Journal of Financial Economics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Is the active fund management industry concentrated enough?[☆]

David Feldman[a,*], Konark Saxena[a,*], Jingrui Xu[b,*]

[a] School of Banking and Finance, UNSW Business School, University of New South Wales, UNSW Sydney, NSW 2052, Australia
[b] Wang Yanan Institute for Studies in Economics (WISE), and School of Economics, Department of Finance, Xiamen University, No. 422, Siming South Road, Xiamen, Fujian 361005, China

ARTICLE INFO

Article history:
Received 13 May 2016
Revised 22 May 2018
Accepted 18 August 2018
Available online xxx

JEL classification:
G11
G23
J24
L11

Keywords:
Active fund management
Market concentration
Effort
Industry size
Alpha

ABSTRACT

We introduce a theoretical model of the active fund management industry (AFMI) in which performance and size depend on the AFMI's competitiveness (concentration). Under plausible assumptions, as AFMI's concentration decreases, so do fund managers' incentives for exerting effort in search of alpha. Consequently, managers produce lower gross alpha, and rational investors, inferring lower expected AFMI performance, allocate a smaller portion of their wealth to active funds. Empirically, we find that a decrease in the US mutual fund industry concentration over our sample period is associated with a decrease in its net alpha and size (relative to stock market capitalization).

1. Introduction

An active area of research in financial economics examines the massive size of the active fund management industry (AFMI) and the high compensation of its mangers, despite its unimpressive historical performance.[1] Recent literature argues that neither its massive size nor its

☆ We thank anonymous referees and Vikas Agarwal, Marie Briére, Prachi Deuskar, Roger Edelen, Gonçalo Faria, Ron Giammarino, Abigail Hornstein, María González Pérez, Gur Huberman, Arati Kale, Chang Mo Kang, Paul Karehnke, Lei Shi, Kai Li, Didier Maillard, Bertrand Maillet, David Nanigian, Linda Pesante, Matthew Spiegel, Yoav Tamir, Thomas Ruf, Khim Veasna, and Russell Wermers for helpful discussions. We also thank seminar participants at McGill University, University of New South Wales, Xiamen University, Australasian Finance and Banking Conference, Midwest Finance Association Annual Meetings, International Conference of the French Finance Association, International Conference of the Financial Engineering and Banking Society, Portuguese Finance Network International Conference, Spanish Finance Association Finance Forum, Center for Analytical Finance, Indian School of Business, Hyderabad, Summer Research Conference in Finance, World Finance Conference, Behavioral Finance and Capital Markets Conference, Finance Research Workshop at Indian Institute of Management, Calcutta, Paul Woolley Center Capital Market Dysfunctionality Conference, Southern Finance Association annual meeting, International Risk Management Conference, the Role of Hedge Funds and Other Collective Investment Funds in the Modern World Conference, the Financial Management Association annual meeting, Greater China Area Finance Conference, and China International Conference in Finance. Jingrui Xu's research was supported by the Fundamental Research Funds for the Central Universities, Grant Number 20720181061.

* Corresponding authors.
E-mail addresses: dfeldman@unsw.edu.au (D. Feldman), k.saxena@unsw.edu.au (K. Saxena), jingrui.xu@xmu.edu.cn (J. Xu).

[1] Studies examining active mutual fund performance include Jensen (1968), Gruber (1996), Carhart (1997), Brown et al. (1996), Daniel et al. (1997), Pastor and Stambaugh (2002), Wermers (2000), Cohen et al. (2005), Kacperczyk et al. (2005), Fama and French (2010), Glode (2011), and Berk and van Binsbergen (2015). Studies examining the relation between active fund performance and size include Berk and Green (2004),

https://doi.org/10.1016/j.jfineco.2019.08.009
0304-405X/© 2019 Published by Elsevier B.V.

performance is puzzling when gross alpha production is subject to decreasing returns to scale (see, for example, Berk and Green, 2004; BG and Pastor and Stambaugh, 2012; PS). Intuitively, as more assets under management (AUM) chase opportunities, prices adjust, making gross alpha harder to find. These insights lead to several interesting questions. For example, do other gross alpha production inputs play a significant role in determining AFMI size and performance?

We posit that incentives of fund managers to exert effort in finding investment opportunities influence gross alpha production and that these incentives depend on AFMI concentration (a measure of its competitive environment).[2] To formally analyze this, we introduce an AFMI model in which active fund managers choose (optimal) costly effort levels when competing over investment funds. In equilibrium, AFMI concentration levels influence optimal effort levels exerted by managers, which, in turn, influence AFMI performance and size. We find evidence, consistent with our model, that decreases in the concentration of the US mutual fund industry are associated with decreases in its performance and size (relative to stock market capitalization).

In our model, gross alpha production by active managers depends on their fund sizes, optimal effort levels, and ability. We refer to ability as a measure of a fund manager's decreasing returns to scale. That is, we ascribe managers with higher ability when they have "lower cost" for managing the same fund size. Competing managers maximize profits by optimally choosing fees and costly effort levels, while offering net alpha (net of management fees) to investors. Mean-variance investors choose optimal portfolios of passive benchmarks and active funds (whose expected net alphas are positive to compensate for active funds' additional risk).[3] Investors' pursuit of net alpha induces a positive relation between fund sizes and managers' abilities. In equilibrium, all fund sizes adjust so that they offer similar expected net alphas to investors at break-even fees (sufficient to cover costs).[4] Any attempt to offer higher (lower) net alpha leads to insolvency (zero AUM allocation by investors). So, higher ability in our model is not associated with higher gross alpha or net alpha. Instead, it is associated with higher value added (gross alpha multiplied by fund size; the Berk and van Binsbergen, 2015 measure of skill).

In this equilibrium, we study the impact of AFMI concentration on managerial costly effort, net alpha production, and AFMI size, all three endogenously determined. A key quantity that determines these relations is the direct benefits of effort, which we define as the difference between productivities of managerial efforts (which measure opportunities to find mispriced assets by exerting effort) and managerial efforts' costs (such as wages and research costs). We find that if higher concentration increases direct benefits of effort, then higher concentration induces higher equilibrium expected net alphas and larger AFMI size.

As the level of AFMI concentration decreases (or competition increases), the direct benefits of fund managers' efforts decrease, and they reduce their efforts. This effort-level reduction captures fund managers' optimal decision to invest less time and expenses, per dollar of AUM, on research and information acquisition, thereby holding less (informed) active positions. As a result, the funds produce lower gross alpha and provide lower net alpha to investors. AFMI size decreases because rational investors infer this lower net alpha and reduce their investment in the AFMI until they are indifferent between leaving an extra dollar in the AFMI and investing it in the passive benchmark. In summary, a decrease in AFMI concentration reduces the incentives of fund managers to exert effort, resulting in lower AFMI performance and smaller AFMI size.

Recent empirical studies find that performance declines as AFMI size increases (Pastor et al., 2015; PST) and that most of the growth in the mutual fund industry is due to the growth in the number of funds, not in the median fund size (Berk and van Binsbergen, 2015). When these empirical findings are combined, they seem to support our theoretical predictions that performance depends positively on AFMI concentration.[5] Our model provides three further predictions for empirical analysis. First, even when controlling for size, higher concentration levels are associated with an increase in AFMI performance. Second, AFMI size and effort levels increase in AFMI concentration.[6] Third, the AFMI expected net alphas and AFMI size are both either concave or convex in AFMI concentration.[7]

Chen et al. (2004), Yan (2008), Pastor and Stambaugh (2012), and Pastor et al. (2015).

[2] We use *concentration* and *competition* as opposites. Also, for simplicity and brevity, we use the term "AFMI concentration" for "AFMI market concentration" and "market concentration."

[3] While AFMI net alphas are positive in our model with risk-averse investors, aggregate net alphas are zero-sum by construction, as they shift wealth between AFMI investors and other (unmodeled) investors. These unmodeled investors could be, for example, individuals with direct equity ownership (see Stambaugh, 2014).

[4] A fund-level decreasing returns scale does not necessarily imply a correlation between fund size and net alpha (Berk and Green, 2004). While Pastor et al. (2015) find strong evidence of net alpha decreasing with size at industry level, they do not find a significant relation at the fund level. We also do not find a significant relation between fund net alpha and fund size (consistent with our model). The fund-level evidence in the literature is mixed (see, for example, Grinblatt and Titman, 1989; Chen et al. 2004; Ferreira et al., 2013a, 2013b; Yan, 2008; Reuter and Zitzewitz, 2013).

[5] Other empirical studies relating mutual fund performance to competition include Wahal and Wang (2011), Khorana and Servaes (2011), Cremers et al. (2016), and Hoberg et al. (2018). Guercio and Reuter (2014) also find evidence consistent with the notion that weaker incentives due to lower competition faced by broker-sold funds lead to their lower performance.

[6] The technical conditions for such an equilibrium are, first, that higher concentration increases the gap between the marginal benefits and marginal costs of gross alpha production and, second, that the (further) sensitivity of this gap to increased effort levels does not reverse this property. See Lemma RA1(3) and Lemma RA1(5).

[7] Our model predicts that if equilibrium fund expected net alphas are concave in AFMI concentration, then AFMI's benefits of effort are concave in AFMI concentration. Consequently, equilibrium AFMI size is also concave in AFMI concentration. On the other hand, if equilibrium AFMI size is convex in AFMI concentration, then the AFMI's benefits of effort are convex in AFMI concentration and, consequently, equilibrium expected fund net alphas are convex in AFMI concentration. Note that the order of statements in the second (convex) case is different from that in the first case for reasons explained following Proposition RA3.

We evaluate three predictions in our empirical counterpart for US AFMI. We find that, consistent with our model, both AFMI size and net alphas are, on average, increasing and concave with measures of AFMI concentration such as the Herfindahl–Hirschman Index (*HHI*).[8] While effort is largely unobservable, we suggest that aggregate managerial effort affects the average AFMI active share and tracking error. The reason is that any effort to outperform the benchmark must involve taking positions that are different from the benchmark (e.g., Cremers and Petajisto, 2009). We find that both the average AFMI active share and tracking error increase in concentration, consistent with our model (assuming these measures proxy for effort).[9]

Our analysis is related to several recent papers, though none of them addresses the question of how AFMI concentration levels affect its performance and size. In particular, our model incorporates key features from BG and PS.

To the PS model, which presents an AFMI size–net alpha relation, we introduce two distinct and novel features. The first is an AFMI concentration–alpha relation, which also exists when controlling for the AFMI size–net alpha relation modeled in PS. The second is an AFMI concentration–size relation (for details, see Section 2.3).

PS identify the AFMI equilibrium elegantly, without the need to specify fund-level size or ability heterogeneity. Ignoring such heterogeneity, our model becomes the one in PS if neither managers' effort levels nor AFMI concentration affect managers' search productivity or costs.[10] Even when, in our model, the search productivity for mispriced assets depends on effort levels, a special case of parameter values leads to a solution in which the optimal allocated effort is zero.[11] In this case as well, our model results become as those in PS, where the AFMI net alpha and size do not depend on AFMI concentration.

Our model also incorporates fund-level decreasing returns to scale, a feature in BG. In our model, this feature facilitates the study of a range of interesting equilibrium concentration levels by allowing fund size heterogeneity. This feature is not directly responsible for the AFMI concentration–alpha relation in our model. For this relation, managerial effort is essential. To see this, consider the following alternative: What if we shut down the effort channel but keep the features of fund-level decreasing returns to scale, as in BG, and industry-level decreasing returns to scale, as in PS? In this case, any effect of fund-level decreasing returns to scale on net alpha would be via AFMI size. In this hypothetical model, concentration would not influence net alpha (controlling for AFMI size). Introducing concentration and its influence on effort levels enables us to model a distinct mechanism that influences net alpha.

The BG equilibrium is compatible with the case in our model in which infinitely many small risk-neutral investors compete. Here, the size of the fund endogenously adjusts to make the gross alpha equal the fee, so that expected net alpha is always zero. Even in this case, significant differences exist between the models. For example, a novel feature of our model is that AFMI size depends on AFMI concentration even when investors are risk-neutral. A higher AFMI concentration incentivizes managers to increase effort levels. So, optimizing risk-neutral investors allocate more to the AFMI until they drive its net alpha to zero.

Another difference, compared with BG, is that managers with more ability have larger fund sizes in our model. More skilled managers receive more AUM until their gross alphas are equal to those of less skilled managers. In equilibrium, gross alpha and fees are the same across funds. So, our model predicts that the cross-sectional distribution of manager ability is reflected in the cross-sectional distribution of fund size and value added but not in gross alpha or net alpha. This is consistent with the evidence in Berk and van Binsbergen (2015). The BG model does not make this prediction: managers with more ability do not necessarily manage larger funds. In BG, competitive pressures do not force managers to choose the same fees, so a more skilled manager is indifferent between more AUM with less fees and less AUM with more fees (as long as their profits stay the same).[12]

The paper proceeds as follows. Section 2 develops the theoretical model; Section 3 describes tests of the model's predictions; Section 4 presents the empirical results; and Section 5 concludes.[13]

2. Theoretical framework

Within PS's world, adopting their notation, we develop a theoretical framework for modeling the effect of AFMI

[8] In the real world, AFMI concentration is likely to be affected by other forces (e.g., macroeconomic, regulatory) that we do not model. For instance, policy can restrict or incentivize certain investors toward a narrowly defined menu of funds, thereby increasing concentration (Hong, 2018). For convenience and parsimony, we assume exogenous concentration levels in our baseline model. In addition, we examine how our main empirical measure of concentration, HHI, can be endogenously determined (see Section 2.4).

[9] Cremers and Petajisto (2009) show that fund-level active share predicts fund performance and that this performance is strongly persistent. Brown and Davies (2017) argue that shirking managers could "jam the signal" in active share by taking uninformed bets to increase their perceived active share, generating a false sense of truly active management. However, such signal jamming behavior is more likely to be an issue if a measure of active share is tied with fund manager incentives, which is not likely for our active share sample period that ends in 2009, the year when Cremers and Petajisto (2009) was published. Also signal jamming is of more concern at the fund level than at the aggregate AFMI level, where information asymmetry and its associated signal jamming is less likely to be a concern.

[10] Analytically, effort levels do not affect alpha production in our model, if the third addend of the right side of Eq. (7) does not exist and if we abandon our cost function, Eq. (18), in favor of defining funds' fees to be net of funds' management costs.

[11] For the technical conditions that lead to this case, see Proposition PS and its corollary. Intuitively, this is the case if, for all concentration levels, costs of efforts producing alphas exceed the benefits of the produced alphas.

[12] The definition of manager ability in BG is different from ours. It corresponds to the magnitude of expected excess return (over the passive benchmark) earned on the first dollar actively managed by a fund. The reason that fund size can be unrelated to ability is, in BG as explained above, managers' indifference between larger fund sizes and higher fee levels. Differences in the definition of ability do not play a role here.

[13] Two online appendices are on the journal webpage.

Please cite this article as: D. Feldman, K. Saxena and J. Xu, Is the active fund management industry concentrated enough? Journal of Financial Economics, https://doi.org/10.1016/j.jfineco.2019.08.009

concentration on fund managers' effort levels, fund fees, fund performance, AFMI size, and potential direct benefits.

2.1. Setting

For brevity and parsimonious notation, we assume that variables and functions are real, continuous, and at least twice differentiable. Within a one-period market, there are two types of agents: fund managers of M funds, $M > 1$, and N investors, $N \geq 1$. Acting competitively, each manager sets a proportional management fee and chooses an effort level to maximize the fund expected net alpha to attract investments. In this section and Section 2.2, we consider the case in which infinitely many small mean-variance risk-averse investors (henceforth, risk-averse investors) allocate their investments to maximize their portfolios' Sharpe ratios. By *infinitely many small investors* we mean that $N \to \infty$ and with investors' finite wealth, their choices do not affect fund sizes. We also consider the case of infinitely many small risk-neutral investors (henceforth, risk-neutral investors).

Our model follows and builds on that of PS. In this partial equilibrium, the passive benchmark portfolio's returns are exogenously given and are unaffected by interactions between investors and managers. Managers' outperformance of the passive benchmark portfolio (i.e., gross alphas), could come at the expense of other investors, who could be noise traders, liquidity seekers, misinformed, or irrational.[14]

2.1.1. Fund alpha and the returns process

Following PS, $\mathbf{r_F}$, a vector of M funds' returns in excess of the riskless rate that investors receive, follows the regression model

$$\mathbf{r_F} = \boldsymbol{\alpha} + \boldsymbol{\beta} r_p + \mathbf{u}, \tag{1}$$

where $\mathbf{r_F}$ is an $M \times 1$ vector with elements $r_{F,i}$, $i = 1, \ldots, M$; $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\mathbf{u}$are $M \times 1$ vectors; $\boldsymbol{\alpha}$ is the vector of fund net alphas received by investors; and $\boldsymbol{\beta}$ is the vector of fund betas. The scalar r_P is the excess return on the passive benchmark portfolio, with mean μ_p and variance σ_p^2, and $\mathbf{u}$ is the residual vector, with elements that follow

$$u_i = x + \varepsilon_i, \quad i = 1, \ldots, M, \tag{2}$$

where ε_is are mean zero and variance σ_ε^2 idiosyncratic risks and are uncorrelated with each other, with x, and with r_p. The common factor x has mean zero and variance of σ_x^2 and is uncorrelated with r_p. The values of μ_p, σ_p^2, σ_ε^2, and σ_x^2 are strictly positive constants that are common knowledge to investors and managers.

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

$$\mathbf{r} \triangleq \boldsymbol{\alpha} + \mathbf{u}. \tag{3}$$

As in PS [see their Eqs. (2) and (3)], the factor structure in Eqs. (1)–(3) means that the benchmark-adjusted returns of AFMI funds are correlated. An economic rationale for a common component x in this factor structure is that similar opportunities are likely to be identified by AFMI funds, resulting in correlated benchmark-adjusted returns (see also PS, pp. 746–747).[15] Technically, this common component x is necessary to guarantee that investors in AFMI portfolios cannot enjoy expected net alphas without increasing their risk, the variance of their portfolios. (This is the case because they can, plausibly, well diversify the ε_is in their AFMI returns.) That is, had the common component x not existed, the risks associated with investing in AFMI funds could be fully diversified away by investing in many of them while retaining the benefits of their positive expected net alphas.

Each element in $\boldsymbol{\alpha}$ has the following structure:

$$\alpha_i = a - b\frac{S}{W} + A(e_i; H) - f_i, \tag{4}$$

where a and b are positive, unknown scalar parameters, where b is the industry level decreasing returns to scale rate; S is the aggregate size of the active management industry and is equal to the sum of all the funds' sizes (i.e., $S = \sum_{i=1}^{M} s_i$); W is the total wealth managed actively and passively and is equal to S plus the amount invested in the passive benchmark; $A(e_i; H)$ is the productivity of manager is proportional effort $e_i \in [0, \infty)$ to increase gross alpha under AFMI concentration H; and f_i is the proportional fee charged by manager i.

The expression for net alpha in PS corresponding to Eq. (4) does not contain the $A(e_i;\, H)$ term, which captures the alpha production function due to extra effort under AFMI concentration H. This is because PS focus on studying how investor beliefs about the unknown parameters a and b influence AFMI size. We build on their findings and study how AFMI concentration influences fund managers' incentives to exert costly effort, thereby influencing AFMI size and alpha.

2.1.2. Productivity of manager effort

We assume that $A(e_i;\, H)$, the productivity of effort under H, is the same across funds and has the following functional characteristics: zero for zero effort, increasing and concave in effort, increasing in AFMI concentration, and positive cross-partial derivatives with respect to effort and AFMI concentration. The assumption that links concentration to gross alpha is that the more concentrated AFMI is, the relatively more investment (mispriced) opportunities there are and the more marginally efficient is the use of industry resources.[16] Thus, managers can generate a higher increment in gross alpha for a given effort level e_i.

2.1.3. AFMI concentration

Our main analysis assumes that H is a known exogenous scalar parameter because it depends mainly on

[14] Please see the detailed discussion in PS (p. 749).

[15] For example, Garvey et al. (2017) decompose fund strategies into a combination of orthogonal and generic insights and suggest that many funds invest partly in orthogonal insights and partly in generic insights that are common across funds. Investing in multiple fund managers acts to concentrate risk into generic ideas.

[16] In a more concentrated market, if a fund manager controls most of the industry resources and develops advanced strategies to produce gross alphas, other funds can mimic this fund's strategy and produce higher gross alphas given a particular effort level. So this assumption is still valid when a dominant fund in the market controls the majority of resources.

some exogenous factors. For example, industrial organization theory posits that AFMI concentration depends not only on the number of incumbents, but also on threats of entry, activity-limiting regulation, and the competitiveness of related industries (see, for example, Claessens and Laeven, 2003). (In Section 2.4, we examine endogenous measures of AFMI concentration.) Without loss of generality, we assume that $H \in [0, 1)$. If $H = 0$, there are infinitely many small funds in the market, and the market is fully competitive. If $H = 1$, the market is monopolistic. If fund managers are competing (the case we consider), H belongs to $[0, 1)$.

2.1.4. Expected alpha and investors' information about unknown parameters

The parameters a and b in Eq. (4) are positive, unknown scalar parameters. The parameter a represents the expected return on an initial small fraction of wealth invested in active management, net of any proportional costs. The parameter b is the industry level decreasing returns to scale rate. As in PS, the first and second conditional moments of a and b are

$$\mathrm{E}\left(\begin{bmatrix} a \\ b \end{bmatrix} \middle| D\right) \triangleq \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} \tag{5}$$

and

$$\mathrm{var}\left(\begin{bmatrix} a \\ b \end{bmatrix} \middle| D\right) \triangleq \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix}, \tag{6}$$

where D denotes investors' information set.[17]

As we do not focus on the effects of σ_{ab} on the equilibrium, we assume that $\sigma_{ab} = 0$.[18] In other words, conditional on current information, we assume that how $\hat{a}$ deviates from a is unrelated to how $\hat{b}$ deviates from b. Finally, with f_i being a proportional management fee charged by manager i, the fund's expected net alpha is[19]

$$\mathrm{E}(\alpha_i|D) = \hat{a} - \hat{b}\frac{S}{W} + \hat{A}(e_i; H) - f_i. \tag{7}$$

2.1.5. Investor's problem

Let $\boldsymbol{\delta}_j$ denote the $M \times 1$ vector of weights that investor j places on the M funds, with elements $\boldsymbol{\delta}_{j,i}$, $i = 1, \ldots, M$. Thus, investor j's excess return is

$$r_j = \boldsymbol{\delta}_\mathbf{j}^\mathbf{T}\mathbf{r}_\mathbf{F} + \left(1 - \boldsymbol{\delta}_\mathbf{j}^\mathbf{T}\boldsymbol{\iota}_\mathbf{M}\right)r_p, \tag{8}$$

where $\boldsymbol{\iota}_\mathbf{M}$ is an $M \times 1$ vector with elements equal to one, and superscript T is a transpose operator. Following PS (p. 750 and footnote 7), we assume that all funds have beta loadings on the benchmark equal to one [i.e., $\boldsymbol{\beta}$, as defined in Eq. (1), fulfills, $\boldsymbol{\beta} = \boldsymbol{\iota}_\mathbf{M}$]. With funds' holding unit beta portfolios, the choice variable $\boldsymbol{\delta}_j$ represents investor j's exposure to the active part in the AFMI in excess of his or her holding of the passive benchmark portfolio. As in PS, this assumption allows parsimonious modeling of the active or passive choice.

Based on Eqs. (1) and (8), we have[20]

$$r_j = r_p + \boldsymbol{\delta}_\mathbf{j}^\mathbf{T}(\boldsymbol{\alpha} + \mathbf{u}). \tag{9}$$

Further, we have

$$\mathrm{E}\left(r_j|D\right) = \mu_p + \boldsymbol{\delta}_\mathbf{j}^\mathbf{T}\mathrm{E}(\boldsymbol{\alpha}|D), \quad \forall j, \tag{10}$$

and

$$\begin{aligned}\mathrm{Var}\left(r_j|D\right) &= \sigma_p^2 + \left[\sigma_a^2 + \sigma_x^2 + \sigma_b^2\left(\frac{S}{W}\right)^2\right]\left(\boldsymbol{\delta}_\mathbf{j}^\mathbf{T}\boldsymbol{\iota}_\mathbf{M}\right)^2 \\ &\quad + \sigma_\varepsilon^2\left(\boldsymbol{\delta}_\mathbf{j}^\mathbf{T}\boldsymbol{\delta}_\mathbf{j}\right), \quad \forall j.\end{aligned} \tag{11}$$

We first focus on the case of infinitely many ($N \to \infty$) small mean-variance risk-averse investors, none of whom can affect fund sizes. We also examine the case of infinitely many small risk-neutral investors, facilitating comparison with BG model. Investors' investment in the AFMI dilutes fund expected returns due to decreasing returns to scale in funds. Mean-variance risk-averse investors face risk-return tradeoffs in marginal allocations. Investor j's objective is to maximize the portfolio's Sharpe ratio by choosing portfolio weights, $\boldsymbol{\delta}_\mathbf{j}$, $j = 1, \ldots M$.

$$\underset{\boldsymbol{\delta}_\mathbf{j}}{\mathrm{Max}} \frac{\mathrm{E}\left(r_j|D\right)}{\sqrt{\mathrm{Var}\left(r_j|D\right)}}, \tag{12}$$

subject to

$$\boldsymbol{\delta}_\mathbf{j}^\mathbf{T}\boldsymbol{\iota}_\mathbf{M} \le 1 \tag{13}$$

and

$$\delta_{j,\,i} \ge 0, \quad \forall i. \tag{14}$$

The argument of the objective function in Eq. (12) is the ratio of Eqs. (10) and (11). Condition (13) is a form of wealth constraint, saying that investors cannot borrow from the passive benchmark to invest in the AFMI. Condition (14) says that there is no short sale of funds. Also, as we assume that there are no marginal diversification benefits across funds, we set the idiosyncratic risk of investor j's portfolio, $\sigma_\varepsilon^2\boldsymbol{\delta}_\mathbf{j}^\mathbf{T}\boldsymbol{\delta}_\mathbf{j}$ to be negligible (that is, zero) when

[17] See PS [p. 747, Eqs. (5) and (6)].

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

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

[20] Eqs. (8) and (9) are similar to Eqs. (10) and (11) in PS. However, our functional forms, represented by variables in these equations (such as $\alpha, \mathbf{r}_\mathbf{F}, \boldsymbol{\delta}_\mathbf{j}, r_j$), are different.

solving the optimization problem (12).[21] Because the equilibrium is symmetric, we have

$$\boldsymbol{\delta}_{\mathbf{j}}^{*\mathbf{T}}\boldsymbol{\iota}_{\boldsymbol{M}} = S/W, \forall j. \tag{15}$$

2.1.6. Fund manager's problem

f_i is the proportional fee charged by manager i. The manager sets this fee considering its effect on the fund's size. The manager also chooses the level of costly proportional effort to exert in order to find mispriced assets and produce additional gross alpha using $A(e_i;\ H)$, which depends on AFMI concentration H. We define manager i's average (per dollar) cost to produce alpha as $C^i(e_i,\ s_i;\ H)$. Therefore, manager i's economic profit is

$$s_i\left(f_i - C^i(e_i, s_i; H)\right) \tag{16}$$

and, for fund i to survive,

$$f_i - C^i(e_i, s_i; H) \geq 0. \tag{17}$$

We assume that average cost functions, $C^i(e_i,\ s_i;\ H)$, contain three independent positive scalar components: $c_{0,\ i}$, the average cost for fund i to operate in the market before receiving investment and before manager i spends effort; $c_{1,i}s_i$, the average cost related to fund i's size, s_i; and $c_{2,i}(e_i;\ H)$, the average cost of manager i's effort under a particular AFMI concentration.[22] That is,

$$C^i(e_i, s_i; H) = c_{0,\,i} + c_{1,\,i}s_i + c_{2,\,i}(e_i; H). \tag{18}$$

Eq. (18) is also manager i's per dollar cost function, which, when multiplied by the fund size, s_i, gives his or her total cost function. The coefficient $c_{1,i}$, then, induces a nonlinear (quadratic) increase in manager i's total cost function, making it convex in s_i and representing the extent of decreasing returns to scale in funds' gross alpha production. This fund cost model is consistent with that of BG, who assume decreasing returns to scale at the fund level.

Simplifying, we assume that $c_{0,\ i}$s, and $c_2(e_i;\ H)$s are the same across funds (we, thus, drop the subscript i) but that $c_{1,\ i}$s are different across funds. Differences in the fund-level decreasing returns to scale parameters $c_{1,i}$ measure differences in the rate at which managers' costs in generating gross alpha increase with size.

We now introduce two terms, an AFMI's individual manager skill and AFMI's aggregate skill. In our model, $c_{1,\ i}{}^{-1}$ is the source of heterogenous manager ability or skill. A more skilled manager is one who has lower total variable costs of active management for the same AUM and gross alpha. We define AFMI aggregate skill as the sum of individual managers' skills, $\sum_{i=1}^{M}(c_{1,\ i}{}^{-1})$. In our model, AFMI is more skilled when the sum of its mangers' skills is higher.

We show that higher AFMI's aggregate skill corresponds to higher AFMI size and that higher individual fund manager skill, relative to other managers, correspond to a higher relative size of their fund. (See Proposition RA1 and the discussion following Lemma RA1.)

We assume that the function $c_2(e_i;\ H)$ has zero for zero effort and is increasing and convex in effort.

The average cost function implies that as fund i's size, s_i, increases, manager i's average cost increases because larger trades are associated with larger price impacts and higher execution costs and because of other factors that create diseconomies of scale in operation. $c_{1,\ i}$ is the average cost sensitivity to fund i's size. Adding the three cost function components shows that the average cost function is increasing and convex in effort.

We do not specify whether costs are increasing or decreasing in concentration. When costs are decreasing in concentration, the advantage of higher concentration is twofold: more opportunities and lower costs. When costs do not change as a function of concentration, the advantage of increasing opportunities due to an increase in concentration is left unmitigated. Moreover, we show in Lemma RA1 that even increasing costs in concentration, for the plausible parameters set, perhaps do not fully mitigate the advantages of the increasing opportunities.

We assume no fixed costs for several reasons. First, fixed costs are lower in comparison with the costs that we model and, we believe, do not affect our analysis.[23] Moreover, the larger funds, with fixed costs that are relatively lower, determine AFMI concentration. In addition, as we focus on modeling decreasing returns to scale in gross alpha production, positive fixed costs could obfuscate this property.[24]

With these assumptions, manager i's problem is

$$\underset{e_i,\, f_i}{\operatorname{Max}}\, s_i\left(f_i - C^i(e_i, s_i; H)\right) \tag{19}$$

subject to $e_i \geq 0$ and $f_i \geq 0$.[25]

2.1.7. Information structure

We follow the information structure of PS when relevant and extend it, in spirit, to the new model structure that we introduce here. Model parameters and functional forms are common knowledge to managers and investors, with the following exceptions. The values of a and b are unknown, but their first two moments specifications are common knowledge. The values of the parameters of managers' cost functions and alpha production functions are private information (manager i's knows his or her cost and production functions). Sensitivities (assumptions on

[21] Here, too, we adopt PS notation. Note that $\sigma_\varepsilon^2\boldsymbol{\delta}_{\mathbf{j}}^{\mathbf{T}}\boldsymbol{\delta}_{\mathbf{j}} = \sigma_\varepsilon^2\boldsymbol{\delta}_{\mathbf{j}}^{\mathbf{T}}\mathbf{I}\boldsymbol{\delta}_{\mathbf{j}}$, where $\mathbf{I}$ is an $M \times M$ identity matrix and $\sigma_\varepsilon^2\mathbf{I}$ stands for the covariance matrix.

[22] To simplify our model, we assume no interaction between effort levels and size in the average cost function because it is unlikely that fund size affects managers' per dollar effort. We also assume no interaction between concentration and size in the average cost function because it is unlikely that concentration affects managers' average cost sensitivities to fund sizes. Even if these interacting effects do exist, they tend to be small in comparison with effects of other terms in the average cost function.

[23] Fixed costs to manage funds, such as registration fees and equipment expenditure, are usually small in comparison with variable costs related to employees' salaries and managers' compensation.

[24] A nonzero fixed cost and decreasing returns to scale in gross alpha production (i.e., costs component that are increasing and convex in fund size) would induce an average cost function that is U-shaped in fund size. Thus, under some cases this can induce instances of increasing returns to scale in gross alpha production.

[25] For simplicity and brevity, we omit the condition in Eq. (17) from the problem statement as it is implied by the optimization and, thus, is not necessary.

derivatives) of cost functions and alpha production functions are common knowledge.

2.2. Equilibrium

For the AFMI equilibrium, we begin by discussing why the manager's optimization problem is equivalent to the problem of maximizing the expected net alpha. The presentation of the latter problem helps to conveniently describe the AFMI equilibrium. Our risk-averse investors invest only in funds that offer the highest expected net alphas. Fund managers, in turn, compete over expected net alphas to attract investments. Manager i's problem becomes

$$\underset{e_i,\, f_i}{\text{Max}}\, \mathrm{E}(\alpha_i|D) \tag{20}$$

subject to $f_i - C^i(e_i, s_i; H) \geq 0, \quad e_i \geq 0, \quad \text{and} \quad f_i \geq 0.$

The online Main Appendix provides proof of the managers' maximization problems equivalence and also shows that solving these problems leads to an AFMI unique Nash equilibrium. The proof intuition is as follows. Under competition, funds that offer lower expected net alphas lose all investments. The possibility (threat) that other managers will improve their expected net alphas induces all managers to maximize expected net alphas to survive. Thus, funds offer similar expected net alphas in a unique Nash equilibrium. We are able to demonstrate that this aspect of the equilibrium, which is similar to that in PS, holds under various concentration levels, endogenous costly effort levels, and endogenous fund sizes.

To further study the equilibrium, we define the direct benefits of effort function of manager i as

$$B(e_i; H) \triangleq A(e_i; H) - c_2(e_i; H), \quad \forall i, \tag{21}$$

$B(e_i;\, H)$ captures the direct benefit from effort exerted in active fund management, in terms of increase in gross alpha production minus the effort cost. We should interpret benefits generally, allowing them to be positive or negative. Whether manager i's marginal direct benefits of initial effort are positive [i.e., $B_{e_i}(0; H) > 0, \forall H$] is an important condition affecting the equilibrium. If this condition is not met, our equilibrium becomes the one in PS (see Proposition PS in Section 2.3). Whether the sensitivity of manager i's direct benefits, at optimal effort, is positive [i.e., $\frac{dB(e_i^*;H)}{dH} > 0$] or not is also an important condition affecting the equilibrium.[26]

Also, the AFMI active search for net alphas could have indirect effects that we do not model here. It could drive security prices toward their true values, induce firms to improve governance and performance and to reduce agency costs, and induce transfer of wealth from less productive firms or investors to more productive ones. Here, as in the literature, gross alphas are zero-sum. (See, for example PS, pp. 748–750, including footnote 6, and references therein, and our footnote 4.) This is the case regardless of whether any manager's direct and or indirect benefits are nonzero or zero.

We are now ready to characterize the AFMI equilibrium for risk-averse investors, induced by managers choosing optimal effort levels and optimal fees. That is, we characterize AFMI equilibrium expected net alphas, Sharpe ratios, effort levels, fee rates, direct benefits of effort, AFMI size, and fund market shares. In Proposition RA0, we formally state the AFMI Nash equilibrium. In Proposition RA1, we describe the qualitative properties of this equilibrium. In Lemma RA1, we describe technical properties of the AFMI equilibrium, used to prove Propositions RA0 and RA1.

In Propositions RA0–RA4, PS, and RN1, Corollaries to Propositions RA2 and PS, and Lemma RA1, we assume infinitely many mean-variance risk-averse investors.

We first define AFMI's equilibrium optimal allocations. Let $\mathbf{e}^*$ be an $M \times 1$ vector with managers' optimal effort allocations, e_i^*, $\mathbf{f}^*$ be an $M \times 1$ vector with managers' optimal fee allocations, f_i^*, and $\boldsymbol{\delta}^*$ be an $M \times N$ matrix with vectors of investors' optimal wealth weights allocations to funds, $\boldsymbol{\delta}_{\mathbf{j}}^*$.

Proposition RA0. Unique Nash equilibrium.

There exists an AFMI unique Nash equilibrium, $\{\mathbf{e}^*, \mathbf{f}^*, \boldsymbol{\delta}^*\}$.

Proof of Proposition RA0. See the online Main Appendix. The proof intuition is below.

Proposition RA1. *For manager* i, $i = 1, \ldots, M$, *if initial effort inputs generate positive direct benefits of effort, then, in the AFMI equilibrium induced by managers choosing optimal effort-fee combinations,* (e_i^*, f_i^*), *AFMI size,* S/W, *and AFMI fund market shares,* s_i/S, $\forall i$, *adjust such that the following eight properties are satisfied.*

1. *Competition drives managers' economic profits to zero, so they can charge only break-even fees.*
2. *Higher managers' aggregate skill results in higher AFMI size.*
3. *Higher manager's relative skill results in higher AFMI fund market share (relative fund size).*
4. *Managers offer the same market competitive expected net alphas.*
5. *Managers offer the same market competitive Sharpe ratios.*
6. *Investors hold the same AFMI portfolio weights (which are proportional to AFMI fund sizes).*
7. *Equilibrium effort levels and fees are the same across funds.*
8. *Equilibrium AFMI's direct benefits of effort are the same across funds.*

Proof of Proposition RA1. See the online Main Appendix. The proof intuition is below.

To prove Proposition RA1, we use the seven results of Lemma RA1, which characterize more specific properties of the AFMI equilibrium.

Lemma RA1. *For manager* i, $i = 1, \ldots, M$, *if initial effort inputs generate positive direct benefits of effort* [i.e., $B_{e_i}(0;\, H) > 0,\ \forall H$]*, the equilibrium, induced by managers, choosing optimal effort-fee combinations* (e_i^*, f_i^*)*, has the following seven properties.*

[26] See, for example, Proposition RA3 and the proof intuition to it.

1. *Fees are equal to costs:*

$$f_i^* - C^i(e_i^*, s_i; H) = 0, \ \forall i. \tag{22}$$

2. *The impact of marginal efforts on gross alpha is set equal to the marginal average costs of effort. Thus manager i's marginal direct benefits of effort under the optimal effort level are zero).*

$$A_{e_i}\left(e_i^*; H\right) - c_{2e_i}\left(e_i^*; H\right) = B_{e_i}\left(e_i^*; H\right) = 0, \forall i. \tag{23}$$

3. *When concentration is higher, equilibrium optimal effort levels are higher (lower) if and only if higher concentration induces a larger (smaller) marginal effort impact on gross alphas than on costs. Or*

$$e_i^{*\prime}(H) \geq 0(<0)\, iff\, A_{e_i,H}\left(e_i^*; H\right) - c_{2e_i,H}\left(e_i^*; H\right) \geq 0(<0), \quad where \quad e_i^{*\prime}(H) \triangleq de_i^*/dH. \tag{24}$$

4. *Whether higher concentrations induce higher equilibrium optimal fees depends on whether they induce an increase in equilibrium AFMI sizes and whether they induce an increase in equilibrium optimal effort levels.*
5. *When concentrations are higher, equilibrium manager i's direct benefits of effort are higher (lower) if and only if higher concentrations induce a larger (smaller) impact on gross alphas than on costs.*
6. *Pairwise relative fund sizes, $s_i/s_j, \forall i,j$, are inversely proportional to their corresponding cost coefficients, $c_{1,i}/c_{1,j}, \forall i,j$ (where $c_{1,i}$ represents the intensity of fund-level decreasing returns to scale in gross alpha production).*
7. *AFMI fund market shares, s_i/S's, are $\frac{s_i}{S} = (c_{1,i}\sum_{j=1}^{M}(c_{1,j}{}^{-1}))^{-1}, \forall i$.*

Proof of Proposition RA1 and Lemma RA1. See the online Main Appendix.

The proof intuition of Propositions RA0, RA1 and Lemma RA1 is as follows.

Competing for investments, managers maximize fund expected net alphas by choosing optimal effort levels and fees, earning zero economic profits (break-even fees) in equilibrium. If managers increase fees, they would lower fund expected net alphas and lose all investments. If managers decrease fees, they would become insolvent, incurring negative cash flows (costs higher than fees). Deviating from equilibrium effort level would also induce a loss of investments (if decreasing effort) or insolvency (if increasing effort). Therefore, managers have no incentive to deviate.

Also, with no diversification benefits across funds, managers who attempt to provide higher expected net alphas attract investments. Consequently, due to decreasing returns to scale in performance, on the one hand, and increasing fund costs, on the other hand, alpha gains are more than mitigated by a (break-even) fees increase, resulting in an overall decrease in expected net alpha. Thus, in equilibrium, the allocation of investments, or fund sizes, set expected net alphas to be equal across funds. If fund managers cannot produce the AFMI highest expected net alpha, even for an infinitesimal fund size, they lose all investments and go out of the market.

In addition, as funds have the same expected net alphas, they have the same expected returns. As the source of fund returns' variance is the same across funds, the fund return variance is the same across funds. Therefore, managers offer the same competitive Sharpe ratio. Because investors cannot obtain a higher Sharpe ratio, they have no incentives to deviate.

These result in a unique Nash equilibrium in which neither investors nor managers have incentives to deviate from their chosen strategies.

If higher concentrations induce a higher (lower) marginal effort impact on gross alphas than a marginal effort impact on costs, managers optimally choose higher (lower) effort levels in producing fund net alphas. If higher concentrations induce higher equilibrium optimal effort levels, managers' costs are driven higher, resulting in higher break-even fees. In addition, higher concentrations have two effects on manager i's direct benefits of effort. First, they directly affect the levels of gross alphas production function and of costs, $A(e_i; H)$ and $c_2(e_i; H)$, being a parameter of each of these functions. Second, they change the equilibrium optimal effort levels, consequently changing the levels of gross alphas and costs. In equilibrium, the latter (net) effect is zero because managers keep increasing effort levels until the marginal effort impact on gross alphas is equal to the marginal effort impact on costs. Thus, the effect of higher concentration through effort on gross alphas is canceled out by its effects on costs. Therefore, in equilibrium (as the net second effect is zero), changes in the concentration level affect gross alphas and costs through the (direct) first effect only. Consequently, if higher concentration levels induce higher direct impacts on gross alphas than on costs, manager i's direct benefits of effort increase in concentration levels.

Managers' different costs of producing gross alphas (skills) induce different fund sizes in equilibrium. A separation exists between the determination processes of AFMI size (that is, AFMI's weight in total wealth, S/W) and AFMI fund market shares (that is, relative fund sizes within AFMI). The former is driven by managers' aggregate skill (cost) and the latter by managers' relative skills (costs). In other words, how investors weight the funds inside the AFMI, or investors' optimal AFMI portfolio, is unaffected by how investors weight the AFMI as a whole relative to the passive benchmark. This separation property facilitates later results.

For convenience in describing the equilibrium in Propositions RA2 and RA3, we define the equilibrium optimal expected net alphas of an initial marginal investment in the AFMI (when $S = 0$) as $X(e_i^*, H)$. Quantitatively,

$$X\left(e_i^*; H\right) \triangleq \hat{a} + A\left(e_i^*; H\right) - \left[c_0 + c_2\left(e_i^*; H\right)\right]. \tag{25}$$

For the AFMI to exist, we must have positive expected net alphas for initial infinitesimal investments into it:[27]

$$X\left(e_i^*; H\right) > 0, \ \forall H. \tag{26}$$

If inequality (26) is violated, investors receive no advantage in diverting funds from the passive index to the AFMI. Also,

[27] The condition in inequality (26) is equivalent to the condition that $a > 0$ in PS. See PS (p. 747) for further discussion and insights.

to offer meaningful results, we assume that initial marginal allocations of effort generate positive AFMI direct benefits of effort; that is,

$$B_{e_i}(0;H) > 0, \quad \forall i, \forall H, \tag{27}$$

such that the optimal effort e_i^* is positive, finite, and attainable, i.e., $B_{e_i}(e_i^*;H) = 0, e_i^* < K, \forall i, \forall H$ for some positive constant K. We focus on the case in which the optimal effort is positive.

As in PS (see their Proposition 2), the explicit analytic solutions for S/W are solutions of a cubic equation and are cumbersome. Proposition RA2 presents the cubic equation and its corollary presents properties of its solution.

Proposition RA2. *Equilibrium optimal allocations.*

For $i = 1, 2, \ldots, M$,

1. $\mathrm{E}(\alpha_i|D)|_{\{\mathbf{e}^*, \boldsymbol{f}^*, \boldsymbol{\delta}^*\}} > 0$; and
2. *when $N \to \infty$, the equilibrium optimal S/W is either one or a real positive solution (smaller than one) of the following first-order condition (a cubic equation) of the investors' problem [Eqs. (12)–(14)]. After we substitute $\boldsymbol{\delta}_\mathbf{j}^{*\mathbf{T}} \boldsymbol{\iota}_\mathbf{M} = S/W$,*

$$-\gamma\sigma_b^2\left(\frac{S}{W}\right)^3 - \left[\gamma\sigma_a^2 + \gamma\sigma_x^2 + \widehat{b} + \left(\sum_{i=1}^{M} c_{1,i}{}^{-1}\right)^{-1} W\right]\frac{S}{W} + X\left(e_i^*;H\right) = 0, \tag{28}$$

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

Proof of Proposition RA2. See the online Main Appendix.

The intuition of Proposition RA2 is as follows. Investors allocate investments to funds based on their risk-return tradeoffs. Investing wealth in the AFMI increases a portfolio's risk, so they choose to limit these investments, leaving $\mathrm{E}(\alpha_i|D)|_{\{\mathbf{e}^*, \mathbf{f}^*, \boldsymbol{\delta}^*\}} > 0$. The risk-return tradeoff of potentially investing the last dollar, the dollar that would drive fund expected net alphas to zero, is in the variance favor. That is, the marginal cost of risk, of investing this last dollar, is higher than the marginal benefit of the gained net alpha. This prevents optimizing risk-averse investors from allocating it to the AFMI, leaving fund expected net alphas to be positive. The properties of the cubic equation guarantee exactly one real positive root. If the positive root is larger than one, then $S/W = 1$.

We can now write the following corollary, characterizing AFMI equilibrium relations between performance and size, and between the rate of returns to scale decrease and size.

Corollary to Proposition RA2. For large enough W, such that $S/W < 1$.

1. *Higher equilibrium optimal expected net alphas of an initial marginal investment in the AFMI induce a larger equilibrium AFMI size relative to total wealth; and*
2. *A higher rate of decrease in aggregate AFMI's returns to scale [fund level and industry level; $\hat{b} + (\sum_{i=1}^{M} c_{1,i}{}^{-1})^{-1}W$] induces a smaller equilibrium AFMI size.*

Proof of Corollary to Proposition RA2. See the online Main Appendix.

The intuition of this corollary is as follows. A higher level of equilibrium optimal expected net alpha of an initial marginal investment, $X(e_i^*, H)$, attracts more investments to the AFMI. Also, $\hat{b}$ is the industry level expected decreasing returns to scale rate at the coming from the alpha production function, based on current information, whereas $(\sum_{i=1}^{M} c_{1,i}{}^{-1})^{-1}W$ can be regarded as the equilibrium decreasing returns to scale factor coming from AFMI managers' costs of alpha production (calculated by aggregating all the fund average cost sensitivities to size, $c_{1,i}$'s). The latter decreasing returns to scale factor, $(\sum_{i=1}^{M} c_{1,i}{}^{-1})^{-1}W$, is inversely proportional to AFMI's aggregate skill. Thus, the factor $\hat{b} + (\sum_{i=1}^{M} c_{1,i}{}^{-1})^{-1}W$ can be regarded as the combined decreasing returns to scale factor. Investors invest less in funds if the effect of decreasing returns to scale is stronger in the AFMI.

Proposition RA3 offers comparative statics, which underlie our main empirical analysis.

Proposition RA3. *AFMI size and expected net alphas sensitivities to concentration.*

When $S/W < 1$, we have the following results.[28]

1. *Higher concentration induces larger (smaller) equilibrium AFMI size and higher (lower) equilibrium expected net alphas if and only if higher concentration induces a larger (smaller) impact on gross alphas than on costs.*
2. *Concave, in concentration, equilibrium direct benefits of effort function indicates concave, in concentration, equilibrium AFMI size. (Convex, in concentration, equilibrium AFMI size indicates convex, in concentration, equilibrium direct benefits of effort function.)*
3. *Concave, in concentration, equilibrium expected net alphas indicates concave, in concentration, equilibrium direct benefit function. (Convex, in concentration, equilibrium direct benefit function indicates convex, in concentration, equilibrium expected net alphas.)*

Proof of Proposition RA3. See the online Main Appendix.

The intuition behind Proposition RA3(1) is as follows. A change of H affects expected net alpha, $\mathrm{E}(\alpha_i|D)|_{\{\mathbf{e}^*, \mathbf{f}^*, \boldsymbol{\delta}^*\}}$, in two stages. In the first stage, if a higher H induces a larger (smaller) impact on gross alphas than on costs, it increases (decreases) managers' ability to produce expected net alphas, thereby increasing (decreasing) the level of expected net alphas produced. In the second stage, investors react to the increase (decrease) in fund expected net alphas by increasing (decreasing) investment levels in funds, consequently decreasing (increasing) expected net alphas, due to decreasing returns to scale. The risk-return tradeoff of risk-averse investors makes their reaction to changes

[28] Where $S/W = 1$, it is the case that S/W is unrelated to industry concentration, higher concentration induces higher (lower) equilibrium expected net alphas if and only if higher concentration induces a larger (smaller) impact on gross alphas than on costs, and equilibrium expected net alphas are concave (convex), in concentration, if and only if the equilibrium direct benefit function is concave (convex), in concentration.

in fund expected net alphas less intense. That is, they subdue their additional investments to funds when inferring higher fund expected net alphas due to risk increase, and they limit their reduction in investments to funds when observing lower fund expected net alphas due to risk decrease.

The first stage and second stage, the latter as affected by risk attitudes, result in a change of AFMI's optimal effort level. AFMI's new optimal effort level, in turn, affects both the level of alpha production and the effort costs producing it. The overall outcome depends on the relative sensitivity, to concentration, of the alpha production function, on the one hand, and of the effort cost function, on the other. We formally show that whether a higher H increases the equilibrium expected net alpha, $\mathrm{E}(\alpha_i|D)|_{\{\mathbf{e}^*,\mathbf{f}^*,\delta^*\}}$, depends on whether it has a larger impact on gross alphas than on the costs producing it [i.e., the sign of $d\mathrm{E}(\alpha_i|D)/dH|_{\{\mathbf{e}^*,\mathbf{f}^*,\delta^*\}}$ depends on the sign of $\frac{dB(e_i^*;H)}{dH} = A_H(e_i^*;H) - c_{2H}(e_i^*;H)$, as shown in Lemma RA1(5)].[29] Thus, a higher H induces a larger expected net alpha if and only if it induces higher equilibrium direct benefits, $B(e_i^*;H)$. This explains the expected net alpha part of Proposition RA3(1).

If a higher H induces a larger (smaller) impact on gross alphas than on costs, then it attracts more (less) investments to the AFMI [if investors have additional wealth to allocate to funds (i.e., $S/W < 1$)]. This explains the size part of Proposition RA3(1).

Examining the second-order effects of concentration on size, we first note that changes in H that induce a larger S/W result in a larger allocation to AFMI funds and, in turn, in a higher investors' overall portfolio risk. Mean-variance risk-averse investors facing risk-return tradeoffs respond to an increase in marginal portfolio risks, holding other parameters constant, by optimally lowering investment in funds. Thus, how changes in H affect changes in equilibrium S/W depend on how changes in H affect this risk-return tradeoff. The implications for the second-order derivative $d^2(S/W)/dH^2$ are in the proof of Proposition RA3, which expresses this tradeoff analytically by identifying $d^2(S/W)/dH^2$ as a sum of two addends. The first addend is negative (positive) if the direct benefits function is concave (convex) in H, and the second one is always negative. This shows that a concave $B(e_i^*;H)$ in H implies an S/W concave in H and that a convex S/W in H implies a convex $B(e_i^*;H)$ in H. This explains Proposition RA3(2).

Examining the second-order effects of concentration on expected net alphas, we show that, as H changes, the change of marginal $\mathrm{E}(\alpha_i|D)|_{\{\mathbf{e}^*,\mathbf{f}^*,\delta^*\}}$ (i.e., $d^2\mathrm{E}(\alpha_i|D)/dH^2|_{\{\mathbf{e}^*,\mathbf{f}^*,\delta^*\}}$) is positively proportional to the change of marginal $B(e_i^*;H)$, i.e., $d^2B(e_i^*;H)/dH^2$, plus an adjustment term that captures the effects of risk. This adjustment term ensures that, holding all other parameters constant, if investors' marginal portfolios risks of investing in funds are higher, investors optimally invest less in funds. In doing so, they exert a smaller negative impact on expected net alphas. A higher H thus induces a higher marginal $\mathrm{E}(\alpha_i|D)|_{\{\mathbf{e}^*,\mathbf{f}^*,\delta^*\}}$. When $d^2B(e_i^*;H)/dH^2$ is positive, $d^2\mathrm{E}(\alpha_i|D)/dH^2|_{\{\mathbf{e}^*,\mathbf{f}^*,\delta^*\}}$ must be positive, whereas if $d^2\mathrm{E}(\alpha_i|D)/dH^2|_{\{\mathbf{e}^*,\mathbf{f}^*,\delta^*\}}$ is negative, $d^2B(e_i^*;H)/dH^2$ must be negative. This explains Proposition RA3(3).

When investors have no additional wealth to allocate to funds, i.e., $S/W = 1$, they exert no impact on marginal $\mathrm{E}(\alpha_i|D)|_{\{\mathbf{e}^*,\mathbf{f}^*,\delta^*\}}$, making the marginal equilibrium optimal expected net alphas depend only on the effect of H on managers' ability to produce net alphas.

2.3. *Relation to Berk and Green (2004) and Pastor and Stambaugh (2012)*

Our model follows the pivotal works of BG and PS in several respects. Central features of our model include industry-level decreasing returns to scale and risk-averse investors, as in PS, as well as fund-level decreasing returns to scale, as in BG. We highlight the main differences between our model and those of PS and BG. We discuss the special cases in which their model and ours overlap and we obtain results similar to theirs.

2.3.1. *Relation to Pastor and Stambaugh (2012)*

While PS model and compare expected net alpha and AFMI size, within two extreme regimes, a fully competitive equilibrium ($M \to \infty$) and a monopolistic equilibrium ($M = 1$), we model and study tradeoffs across a continuum of AFMI concentration levels for any given M, where $M > 1$. This concentration–alpha relation in our model is a mechanism distinct from and additional to the AFMI size–alpha relation in PS. In our model, the concentration–alpha relation also exists when controlling for the size of the industry (or the growth of the industry). Analytically, this is the case because concentration affects the optimal level of effort by affecting effort productivity [the third addend of the right side of Eq. (7)] and the cost of effort [Eq. (18)].

In our model, heterogeneous fund-level decreasing returns to scale are required for making the AFMI concentration nontrivial by allowing funds to have heterogeneous sizes. (PS's model identifies the AFMI equilibrium elegantly, without the need to specify fund-level heterogeneity in fund size.) Fund-level decreasing returns does not directly influence alpha in our model. It influences alpha only via effort, concentration, and size. If we do not model effort, our model with only fund-level decreasing returns will not generate the concentration–alpha relation. The gist of the argument is that introducing fund heterogeneity would affect AFMI size and expected net alphas only if, in aggregate, it affects the industry's alpha production. Such aggregate effects are fully captured within the industry returns to scale structure (as in PS). The only difference is that in PS the industry returns to scale parameter is $\hat{b}$, while in our model the aggregate industry returns to scale parameter is $\hat{b} + (\sum_{i=1}^{M} c_{1,i}{}^{-1})^{-1}W$ (see the Corollary to Proposition RA2). Introducing fund-level heterogeneity that does not have industry aggregate effect would not affect the AFMI size and expected net alphas. In other words, changes in AFMI concentration due to introducing fund-level heterogeneity, without all effort effects, either are captured by industry returns to scale effects or have no effect.

[29] This total derivative of AFMI direct benefits with respect to H is the same as its partial derivative with respect to H.

Our model becomes similar to the one in PS in our special case in which neither AFMI concentration nor managers' effort affects managers' search productivity for mispriced assets. Analytically, effort does not affect alpha production in our model if the third addend of the right side of Eq. (7) does not exist and if we abandon our cost function, Eq. (18), in favor of defining fund fees to be the net of fund management costs. Even when, in our model, the search productivity for mispriced assets depends on effort, a special case of parameter values leads to a solution in which optimal allocated effort is zero. We report the conditions for this special case in Proposition PS and its corollary. Intuitively, this is the case if, for all concentration levels, costs of efforts to improve alpha production exceed the benefits of the resulting increase in alphas. That is, market conditions are insufficiently conducive to launching a costly search for favorable investment opportunities. Thus, optimally, no extra effort is exerted, and our model results resemble those in PS.

Proposition PS. *For manager i, $i = 1, 2, \ldots, M$ if initial effort inputs generate non-positive AFMI direct benefits of effort [*i.e., *equilibrium optimal proportional effort levels e_i^* are zero [*i.e., *$e_i^* = 0$, $\forall i$], and the optimal proportional fee f_i^* equals the average cost of operating funds $c_0 + c_{1,i}s_i$* (i.e., *$f_i^* = c_{0,i} + c_{1,i}s_i$, $\forall i$].*

Corollary to Proposition PS. *Under the conditions in Proposition PS, the equilibrium here resembles the one in PS. That is, effort is not exerted, and managers optimally choose not to charge fees above break-even costs. Eq. (4) becomes*

$$\alpha_i = a - b\frac{S}{W} - f_i, \tag{29}$$

identical to Eq. (8) in PS.

Proof of Proposition PS. See the online Main Appendix.

This corollary says that AFMI concentration will not influence the AFMI size or AFMI expected net alpha if, for all concentration levels, the optimal effort of fund managers is zero, given their trade-off between productivity and costs. The industry-level decreasing returns to scale mechanism of PS will still function and generate a negative relation between AFMI size and AFMI expected net alpha. While the latter effect represents how managers' ability to outperform passive benchmarks declines with AFMI size, the former effect represents how a manager's incentives to exert individual effort to outperform passive benchmarks are influenced by AFMI's concentration (for the same AFMI size and number of managers).

2.3.2. Relation to Berk and Green (2004)

As in PS, our baseline model assumes risk-averse investors. This assumption produces positive AFMI expected net alpha. When investors are risk-averse, expected net alphas are positive because investors require compensation to bear the risk of investing in active funds. BG do not solve the investor' optimization problem and fix expected net alpha to be zero by invoking the assumption that non-benchmark risk can be completely diversified away across many funds (see also the discussion in PS, p. 775). This feature of the BG equilibrium is compatible with the case, in our model, in which infinitely many small risk-neutral investors compete, and the size of the fund endogenously adjusts to make the gross alpha equal to the fee so that expected net alpha is always zero. In the case of risk-neutral investors, our model, and the corresponding model in PS, produces zero AFMI expected net alpha. This is formally stated in Proposition RN1.

In the case of risk-neutral investors, the AFMI Nash equilibrium (Proposition RA0) and the equilibrium characterizations of Proposition RA1 and Lemma RA1 hold as well. As the proofs are highly similar, for brevity, we omit them here.[30] We describe the different equilibrium characteristics for which investors are risk-neutral.

Proposition RN1. *For $N \to \infty$ risk-neutral investors, equilibrium optimal allocations induce AFMI size to be*

$$\boldsymbol{\delta}_{\mathbf{j}}^{*\mathbf{T}}\boldsymbol{\iota}_{\mathbf{M}} = S/W = \min\left\{\frac{X\left(e_i^*, H\right)}{\hat{b} + \left(\sum_{i=1}^M c_{1,i}^{-1}\right)^{-1} W}, 1\right\}, \forall j, \tag{30}$$

and equilibrium expected net alphas to be

$$\mathrm{E}(\alpha_i|D)|_{\{\mathbf{e}^*, \boldsymbol{\delta}^*\}} = 0, \quad \textit{where} \quad S/W < 1, \tag{31}$$

and

$$\mathrm{E}(\alpha_i|D)|_{\{\mathbf{e}^*, \mathbf{f}^*, \boldsymbol{\delta}^*\}} = X\left(e_i^*, H\right) - \left[\hat{b} + \left(\sum_{i=1}^M c_{1,i}^{-1}\right)^{-1} W\right] \geq 0,$$
$$\textit{where} \quad S/W = 1. \tag{32}$$

Proof of Proposition RN1. See the online Main Appendix.

Risk-neutral investors keep investing in the AFMI as long as they expect to earn positive net alphas. Eventually, either they drive alphas to zero and have $\mathrm{E}(\alpha_i|D) = 0$ and $S/W \leq 1$ or they run out of funds and have $\mathrm{E}(\alpha_i|D) \geq 0$ and $S/W = 1$. If some of the wealth is passively managed ($S/W < 1$), then, irrespective of AFMI concentration or AFMI size, equilibrium expected net alpha will be zero [$\mathrm{E}(\alpha_i|D) = 0$]. This result parallels the results of BG and the risk-neutral case (with perfect competition) of PS. Our additional result is that, even in the risk-neutral case, the AFMI size will depend on AFMI concentration through its effect on $X(e_i^*, H)$. The intuition is that, even though the AFMI expected net alphas are driven to zero, higher AFMI concentration incentivizes managers to invest more effort for finding mispriced assets. (Expected net alphas are driven to zero along a path of search for investment opportunities when managers exert more effort.) This increase in optimal effort increases the AFMI size at which investors are indifferent between investing an additional dollar with the AFMI and the passive benchmark.

Another difference between BG and our model is the source of heterogenous manager ability. In BG, the source of heterogenous manager ability is the expected excess return (over the passive benchmark) earned on the first dollar actively managed by a fund. In our model, this quantity is the same across funds. The source of heterogenous manager ability in our model is the fund-level decreasing

[30] For risk-neutral investors, Proposition RA1(2) follows directly from Eq. (A49).

returns to scale parameter $c_{1,i}$, which measures the rate at which the manager's costs in generating gross alpha increase with size. A more skilled manager in our model is one who has lower total variable costs of active management for the same AUM and gross alpha. Therefore, in our equilibrium, managers with the lower fund-level decreasing returns to scale parameter ($c_{1,i}$) have more AUM. While the fund-level decreasing returns to scale parameter in BG influences the fund's AUM in the same way [large parameter corresponds to a smaller fund, see their Eq. (27)], they assume this parameter to be the same across funds.

Our choice of modeling heterogeneity in $c_{1,i}$ (as a source of heterogenous manager ability) enables us to obtain heterogeneity in equilibrium fund sizes as well as a positive equilibrium fee charged by managers (see Lemma RA1). In the competitive equilibrium of PS, the fee is zero. If the fee (net of costs) were, instead, equal to some positive value in PS, then any fund manager would set an infinitesimally lower fee to attract all investment from other funds. We model costs explicitly, and in equilibrium fees compensate managers for their costs [fund managers charge (positive) break-even fees], which include a component related to size and a component related to effort (see Lemma RA1).

Fund managers in the BG model are indifferent to the fee they charge as long as two conditions are met (see their Section II.A): (1) this fee is less than the hypothetical fee they could charge to maximize their compensation, and (2) they can expand their fund by investing in the passive benchmark (i.e., closet indexing). They show that, under their assumptions, managers are indifferent between large AUM with a small fee and small AUM with a large fee as long as their profits stay the same. In BG's framework, fund managers can choose their AUM independently of competing fund managers' skills. This assumption allows fund size to be arbitrary and unrelated to skill.

We model competition between managers with different returns to scale parameters. In our equilibrium, this competition for finite AUM results in zero profits and break-even fees charged by managers, as well as in relative fund sizes that correspond to the relative rates at which fund-level returns to scale decrease (our measure of ability). In other words, the AFMI fund size distribution reflects the distribution of ability. This implies a tight link between skill of an AFMI fund manager ($c_{1,i}{}^{-1}$), the manager's fund size, the manager fund's market share, and the manager fund's net alpha. In Proposition RA4, we describe how a decrease in skill (increase in the cost, $c_{1,i}$) of one fund manager leads to a decrease in the manager's fund size and an increase in the fund sizes of competing managers. Such a decrease in an AFMI manager skill will also have an industry-wide effect of a decreasing the AFMI net alpha. Such effects of changes in skill of one fund manager on the fund sizes and alphas of other fund managers are absent in the BG equilibrium.

Proposition RA4. Relation between skill, market share and net alpha.

When $S/W \le 1$, a decrease (increase) in manager i's skill, $c_{1,i}{}^{-1}$, while manager j's skill, $c_{1,j}{}^{-1}$, $\forall j \neq i$ is unchanged, induces

1. *A decrease (increase) in s_i/S and an increase (decrease) in s_j/S, $\forall j \neq i$ and*
2. *A decrease (increase) in $\mathrm{E}(\alpha_i|D)|_{\{\mathbf{e}^*,\mathbf{f}^*,\delta^*\}}$ and a decrease (increase) in $\mathrm{E}(\alpha_j|D)|_{\{\mathbf{e}^*,\mathbf{f}^*,\delta^*\}}$, $\forall j \neq i$.*

Proof of Proposition RA4. See the online Main Appendix.

According to Proposition RA4, a decrease in manager *i*'s skill leads to a decrease in *i*'s market share, s_i/S. Some of the assets that fund *i* loses are invested in all other funds, thereby increasing the market share of all other funds.

A higher skill (lower $c_{1,i}$) affects $\mathrm{E}(\alpha_i|D)|_{\{\mathbf{e}^*,\mathbf{f}^*,\delta^*\}}$ in two stages. In the first stage, it decreases manager *i*'s average cost and, thus, induces higher fund expected net alphas. As manager *i* offers a higher fund expected net alpha, investments shift into fund *i* from other funds, making all those fund expected net alphas higher due to decreasing returns to scale at fund level. At the second stage, an increase in fund expected net alphas attracts investments into the AFMI, which drives down fund expected net alphas due to decreasing returns to scale at industry level. When $N \to \infty$ and $S/W < 1$, investors' portfolio risks increase (decrease) when they invest more (less) in the AFMI. Thus, they subdue AFMI investments increases when observing an increase in fund expected net alphas, and they limit investment reductions when observing a decrease in fund expected net alphas. Investors' risk aversion mitigates the countered effect at the second stage and makes the first stage's effect dominant.

When $S/W = 1$, investors have no additional wealth to allocate to funds, so their investments have no impact on marginal equilibrium optimal expected net alphas, causing the first stage's effect to dominate.

In summary, this proposition describes novel results, compared with BG, arising from modeling competition between managers with different fund-level returns to scale parameters.

2.4. Endogeneity in measures of AFMI concentration

Our model allows for an endogenous measure of AFMI concentration. Modeling an endogenous measure of concentration facilitates the use of available and prevalent empirical measures. If we define concentration to be the Herfindahl–Hirschman Index (*HHI*), which is the sum of market shares squared, then, *HHI*is endogenous to our model.[31] Using funds' equilibrium market share, as identified in Lemma RA1, we can write the equilibrium AFMI concentration HHI^* as

$$HHI^* \triangleq \sum_{i=1}^{M}\left(\frac{s_i}{S}\right)^2 = \sum_{i=1}^{M}\left(c_{1,i}\sum_{j=1}^{M}\left(c_{1,j}{}^{-1}\right)\right)^{-2}. \quad (33)$$

[31] In an *M*-fund AFMI, for example, the HHI could have values between the highest concentration, 1, where one of the funds captures practically all the market share, and the lowest concentration, $1/M$, where market shares are evenly divided. That is, in an *M*-funds' market, HHI $\in [\frac{1}{M}, 1)$.⟨*/END*⟩In our model, predictions 1 and 2 follow if and only if direct benefits of effort increase with concentration. Therefore, empirical support for these predictions also provides additional evidence in support of our assumption that direct benefits of effort increase with concentration. Prediction 3 relies on an alternative assumption described in Footnote 6.

HHI^* is determined by $c_{1,i}$s. Depending on the size of $c_{1,i}$ relative to that $c_{1,j}$, $\forall j \neq i$, an increase in $c_{1,i}$, holding $c_{1,j}$, $\forall j \neq i$ constant, increases or decreases HHI^*. For further analysis and discussion of the endogenous measure of AFMI concentration, see the online Main Appendix.

In general, we expect the theoretical concentration level in our framework to be influenced by industry characteristics such as regulations, transaction costs, tax rates, barriers to entry, and funds' idiosyncratic outcomes, in addition to funds' cost sensitivity to size (i.e., $c_{1,i}$s). For example, Hong (2018) finds that a policy reform in Hong Kong (the 2012 Employee Choice Arrangement) substantially increased competition in the fund management industry by dramatically expanding the choices for pension plan participants from an average of 11 funds to more than four hundred funds. In these cases, the concentration level can change even when all the cost sensitivities (or fund manager skill) are constant. We do not model the various determinants of concentration levels and simply assume them to be exogenous. As long as real-world concentration is not exactly determined by the $c_{1,i}$s (or any other exogenous parameter of our model), we are back to the case that when concentration is exogenous (that is, has an exogenous component), our predictions remain unaltered regarding the relation between changes in exogenous AFMI concentration level, the equilibrium fund expected net alphas, and AFMI size.

3. Empirical predictions and test method

In this section, we describe key empirical predictions that our theoretical model generates, followed by our data and methodology to test these predictions.

3.1. Empirical predictions

Underlying our empirical predictions is the theoretical scenario in which an increase in the AFMI concentration has a larger effect on the availability of mispriced investment opportunities than on any associated costs of exploiting these opportunities (higher concentration induces a larger marginal effort impact on gross alpha than on costs). For instance, costs associated with an increase in effort can be staff's increase in compensation (endogenously determined). We assume that these costs are less than the value added to the firm due to the increase in effort. We claim this is a reasonable assumption. [See, for example, Ibert et al. (2017), who find concavity of managerial compensation in firm revenue and weak sensitivity of pay to performance.] This assumption means that the direct benefits of effort [as defined in Eq. (21)] increase with concentration. Based on this scenario, we predict that a higher concentration level is associated with larger AFMI size [Proposition RA3(1)], higher AFMI net alpha [Proposition RA3(1)], and higher AFMI effort [Lemma RA1(3)].[32]

Effort is largely unobservable. Even the salaries of managers are difficult to observe. Therefore, it is difficult to test the third prediction directly. Instead, we provide indirect suggestive evidence in favor of our model. We use aggregate AFMI active share and tracking error as proxies for effort.[33] These proxies are likely to be correlated with effort because any attempt to outperform the benchmark must involve taking positions that are different from the benchmark (Cremers and Petajisto, 2009). One reason active share can be uncorrelated with effort is that fund managers could jam the signal in active share by taking uninformed bets to increase their perceived active share, generating a false sense of truly active management (Brown and Davies, 2017). Such signal jamming behavior is more likely to be an issue if a measure of active share is tied with fund manager incentives, which is not likely in our sample period (our sample for active share tests ends in 2009, the same year as the publication of Cremers and Petajisto, 2009. Also, information asymmetry and its associated signal jamming in Brown and Davies (2017) pertains to fund-level active share. Our prediction is for aggregate AFMI-level effort (not relative effort of funds), where information asymmetry and signal jamming" are less important.

Our model also has a second-order prediction that we test, that is, the AFMI net alphas and AFMI size are both either concave or convex in AFMI concentration [Proposition RA3(2) and Proposition RA3(3)].

3.2. Data

We obtain our active fund data from Morningstar Direct. Our sample contains 1374 actively managed US (domestic) equity-only mutual funds from January 1979 to December 2014. We include both open-ended and closed-ended funds. We exclude index funds, enhanced index funds, funds of funds, international funds, industry sector funds, real estate funds, and other non-equity funds. While we analyze fund-level data, the Morningstar data are often at the fund share class level. We use the fund identification provided by Morningstar to aggregate fund share class-level information to fund-level information (because many mutual funds offer multiple share classes, which represent claims on the same underlying assets but have different fee structures; see also PST, footnote 11). Because we use a five-year rolling window to estimate fund net alphas, we require each of our active equity mutual funds to have at least ten years of monthly return observations.[34] The online Empirical Appendix details the keywords and filters

[32] We also omit some rare cases in which more than five years' return observations are missing in a ten-year window. This improves the estimation of our style-matching model with a five-year rolling estimation window.

[33] More generally, these two measures have different possible concentration-level intervals. Independent of the number (greater than one) of industry firms, under the $NHHI$ all concentration distributions are on the interval [0, 1). Under the HHI, an M-firm industry ($M > 1$) has a concentration distribution on the interval $[\frac{1}{M}, 1)$. At concentration levels near one, one of the funds captures practically all the market share, and at the lowest concentration, $1/M$, market shares are evenly divided. A three-firm industry would induce a concentration-level distribution on [0.33, 1), and a four-firm industry would induce a concentration distribution on [0.25, 1).

[34] As in PST, we do not use the Fama–French factors (Fama and French, 1993) as our benchmark. PST (p. 31) note, "The Fama–French factors are popular in mutual fund studies because their returns are freely avail-

used to construct our sample of actively managed US equity mutual funds.

We also obtain data on index funds from Morningstar. These index funds, which we use to benchmark the performance of active mutual funds, include those from the Morningstar Institutional Categories of Small Core (Vanguard Small Cap Index), Large Core (EQ/Common Stock Index Portfolio), and Standard & Poor's 500 Tracking (Vanguard 500 Index). We require index funds to have no missing observations in our sample period. All the fund returns are net of administrative and management fees and other costs taken out of fund assets.

We obtain quarterly data on fund-level active share and tracking error from the Antti Petajisto website (www.petajisto.net/data.html). Petajisto (2013) contains a description of how these data are constructed.

3.3. Variable definitions

We now define how we measure key variables in our analysis.

AFMI Size represents the active equity mutual fund AFMI size relative to total stock market capitalization. Our measure of AFMI size (SoW_t, size over wealth) is the sum of the net AUM of US active equity funds in our sample, divided by stock market capitalization, which is defined as the sum of all individual stocks' market capitalization in Center for Research in Security Prices (CRSP) (share code of 10 or 11) in the same month.

AFMI Active Share is the quarterly average active share of active funds (in Petajisto's database). We exclude index funds and enhanced index funds in the database when identifying active funds.

AFMI Tracking Error is the AFMI tracking error as the average tracking error of active funds (excluding index funds and enhanced index funds) in that quarter.

$MS_{i,t}$ is the *market share* of fund i at time t, measured by the fund's AUM at time t over the total AFMI's AUM at time t. We use fund net asset value data at a monthly frequency to calculate $MS_{i,t}$.

AFMI Concentration, following the literature, is measured using three indices (see, for example, Berger and Hannan, 1989; Geroski, 1990; Berger, 1995; Goldberg and Rai, 1996; Nickell, 1996; Berger et al., 1999; Cremers et al., 2008; Giroud and Mueller, 2011). All three indices are constructed using two variables: $MS_{i,t}$ and m_t, where $m_t$is the number of funds at time t. We calculate m_t at a monthly frequency.

Our three measures of AFMI concentration are *HHI*, the normalized *HHI* (*NHHI*), and the sum of the first five largest funds' market shares (5*FI*).

able. Yet the Fama–French factors are not obvious choices because they are long-short portfolios whose returns cannot be costlessly achieved by mutual fund managers or investors." In addition, Cremers, Petajisto, and Zitzewitz (2012) and Grinblatt and Saxena (2017) argue that the Fama–French model produces biased assessments of alpha. To avoid such problems and remain consistent with our model in which investors compare active funds with a traded passive benchmark, we use Sharpe style analysis and identify an appropriate traded benchmark for each mutual fund in our sample.

For the *HHI*,

$$HHI_t = \sum_{i}^{m_t} MS_{i,t}^2. \tag{34}$$

The *HHI* is a commonly used measure of concentration (see, for example, Cremers et al., 2008 and Giroud and Mueller, 2011) and is well grounded in theory (see Tirole, 1988, pp. 221–223). As the value of the *HHI* is related to the number of funds (m_t), for a robustness check, we also use two measures not related to the number of funds to measure AFMI concentration: the *NHHI* used by Cremers et al. (2008) and the 5*FI*, another common measure of AFMI concentration.

For the *NHHI*,

$$NHHI_t = \frac{HHI_t - \frac{1}{m_t}}{1 - \frac{1}{m_t}}. \tag{35}$$

The *NHHI* induces similarity in possible concentration-level distributions. For example, it is zero for an industry in which all firms have equal market shares, regardless of whether it has three or four firms. In contrast, the *HHI* is 0.33 for the three-firm industry with equal market shares and 0.25 for the four-firm industry.[35]

For the five-fund index,

$$5FI_t = \sum_{i=1}^{5} MS_{i,t}. \tag{36}$$

Our model provides a relation between these measures of AFMI concentration and fund-level decreasing returns to scale parameters. Using the *HHI* as the measure of AFMI concentration *HHI** and fund equilibrium market shares, as identified in Lemma RA1, we can write an expression for equilibrium AFMI concentration *HHI** [Eq. (33)]. (Similar relations can be obtained for *NHHI* and five-fund index.) This expression shows a relation between *HHI** and the $c_{1,i}$'s. The question of which of the two quantities is exogenous, or whether both are determined together in equilibrium, is a complex one that is beyond the scope of this paper (see Section 2.4 for a discussion of this issue). We simply use these empirical measures as proxies for the true level of AFMI concentration (or competition) and do not enforce restrictions between cost parameters and *HHI*.

Fund net alpha. Our measure of net alpha ($\alpha_{i,t}$) is the difference between a fund's net return and the net return on the benchmark we assign to the fund. The benchmark against which we judge a fund's net alpha is a set of (traded) index funds selected using style analysis (Sharpe, 1992).[36] These index funds are intended to represent the next-best investment opportunity available to investors as

[35] To be a valid instrument of $\overline{MS_{i,t-1}}$, $\underline{MS_{i,t-1}}$ must satisfy the relevance and exclusion conditions. The relevance condition is likely to hold because both $\overline{MS_{i,t-1}}$ and $\underline{MS_{i,t-1}}$ are derived from $MS_{i,t-1}$ and are, thus, likely to be closely related. The exclusion condition is also likely to hold because the backward-looking information in $\underline{MS_{i,t-1}}$ is unlikely to be helpful in predicting the forward-looking net alpha information in $\overline{\varepsilon_{i,t}}$, where $\overline{\varepsilon_{i,t}}$ is the residual in the RD method. We correct the second-stage standard error estimates of β_1 by incorporating the estimation errors from the first-stage regression.

[36] To be a valid instrument of $\overline{MS_{i,t-1}}$, $\underline{MS_{i,t-1}}$ must satisfy the relevance and exclusion conditions. The relevance condition is likely to hold because

a tradable passive index (Berk and van Binsbergen, 2015). In our model, we assume that a single passive benchmark exists and is common knowledge to investors and managers. In the theoretical analysis, we make this assumption for parsimony. Relaxing it does not alter the key insights from our model. In Section 4, we allow for multiple benchmarks and match each active equity mutual fund to a set of tradable index funds that reasonably replicate passive alternatives available to an average mutual fund investor. We estimate the equation

$$R_{i,t} = \alpha_{i,t} + b^1_{i,t}F^1_t + b^2_{i,t}F^2_t + \cdots + b^n_{i,t}F^n_t, \quad (37)$$

where the indices i and t represent the fund and time indices, and n indicates the number of tradable index funds in the market. $R_{i,t}$ is the return net of management fee of a fund, and F^1_t through F^n_t are the returns net of management fees of tradable index funds in different asset classes. We also allow for a risk-free fund by including the CRSP Fama–French risk-free rate as a potential benchmark. We treat the index fundsF^1_t through F^n_tas a basis fund set that can be used to replicate the returns on any passive benchmarks used by mutual fund investors.

We perform this analysis on a rolling basis, using returns from months $(t-60)$ to $(t-1)$, in order to avoid a look-ahead bias. For each active fund in our sample, we identify coefficients, $b^1_{i,t}$ to $b^n_{i,t}$, that minimize the variance of the residuals, which are also the tracking errors between the active fund return and a corresponding passive benchmark portfolio return (Sharpe, 1992). These coefficients are constrained to be between zero and one (we do not allow short selling), and their sum is constrained to be one. These coefficients identify the portfolio weights, on our basis index fund set, that provide the estimated passive benchmark portfolio for a fund. This portfolio gives the minimum tracking error. Our empirical design of identifying passive benchmarks using matching tradable index funds fits our theoretical structure, which assumes the appropriate passive benchmarks for each fund.

To calculate a fund's net alphas in month t, we subtract the returns on the identified set of passive portfolios (the style benchmark) for month t from the active equity fund's returns in month t [see Eq. (37)]. This provides us with fund net alphas in each month for each fund.

To evaluate the robustness of our results, we use an alternative method to measure fund net alphas. This method addresses the possibility that traded index funds do not capture unobserved risk factors that drive excess returns. Errors in our set of passive benchmarks or our matching strategy can result in net alphas that measure exposure to unobserved common risk factors instead of fund manager performance. Using the method developed by Connor and Korajczyk (1988), we estimate unobserved common factors in our estimated fund net alphas using the principal components (PC) of our estimated fund net alphas series. We use these estimated PC factors to control for unobserved common factors in fund net alphas. We regress (without a constant term) each fund's net alphas on the first two PC factors. We refer to the residuals of these regressions as PC-adjusted fund net alphas and use them as the dependent variable in our robustness analysis. See the online Empirical Appendix for details on how we calculate PC-adjusted fund net alphas.

We control for fund-level decreasing returns to scale by adding lagged fund size as a control. Following PST, we measure *Fund size* as the fund's AUM at the end of the previous month, inflated to December 2014 dollars by using the ratio of the total stock market capitalization in December 2014 to its value at the end of the previous month. They argue that this is a reasonable way to measure the limitations on a fund due to its size. It captures the size of the fund relative to the universe of stocks that the fund can buy.

3.4. Methodology

We analyze the impact of AFMI concentration on AFMI size, AFMI active share, and tracking error (predictions 1 and 3) at industry level. We analyze the impact of AFMI concentration on AFMI alpha (prediction 2) using fund-level data to control for potential effects of market share (fund size divided by total AFMI size) on performance and an associated omitted-variable and finite sample bias (see PST). We study second-order predictions using both industry-level data and fund-level data as prediction 4 requires that AFMI net alphas (measured using fund-level data) and AFMI size (measured using industry-level data) are both either concave or convex in AFMI concentration.

3.4.1. Industry-level analysis

Using monthly data, in analyzing the relation between AFMI size and AFMI concentration, we use vector autoregression (VAR). The main equation in the VAR system is

$$SoW_t = b_0 + b_1 SoW_{t-1} + b_2 HHI_{t-1} + b_3 HHI^2_{t-1} + e_t, \quad (38)$$

Where SoW_t is AFMI size and $e_{\ t}$ represents regression residuals. In the VAR system, we also have equations in which HHI_t depends on HHI_{t-1} and SoW_{t-1} and HHI^2_t depends on HHI^2_{t-1}.

In addition, we use two effort proxies, active share (AS_t) and tracking error (TE_t), which are likely to represent different dimensions of effort. According to Cremers and Petajisto (2009), the active share measure emphasizes stock selection, and tracking error volatility emphasizes bets on systematic risk factors. Therefore, we include the effects of both these dimensions in testing prediction 3, using VAR:

$$AS_t = b_{01} + b_{11} AS_{t-1} + b_{11} TE_{t-1} + b_{31} HHI_{t-1} + e_{1t} \quad (39)$$

$$TE_t = b_{02} + b_{12} AS_{t-1} + b_{12} TE_{t-1} + b_{32} HHI_{t-1} + e_{2t}, \quad (40)$$

where $e_{\ 1t}$ and e_{2t} represent regression residuals. In the VAR system, we also have the equation in which HHI_t depends on HHI_{t-1}, AS_{t-1}, and TE_{t-1}.

both $\overline{MS_{i,t-1}}$ and $MS_{i,t-1}$ are derived from $MS_{i,t-1}$ and are, thus, likely to be closely related. The exclusion condition is also likely to hold because the backward-looking information in $\overline{MS_{i,t-1}}$ is unlikely to be helpful in predicting the forward-looking net alpha information in $\overline{\varepsilon_{i,t}}$, where $\overline{\varepsilon_{i,t}}$ is the residual in the RD method. We correct the second-stage standard error estimates of β_1 by incorporating the estimation errors from the first-stage regression.

3.4.2. Fund-level analysis

We follow PST's methodology to control for omitted-variable and finite-sample bias in our alpha analysis. The omitted-variable problem arises from the cross-sectional variation in performance that is due to differences in skill across funds. PST note that fund fixed effects can control for this heterogeneity as long as fund skill is time-invariant. However, adding fund fixed effects introduces finite-sample bias due to the positive contemporaneous correlation between changes in fund size and unexpected fund returns. To avoid these biases, we use the PST recursive demeaning (RD) estimator. We estimate the effects of a fund's market share (β_1), AFMI concentration (β_2 and β_3), and AFMI size (β_4) on fund net alphas using the panel regression

$$\overline{\alpha_{i,t}} = \beta_1 \overline{MS_{i,t-1}} + \beta_2 \overline{H_{t-1}} + \beta_3 \overline{H^2_{t-1}} + \beta_4 \overline{SoW_{t-1}} + \overline{\varepsilon_{i,t}}. \tag{41}$$

The bar above the variables denotes forward-demeaned variables:

$$\overline{\alpha_{i,t}} = \alpha_{i,t} - \frac{1}{T_i - t + 1} \sum_{s=t}^{T_i} \alpha_{i,s}, \tag{42}$$

$$\overline{MS_{i,t}} = MS_{i,t} - \frac{1}{T_i - t + 1} \sum_{s=t}^{T_i} MS_{i,s}, \tag{43}$$

$$\overline{SoW_t} = SoW_t - \frac{1}{T_i - t + 1} \sum_{s=t}^{T_i} SoW_s, \tag{44}$$

$$\overline{HHI_t} = HHI_t - \frac{1}{T_i - t + 1} \sum_{s=t}^{T_i} HHI_s, \tag{45}$$

and

$$\overline{HHI^2_t} = HHI^2_t - \frac{1}{T_i - t + 1} \sum_{s=t}^{T_i} HHI^2_t, \tag{46}$$

where T_i is the number of time series observations of fund i. We run robustness checks by replacing HHI with $NHHI$ and with $5FI$.

The RD method in Eq. (41) can control for the fund fixed effect. We include market share as a control, not only because the equilibrium market share provides information on a fund's cost sensitivity to fund size (Lemma RA1), but also because empirical studies show a linear relation between changes in market share and fund performance (Spiegel and Zhang, 2013) and use it as a firm-level market power measure (e.g., Berger et al., 1999 and Nickell, 1996). Endogeneity (reverse causality) could exist between AFMI shares and fund net alphas because when fund net alphas are higher, corresponding asset values increase and funds attract investments, both leading to a higher market share. This endogeneity issue could bias our results. Following PST, we address this issue using an instrumental variable method.[37] In the first stage, we regress $\overline{MS_{i,t-1}}$ (recursively forward-demeaned market share) on $\underline{MS_{i,t-1}}$ (recursively backward-demeaned market share) without a constant term. In the second stage, we use the fitted value from the first stage to run Eq. (41), where

$$\underline{MS_{i,t}} = MS_{i,t} - \frac{1}{t-1} \sum_{s=1}^{t-1} MS_{i,s}. \tag{47}$$

There is no reason to believe that individual fund net alphas, which are fund-level variables, are endogenous to industry-level measures such as AFMI concentration ratios (see, for example, footnote 17 of PST). Thus, to test the concentration–alpha relation, we do not use a backward-demeaned instrument. We just use the recursive forward-demeaned AFMI concentration ratios in Eq. (41).

4. Empirical results

Table 1 reports the summary statistics. Monthly fund net alphas are positive on average but exhibit a wide variation. We also report summary statistics of the fit of our passive benchmark-matching method using R-squared, which is measured as

$$Rsqr_{i,t} = 1 - \frac{Var(\alpha_{i,t})}{Var(R_{i,t})}, \tag{48}$$

where $Var(.)$ denotes variance. On average, our style-matching model fits well with an average R-squared of 0.86 and a standard deviation of about 0.12. The summary statistics of AFMI size (total AFMI funds' net assets divided by stock market capitalization) and fund sizes in December 2014 dollars (funds' net assets divided by stock market capitalization in the same month, multiplied by the stock market capitalization in December 2014) are similar to the sample in PST.

The number of active equity mutual funds in our sample increases over time and the AFMI concentration measures, such as the HHI, $NHHI$, and $5FI$, with fluctuations, tend to decrease over time. Fig. 1 shows the HHI value from January 1984 to December 2014. Before 1990, the HHI value was relatively high, fluctuating from 0.02 to 0.03. After that, it continued decreasing. In the most recent years, it has reached 0.006, which is around a quarter of the values before 1990. This figure shows that the concentration of the US active equity mutual fund market decreased substantially. Alternative AFMI concentration measures, such as $NHHI$ and $5FI$, show similar trends.

Because our sample differs from PST, we check for any alarming systematic differences by evaluating the returns to scale relation in our sample. In unreported results, we find results consistent with PST's: fund net alpha is significantly negatively associated with lagged AFMI size and is negatively (but insignificantly) associated with lagged fund size. The results suggest decreasing returns to scale at industry level.

[37] To be a valid instrument of $\overline{MS_{i,t-1}}$, $\underline{MS_{i,t-1}}$ must satisfy the relevance and exclusion conditions. The relevance condition is likely to hold because both $\overline{MS_{i,t-1}}$ and $\underline{MS_{i,t-1}}$ are derived from $MS_{i,t-1}$ and are, thus, likely to be closely related. The exclusion condition is also likely to hold because the backward-looking information in $\underline{MS_{i,t-1}}$ is unlikely to be helpful in predicting the forward-looking net alpha information in $\overline{\varepsilon_{i,t}}$, where $\overline{\varepsilon_{i,t}}$ is the residual in the RD method. We correct the second-stage standard error estimates of β_1 by incorporating the estimation errors from the first-stage regression.

Please cite this article as: D. Feldman, K. Saxena and J. Xu, Is the active fund management industry concentrated enough? Journal of Financial Economics, https://doi.org/10.1016/j.jfineco.2019.08.009

Table 1
Statistical summary.
Our sample period is from January 1979 to December 2014, and monthly data are used. Panel A reports the summary statistics for fund-level data, and Panel B reports those for industry-level data. *Fund net return* and *fund net alpha* are in percentages, and both are net of administrative and management fees and other costs taken out of fund assets. *HHI, NHHI*, and *5FI* are Herfindahl–Hirschman index, normalized Herfindahl–Hirschman index, and five-fund index, respectively. *MS* is fund market share, calculated as a fund's net assets under management (AUM) divided by the sum of all funds' net AUM in the same month. *SoW* is AFMI size, calculated as the sum of funds' net AUM divided by the stock market capitalization in the same month. *AS* is AFMI active share, measured as the average active share of active funds (in Antti Petajisto's database, www.petajisto.net/data.html) in a quarter. *TE* is AFMI tracking error, calculated as the average tracking error of active funds in a quarter. *Style-matching model R-squared, MS, HHI, NHHI, 5FI, AS*, and *TE* are in decimals. *Fund size* is measured in $100 million and is equal to the fund's total net AUM, divided by the stock market capitalization in the same month and multiplied by the stock market capitalization in December 2014. *Number of funds* is in units.

Variable	Number of observations	Mean	Standard deviation	Percentile 1st	25th	50th	75th	99th
Panel A: Fund-level data								
Fund net return (percent)	321,456	0.8736	5.1508	−14.4922	−1.7976	1.2998	3.8907	13.0053
Fund net alpha (percent)	246,553	0.0349	1.9499	−5.4465	−0.8570	0.0215	0.9156	5.5982
Style-matching model R-squared (decimal)	246,557	0.8607	0.1175	0.4223	0.8178	0.8953	0.9408	0.9894
Fund size (hundreds of millions of December 2014 dollars)	314,083	28.7796	95.3306	0.0399	1.3833	5.5718	20.1835	416.9203
MS (decimal)	314,083	0.0012	0.0041	0.0000	0.0000	0.0002	0.0007	0.0185
Panel B: Industry-level data								
SoW (decimal)	432	0.0982	0.0591	0.0200	0.0389	0.1035	0.1638	0.1801
Number of funds	432	850.2	659.5	86.0	249.0	677.5	1468.5	2126.0
HHI (decimal)	432	0.0191	0.0230	0.0061	0.0101	0.0157	0.0243	0.0382
NHHI (decimal)	432	0.0157	0.0139	0.0057	0.0094	0.0141	0.0201	0.0269
5FI (decimal)	432	0.2166	0.0765	0.1240	0.1640	0.1986	0.2650	0.3438
AS (decimal)	119	0.8349	0.0416	0.7620	0.7980	0.8440	0.8740	0.8940
TE (decimal)	119	0.0755	0.0260	0.0382	0.0606	0.0707	0.0868	0.1954

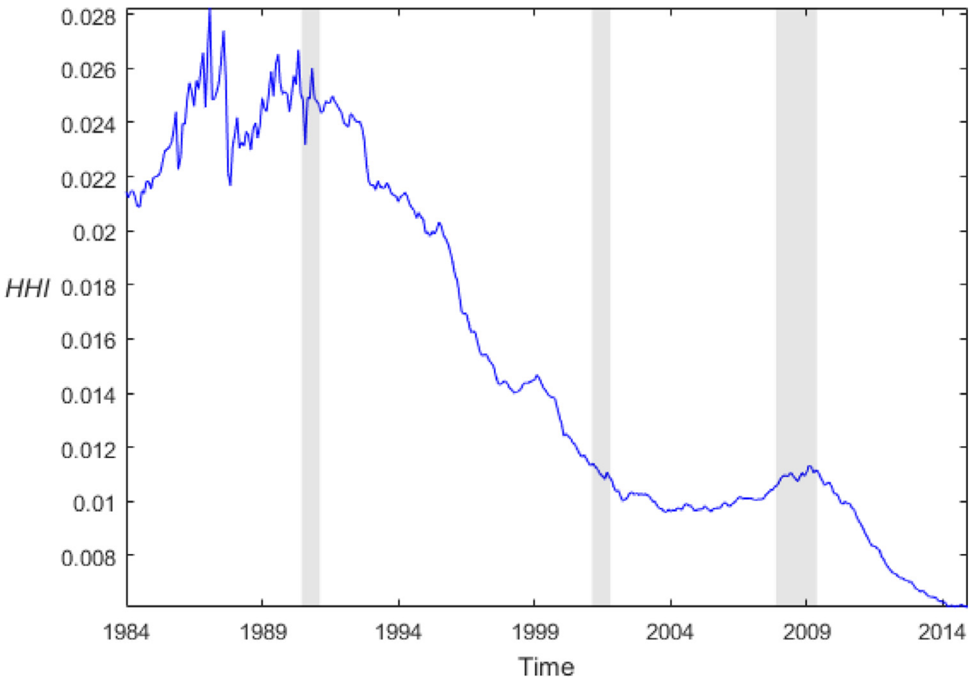

Fig. 1. Herfindahl–Hirschman Index (*HHI*) value January 1984–December 2014.
This figure shows the monthly *HHI* values from January 1984 to December 2014. The *HHI* is calculated as the sum of the funds' market shares squared, where each fund's market share is calculated as the fund's net assets under management (AUM) divided by the sum of all the funds' net AUM. The *HHI*'s value is in decimals. The gray bars represent the recession periods.

Table 2
Industry-level analysis: active fund management industry (AFMI) size and AFMI concentration.

This table reports the results of the main equations in various vector autoregression models, in which AFMI size, *SoW*, is the dependent variable. The sample period is from January 1979 to December 2014, and monthly data are used. *SoW* is the sum of funds' net assets under management divided by the stock market capitalization in the same month. *HHI, NHHI*, and 5*FI* are Herfindahl–Hirschman index, normalized Herfindahl–Hirschman index, and five-fund index, respectively, and HHI^2, $NHHI^2$, and $5FI^2$ are their squared terms. Panels A, B, and C report the results of using *HHI, NHHI*, and 5*FI* as the concentration measures, respectively. *Time trend* is set to be one for January 1979 and to increase by one each month. Small sample-adjusted standard errors are used and presented in parentheses. ***, **, and * represent the 1%, 5%, and 10% significant level in a two-tail *t*-test, respectively.

	SoW (1)	*SoW* (2)	*SoW* (3)	*SoW* (4)
Panel A: HHI results				
Lagged SoW	1.0033***	1.0052***	0.9952***	0.8772***
	(0.0014)	(0.0035)	(0.0063)	(0.0293)
Lagged HHI	0.0287***	0.0491*	0.0754**	0.9341***
	(0.0037)	(0.0278)	(0.0310)	(0.1126)
Lagged HHI^2		−0.1361*	−0.2044**	−2.4601***
		(0.0761)	(0.0846)	(0.3003)
Time trend			0.0000*	
			(0.0000)	
Constant	−0.0005**	−0.0009	−0.0017*	0.0155***
	(0.0002)	(0.0008)	(0.0009)	(0.0054)
Year dummies	No	No	No	Yes
Number of observations	431	431	431	431
R-squared	0.999	0.999	0.999	0.999
Panel B: NHHI results				
Lagged SoW	1.0040***	1.0097***	0.9999***	0.8660***
	(0.0015)	(0.0029)	(0.0064)	(0.0286)
Lagged NHHI	0.0497***	0.1328***	0.1425***	1.0085***
	(0.0063)	(0.0335)	(0.0341)	(0.1148)
Lagged $NHHI^2$		−0.6050***	−0.6446***	−4.2975***
		(0.1493)	(0.1515)	(0.4981)
Time trend			0.0000	
			(0.0000)	
Constant	−0.0008***	−0.0024***	−0.0025***	0.0175***
	(0.0002)	(0.0007)	(0.0008)	(0.0052)
Year dummies	No	No	No	Yes
Number of observations	431	431	431	431
R-squared	0.999	0.999	0.999	0.999
Panel C: 5FI results				
Lagged SoW	1.0090***	1.0052***	0.9961***	0.8526***
	(0.0020)	(0.0032)	(0.0062)	(0.0291)
Lagged 5FI	0.0109***	0.0122**	0.0149***	0.1353***
	(0.0015)	(0.0052)	(0.0055)	(0.0208)
Lagged $5FI^2$		−0.0114***	−0.0132***	−0.0977***
		(0.0043)	(0.0045)	(0.0156)
Time trend			0.0000	
			(0.0000)	
Constant	−0.0029***	−0.0022*	−0.0028**	0.0101
	(0.0005)	(0.0012)	(0.0013)	(0.0062)
Year dummies	No	No	No	Yes
Number of observations	431	431	431	431
R-squared	0.999	0.999	0.999	0.999

We begin our empirical analysis by evaluating the relation between AFMI size and AFMI concentration. The results of the main equations of the VARs are shown in Table 2. The first column of each model specification shows how AFMI size is positively associated with *HHI*. The result of interest in this table is that AFMI size is significantly positively associated with lagged *HHI* (Model Specification 1) and is significantly negatively associated with the second order of lagged *HHI* (Model Specification 2). If we further include a time trend or year dummies into the model, we find consistent results (Model Specifications 3 and 4). That is, AFMI size is increasing and concave in AFMI concentration. In Panels B and C, we analyze the sensitivity of our results to alternative measures of AFMI concentration: *NHHI* and 5*FI*. We generally find consistent results. Thus, we conclude that the data supports prediction 1. From our model's perspective, the positive relation between AFMI size and AFMI concentration indicates that higher AFMI concentration levels, on average, increase gross alphas more than they increase effort costs.

Next, we evaluate the relation between fund net alphas and AFMI concentration. The results using the RD method are shown in Table 3. Panel A reports the results using fund net alpha as the dependent variable. In the first two columns, we find that the coefficient of the first-order term of lagged *HHI* is significantly positive and the coefficient

Table 3
Fund-level analysis: fund net alpha and active fund management industry (AFMI) concentration.

This table reports the results of our recursive demeaning panel regression model. Results are presented using *fund net alpha* in columns (1)–(5) and *PC-adjusted fund net alpha* (adjusted by the first two principal components of fund net alphas) in columns (6)–(8), as the dependent variables. *SoW* is AFMI size, calculated as the sum of funds' net assets under management (AUM) divided by the stock market capitalization in the same month. *MS* is fund market share, calculated as a fund's net AUM divided by the sum of all funds' net AUM in the same month. *HHI, NHHI*, and 5*FI* are Herfindahl–Hirschman index, normalized Herfindahl–Hirschman index, and five-fund index, respectively, and HHI^2, $NHHI^2$, and $5FI^2$ are their squared terms. Panels A, B, and C report the results of using *HHI, NHHI*, and *5FI* as the concentration measures, respectively. The unit of coefficients is percentage. Standard errors are clustered by fund and presented in parentheses. ***, **, and * represent the 1%, 5%, and 10% significant level in a two-tail *t*-test, respectively.

	Fund net alpha					*PC-adjusted fund net alpha*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: HHI results								
Lagged HHI	6.5277***	40.0796***	39.1271**	35.3591***	34.8497**	2.5033***	11.8816***	9.4626**
	(1.0362)	(4.8085)	(17.7560)	(4.6913)	(17.4796)	(0.8485)	(4.3336)	(4.1665)
Lagged HHI^2		−1110.4260***	−1081.2070*	−1402.0231***	−1394.8592**		−310.3817**	−459.8112***
		(156.2080)	(589.6182)	(184.4615)	(700.8128)		(142.7130)	(161.0837)
Lagged MS			−12.0701		−15.1453			
			(23.6434)		(24.8670)			
Lagged SoW				−1.8946***	−1.9440			−0.9709***
				(0.3977)	(1.4847)			(0.2800)
Number of observations	245,178	245,178	239,537	245,178	239,537	245,179	245,179	245,179
R-squared	0.0002	0.0004	0.0004	0.0006	0.0006	0.0000	0.0001	0.0002
Adjusted *R*-squared	0.0002	0.0004	0.0004	0.0006	0.0006	0.0000	0.0001	0.0002
Panel B: NHHI results								
Lagged NHHI	7.8880***	46.4840***	45.7624**	44.3478***	44.2149**	2.9822***	12.4672**	11.4107**
	(1.2448)	(5.4655)	(20.1084)	(5.4142)	(20.0634)	(0.9912)	(4.9857)	(4.9085)
Lagged $NHHI^2$		−1468.6397***	−1452.2241*	−1884.2091***	−1917.8119**		−360.9219*	−566.4512***
		(204.1014)	(766.9316)	(244.4729)	(922.9258)		(187.9899)	(213.4361)
Lagged MS			−11.3222		−14.6674			
			(23.4579)		(24.8045)			
Lagged SoW				−1.6200***	−1.7313			−0.8012***
				(0.3354)	(1.2683)			(0.2422)
Number of observations	245,178	245,178	239,537	245,178	239,537	245,179	245,179	245,179
R-squared	0.0002	0.0004	0.0004	0.0006	0.0006	0.0000	0.0001	0.0002
Adjusted *R*-squared	0.0002	0.0004	0.0004	0.0006	0.0006	0.0000	0.0001	0.0001
Panel C: 5FI results								
Lagged 5FI	0.7589***	5.1426***	5.1090	6.9152***	7.0271**	0.2624**	1.3057*	2.1864***
	(0.1499)	(0.8780)	(3.2665)	(0.9451)	(3.5546)	(0.1118)	(0.7587)	(0.8358)
Lagged $5FI^2$		−11.3079***	−11.2617	−20.0470***	−20.4280*		−2.6911	−7.0332***
		(2.2138)	(8.3831)	(2.8461)	(10.8268)		(1.9786)	(2.4875)
Lagged MS			−9.9791		−14.8748			
			(23.0337)		(24.8087)			
Lagged SoW				−1.8969***	−1.9487			−0.9425***
				(0.3570)	(1.3386)			(0.2559)
Number of observations	245,178	245,178	239,537	245,178	239,537	245,179	245,179	245,179
R-squared	0.0001	0.0002	0.0002	0.0006	0.0006	0.0000	0.0000	0.0002
Adjusted *R*-squared	0.0001	0.0002	0.0002	0.0006	0.0006	0.0000	0.0000	0.0002

of the second-order term is significantly negative. This result is robust to including lagged market share and lagged AFMI size as controls. This suggests that the effect of concentration is distinct from the effect of decreasing returns to scale at the fund and industry levels. To control for the possibility of unaccounted common factors in the estimated net alphas, we also use PC-adjusted fund net alphas as the dependent variable (Panel B) and find similar results.

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

Table 4 analyzes the relation between AFMI active share, AFMI tracking error, and AFMI concentration. We find support for prediction 3, in that active share increases with all measures of concentration we consider. The relation with tracking error is less robust. This could reflect that the relation between concentration and effort is more due to the effort involved in stock picking (as measured by active share), rather than the effort involved in factor timing (as measured by tracking error). Given the difficulties in measuring effort, we leave a more fuller analysis of the relation between effort and concentration for future research.

4.1. Robustness

In addition to the reported tables, we examine the sensitivity of the results in Table 3 by using fund fixed effect regressions instead of the RD method. Most of the results are consistent, except when regressing the PC-adjusted

Please cite this article as: D. Feldman, K. Saxena and J. Xu, Is the active fund management industry concentrated enough? Journal of Financial Economics, https://doi.org/10.1016/j.jfineco.2019.08.009

Table 4
Industry-level analysis: active fund management industry (AFMI) active share, tracking error, and AFMI concentration.
This table reports the results of the main equations in various vector autoregression models, in which AFMI active share, *AS*, and AFMI tracking error, *TE*, are dependent variables. The sample period is from 1980 to 2009, and the frequency is quarterly. *AS* is measured as the average active share of active funds (in Antti Petajisto's database, www.petajisto.net/data.html) in a quarter. *TE* is calculated as the average tracking error of active funds in a quarter. *HHI, NHHI*, and 5*FI* are Herfindahl–Hirschman index, normalized Herfindahl–Hirschman index, and five-fund index, respectively. Eqs. (1) and (2)–(4), and (5) and (6) report the results of the three concentration measures, *HHI, NHHI*, and 5*FI*, respectively. Small sample-adjusted standard errors are used and presented in parentheses. ***, **, and * represent the 1%, 5%, and 10% significant level in a two-tail *t*-test, respectively.

	AS (1)	*TE* (2)	*AS* (3)	*TE* (4)	*AS* (5)	*TE* (6)
Lagged HHI	0.6279***	0.7652*				
	(0.1865)	(0.4302)				
Lagged NHHI			0.4732**	0.7533		
			(0.2365)	(0.5321)		
Lagged 5FI					0.0872***	0.0822
					(0.0284)	(0.0654)
Lagged AS	0.8979***	−0.0971	0.9430***	−0.0590	0.9023***	−0.0680
	(0.0315)	(0.0728)	(0.0283)	(0.0637)	(0.0326)	(0.0751)
Lagged TE	−0.0299	0.7111***	−0.0111	0.7297***	−0.0165	0.7319***
	(0.0280)	(0.0645)	(0.0279)	(0.0627)	(0.0272)	(0.0628)
Constant	0.0753***	0.0884	0.0406*	0.0577	0.0631***	0.0587
	(0.0240)	(0.0554)	(0.0212)	(0.0476)	(0.0224)	(0.0516)
Number of observations	118	118	118	118	118	118
R-squared	0.969	0.574	0.967	0.570	0.968	0.569

fund net alpha on AFMI concentration measures. We find that the significance of AFMI concentration measure is reduced. We also analyze whether our results are driven by small funds. We redo our main analyses using observations after restricting our sample to funds with a net asset value above $15 million in any month of our sample period. Again, we find consistent results. To test whether our main results are stable across subsamples, we redo our analyses for three subperiods. We find a significantly positive relation between fund net alphas and lagged *HHI* in all three subperiods.

5. Conclusion

We introduce a model in which optimal fund manager effort to find mispriced assets influences AFMI size and performance for a continuum of AFMI concentration levels. While models that focus on decreasing returns to scale suggest that AFMI performance must be low in an economy with a massive AFMI, our model says that, even in such an AFMI, if managers exert more effort, they can achieve higher net alpha.

Increased effort productivity and lower effort costs incentivize managers to exert more effort. We identify an equilibrium in which less competition (or more concentration) leads to improved productivity and lowers costs. If a higher concentration results in lower search costs or more unexplored investment opportunities, per manager and per dollar of AUM, it provides stronger managerial incentives to exert effort. According to our model, this higher concentration could be due to a wider distribution of fund manager skills, with some highly skilled managers allocated a sizable proportion of the AFMI AUM. (We allow for exogenous AFMI concentration to accommodate other forces that influence the competitive environment or increases optimal effort such as technological advances or regulatory changes.) In our model, increases in fund managers' efforts improve AFMI performance and increase its size, even at elevated levels of AUM.

This model reproduces several empirical regularities and makes new predictions. We test these new predictions using US mutual fund data and find that, on average, AFMI net alphas, size, and effort (proxied by active share and tracking error) are increasing concave with AFMI concentration. Additional tests using natural experiments (with exogenous changes in concentration levels) and better data (e.g., direct measures of effort) are left for future research.

While our findings identify the AFMI's performance and direct benefits sensitivities to concentration, because we model a partial equilibrium, statements regarding general societal benefits will have to wait for future research. If we hypothesize AFMI's gains to be coming mainly from noise or liquidity traders and from disciplining firms toward higher managerial productivity, one would have to model those to be confident about policy implications with respect to general societal welfare. In view of our findings, we suggest that regulators act judiciously when regulating AFMI concentration. Future research could also extend our analysis to international fund markets, pension funds, and hedge funds.

References

Berger, A.N., 1995. The profit-structure relationship in banking–tests of market-power and efficient-structure hypotheses. J. Money Credit Bank. 27, 404–431.
Berger, A.N., Bonime, S.D., Covitz, D.M., Hancock, D., 1999. Why are bank profits so persistent? The roles of product market competition, informational opacity, and regional/macroeconomic shocks. J. Bank. Financ. 24, 1203–1235.
Berger, A.N., Hannan, T.H., 1989. The price-concentration relationship in banking. Rev. Econ. Stat. 71, 291–299.
Berk, J.B., Green, R.C., 2004. Mutual fund flows and performance in rational markets. J. Polit. Econ. 112, 1269–1295.
Berk, J.B., van Binsbergen, J.H., 2015. Measuring skill in the mutual fund industry. J. Financ. Econ. 118, 1–20.
Brown, D.C., Davies, S.W., 2017. Moral hazard in active asset management. J. Financ. Econ. 125, 311–325.

Brown, K.C., Harlow, W.V., Starks, L.T., 1996. Of tournaments and temptations: an analysis of managerial incentives in the mutual fund industry. J. Financ. 51, 85–110.
Carhart, M.M., 1997. On persistence in mutual fund performance. J. Financ. 52, 57–82.
Chen, J., Hong, H., Huang, M., Kubik, J., 2004. Does fund size erode mutual fund performance? Am. Econ. Rev. 94, 1276–1302.
Claessens, S., Laeven, L., 2003. What drives bank competition? Some international evidence. J. Money Credit Bank. 36, 563–583.
Cohen, R., Coval, J., Pastor, L., 2005. Judging fund managers by the company they keep. J. Financ. 60, 1057–1096.
Connor, G., Korajczyk, R.A., 1988. Risk and return in an equilibrium APT application of a new test methodology. J. Financ. Econ. 21, 255–289.
Cremers, M., Ferreira, M.A., Matos, P., Starks, L., 2016. Indexing and active fund management: international evidence. J. Financ. Econ. 120 (3), 539–560.
Cremers, M., Nair, V.B., Peyer, U., 2008. Takeover defenses and competition: the role of stakeholders. J. Empir. Leg. Stud. 5, 791–818.
Cremers, M., Petajisto, A., 2009. How active is your fund manager? A new measure that predicts performance. Rev. Financ. Stud. 22 (9), 3329–3365.
Cremers, M., Petajisto, A., Zitzewitz, E., 2012. Should benchmark indices have alpha? Revisiting performance evaluation. Crit. Financ. Rev. 2, 1–48.
Daniel, K., Grinblatt, M., Titman, S., Wermers, R., 1997. Measuring mutual fund performance with characteristic-based benchmarks. J. Financ. 52, 1035–1058.
Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 33 (1), 3–56.
Fama, E.F., French, K.R., 2010. Luck versus skill in the cross section of mutual fund returns. J. Financ. 65, 1915–1947.
Ferreira, M.A., Keswani, A., Miguel, A.F., Ramos, S., 2013a. Testing the Berk and Green model Around the World. Nova School of Business and Economics, Lisbon, Portugal Unpublished Working Paper.
Ferreira, M.A., Keswani, A., Miguel, A.F., Ramos, S.B., 2013b. The determinants of mutual fund performance: a cross-country study. Rev. Financ. 17 (2), 483–525.
Garvey, G., Kahn, R.N., Savi, R., 2017. The dangers of diversification: managing multiple manager portfolios. J. Portf. Manag. 43 (2), 13–23.
Geroski, P.A., 1990. Innovation, technological opportunity, and market structure. Oxf. Econ. Pap. 42, 586–602.
Giroud, X., Mueller, H.M., 2011. Corporate governance, product market competition, and equity prices. J. Financ. 2, 563–600.
Glode, V., 2011. Why mutual funds “underperform.”. J. Financ. Econ. 99 (3), 546–559.
Goldberg, L.G., Rai, A., 1996. The structure-performance relationship for European banking. J. Bank. Financ. 20, 745–771.
Grinblatt, M., Titman, S., 1989. Mutual fund performance: an analysis of quarterly portfolio holdings. J. Bus. 62, 393–416.
Grinblatt, M., Saxena, K., 2017. When factors don’t span their basis portfolios. J. Financ. Quant. Anal. 53 (6), 2335–2354 (Dec. 2018).
Gruber, M.J., 1996. Another puzzle: the growth in actively managed mutual funds. J. Financ. 51, 783–810.
Guercio, D.D., Reuter, J., 2014. Mutual fund performance and the incentive to generate alpha. J. Financ. 69, 1673–1704.
Hoberg, G., Kumar, N., Prabhala, N.R., 2018. Mutual fund competition, managerial skill, and alpha persistence. Rev. Financ. Stud. 31, 1896–1929.
Hong, C., 2018. Freedom of Choice in Pension Plans: Evidence From a Quasi-Natural Experiment. Shanghai Advanced Institute of Finance, Shanghai, China Unpublished Working Paper.
Ibert, M., Kaniel, R., Van Nieuwerburgh, S., Vestman, R., 2017. Are Mutual Fund Managers Paid for Investment Skill?. National Bureau of Economic Research, Cambridge, MA Unpublished Working Paper.
Jensen, M.C., 1968. The performance of mutual funds in the period 1945–1964. J. Financ. 23, 389–416.
Kacperczyk, M., Sialm, C., Zheng, L., 2005. On the industry concentration of actively managed equity mutual funds. J. Financ. 60, 1983–2011.
Khorana, A., Servaes, H., 2011. What drives market share in the mutual fund industry? Rev. Financ. 16 (1), 81–113.
Nickell, S.J., 1996. Competition and corporate performance. J. Polit. Econ. 104, 724–746.
Pastor, L., Stambaugh, R.F., 2002. Mutual fund performance and seemingly unrelated assets. J. Financ. Econ. 63, 315–349.
Pastor, L., Stambaugh, R.F., 2012. On the size of the active management industry. J. Polit. Econ. 120, 740–781.
Pastor, L., Stambaugh, R.F., Taylor, L.A., 2015. Scale and skill in active management. J. Financ. Econ. 116, 23–45.
Petajisto, A., 2013. Active share and mutual fund performance. Financ. Anal. J. 69 (4), 73–93.
Reuter, J., Zitzewitz, E., 2013. How Much Does Size Erode Mutual Fund performance? A Regression Discontinuity Approach. Boston College, Chestnut Hill, MA Unpublished Working Paper.
Sharpe, W.F., 1992. Asset allocation: management style and performance measurement. J. Portf. Manag. 18, 7–19.
Spiegel, M., Zhang, H., 2013. Mutual fund risk and market share-adjusted fund flows. J. Financ. Econ. 108, 506–528.
Stambaugh, R.F., 2014. Presidential address: investment noise and trends. J. Financ. 69, 1415–1453.
Tirole, J., 1988. The Theory of Industrial Organization. MIT Press, Cambridge, MA.
Wahal, S., Wang, A.Y., 2011. Competition among mutual funds. J. Financ. Econ. 99, 40–59.
Wermers, R., 2000. Mutual fund performance: an empirical decomposition into stock-picking talent, style, transactions costs, and expenses. J. Financ. 55, 1655–1695.
Yan, X., 2008. Liquidity, investment style, and the relation between fund size and fund performance. J. Financ. Quant. Anal. 43, 741–768.

Please cite this article as: D. Feldman, K. Saxena and J. Xu, Is the active fund management industry concentrated enough? Journal of Financial Economics, https://doi.org/10.1016/j.jfineco.2019.08.009