-level HHI, dif_HHI, which is in decimal. The independent variables are lagged by one month. The columns (2) to (4) report the results of the model using the measures of fund performance and performance variation estimated by the FF5 model, and columns (5) to (7) report the results of the model using the measures estimated by the FFC4 model. The detailed definitions of the variables are in the Data Appendix. The standard errors are presented in parentheses, which are estimated by the Newey-West estimator, with the maximum lag of 12 to be considered in the autocorrelation structure of the regression error. The symbols ***, **, and * represent the 1%, 5%, and 10% significance levels, respectively, in a two-tail t-test.

		FF5			FFC4		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dif_VIX	-0.0185***	-0.0128***	-0.0072***	-0.0071***	-0.0148***	-0.0063***	-0.0062***
	(0.0061)	(0.0046)	(0.0018)	(0.0018)	(0.0046)	(0.0019)	(0.0019)
Dif_MarketShare ^{L}	1.1648***	1.8136***	1.8950***	1.8854***	1.7348***	1.8431***	1.8356***
	(0.3545)	(0.3655)	(0.1312)	(0.1323)	(0.3926)	(0.1166)	(0.1179)
Dif_MarketShare ^{S}	-10.4878	-38.9940	-11.7915	-13.3416	-47.3898*	-14.2290	-15.0738
	(10.4128)	(24.7941)	(11.6700)	(11.5912)	(28.2525)	(11.5485)	(11.4543)
*Dif_MarketShare ^{L} *OMR2 ^{L}*		-13.7309***			-11.1934***		
		(3.3880)			(2.9644)		
OMR2 ^{L}		-0.0041			-0.0030		
		(0.0035)			(0.0028)		
*Dif_MarketShare ^{S} *OMR2 ^{S}*		223.3505*			240.3995*		
		(128.9604)			(136.4108)		
OMR2 ^{S}		-0.0091			-0.0084		
		(0.0077)			(0.0078)		
*Dif_MarketShare ^{L} *NetAlpha_Std ^{L}*			-146.8074***			-142.0496***	
			(18.4656)			(16.9828)	
NetAlpha_Std ^{L}			-0.0039			0.0031	
			(0.0706)			(0.0752)	
*Dif_MarketShare ^{S} *NetAlpha_Std ^{S}*			1,457.3450*			1,700.5736**	
			(870.1754)			(862.9170)	
NetAlpha_Std ^{S}			-0.0197			-0.0268	
			(0.0332)			(0.0327)	
*Dif_MarketShare ^{L} *GrossAlpha_Std ^{L}*				-146.1784***			-141.6057***
				(18.3745)			(16.9451)
GrossAlpha_Std ^{L}				-0.0046			0.0035
				(0.0714)			(0.0767)
*Dif_MarketShare ^{S} *GrossAlpha_Std ^{S}*				1,634.9686*			1,832.8401**
				(883.6655)			(875.0814)
GrossAlpha_Std ^{S}				-0.0177			-0.0264
				(0.0366)			(0.0349)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	336	336	336	336	336	336	336

Table 5. Dynamics of Family-Level HHI, Changes in Stock Market Volatility and Fund Performance, and Performance Variation

Table 5 reports the results of the model in Equation **Error! Reference source not found.** where the dependent variable is the change in family-level HHI, dif_HHI, which it is in decimal. The independent variables are lagged by one month. The columns (2) to (4) report the results of the model using the measures of fund performance and performance variation estimated by the FF5 model, and columns (5) to (7) report the results of the model using the measures estimated by the FFC4 model. The detailed definitions of the variables are in the Data Appendix. The standard errors are presented in parentheses, which are estimated by the Newey-West estimator, with the maximum lag of 12 to be considered in the autocorrelation structure of the regression error. The symbols ***, **, and * represent the 1%, 5%, and 10% significance levels, respectively, in a two-tail t-test.

		FF5			FFC4		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dif_VIX	-0.0478***	-0.0332***	-0.0190***	-0.0187***	-0.0385***	-0.0167***	-0.0165***
	(0.0157)	(0.0120)	(0.0047)	(0.0047)	(0.0119)	(0.0050)	(0.0050)
Dif_MarketShare ^{L}	2.9809***	4.6419***	4.8479***	4.8233***	4.4473***	4.7160***	4.6966***
	(0.9067)	(0.9354)	(0.3371)	(0.3399)	(1.0027)	(0.2985)	(0.3018)
Dif_MarketShare ^{S}	-25.9133	-103.2295	-27.7098	-32.0121	-124.7267*	-33.6284	-36.0972
	(26.8073)	(63.6037)	(28.8948)	(28.8223)	(72.0076)	(28.4053)	(28.2913)
Dif_MarketShare ^{L} **OMR2* ^{L}		-35.1589***			-28.7937***		
		(8.6739)			(7.5898)		
OMR2 ^{L}		-0.0102			-0.0071		
		(0.0095)			(0.0075)		
Dif_MarketShare ^{S} **OMR2* ^{S}		601.0876*			642.9683*		
		(330.8901)			(347.9415)		
OMR2 ^{S}		-0.0267			-0.0258		
		(0.0192)			(0.0194)		
Dif_MarketShare ^{L} **NetAlpha_Std* ^{L}			-374.9015***			-362.5442***	
			(47.9735)			(44.1819)	
NetAlpha_Std ^{L}			0.0258			0.0667	
			(0.1886)			(0.2029)	
Dif_MarketShare ^{S} **NetAlpha_Std* ^{S}			3,601.9172*			4,189.1962**	
			(2,150.3450)			(2,125.7289)	
NetAlpha_Std ^{S}			-0.0564			-0.0778	
			(0.0930)			(0.0890)	
Dif_MarketShare ^{L} **GrossAlpha_Std* ^{L}				-373.2812***			-361.4011***
				(47.7271)			(44.0714)
GrossAlpha_Std ^{L}				0.0250			0.0693
				(0.1900)			(0.2061)
Dif_MarketShare ^{S} **GrossAlpha_Std* ^{S}				4,078.8145*			4,546.4599**
				(2,193.0905)			(2,165.1525)
GrossAlpha_Std ^{S}				-0.0527			-0.0784
				(0.1013)			(0.0944)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	336	336	336	336	336	336	336

Figure A1. AFMI Equilibrium HHI and Relative Inferred Abilities

Figure A1 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.

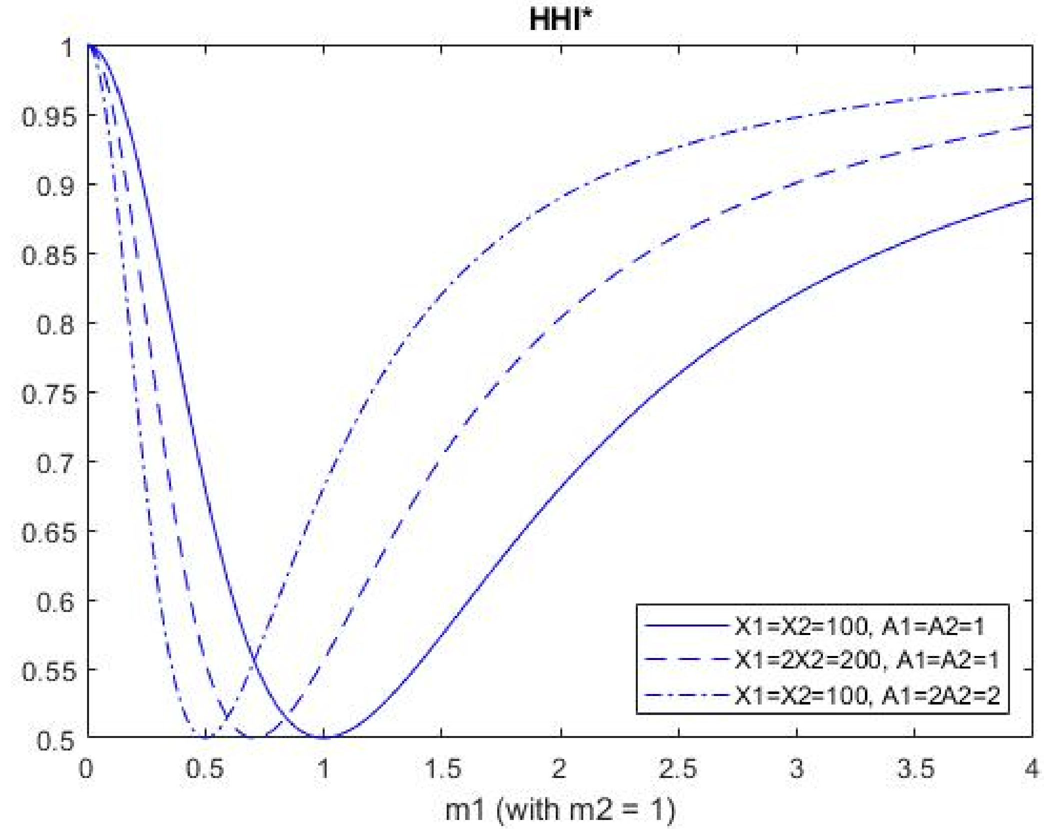
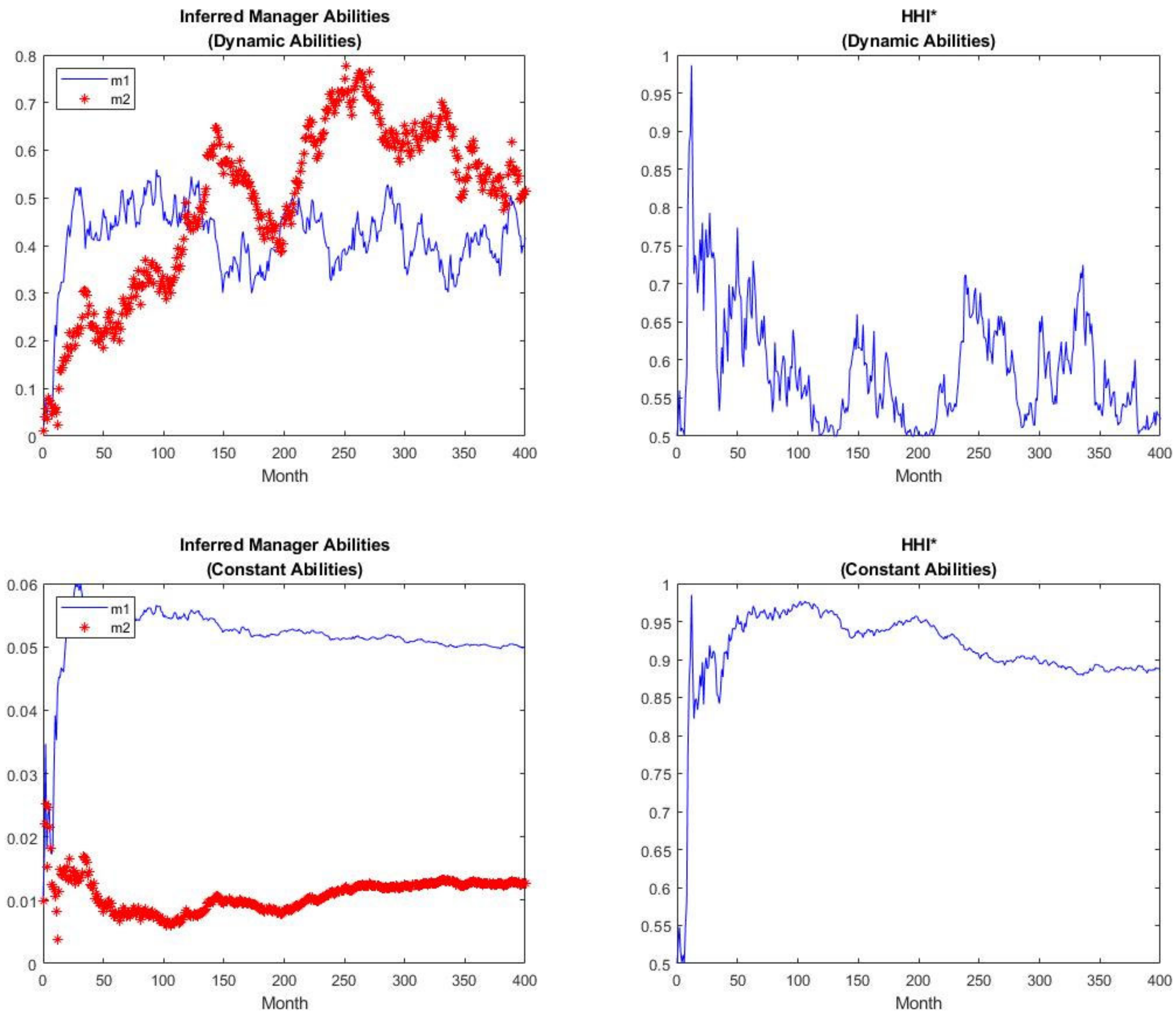

HHI*
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
m1 (with m2 = 1)
X1=X2=100, A1=A2=1
X1=2X2=200, A1=A2=1
X1=X2=100, A1=2A2=2

Figure A2. AFMI Equilibrium HHI and Inferred Abilities with Dynamic Abilities and Constant Abilities

Figure A2 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.


Inferred Manager Abilities
(Dynamic Abilities)
m1
m2
Month
HHI*
(Dynamic Abilities)
Month
Inferred Manager Abilities
(Constant Abilities)
m1
m2
Month
HHI*
(Constant Abilities)
Month

Conclusion

We introduce continuous-time rational models of dynamics of AFMI HHI in which unobservable fund manager abilities are heterogeneous and dynamic. In equilibrium, managers with higher inferred abilities receive larger fund sizes, so their relative inferred abilities determine HHI. Our model predicts that if a manager's inferred ability is sufficiently larger (smaller) than those of others, then an increase in this manager's inferred ability exerts positive (negative) impact on HHI. Also, if a manager has sufficiently large (small) inferred ability relative to those of others, then HHI is concave in this manager's inferred ability, and the concavity has negative impact on HHI. If all funds' inferred abilities are sufficiently close, then HHI is convex in a manager's inferred ability, and this convexity has positive impact on HHI.

Our model also shows that when funds' performance variations are larger, investors rely less on the shocks of managers' relative performances to infer manager abilities, making investment flows less sensitive to these shocks. Consequently, the positive (negative) impacts of higher relative performances of sufficiently large (small) funds on HHI are mitigated and have smaller absolute magnitudes.

In addition, in our nonlinear framework where sensitivities of gross alphas to manager abilities decrease with stock market volatility, we find that higher stock market volatility decreases all funds' sizes. If there are some extremely large funds in the market, then the effect of higher stock market volatility on these funds dominates that of other funds, inducing a negative aggregate effect on HHI. Linear frameworks of manager abilities and gross alphas that are used in the current literature cannot directly model this effect and effects of other economic factors on the dynamics of HHI, as we do in our nonlinear frameworks.

We also show a special case in which unobservable fund manager abilities are constant in a linear framework. In this case, as time goes to infinity, managers' inferred abilities converge to their true ability levels and do not change, making both equilibrium fund sizes and 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 empirical results are consistent with our theoretical findings. In particular, the flow–net alpha sensitivity significantly decreases with our measures of stock market volatility and fund performance variation, implying the effectiveness of these measures. Also, an increase in stock market volatility significantly decreases HHI. An increase in the large-fund group's market share, which proxies this group's relative performance, exerts a significantly positive impact on HHI; and a larger performance variation in this group significantly decreases such positive impact. An increase in the small-fund group's market share tends to exert a negative effect on HHI, although this effect is insignificant. However, we find evidence that a larger performance variation in this group mitigates the effect of the group's change in market share on HHI.

Moreover, the fluctuation of the empirical HHI over time is consistent with our theoretical results in which manager abilities are dynamic and unobservable, but it is inconsistent with a model with constant unobservable manager abilities in a linear framework. Also, the fact that the empirical HHI moves more closely with large funds' market shares than the inverse of the number of funds shows the importance of modeling heterogeneous managers, where HHI captures managers' relative inferred abilities, instead of homogeneous managers, where HHI is simply the inverse of the number of competitors. In addition, our model explains the following literature findings in a compatible way: 1) from the 1990s to early 2000s, new entrants who have portfolio holdings similar to those of incumbents decrease fund performances and fund flows, 2) outperforming managers are scarce, and 3) HHI decreases during this period.

Our paper sheds light on future research on the dynamics of AFMI concentration. In particular, future research in this area can focus on factors that affect fund managers' relative inferred abilities. For example, current literature finds that fund family members can compete or cooperate with each other [see, for example, Evans, Prado, and Zambrana (2020), Eisele, Nefedova, Parise, and Peijnenburg (2020), and Xu (2022)]. Other literature shows that mutual funds compete in different dimensions, such as by trading assets in specific industries and style markets (defined by, for example, stock's total capitalization and book-to-market-ratio), by selling fund shares in specific retail market segments (such as direct-sold and broker-sold), by concentrating research on stocks that are informationally intense, and by offering unique products [see, for example, Kacperczyk, Sialm, and Zheng (2005), Guercio and Reuter (2014), Hoberg, Kumar, and Prabhala (2018), Jiang, Shen, Wermers, and Yao (2018), and Kostovetsky and Warner (2020)]. Because the methods that fund managers use to compete in the market affect managers' relative inferred abilities, these methods would consequently exert impacts on

AFMI concentration. Our study also suggests that a nonlinear framework of gross alphas and manager abilities can directly model the effects of these factors and offer more insights to the market equilibrium.

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.