FUND FLOWS, PERFORMANCE, AND EXIT UNDER DYNAMIC UNOBSERVABLE MANAGING ABILITY

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

Latest revision, January 11, 2022

Abstract. We introduce continuous-time models of dynamic unobservable fund manager abilities under a nonlinear framework, with risk-neutral or risk-averse investors. Our framework incorporates effects of factors, such as cross-fund subsidization, manager replacement, and competition by entrants, on manager abilities and performances and generates time-nonmonotonic flow-performance sensitivities and convexities. Funds' exit probabilities change with time and fund sizes at any fund age. Our empirical evidence of time-nonmonotonic flow-performance sensitivities and convexities, and old funds' increasing exit probability with fund age support our framework, rather than current (linear) frameworks. We also facilitate resolving the controversy of whether flow-performance relations are linear or convex.

JEL Codes: G11, G14, G23

Keywords: Fund flows, Fund performance, Fund exit probabilities, Dynamic unobservable manager abilities, Learning, Nonlinear Filtering

---

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

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

[c] We thank Daniel Andrei, Doron Avramov, Fan Chen, David Colwell, Hugues Dastarac, Wenjie Ding, Juan Pedro Gómez, Idan Hodor, Chang-Mo Kang, Paulo Leite, Abhiroop Mukherjee, Stefano Pegoraro, Linda Pesante, Carolina Salva, Wing Wah Tham, Nhu Tran, Youchang Wu, and Terry Zhang for helpful discussions, and participants at the Workshop on Asset Pricing and Risk Management, National University of Singapore; Monash/Xiamen Workshop in Finance, Econometrics, Economics and Statistics (FEES); Shanghai Jiao Tong University; The University of New South Wales; Yeshiva University; Southern Finance Association Annual Meeting; Greater China Area Finance Conference; Midwest Finance Association Annual Meeting; International Conference of the French Finance Association; SDU Finance Workshop; The Econometric Society North American Summer Meeting; The European Financial Management Association Annual Meeting; Spanish Association of Finance (AEFIN), Finance Forum; Portuguese Finance Network International Conference; The Econometric Society Australasian Meetings; The Econometric Society European Meetings; The Finance Symposium; FIRN Annual Conference; Australasian Finance and Banking Conference, and Curtin University.

1 Introduction

Fund flows, performance, and exit are main topics of fund literature [see, for example, the innovative Berk and Green (2004) (BG), Berk (2005), and Berk and van Binsbergen (2015)]. However, current models are not able to explain or predict many stylized facts of these topics. For example, why are the empirical flow-performance sensitivity and convexity nonmonotonic over time? Why is the probability of exiting the market still high for very old funds? Also, why are both convex and linear empirical flow-performance relations possible? Those models lack the power to explain and predict the above stylized facts because they assume (unobservable) constant manager abilities or dynamic manager abilities in a linear framework.[1] In this paper, for the first time, we explain and predict these stylized facts by allowing dynamic manager abilities in a nonlinear framework. We demonstrate these both, for risk-neutral investors and for risk-averse ones.

Current empirical evidence supports dynamic manager abilities that evolve under complex structures. Researchers find that fund performance persists only in the short term [see, for example, Carhart (1997), Berk and Tonks (2007), Mamaysky, Spiegel, and Zhang (2007), and Wang (2014)]. Besides the effects of the fund industry's decreasing returns to scale,[2] the dynamics of unobservable manager abilities are also likely to contribute to the lack of long-term persistence in fund performance. Further, current studies show that manager ability to outperform passive benchmarks is affected by different factors, such as fund family activities [see, for example, Gaspar, Massa, and Matos (2006), Evans (2010), Brown and Wu (2016), Eisele, Nefedova, Parise, and Peijnenburg (2020), and Xu (2021)]; changing attention allocation [Kacperczyk, Nieuwerburgh, and Veldkamp (2016)]; changing investment strategies

---

[1] Regarding models with constant manager abilities, see for example, Lynch and Musto (2003), BG, Huang, Wei, and Yan (2007), Brown and Wu (2016), and Choi, Kahraman, and Mukherjee (2016). Regarding models with dynamic manager abilities in a linear framework, see for example, Dangl, Wu, and Zechner (2008), Brown and Wu (2013) which is a working paper version of Brown and Wu (2016), and Roussanov, Ruan, and Wei (2020).

[2] Theoretical models, such as those of BG and Pastor and Stambaugh (2012), show that investors invest more (less) in the funds that perform better (worse), and this larger (smaller) amount invested increases (decreases) fund costs due to decreasing returns to scale, driving down (up) the fund performance in the future. Thus, fund performance does not persist.

[Lynch and Musto (2003)]; managers' replacements [Dangl, Wu, and Zechner (2008)]; and macroeconomic conditions [see, for example, Ferreira, Keswani, Miguel, and Ramos (2012, 2013), Kacperczyk, Nieuwerburgh, and Veldkamp (2014), Feldman, Saxena, and Xu (2020, 2021), and Feldman and Xu (2021)]. Because these factors are dynamic with complex patterns, they drive fund manager abilities to change with complex patterns over time. Therefore, to offer more insights to market equilibria, we introduce a more general framework when modeling manager abilities.

We develop a continuous-time framework and model representative (identical) active funds with a passive benchmark portfolio. In our baseline model, the active funds' observable gross alphas follow Itô processes, in which instantaneous expected gross alphas (the drift terms) depend on dynamic unobservable manager ability levels. These ability levels also follow Itô processes, and their diffusions are (locally, imperfectly) correlated with those of funds' gross alpha processes. New to the literature, we allow the coefficient parameters to change with time and fund prices, introducing nonlinearity into the framework. This new feature allows us to incorporate several effects when modeling evolutions of manager abilities and gross alphas, effects such as those of fund family activities, changing attention allocation, changing investment strategies, managers' replacements, and macroeconomic factors. This is a nonlinear framework of dynamic abilities, more general and offering more insights than current linear frameworks, in which the coefficient parameters are constants.[3] Both managers and investors estimate manager abilities by observing gross alphas. Due to our framework's nonlinear structure, estimation error (or precision) of inferred manager abilities changes over time nonmonotonically. Consequently, sensitivities of inferred manager abilities to new observations of fund gross alphas change over time nonmonotonically.

Other features of our model are similar to those in classical models.[4] In particular, we

[3] Linear filtering techniques are used in solving the learning processes of linear frameworks of abilities, whereas nonlinear filtering techniques are needed in solving the learning processes of nonlinear frameworks of dynamic abilities.
[4] See, for example, BG, Brown and Wu (2016), and Choi, Kahraman, and Mukherjee (2016).

assume decreasing returns to scale;[5] we allow managers to set constant management fees and choose the size of wealth they actively manage; and we assume that fund managers and investors are rational and symmetrically informed. In our baseline model, we assume risk-neutral investors.[6] Then, we model the equilibrium for mean-variance risk-averse investors, which is also new to the literature.

Our nonlinear framework of dynamic abilities derives equilibrium flow-performance sensitivity and convexity that change nonmonotonically over time.[7] Thus, our framework can explain and predict nonmonotonicities in empirical flow-performance sensitivity and convexity due to different economic reasons. We specialize our framework in three ways and demonstrate how each—cross-fund subsidization, manager replacement, and competition by new entrants—affects manager abilities and gross alpha productions and, consequently, induces nonmonotonicities in equilibrium.[8,9] Linear frameworks of manager abilities, used in current literature, cannot generate these results. In these linear frameworks, sensitivities of inferred manager abilities to new realizations of fund returns change over time monotonically, driving the flow-performance sensitivity and convexity to change monotonically over time.

We show that, for specific parameter values, our model degenerates to a continuous-time analog of the BG model, recreating all BG model's insights in a dynamic context. Also, for specific parameter values, the flow-performance relation in our model degenerates to the single-fund versions in Brown and Wu (2016) and Choi, Kahraman, and Mukherjee (2016).

We also study the case of mean-variance risk-averse investors who maximize their portfolios' instantaneous Sharpe ratios. These investors' optimal portfolios are the same as

---

[5] That is, funds' total costs are increasing and convex in the size of assets under active management.
[6] That is, investors supply capital with infinite elasticity to funds that have positive expected net alphas.
[7] Please see discussions of equilibrium flow-performance relations in Section 2.3.
[8] Please see implementations in Section 4 and how they relate to current literature, such as Gaspar, Massa, and Matos (2006), Evans (2010), Eisele, Nefedova, Parise, and Peijnenburg (2020), and Xu (2021) regarding cross-fund subsidization; Dangl, Wu, and Zechner (2008) regarding manager replacement; and Wahal and Wand (2011) regarding competition induced by new mutual funds.
[9] We choose these factors, suggested in current literature, and specialize our more general framework with each of them to gain corresponding insights. Our framework can incorporate effects of other economic factors into the evolutions of manager ability and gross alpha production. We leave these implementations for future studies.

those of investors with Bernoulli logarithmic preferences and are "growth optimal" [see, for example, Feldman (1992)].[10] We show that investors' risk aversion affects the dollar amounts of fund flows; when fund flows are calculated as flow percentages as we do here, effects of investors' risk aversion cancel out.[11] Thus, equilibrium flow-performance sensitivities and convexities when investors are mean-variance risk-averse are similar to those when investors are risk-neutral.

We derive funds' exit probability in the next instant and at any future time. A fund exits the market if the manager's inferred ability falls below an exogenous "survival level." Our nonlinear framework of dynamic ability predicts that a fund's exit probability changes with both time and fund size (as equilibrium fund size is a function of inferred ability), at any fund age level. In contrast, in a linear framework, when a fund gets old (time goes to infinity) and precision of inferred ability achieves the steady state, the fund's exit probability changes only with inferred ability and, consequently, changes only with fund size. As old funds usually have large sizes, a linear framework would predict a very low exit probability for old funds. Further, under constant manager ability, as in BG for example, an old fund's exit probability is zero because the inferred ability converges to the constant true ability, above the survival level. Therefore, our nonlinear framework better explains the real-world large exit probability of funds and the patterns of exit probability, especially for old funds.

Empirical study

Using the U.S. active equity mutual fund data of the Center for Research in Security Prices (CRSP), we empirically test our model's predictions, where we set funds' "time" to be

[10] This "growth optimal" result was independently discovered by Bernoulli in 1738 [Bernoulli (1954)] and by the "Kelly Criterion" [Kelly (1956)]. Further, this criterion might be seen as active managers' "horizon" choice for investors with potentially heterogeneous horizons and as resolving/avoiding the time inconsistency of mean-variance preferences. See Basak and Chabakauri (2010) and Feldman and Leisen (2021).

[11] Risk-averse investors maximize the instantaneous Sharpe ratios of their portfolios, which contain active funds and passive benchmark portfolios. Investors' risk aversion affects the equilibrium portfolio weights allocated to the active funds and, consequently, affect the equilibrium fund sizes and changes in fund sizes. The flow percentage is the change with fund sizes divided by fund sizes, and the effects of investors' risk aversion on the numerator and on the denominator cancel out.

their age. We find that flow–net alpha sensitivity and convexity, on average, change with fund age nonmonotonically: decrease first, then increase, decrease, and finally increase. The detected decrease in flow–net alpha sensitivity and convexity in the early years for an average fund arises mainly because investors have more and more precise estimates of manager abilities as the number of observations increases, making inferred manager abilities less and less sensitive to funds' performances. The fluctuations of flow–net alpha sensitivity and convexity in the later years might be due to multiple forces. For instance, Gaspar, Massa, and Matos (2006) and Eisele, Nefedova, Parise, and Peijnenburg (2020) show that fund families transfer returns from old funds to young funds to improve family profit.[12] Thus, a fund might receive return transfer when it is young and provide such transfer when it gets older, making its gross alpha sensitivity to manager ability increase and then decrease.[13] Also, Dangl, Wu, and Zechner (2008) show that a manager replacement should be preceded by a portfolio risk increase and followed by a portfolio risk decrease, and these patterns are also affected by the length of the manager's tenure. In reality, a fund replaces and tenures managers over time, making its gross alpha volatility fluctuate. The above forces, and other economic forces that create time-nonmonotonic gross alpha sensitivity to manager ability, gross alpha volatility, and sensitivity of inferred manager ability to fund performance, make the flow–net alpha sensitivities and convexities change nonmonotonically with fund age.

We also estimate the flow–net alpha sensitivity and convexity for each individual fund and find nonmonotonicities in these sensitivities and convexities for many funds. The above empirical results are sufficient to support our nonlinear framework rather than linear ones.

Our results of nonmonotonic flow–net alpha sensitivities and convexities are robust after controlling for factors affecting flow–net alpha relations shown in the literature, such as

---

[12] These papers argue that a fund that receives (provides) return transfer performs better (worse), inducing larger (smaller) fund size and profit. If the increase in the receiver's profit overwhelms the decrease in the provider's profit, the family's total profit is improved.

[13] If a fund receives (provides) return transfer, for the same level of manager ability, it produces larger (smaller) gross alpha to investors.

fund volatility [Huang, Wei, and Yan (2012)]; states of market return [Franzoni and Schmalz (2017)]; cross-sectional net alpha dispersion [Harvey and Liu (2019)]; and economic policy uncertainty [Jiang, Starks, and Sun (2021)]. Therefore, we demonstrate that besides these factors, dynamics of manager ability and nonlinear association of manager ability and gross alpha induced by unexplored factors and latent factors are relevant forces driving the flow-performance relation.

We also empirically analyze funds' exit probabilities, finding that survival rates for old funds decrease with their age. For example, the probability for a fund to survive in the next two years when it is 25 years old is 97.25%, but only 95.56% when it is 40 years old. Also, in our logit model, we find that the probability that a fund exits in the next month decreases with fund size and increases with fund age, not only in the whole sample but also in subsamples of old funds. Thus, older funds are more likely to exit the market. The reason might be that, over time, new entrants use portfolio strategies similar to those of incumbents, intensifying market competition [see, for example, Wahal and Wand (2011)] and exerting negative impacts on the tendency of incumbents' abilities to outperform the market. This finding can be systematically explained by our nonlinear framework, and our study complements the literature of fund exit.

More on the literature and our findings

There is active discussion in current literature on the curvature of flow-performance relations. For example, Lynch and Musto (2003), BG, and Brown and Wu (2016) suggest that this relation is convex, whereas Spiegel and Zhang (2013) suggest that this relation is linear. Our study complements this discussion by showing that the intercept, slope, and curvature of the flow-performance relation can change over time nonmonotonically. Consequently, empirical flow-performance relations may exhibit linearity or convexity. In the real world, both cross-sectional heterogeneity and time dynamics of flow-performance relations affect empirical results; thus, both patterns should be considered.

Franzoni and Schmalz (2017) find that the flow-performance sensitivity is less steep when the market excess return is more extreme. They theoretically show that if investors need to learn both the manager skill and the loading of the fund portfolio on the market factor, then fund performance is less informative about manager skill when market factor realizations are larger in absolute value, resulting in "hump-shape" flow-performance sensitivity. However, their model can predict this sensitivity for funds only in their earliest ages because, as Franzoni and Schmalz (2017) point out, the posterior estimates of skill and factor loading become correlated after some periods, eliminating and even reversing the "hump shape." In our model, if fund gross alphas are less sensitive to manager abilities under extreme market conditions,[14] then investors reduce their reactions to fund returns, decreasing flow-performance sensitivities. This result holds at any fund age level. Therefore, our model explains and predicts the "hump-shape" flow-performance sensitivity in a more consistent way, and more so for older funds.

There is also a discussion in the literature of how funds' marketing activities affect flow-performance relations. For example, Huang, Wei, and Yan (2007) find that funds with marketing activities exhibit a less convex flow-performance relation. They theoretically show that as funds with marketing activities reduce investors' participation costs, new investors' requirements for fund performance are lower, making fund flows more sensitive to low or medium fund performance. We show, different from Huang, Wei, and Yan's (2007) insights, that if marketing activities induce higher management fees and/or improve investors' estimates of manager abilities over time, then equilibrium flow-performance convexities are lower.

Our paper also relates to other recent papers that study fund flows, fund performance, fund sizes, and fund asset classes, such as Bollen (2007), Chen, Goldstein, and Jiang (2010), Chen, Hong, Huang, and Kubik (2004), Rakowski (2010), and Yan (2008).

Contribution

We contribute to the literature, first, by introducing a model of dynamic unobservable

[14] This assumption seems realistic because during periods with extreme market conditions, the security market liquidity and volatility, which are relevant to alpha production, are more uncertain, making funds' gross alphas less sensitive to manager abilities but more dependent on "luck".

manager abilities under a nonlinear framework, which better explains the real-world nonmonotonic time-varying flow-performance sensitivities and convexities, and better explains effects of fund size and fund age on exit probabilities, especially for old funds. Current theories model manager abilities in linear frameworks; thus, they cannot predict these results, lacking explanatory power.[15] Second, we provide empirical evidence that supports our nonlinear framework of dynamic unobservable manager abilities. Third, our model offers new insights into empirical findings in the current literature, including the complex curvature of flow-performance relations, the "hump-shape" flow-performance sensitivity, and the findings that marketing activities induce smaller flow-performance convexity.

Finally, our findings imply that the innovative insights of BG regarding the myth of active portfolio management,[16] which were demonstrated within a parsimonious model, hold in a wider class of equilibria in terms of production structure (dynamic abilities with nonmonotonic estimation precision), information structure (unobservable processes, including nonlinear ones), and preferences structure (risk-neutral or risk-averse investors).

Section 2 introduces our model and derives flow-performance relations. Section 3 analyzes funds' exit probabilities. Section 4 provides implementations of nonlinear frameworks that incorporate specific economic factors. Section 5 illustrates our empirical study. Section 6 discusses our model's insights into current empirical phenomena, and Section 7 concludes.

2 The Model

We introduce a rational equilibrium framework, studying how nonlinear dynamics of unobservable manager abilities affect equilibrium flow-performance relations and funds' exit probabilities. Some of our settings are similar to those of BG, Brown and Wu (2016), and Choi,

[15] For example, Brown and Wu (2013), an earlier working paper version of Brown and Wu (2016), and Dangl, Wu and Zechner (2008) model dynamic unobservable managing ability under linear frameworks. Their models can explain only a monotonic time-varying pattern of flow-performance sensitivity and convexity and can predict only the relation of the probability of fund exit and fund size, for old funds.

[16] See also Berk (2005).

Kahraman, and Mukherjee (2016).[17] We use a two-fund setting, i.e., investors can invest in a representative active fund that has one manager and in a passive benchmark portfolio.[18] This setting is also similar to those in Wei, and Yan (2007), and Lynch and Musto (2003). Within a continuous-time framework, we study the market over a time interval, at times $t,\ t \in [0, T]$, where $T,\ T > 0$, is a constant.

2.1 Observable Returns and Unobservable Manager Ability: Nonlinear Filtering

Let $\xi_t,\ 0 \le t \le T$ be the active fund's gross share price, before fund costs and fees,[19] so $d\xi_t/\xi_t$ is the instantaneous fund gross return. For simplification, we assume that this active fund has a beta loading of one on a passive benchmark portfolio. Focusing on the active fund's return, as previous models,[20] we normalize the passive benchmark portfolio's return to zero, so funds' instantaneous gross returns in excess of the passive benchmark is $d\xi_t/\xi_t - 0 = d\xi_t/\xi_t$. Hereafter, we briefly call $d\xi_t/\xi_t$ gross alpha.

The active fund's gross alphas depend on the fund manager's instantaneous ability, θ_t, $0 \le t \le T$, to beat the benchmark. We briefly call it manager ability. Manager abilities are unobservable for both managers and investors. Managers and investors learn about θ_t by observing evolutions of gross alphas $d\xi_s/\xi_s,\ 0 \le s \le t$ (or equivalently by observing gross fund share prices $\xi_s,\ 0 \le s \le t$). We assume a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with filtration $\{\mathcal{F}_t\}_{0 \le t \le T}$. Two independent Wiener processes, $W_{1,t}$ and $W_{2,t},\ 0 \le t \le T$, are adapted to this filtration. The unobservable θ_t and the observable ξ_t evolve as follows.

$$d\theta_t = [a_0(t, \xi_t) + a_1(t, \xi_t)\theta_t]dt + b_1(t, \xi_t)dW_{1,t} + b_2(t, \xi_t)dW_{2,t}, \tag{1}$$

$$\frac{d\xi_t}{\xi_t} = A(t, \xi_t)\theta_t dt + B(t, \xi_t)dW_{2,t}, \tag{2}$$

---

[17] Similar to BG, Brown and Wu (2016) and Choi, Kahraman, and Mukherjee (2016), we assume that participants in the model are symmetrically informed. Also, the model is partial equilibrium. Managers' actions do not affect the passive benchmark returns, and we do not model the source of managers' abilities to outperform the passive benchmark portfolio.

[18] This two-fund model can be extended to a multiple-fund model in which investors invest in n $(n \ge 2)$ active funds and a passive benchmark portfolio.

[19] In the real world, fund costs and fees are usually paid separately when investors buy and/or redeem fund shares.

[20] See, for example, Huang, Wei, and Yan (2007).

with initial conditions θ_0 and ξ_0, respectively. The parameters $a_0(t,\xi_t)$, $a_1(t,\xi_t)$, $b_1(t,\xi_t)$, $b_2(t,\xi_t)$, $A(t,\xi_t)$, and $B(t,\xi_t)$ are functions of t and ξ_t. To make economic sense, we assume that $A(t,\xi_t) > 0$ (otherwise the ability becomes a "disability"). For simplicity and without loss of generality, we assume $b_1(t,\xi_t) > 0$, and $B(t,\xi_t) > 0$. The evolution processes ("laws of motion") and all parameter values are common knowledge.

The above setting implies that, first, abilities, θ_t, to beat the benchmark follow a nonlinear dynamic process. Second, funds' gross alphas, $d\xi_t/\xi_t$, depend on manager abilities and on random shocks. As $A(t,\xi_t) > 0$, managers with higher ability tend to create higher fund gross alphas, and the larger $A(t,\xi_t)$ is, the higher the gross alpha sensitivities to abilities. $B(t,\xi_t)$ positively corresponds to gross alpha volatility. Third, where $b_2(t,\xi_t)$ is strictly positive (negative), the shock $W_{2,t}$ affects both ability and fund gross alpha, making them instantaneously positively (negatively) correlated as $b_2(t,\xi_t)B(t,\xi_t) > 0$ ($b_2(t,\xi_t)B(t,\xi_t) < 0$) Where $b_2(t,\xi_t) = 0$ and $b_1(t,\xi_t) > 0$, ability and gross alpha are affected by independent shocks, so they are instantaneously uncorrelated. A larger absolute value of $b_2(t,\xi_t)$ relative to that of $b_1(t,\xi_t)$ implies that gross alpha and ability are more highly correlated in absolute sense.

Allowing coefficient parameters to be functions of t and ξ_t is a new feature that differentiates our model from those in the current literature. Due to this setting, we can incorporate effects of economic factors when modeling manager abilities. For example, dynamic factors, such as fund family activities, changing attention allocation, changing investment strategies, managers' replacements, and macroeconomic factors, exert impacts on manager abilities and gross alpha productions, changing the tendency of fund manager abilities, the correlations of manager abilities and gross alphas, and the volatility of manager performances. These effects can be modeled in the coefficient parameters in Equations (1) and (2), and consequently influence the market equilibrium.

To facilitate our discussions, we keep using the general notations of Equations (1) and

(2) in Sections 2 and 3 when deriving our equilibrium results. In Section 4, we specialize our framework and incorporate effects of economic factors, such as cross-fund subsidization, manager replacement, and competition induced by new entrants on manager abilities and gross alpha productions and offer additional economic rationale for our framework.

To facilitate our analysis, we define the following terms:

- $\mathcal{F}_t^{\xi} \triangleq$ the σ-algebras generated by $\{\xi_s, 0 \le s \le t\}$, with $\left\{\mathcal{F}_t^{\xi}\right\}_{0 \le t \le T}$ as the corresponding filtration over $0 \le t \le T$.
- $m_t \triangleq$ the mean of θ_t conditional on the observations ξ_s, $0 \le s \le t$, i.e.,
$$m_t \triangleq \mathrm{E}\left(\theta_t | \mathcal{F}_t^{\xi}\right).$$
- $\gamma_t \triangleq$ the variance of θ_t conditional on the observations ξ_s, $0 \le s \le t$, i.e.,
$$\gamma_t \triangleq \mathrm{E}\left[(\theta_t - m_t)^2 | \mathcal{F}_t^{\xi}\right].$$

We assume that the conditional distribution of $\theta_{0,}$ given ξ_0 (the prior distribution) is Gaussian, $N(m_0, \gamma_0)$, with finite values of ξ_0, m_0, and γ_0.

Observing ξ_t, managers and investors update their estimates of θ_t in a Bayesian fashion. Such a model is presented in Liptser and Shiryaev (2001a, Ch. 8; 2001b, Ch. 12). These techniques are called optimal filtering and are used in numerous previous studies [see, for example, Dothan and Feldman (1986), Detemple (1986), Feldman (1989, 2007), Berk and Stanton (2007), Dangl, Wu, and Zechner (2008), and Brown and Wu (2013, 2016)]. The following describes how managers and investors form and update their estimates of θ_t.

Let $\mathcal{F}_t^{\xi_0, \overline{W}}$, $0 \le t \le T$, be the σ-algebras generated by $\{\xi_0, \overline{W}_s, 0 \le s \le t\}$, where

$$\overline{W}_t = \int_0^t \frac{d\xi_s(s, \xi_s)/\xi_s(s, \xi_s) - A(s, \xi_s) m_s(s, \xi_s) ds}{B(s, \xi_s)} \tag{3}$$

is a Wiener process with respect the filtration $\left\{\mathcal{F}_t^{\xi}\right\}_{0 \le t \le T}$, with $\overline{W}_0 = 0$. Then, the σ-algebras $\mathcal{F}_t^{\xi}$ and $\mathcal{F}_t^{\xi_0, \overline{W}}$ are equivalent. $\overline{W}_t$ innovates the observable conditional mean, m_t, of the unobservable ability, θ_t. The variables m_t, ξ_t, and γ_t are the unique, continuous, $\mathcal{F}_t^{\xi}$-measurable solutions of the system of equations

$$dm_t = [a_0(t,\xi_t) + a_1(t,\xi_t)m_t]dt + \sigma_m(\gamma_t)d\overline{W}_t, \tag{4}$$

$$\frac{d\xi_t}{\xi_t} = A(t,\xi_t)m_t dt + B(t,\xi_t)d\overline{W}_t, \tag{5}$$

$$d\gamma_t = [b_1^2(t,\xi_t) + b_2^2(t,\xi_t) + 2a_1(t,\xi_t)\gamma_t - \sigma_m^2(\gamma_t)]dt, \tag{6}$$

where

$$\sigma_m(\gamma_t) \triangleq \frac{b_2(t,\xi_t)B(t,\xi_t) + A(t,\xi_t)\gamma_t}{B(t,\xi_t)}, \tag{7}$$

with initial conditions ξ_0, m_0, and γ_0. The random process (θ_t, ξ_t), $0 \leq t \leq T$ is conditionally Gaussian given $\mathcal{F}_t^{\xi}$.

Equations (3)–(7) are demonstrated in Theorem 8.1 of Liptser and Shiryaev (2001a), Theorems 11.1, and 12.5 of Liptser and Shiryaev (2001b). These results require regular technical conditions regarding the smoothness and boundedness of the parameter values over the period $0 \leq t \leq T$.[21]

The Wiener process $\overline{W}_t$ represents the innovation shocks to estimates of managers' unobservable abilities. The processes $(\xi_t, \overline{W}_t)$ or equivalently (ξ_t, m_t, γ_t) provide same information as (ξ_t, θ_t) over $0 \leq t \leq T$. Hence, investors' original non-Markovian problem can be stated as an equivalent Markovian one, which allows state vector solution.[22] To make economic sense, we assume a nonnegative $b_2(t,\xi_t)$, which ensures a positive $\sigma_m(\gamma_t)$ in Equation (4). That is, a positive (negative) corresponding shock in fund gross alpha induces an increase (a decrease) in inferred manager ability.

Then, in our model, investors make their optimal decisions in two steps. First, by observing the evolution of the fund's share price ξ_t, they generate Markovian posterior conditional moments of the fund manager ability θ_t, converting the problem from a non-

---

[21] See these technical requirements in the corresponding theorems in Liptser and Shiryaev (2001a, 2001b).

[22] While there was wide use of a linear version of this filter [see, for example, the literature review in Feldman (2007)], to the best of our knowledge, this is the first use of this nonlinear filter.

Markovian one to an equivalent tractable Markovian one. Second, they use their posterior estimate m_t to predict future funds' gross alpha, as shown by Equation (5). They use these predictions in solving their investors' problem, as shown in the next sections. Notice that in these optimization processes, the unobservable manager ability θ_t is replaced by its observable inferred estimate m_t, which is also updated as a function of the stochastic conditional variance γ_t, representing the imprecision of the estimate. Further, all decisions involving the inferred abilities, m_t, become functions of their imprecision, γ_t, as well.

The intuition regarding our ability to solve this nonlinear system and the nature of its equilibrium is as follows. Because time and fund share price levels are observable, at each time point the coefficient parameters, $a_0(t,\xi_t)$, $a_1(t,\xi_t)$, $b_1(t,\xi_t)$, $b_2(t,\xi_t)$, $A(t,\xi_t)$, and $B(t,\xi_t)$, conditional on ξ_t realizations become known constants, thus induce a Gaussian distribution for manager ability. In the next time instant, realized ξ_t stochastically change. Consequently, the coefficient parameters stochastically change as well, inducing a "new" Gaussian conditional distribution of manager ability.

We note that each set of ability distribution parameter values, valid at a certain time point, induces a future evolution path. As the moments of the ability-conditional distributions evolve stochastically, we can think of the equilibrium as stochastically time traveling among these evolution paths, each induced by one set of parameter values. We proceed by, first, characterizing one evolution path induced by a set of constant parameters.

Inferred Abilities and Their Precision Under Constant Parameters

For the special case in which the parameters a_0 a_1, b_1, b_2, A, and B are constants, Equations (3)–(7) construct a linear framework that can be solved by linear filtering techniques. Given γ_0, γ_t becomes deterministic as shown by Equation (6). Consequently, $\sigma_m(\gamma_t)$, the sensitivity of expected manager ability to innovation shocks in fund gross alpha is dynamic but deterministic. Also, m_t is stochastic.[23] Therefore, investors know the precision of their future estimates of manager ability in advance but do not know the future inferred abilities. In this case, depending on parameter values, γ_t monotonically increases or decreases to a nonnegative steady state,[24] or stays unchanged in it over time.[25] Consequently, depending on parameter values, $\sigma_m(\gamma_t)$ monotonically increases or decreases to $\sigma_m(\gamma_t)$'s steady state, or stays unchanged in this steady state over time.[26]

Inferred Abilities and Their Precision Under Stochastic Parameters

Stochastic parameters induce inferred abilities and their precision to switch at every time point among evolution paths. Each of these evolution paths is induced by a set of constant parameters/linear system, as described above. In contrast to the case of constant parameters, this switching induces nonmonotonicity in the evolutions of inferred ability precision, γ_t, which, in turn, induces nonmonotonicity in the sensitivities of inferred abilities to innovation shocks, $\sigma_m(\gamma_t)$. Then, under this framework, there are no steady states for γ_t and $\sigma_m(\gamma_t)$, which are forever dynamic.

The dynamics of γ_t is one of the key differences between our model and BG's and models subsequent to theirs. In those models, the observable fund gross alpha equals the unobservable manager ability, an unknown constant, plus a Gaussian noise term. When investors update their estimates of manager ability with a Gaussian prior, the precision of their posterior estimates of that ability consistently increases over time as more observations are realized. In this case, over time, investors' estimates of manager ability become less sensitive

---

[23] The fact that the random process (θ_t, ξ_t), $0 \leq t \leq T$ is conditionally Gaussian, given $\mathcal{F}_t^{\xi}$, facilitates the generation of the posterior estimate of gross alphas in closed form.

[24] As we study the processes in the period $0 \leq t \leq T$, we allow T to be sufficiently large (i.e., $T \to \infty$) so that γ_t can achieve its steady state.

[25] The expression of $d\gamma_t$ with constant parameter values implies that γ_t follows a Riccati equation and γ_t has a steady state. See Corollary 1.2 in Feldman (1989).

[26] We note that even under the constant parameter values, there is a "knife edge" case in which the dynamics of γ_t might induce a transient nonmonotonic time pattern of the local variance of inferred abilities to innovation shocks, $\sigma_m^2(\gamma_t)$. This can be seen from Equation (7). A negative b_2 induces a negative instantaneous/idiosyncratic correlation, $(b_2 B)$. Then, the dynamic weight γ_t may induce the expression $b_2 B + A\gamma_t$ to change sign, inducing a decreasing-increasing $\sigma_m^2(\gamma_t)$. Detailed analysis of this nonmonotonicity is in Feldman (1989, Proposition 4).

to the innovation shocks in fund returns. In contrast to those studies, within our more general structure, precision of investors' estimates of manager abilities can be nonmonotonic over time. Then, in turn, sensitivities of inferred manager ability to the innovation shocks in fund returns can be nonmonotonic over time. These features generate a framework that has stronger theoretical and empirical explanatory and predictive powers in studying the flow-performance relation.

In the following discussions, we call our framework *a nonlinear–dynamic ability framework,* call the framework in which a_0 a_1, b_1, b_2, A, and B are constants *a linear–dynamic ability framework*, and call the BG case, *a linear–constant ability framework*.

2.2 Investors' Optimization and the Fund Manager's Optimization

Using the above filter to re-represent the state space $\{\theta_t, \xi_t\}$ in terms of observable variables $\{\xi_t, m_t, \gamma_t\}$, we can solve investors' and the fund manager's optimization problems.

There are infinitely many small risk-neutral investors in the market and each investor's investment decision does not affect funds' returns and sizes, although all investors together do affect these. Investors' portfolio returns depend on three components: gross alphas, fees, and fund costs. BG show that the case in which fund managers actively manage funds choosing management fees f_t at each time t is equivalent to fund managers choosing sizes of funds they actively manage at each time t, charging fixed management fees f. As the latter case is more realistic, we focus on it to conduct our analysis.

At time t, a fund's costs variable $C(q_t^a)$ is a function of the fund's actively managed amount, q_t^a. Out of the q_t, the total fund assets, the amount $q_t - q_t^a$ $(q_t - q_t^a \geq 0)$ is invested in the passive index, earning the passive benchmark portfolio return and inducing no costs. There are decreasing returns to scale at the fund level, similar to BG and Feldman, Saxena, and Xu (2020, 2021). Thus, $C(q_t^a)$ is increasing and convex in q_t^a, and we assume

$$C(q_t^a) = c{q_t^a}^2, \tag{8}$$

where a known constant c, $c > 0$, is the fund cost sensitivity to size.

At time t, let price of the active fund's asset under management, net of fund costs and fees, be S_t, $0 \leq t \leq T$. Then, the active fund's net return is dS_t/S_t. As we normalize the passive benchmark portfolio's return to zero, the active fund's net return in excess of the passive benchmark is $dS_t/S_t - 0 = dS_t/S_t$. Hereafter, we briefly call dS_t/S_t net alpha. Then,

$$\frac{dS_t}{S_t} = \frac{q_t^a}{q_t}\frac{d\xi_t}{\xi_t} - \frac{C(q_t^a)}{q_t}dt - fdt. \tag{9}$$

Similar to BG, we assume that risk-neutral investors supply capital with infinite elasticity to funds that have positive excess expected returns. With sufficient capital, investors' fund allocations drive the conditional expectation of fund net alpha to zero at each time t. Thus,

$$\mathrm{E}\left[\frac{dS_t}{S_t}\middle|\mathcal{F}_t^{\xi}\right] = 0,\ \forall t. \tag{10}$$

Taking conditional expectation on Equation (9) and setting it to zero, we have

$$\frac{q_t^a}{q_t}A(t,\xi_t)m_t - \frac{c{q_t^a}^2}{q_t} - f = 0. \tag{11}$$

Rearranging,

$$fq_t = A(t,\xi_t)m_t q_t^a - c{q_t^a}^2. \tag{12}$$

Managers maximize fund profit fq_t by choosing q_t^a. Then, managers' problem is

$$\max_{q_t^a} fq_t = \max_{q_t^a} A(t,\xi_t)m_t q_t^a - c{q_t^a}^2 \tag{13}$$

subject to

$$0 \leq q_t^a \leq q_t. \tag{14}$$

Solving investors' and managers' problems, we obtain the equilibrium flow-performance relation.

2.3 The Flow-Performance Relation

As in BG, we define the lowest level of conditional expected manager ability that makes the fund survive, $\underline{m_t}$. If $m_t < \underline{m_t}$, the fund receives no investments and exits the market. We call $\underline{m_t}$ the survival ability level and assume $\underline{m_t} \geq 0$, although our results hold for unrestricted $\underline{m_t}$ values.[27] We assume $\underline{m_t} \geq 0$ because given updated information, expected instantaneous gross alpha accumulated in dt is $\mathrm{E}\left(d\xi_t/\xi_t|\mathcal{F}_t^{\xi}\right) = A(t,\xi_t)m_t dt$, with $A(t,\xi_t) > 0$. If $m_t < 0$, the expected instantaneous gross alpha is negative. With positive fund costs and fees, the expected instantaneous net alpha earned by investors in dt would be substantially smaller than zero, so investors would switch their investments to the passive benchmark portfolio. The optimal amount under active management and the optimal total assets under management, q_t^{a*} and q_t^*, are not trivial where $m_t > \underline{m_t} \geq 0$; otherwise, they both are zero. Also, we assume that the manager would set the fee f so that $q_t^{a*} \leq q_t^*$, as BG assume. Then, q_t^{a*} and q_t^* are

$$q_t^{a*} = \frac{A(t,\xi_t)m_t}{2c} \tag{15}$$

$$q_t^* = \frac{[A(t,\xi_t)m_t]^2}{4cf}. \tag{16}$$

To characterize the flow-performance relation, we apply Itô's Lemma to calculate dq_t^* and divide it by q_t^* to get the equilibrium percentage fund flows.[28] The following proposition shows the equilibrium flow-performance relation.

Proposition RN. Flow-Performance Relations with Risk-Neutral Investors. If $m_t \leq \underline{m_t}$, then the fund receives no investments. If $m_t > \underline{m_t}$, where investors are risk-neutral, then the equilibrium flow-performance relation is

$$\frac{dq_t^*}{q_t^*} = \frac{A(t,\xi_t)\sigma_m(\gamma_t)}{fB(t,\xi_t)}\left(\frac{dS_t}{S_t}\right) + \frac{A^2(t,\xi_t)\sigma_m^2(\gamma_t)}{4f^2B^2(t,\xi_t)}\left(\frac{dS_t}{S_t}\right)^2 \tag{17}$$

---

[27] That is, $\underline{m_t}$ can be positive, zero, or negative. In practice, usually $m_t \geq 0$, as explained above. However, if $\underline{m_t} < 0$, then when $\underline{m_t} < m_t < 0$, funds do not exit the market, and managers need to short their portfolios and invest in the benchmark portfolios, making $q_t^a < 0$. When $m_t \geq 0$ they stop shorting.

[28] We are interested in next instant change with fund flows relative to current fund flows. Conditional on observations, $A(t,\xi_t)$ is a constant, and we treat it as such when applying Itô's Lemma to calculate dq_t^*.

$$+2\left[\frac{a_0(t,\xi_t)}{m_t}+a_1(t,\xi_t)\right]dt.^{29}$$

Proof. See the Internet Appendix. □

Similar to the flow-performance relations found in the literature, our equilibrium fund flows are increasing with and convex in fund performance, and the flow-performance sensitivity and convexity decrease with fund fees.[30] Equation (17) shows the following new features of the flow-performance relation under our nonlinear–dynamic ability framework, compared with those found in the literature. First, higher sensitivity of expected manager ability to innovation shocks in fund gross alpha, $\sigma_m(\gamma_t)$, induces higher flow-performance sensitivity. This is because a higher $\sigma_m(\gamma_t)$ implies that shocks in fund gross alphas contain more information about manager ability (thus have more impact on the expectation of manager ability), making fund flows more sensitive to fund gross alphas. Second, higher fund gross alpha volatility $B(t,\xi_t)$ induces lower flow-performance sensitivity. This is because a higher $B(t,\xi_t)$ implies that the fund performance contains less information about manager ability, so investors rely less on fund performance when learning manager ability, making fund flows less sensitive to fund performance. Third, higher sensitivity of fund gross alpha to manager ability, $A(t,\xi_t)$, induces higher flow-performance sensitivity. This is because a higher $A(t,\xi_t)$ implies that fund performance is more highly correlated with manager ability in the long term. Consequently, fund return observations are more informative for manager ability and future returns, making fund flows more sensitive to fund performance. In addition, higher $\sigma_m(\gamma_t)$, $B(t,\xi_t)$, and $A(t,\xi_t)$ have stronger effects on fund flows when fund performance is high, so they affect the flow-performance convexity in the same directions as they affect the flow-

[29] The term $\left(\frac{dS_t}{S_t}\right)^2$ in Equation (17) in its continuous-time limit (the quadratic variation) is, in equilibrium, $\left(\frac{2fB(t,\xi_t)}{A(t,\xi_t)m_t}\right)^2 dt$ [see Equation (A6) in the Internet Appendix], suggesting an instantaneous linear flow-performance sensitivity. However, as investors allocate wealth to funds discretely, Equation (17) implies that flow-performance sensitivities are convex.

[30] The intuitions of these results are discussed in classical theoretical papers, such as BG.

performance sensitivity.

More importantly, as $\sigma_m(\gamma_t)$, $B(t,\xi_t)$, and $A(t,\xi_t)$ change over time nonmonotonically under our nonlinear–dynamic ability framework, Equation (17) implies that the equilibrium flow-performance sensitivity and convexity in this framework change with time nonmonotonically. Thus, if dynamic economic factors affect the coefficient parameters in our nonlinear framework, Equations (1) and (2), then our model explains and predicts the nonmonotonicity of the equilibrium flow-performance sensitivities and convexities induced by these economic factors. Section 4 provides implementations of our framework to incorporate some of these economic factors. More discussions of our model's explanatory and predictive power are in Section 5, where we illustrate our empirical evidence of flow-performance relations, and in Section 6 ,where we offer insights into empirical findings of flow-performance relations in the current literature.

We have the immediate results of the flow-performance relations under a linear–dynamic ability framework where parameters are constants and $\sigma_m(\gamma_t)$ changes monotonically to its steady state, as shown in the following corollary.

Corollary RN1. Flow-Performance Relations Under a Linear–Dynamic Ability Framework. Under a linear–dynamic ability framework, where a_0 a_1, b_1, b_2, A, and B are constants, equilibrium flow-performance sensitivities and convexities monotonically increase or decrease to their steady-state values or stay unchanged in the steady-state values.

□

2.4 Relation to Berk and Green (2004), Brown and Wu (2016), and Choi, Kahraman, and Mukherjee (2016)

BG provide one of the earliest discrete-time models that studies flow-performance relations and offers relevant insights. In their model, manager abilities are unknown constants that investors and the fund manager learn by observing fund returns. Our model nests BG in the sense that we can degenerate it to a continuous-time analog of it.

To make the manager ability θ_t an unobservable constant θ, we assign the following parameter values. In Equation (1), we set $a_0 = a_1 = b_1 = b_2 = 0$. In Equation (2), we set $A = 1$, i.e., the sensitivity of fund gross alpha to manager ability is one, to further simplify our model and match it with BG's. Then, Equations (7), (4), and (6) become

$$\sigma_m(\gamma_t) = \frac{\gamma_t}{B} \tag{18}$$

$$dm_t = \frac{\gamma_t}{B^2}\left(\frac{d\xi_t}{\xi_t} - m_t dt\right) \tag{19}$$

$$\gamma_t = \frac{\gamma_0 B^2}{B^2 + \gamma_0 t}. \tag{20}$$

The equilibrium flow-performance relation becomes

$$\frac{dq_t^*}{q_t^*} = \frac{1}{f}\left(\frac{\gamma_0}{B^2 + \gamma_0 t}\right)\left(\frac{dS_t}{S_t}\right) + \frac{1}{4f^2}\left(\frac{\gamma_0}{B^2 + \gamma_0 t}\right)^2\left(\frac{dS_t}{S_t}\right)^2. \tag{21}$$

This result is valid only if $m_t > \underline{m_t}$. Otherwise, the fund receives zero investments and $dq_t^*/q_t^* = 0$.

Proof. See the Internet Appendix. □

The flow-performance relation in Equation (21) is a continuous-time analog of Equation (30) in BG[31] and is also a special case of our equilibrium flow-performance relation shown in Equation (17). In this case, $\sigma_m(\gamma_t)$ monotonically decreases to the steady state, zero. Then, we have the immediate results of how the equilibrium flow-performance relation changes over time under a linear–constant ability framework.

Corollary RN2. Flow-Performance Relations Under a Linear–Constant Ability Framework. Under a linear–constant ability framework, such that $a_0 = a_1 = b_1 = b_2 = 0$ and A and B are constants, the equilibrium flow-performance sensitivity and convexity monotonically decrease to zero. □

As the above result shows, the equilibrium flow-performance relation is transient under

[31] See the discussion below Equation (30) of BG.

a linear–constant ability framework. This result also applies to Brown and Wu (2016) and Choi, Kahraman, and Mukherjee (2016), which use linear–constant ability frameworks to model cross-fund learning within fund families.[32] These two papers also find that sensitivities of fund flows to fund performances decrease monotonically over time, similar to our Corollary RN2.

However, the result of a linear–constant ability framework does not match with empirical data because empirically, even for very old funds, fund flows are still sensitive to performance and can become more sensitive as fund age increases. Our nonlinear–dynamic ability framework does not have this limitation because it allows the equilibrium flow-performance sensitivity and convexity to change nonmonotonically over time.

For illustration, we offer simulation results of equilibrium flow-performance sensitivities under different frameworks in the Internet Appendix.

2.5 Mean-Variance Risk-Averse Investors and the Flow-Performance Relation

In studying how investors' risk aversion affects the equilibrium flow-performance relation, we assume that investors are mean-variance risk-averse who maximize their portfolios' instantaneous Sharpe ratios. These investors' optimal portfolios are the same as those of investors with Bernoulli logarithmic preferences, who maximize expected utility [see, for example, Feldman (1992)]. Moreover, these portfolios are "growth optimal," as independently discovered by Bernoulli (in 1738) [see Bernoulli (1954)] and the "Kelly Criterion" [see Kelly (1956)]. This setting is also similar to the one of Pastor and Stambaugh (2012), Feldman, Saxena, and Xu (2020, 2021), and Feldman and Xu (2021).

Sharpe ratio maximization is a common feature while modeling mean-variance risk-averse investors' behavior. Current literature shows that if a fund manager's compensation is related to his/her portfolio's Sharpe ratio for a particular period, then that manager has

[32] In particular, if there is only one fund in the Brown and Wu (2016) model and in the Choi, Kahraman, and Mukherjee (2016) model, i.e., there is no cross-fund learning. Equations (5) and (6) in Brown and Wu (2016) become our Equations (19) and (20), whereas Equation (10) in Choi, Kahraman, and Mukherjee (2016) becomes a discrete-time analog of our Equation (19).

incentives for manipulation. The manager can increase (decrease) risk in the later part of the period if the return in the early part of the period is low (high) in order to improve the whole period's Sharpe ratio. Alternatively, he/she can trade off the tails of returns' distributions. In our model, as investors act on their own interests, they have no incentives to manipulate their portfolios' Sharpe ratios. Our assumption that investors maximize instantaneous portfolio Sharpe ratios prevents manipulation in our framework. See, for example, Ingersoll, Spiegel, and Goetzmann (2007), and Cvitanic, Lazrak, and Wang (2008).

Risk-averse investors, requiring compensation for increased risk that comes with excess return, do not drive alphas to zero. Thus, we need to change the model to incorporate this. First, we cannot normalize the passive benchmark portfolio return to be zero, as the level of this return is relevant.[33] Instead, we define the share price of the passive benchmark portfolio at time t, $0 \leq t \leq T$, as η_t. We also need to redefine net and gross alphas. The passive benchmark portfolio returns, $d\eta_t/\eta_t$, then follow

$$\frac{d\eta_t}{\eta_t} = \mu_p dt + \sigma_p dW_{p,t}, \tag{22}$$

where μ_p and σ_p are positive known constants and $W_{p,t}$ is a Wiener Process. Second, we still define $d\xi_t/\xi_t$ as the fund gross alpha, which follows the process defined in Equations (1) and (2), and define dS_t/S_t as the fund net alpha. As the active fund has beta loading of one on the passive benchmark portfolio, the fund gross return is $d\xi_t/\xi_t + d\eta_t/\eta_t$ and the fund net return is $dS_t/S_t + d\eta_t/\eta_t$. Also, we assume that the risk source of the benchmark return, $W_{p,t}$, is independent of that of gross alphas, so

$$dW_{p,t} d\overline{W}_t = 0. \tag{23}$$

Third, to simplify, we normalize the risk-free rate to zero.[34] All other terms stay the same.

[33] As risk-averse investors preferences are defined over their whole portfolios, they do not form their decision based on a marginal analysis of the active funds' risk alone. [See, for example, Equations **Error! Reference source not found.** and **Error! Reference source not found.** below, which collapse if the passive benchmark return is normalized to zero.]

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

Then, the investor's problem is to maximize the portfolio's instantaneous Sharpe ratio,

$$\max_{w_t} \frac{\mathrm{E}\left[\frac{dp_t}{p_t}\middle|\mathcal{F}_t^{\xi}\right]}{\sqrt{\mathrm{Var}\left[\frac{dp_t}{p_t}\middle|\mathcal{F}_t^{\xi}\right]}}, \tag{24}$$

subject to

$$0 \leq w_t \leq 1, \tag{25}$$

where w_t is the weight allocation to the active fund,[35] p_t is the portfolio's value, and dp_t/p_t is the portfolio's instantaneous return. The investor's portfolio's instantaneous return is

$$\frac{dp_t}{p_t} = w_t\left(\frac{dS_t}{S_t} + \frac{d\eta_t}{\eta_t}\right) + (1 - w_t)\frac{d\eta_t}{\eta_t} = w_t\frac{dS_t}{S_t} + \frac{d\eta_t}{\eta_t}. \tag{26}$$

Solving the investor's problem, we have the optimal weight allocation w_t^*. As investors face the same risk-return tradeoff and have the same objective function, they all make the same optimal decision of w_t^*. We define the part of the total wealth of all investors that is allocated to financial assets (i.e., allocated to the active fund and the passive benchmark portfolio) as V, $V \in (0, +\infty)$. In reality, this amount of wealth not only depends on the returns from financial assets, but also depends on production activities, research and development expenditures, consumptions, taxes, and many other aspects of the economy that we do not model here. To simplify our analysis, we assume that V is exogenous to both investors and managers and is a constant.[36] Here, the amount of wealth allocated to the fund or fund size is $q_t = w_t^* V$. As in the risk-neutral case, we can write the fund manager's profit as a function of q_t^a, i.e., $g(q_t^a)$, where g is some (smooth, increasing, concave) function.

Then, the manager's problem is

$$\max_{q_t^a} f q_t = \max_{q_t^a} g(q_t^a) \tag{27}$$

subject to

---

[35] As the risk-return tradeoff is the same for all investors, they make the same optimal decision in equilibrium, so we do not differentiate w_t across investors to simplify the notations.

[36] Even if we assume that V changes over time, i.e., V_t, the dynamics of V_t, does not affect the flow-performance sensitivity and convexity. Thus, assuming a dynamic V_t does not affect our model's insights.

$$0 \leq q_t^a \leq q_t. \tag{28}$$

By solving investors' and managers' problems,[37] we find q_t^{a*} and q_t^*

$$q_t^{a*} = \frac{A(t,\xi_t) m_t V \sigma_p^2}{2\left[B^2(t,\xi_t)\mu_p + cV\sigma_p^2\right]} \tag{29}$$

$$q_t^* = \frac{[A(t,\xi_t) m_t]^2 V \sigma_p^2}{4f\left[B^2(t,\xi_t)\mu_p + cV\sigma_p^2\right]}. \tag{30}$$

Substituting the equilibrium values q_t^{a*} and q_t^* into the fund net alpha in Equation (9), we have

$$\frac{dS_t}{S_t} = \frac{fB^2(t,\xi_t)\mu_p}{B^2(t,\xi_t)\mu_p + cV\sigma_p^2} dt + \frac{2fB(t,\xi_t)}{A(t,\xi_t) m_t} d\overline{W}_t. \tag{31}$$

Examining drift term on the right-hand side of Equation (31), we can see that, on average, funds' net alphas increase with management fees,[38] f, funds' gross alphas volatility, $B(t,\xi_t)$, and benchmark mean return, μ_p. On the other hand, funds' net alphas, on average, decrease with funds' cost sensitivity to size, c, and benchmark volatility, σ_p^2. Further, Equation (31) shows that expected funds' net alphas (conditional on current information) is positive where investors are risk-averse because all the parameters in the drift term are positive. This is because compared with the passive benchmark portfolio, the active fund is a riskier asset, thus, has to provide a higher expected return to attract investments. This result is consistent with the one in Pastor and Stambaugh (2012) and Feldman, Saxena, and Xu (2020, 2021).

Analysis similar to the one in the risk-neutral case yields the equilibrium flow-performance relation shown in the following proposition.

Proposition RA. Flow-Performance Relations with Mean-Variance Risk-Averse Investors.

---

[37] We assume that managers choose f such that the constraint $0 \leq q_t^{a*} \leq q_t^*$ is satisfied, so this constraint does not affect the manager's optimization process. Also, we assume that μ_p is sufficiently large or σ_p^2 is sufficiently small so that $0 \leq w_t^* \leq 1$, is satisfied, so this constraint does not affect the investors' optimization processes. See the proof in the Internet Appendix.

[38] A higher fee discourages investments to the fund, decreasing the fund's size. At a lower fund size, due to decreasing returns to scale, a manager is able to produce higher returns to investors.

If $m_t \leq \underline{m_t}$, then funds receive no investments from investors. If $m_t > \underline{m_t}$, where investors are mean-variance risk-averse, the equilibrium flow-performance relation is

$$\frac{dq_t^*}{q_t^*} = \frac{A(t,\xi_t)\sigma_m(\gamma_t)}{fB(t,\xi_t)}\left(\frac{dS_t}{S_t}\right) + \frac{A^2(t,\xi_t)\sigma_m^2(\gamma_t)}{4f^2B^2(t,\xi_t)}\left(\frac{dS_t}{S_t}\right)^2 + Y_t dt, \quad (32)$$

where

$$Y_t = \frac{2a_0(t,\xi_t)}{m_t} + 2a_1(t,\xi_t) - \frac{A(t,\xi_t)\sigma_m(\gamma_t)B(t,\xi_t)\mu_p}{B^2(t,\xi_t)\mu_p + cV\sigma_p^2}. \quad (33)$$

Here Y_t is independent of dS_t/S_t.

Proof. See the Internet Appendix. □

From Equation (32), we can see that investors' risk aversion does not affect the flow-performance sensitivity and convexity. Investors' risk aversion affects only the components of the fund flows that are unrelated to fund performance, as shown in Equations (32) and (33). The intuition is that investors' risk aversion affects investment amounts allocated to the risky active fund, q_t^*, so it affects the dollar amount of fund flows, dq_t^*. However, when the fund flows are calculated as percentage flow, dq_t^*/q_t^*, the effects of risk-aversion cancel out. Therefore, the flow-performance sensitivity and convexity when investors are mean-variance risk averse are similar to those when investors are risk neutral.

3 Funds' Exit Probability

Probability of fund exit is important, and our framework facilitates analyzing it. As mentioned earlier, we assume that a fund exits the market if the manager's inferred ability is below a survival ability level, i.e., $m_t < \underline{m_t}$. To simplify our discussion, we assume that $\underline{m_t}$ is exogenous and constant, $\underline{m_t} = \underline{m}$. The following proposition gives the probability that a fund exits the market in the forthcoming period.

Proposition FE1. Funds' Exit Probability in the Forthcoming Period. At time t, for $m_t >$

$\underline{m}$, the conditional probability for a fund to exit the market at $t + dt$ is

$$\text{Prob}\left(dm_t < \underline{m} - m_t \middle| \mathcal{F}_t^{\xi}\right) = \int_{-\infty}^{\underline{m}-m_t} \frac{e^{-\frac{[x-(a_0(t,\xi_t)+a_1(t,\xi_t)m_t)dt]^2}{2\sigma_m^2(\gamma_t)dt}}}{\sqrt{2\pi\sigma_m^2(\gamma_t)dt}} dx. \tag{34}$$

Proof. This is directly from the Gaussian distribution of dm_t shown in Equation (4). □

The probability that a fund will exit at a particular future time is also of interest. The time that the fund exits is the first future time when an existing fund's inferred ability deteriorates to the exit value $\underline{m}$. Mathematically, the problem is considerably more challenging than funds' exit probability in the next instant, shown above. We adapt results of Alili, Patie, and Pedersen (2005) (APP) to our framework and provide the probability of a fund's exit at any future time in the next proposition.

Proposition FE2. Funds' Exit Probability at a Future Time. At time t, for $m_t > \underline{m}$ and any time u, $u > t$, the probability of a fund's exit at time $u + du$, assuming current parameter values prevail, is $P_{m_t'-\underline{m'}}^{-a_1(t,\xi_t)/\sigma_m(\gamma_t)}(u-t)du$, where the density

$$P_{m_t'-\underline{m'}}^{-a_1(t,\xi_t)/\sigma_m(\gamma_t)}(u-t) = e^{-\frac{a_1(t,\xi_t)}{2\sigma_m(\gamma_t)}\left(\underline{m'}^2 - m_t'^2 - (u-t)\right)} P_{m_0'-\underline{m'}}^{0}(u-t) \times \mathrm{E}_t^{m_t'-\underline{m'}}\left[\exp\left(-\frac{a_1^2(t,\xi_t)}{2\sigma_m^2(\gamma_t)}\int_t^u \left(r_l + \underline{m'}\right)^2 dl\right)\right] \tag{35}$$

with

$$m_t' = \frac{1}{\sigma_m(\gamma_t)}\left(m_t + \frac{a_0(t,\xi_t)}{a_1(t,\xi_t)}\right) \quad \text{and} \quad \underline{m'} = \frac{1}{\sigma_m(\gamma_t)}\left(\underline{m} + \frac{a_0(t,\xi_t)}{a_1(t,\xi_t)}\right), \tag{36}$$

and r_l is a 3-dimensional Bessel bridge, over the interval $[t, u]$, between 0 and $m_0' - \underline{m'}$, with $r_t = m_t'$ and

$$dr_l = \left(\frac{\underline{m'} - r_l}{u - l} + \frac{1}{r_l}\right)dl + d\overline{W}_l, \tag{37}$$

under the density

$$P^0_{m'_0-\underline{m'}}(u-t)=\frac{\left|m'_t-\underline{m'}\right|}{\sqrt{2\pi(u-t)^3}}\exp\left(-\frac{\left(m'_t-\underline{m'}\right)^2}{2(u-t)}\right), \tag{38}$$

and $\mathrm{E}_t^{m'_0-\underline{m'}}(\cdot)$ is the expectation operator with this density. The density $P^0_{m'_0-\underline{m'}}(u-t)$ is the probability density of a fund's exit at $u+du$ if $a_0(t,\xi_t)=a_1(t,\xi_t)=0$.

Proof. We adapted APP theorem 5.1, to our process parameters $a_0(t,\xi_t)$, $a_1(t,\xi_t)$, $\sigma_m(\gamma_t)$, the initial condition m_t, and the exit barrier $\underline{m}$. □

Although Propositions FE1 and FE2 involve considerable mathematical notations, their implications are straightforward. First, if a manager's inferred ability m_t is farther away from the survival ability level $\underline{m}$, then it is more unlikely for the fund to exit the market. As a fund's equilibrium size is determined by m_t, the above results also imply that if a fund is larger, then the probability for it to exit the market is lower. Second, sensitivity of expected manager ability to innovation shocks in fund gross alpha, $\sigma_m(\gamma_t)$, also affects the fund's probability to exit the market. This is because a higher $\sigma_m(\gamma_t)$ implies that the same innovation shock in fund performance induces a larger change in m_t, consequently affecting m_t's relative size to $\underline{m}$. Third, parameters, such as $a_0(t,\xi_t)$ and $a_1(t,\xi_t)$, change over time, inducing more complex time patterns of the probability of exit.

We next write a fund's exit probability under a special case of survival ability levels. While economically, survival ability levels are a matter of scaling; mathematically, the solutions for this special case happen to fit existing functional forms and become substantially simpler and in full closed form.

Corollary FE2. Fund Exit Probability at a Future Time Under Zero Survival Level. Under the conditions of Proposition FE2, for the survival level, $\underline{m'}=0$, we have

$$P_{m_t'-0}^{-a_1(t,\xi_t)/\sigma_m(\gamma_t)}(u-t) = \frac{|m_t'|}{\sqrt{2\pi}}\left[\frac{-\frac{a_1(t,\xi_t)}{\sigma_m(\gamma_t)}}{\sinh\left(-\frac{a_1(t,\xi_t)}{\sigma_m(\gamma_t)}(u-t)\right)}\right]^{3/2}$$
$$\times \exp\left(\frac{\frac{a_1(t,\xi_t)m_t'^2}{2\sigma_m(\gamma_t)} \times e^{\frac{a_1(t,\xi_t)}{\sigma_m(\gamma_t)}(u-t)}}{\sinh\left(-\frac{a_1(t,\xi_t)}{\sigma_m(\gamma_t)}(u-t)\right)} - \frac{a_1(t,\xi_t)}{2\sigma_m(\gamma_t)}(u-t)\right). \tag{39}$$

Proof. As in the proof of Proposition FE2, we now adapted APP Equation 2.8. □

Based on the results of Propositions FE1 and FE2, we have some immediate results of the probability of exit for old funds, i.e., $t \to \infty$, as shown in the next corollary.

Corollary FEOF. Old Funds' Exit Probability. As a fund gets very old, i.e., $t \to \infty$, for $m_t > \underline{m}$ and $u > t$, we have the following results.

- Under a nonlinear–dynamic ability framework, both $\text{Prob}\left(dm_t < \underline{m} - m_t \middle| \mathcal{F}_t^{\xi}\right)$ and $P_{m_t'-\underline{m}'}^{-a_1(t,\xi_t)/\sigma_m(\gamma_t)}(u-t)$ still change with time and fund size.
- Under a linear–dynamic ability framework, both $\text{Prob}\left(dm_t < \underline{m} - m_t \middle| \mathcal{F}_t^{\xi}\right)$ and $P_{m_t'-\underline{m}'}^{-a_1/\sigma_m(\gamma_t)}(u-t)$ change only with fund size.
- Under a linear–constant ability framework, both $\text{Prob}\left(dm_t < \underline{m} - m_t \middle| \mathcal{F}_t^{\xi}\right)$ and $P_{m_t'-\underline{m}'}^{-a_1/\sigma_m(\gamma_t)}(u-t)$ are zero.

The intuitions of the above results are as follows. Under a linear–constant ability framework, as a fund gets old and investors have observed sufficiently many return realizations, the manager's constant ability is perfectly estimated, making the inferred ability stay unchanged and, consequently, a constant equilibrium fund size. Given that this old fund survives in the market (with inferred ability higher than the survival level), its inferred ability will be higher than the survival level for sure, so it will not exit the market. Under a linear–

dynamic ability framework, after the fund gets very old, the estimation precision achieves its steady state and will not change anymore. However, the manager's inferred ability, implied by fund size, will still change over time. Consequently, the probability of exit for this old fund depends on how far its inferred ability is from the survival level, making this probability a function of the manager's inferred ability level and, consequently, a function of fund size. As old funds tend to have a large size, a linear–dynamic ability framework would predict a very low exit probability for old funds. Under a nonlinear–dynamic ability framework, the coefficient parameters can be driven by dynamic economic factors, so they change over time. Consequently, when a fund gets very old, the probability of its manager's inferred ability to be below the survival level not only changes with its current inferred ability level (implied by size), but also changes with time.

Current mutual fund data shows that the probability for old funds to exit the market is nonnegligible and changes with time (we also show these results in our empirical study). While the linear frameworks, such as those used by BG, Dangl, Wu, and Zechner (2008), and Brown and Wu (2013, 2016), cannot explain these phenomena, our nonlinear–dynamic ability framework explains these phenomena in a systematic way.

For illustration, we offer simulation results of probabilities of fund exit under different frameworks in the Internet Appendix.

4 Implementations of Nonlinear–Dynamic Ability Frameworks

This section provides implementations of nonlinear–dynamic ability frameworks that incorporate effects of different economic factors into evolutions of manager abilities and gross alphas. Consequently, we show how these factors affect equilibrium flow-performance relations and old funds' exit probability. To simplify our discussions, we focus on risk-neutral investors and set some parameters constant or zero. We directly apply the corresponding results shown in Propositions RN, FE1, and FE2.

4.1 Fund Family Activity

Current literature demonstrates that fund families optimize family values by cross-fund subsidization (i.e., performance transfer among affiliated funds).[39] Consequently, given the same level of manager ability, the subsidization receiver's (provider's) fund gross alpha increases (decreases) with the intensity of this subsidization. To model this effect in a simplified way, we specialize Equations (1) and (2) as follows:

$$d\theta_t = b_1 dW_{1,t}, \tag{40}$$

$$\frac{d\xi_t}{\xi_t} = A(t)\theta_t dt + B dW_{2,t}. \tag{41}$$

Consider a simple case where cross-fund subsidization is from old, affiliated funds to young ones, as suggested by Gaspar, Massa, and Matos (2006) and Eisele, Nefedova, Parise, and Peijnenburg (2020). Then, a fund would receive return transfer when it is young and provide such transfer when it gets older. With the same level of manager ability, a fund's gross alpha is higher when it is young, on average. Thus, we set $A'(t) < 0$. By the earlier results, we have $\sigma_m(\gamma_t) \triangleq A(t)\gamma_t/B$. That is, the decrease in $A(t)$ exerts negative impact on $\sigma_m(\gamma_t)$ because gross alpha is less sensitive to manager ability, making inferred ability less sensitive to innovation shocks. However, depending on parameter values, the decrease in $A(t)$ also increases or decreases estimation error of manager ability, γ_t, which consequently exerts positive or negative impact on $\sigma_m(\gamma_t)$.

If the subsidization effect on $A(t)$ is so strong that $A(t)$ drives $\sigma_m(\gamma_t)$, then $\sigma_m(\gamma_t)$ also decreases over time. Consequently, in equilibrium, flow-performance sensitivity and convexity decrease with fund age because both $A(t)$ and $\sigma_m(\gamma_t)$ decrease over time. If γ_t is increasing over time and its effect is stronger than $A(t)$ in some periods and weaker in others, then $\sigma_m(\gamma_t)$ changes with time nonmonotonically. Consequently, equilibrium flow-performance sensitivity and convexity can change with fund age nonmonotonically. In these two cases, an old fund's exit probability still changes with time as $\sigma_m(\gamma_t)$ has no steady state.

[39] See, for example, Gaspar, Massa, and Matos (2006), Evans (2010), Eisele, Nefedova, Parise, and Peijnenburg (2020), and Xu (2021).

4.2 Manager Replacement

Dangl, Wu, and Zechner (2008) show that a fund's portfolio risk increases before a manager replacement and decreases after this replacement. The changes in portfolio risk would induce changes in the volatility of fund gross alpha. To model this effect, we simply specialize Equations (1) and (2) as follows:

$$d\theta_t = b_1 dW_{1,t}, \tag{42}$$

$$\frac{d\xi_t}{\xi_t} = A\theta_t dt + B(t) dW_{2,t}. \tag{43}$$

Assume that the time of a manager replacement is $\bar{t}$, and $B'(t) > 0$ when $t < \bar{t}$ and $B'(t) < 0$ when $t > \bar{t}$.[40] The above framework also implies that the change in manager ability induced by manager replacement is captured by the shock $dW_{1,\bar{t}}$. By the earlier results, we have $\sigma_m(\gamma_t) \triangleq A\gamma_t / B(t)$. That is, the increase (decrease) in $B(t)$ over time exerts negative (positive) impact on $\sigma_m(\gamma_t)$ because the higher (lower) gross alpha volatility induced by the future (previous) manager replacement makes return shocks less (more) informative, so that inferred ability is less (more) sensitive to these shocks. However, depending on parameter values, the change in $B(t)$ over time also increases or decreases γ_t over time. Therefore, putting all these effects together, it is plausible that $\sigma_m(\gamma_t)$ change over time nonmonotonically.

If the effect of manager replacement on $B(t)$ is sufficiently strong so that $B(t)$ drives $\sigma_m(\gamma_t)$, then $\sigma_m(\gamma_t)$ would decrease (increase) with time before (after) $\bar{t}$. Consequently, in equilibrium, flow-performance sensitivity and convexity decrease (increase) with fund age before (after) $\bar{t}$. Otherwise, the evolution of $B(t)$ due to manager replacement induces more complex time patterns of equilibrium flow-performance sensitivity and convexity. In all these cases, an old fund's exit probability changes with time because $\sigma_m(\gamma_t)$ has no steady state.

[40] Here we assume that the time of manager replacement, $\bar{t}$, is known and fixed. If $\bar{t}$ is random, it introduces complexity in the learning processes, but the equilibrium flow-performance sensitivity and convexity are still nonmonotonic over time, and an old fund's exit probability still changes with time. Also, multiple replacements would enhance these insights.

4.3 Competition by New Entrants

Competition induced by new funds affects the performance of incumbent funds. For example, Wahal and Wand (2011) show that if new funds have portfolios similar to those of incumbents, then incumbents' performances, fund fees, and investment flows are negative affected. To model this effect, we specialize Equations (1) and (2) as follows.

$$d\theta_t = [a_0(t) + a_1(t)\theta_t]dt + b_1 dW_{1,t}, \tag{44}$$

$$\frac{d\xi_t}{\xi_t} = A\theta_t dt + BdW_{2,t}, \tag{45}$$

where the ability tendency parameters, $a_0(t)$ and $a_1(t)$, decrease if there are a larger number of new funds with similar portfolios enter the market. As the number of these new entrants varies over time, $a_0(t)$ and $a_1(t)$ change over time nonmonotonically. By our earlier results, the changes in $a_0(t)$ and $a_1(t)$ affect the level of the fund flow directly, as ability tendency determines the expected return and consequently influences investment flows. Also, a higher $a_1(t)$ induces a larger γ_t because the inferred ability is less precise if changes in true unobservable ability are more corelated with previous unobservable true ability levels. Consequently, $\sigma_m(\gamma_t)$ is larger. Therefore, the nonmonotonicity of $a_1(t)$ induces nonmonotonicity in $\sigma_m(\gamma_t)$, making equilibrium flow-performance sensitivity and convexity change over time nonmonotonically. Moreover, old funds' exit probability changes over time nonmonotonically as $a_0(t)$, $a_1(t)$, and $\sigma_m(\gamma_t)$ change over time nonmonotonically.

The real-life situation can be more complex than those in the above implementations. For instance, fund families can also direct cross-fund subsidization from low-value funds to high-value funds, the effect of manager replacement on fund portfolio risk can vary across fund value, and competition induced by new entrants might have different impact on funds of different values. Consequently, the parameters shown in Sections 4.1, 4.2, and 4.3 also change with fund share value, i.e., $A(t,\xi_t)$, $B(t,\xi_t)$, $a_0(t,\xi_t)$ and $a_1(t,\xi_t)$, and other parameters can also be functions of t and ξ_t due to the effects of these factors. Further, besides the factors discussed above, there can be many other economic factors affecting these parameters.

More importantly, different economic factors can simultaneously affect the evolutions of gross alpha and manager ability and, thus, simultaneously exert impacts on the market equilibrium. As some economic factors (or forces) are unobservable in reality, even though we control for all observable variables, empirically we still expect to observe nonmonotonic flow-performance sensitivity and convexity, and old funds' dynamic exit probability over time.

5 Empirical Study

Based on our theoretical findings shown in the propositions and corollaries, we have the following empirical predictions.

- Under a nonlinear–dynamic ability framework, the flow-performance sensitivities and convexities change with time nonmonotonically, and the probability of exit for old funds changes with both time and fund size.
- Under a linear–dynamic ability framework, the flow-performance sensitivities and convexities change with time monotonically, and the probability of exit for old funds changes with fund size only.
- Under a linear–constant ability framework, the flow-performance sensitivities and convexities decrease with time monotonically, and the probability of exit for old funds is zero.

The goal of our empirical study is to show which framework is supported by empirical evidence.

5.1 Methodology

We first analyze the flow-performance relation using methods common in the literature. The (percentage) fund flow, $Flow_{i,t}$, is calculated as

$$Flow_{i,t} = \frac{TNA_{i,t} - TNA_{i,t-1}\left(1 + Ret_{i,t}\right)}{TNA_{i,t-1}}, \tag{46}$$

where, i is the fund index, t is the time (month) index, $TNA_{i,t}$ is the fund's total net assets under management, and $Ret_{i,t}$ is the fund net return. We use the fund net alpha, $\alpha_{i,t}$, to measure fund performance. Following Feldman, Saxena, and Xu (2020, 2021), we estimate the following style-matching model:

$$Ret_{i,t} = \alpha_{i,t} + b_{i,t}^1 F_t^1 + b_{i,t}^2 F_t^2 + \cdots + b_{i,t}^n F_t^n, \tag{47}$$

where F_t^1 through F_t^n are the net returns of tradable index funds of different asset classes. Following Berk and van Binsbergen (2015), we use tradable index funds as factors in this model because they represent the next-best investment opportunity available to investors as tradable assets. Among F_t^1 through F_t^n, we also allow for a “risk-free fund” by including the CRSP Fama-French risk-free rate as a potential benchmark. We perform this analysis on a rolling basis using returns from months $(t-60)$ to $(t-1)$ to avoid look-ahead bias. In particular, we estimate coefficients $b_{i,t}^1$ to $b_{i,t}^n$ to minimize the variance of the residual using observations in the previous 60 months and then subtract Ret_t by $b_{i,t}^1 F_t^1 + b_{i,t}^2 F_t^2 + \cdots + b_{i,t}^n F_t^n$ to calculate $\alpha_{i,t}$. Coefficients $b_{i,t}^1$ to $b_{i,t}^n$ are constrained to be between zero and one and to be summed up to one, as investors cannot short sell assets.

Our main purpose is to show how the flow–net alpha relations change over time. We use a fund’s age to represent its “time,” and analyze how the flow–net alpha sensitivity and convexity change with fund age. We use the model

$$\begin{aligned} Flow_{i,t} = \delta_0 + \delta_1 \alpha_{i,t-1} + \sum_j \beta_j \alpha_{i,t-1} \left(Age_{i,t-1}\right)^j + \sum_j d_j \left(Age_{i,t-1}\right)^j \\ + \delta * Controls_{i,t-1} + \varepsilon_{i,t}, \end{aligned} \tag{48}$$

where $Age_{i,t-1}$ is the lagged values of fund age. The coefficients β_j’s show how the flow–net alpha sensitivity changes over fund age. The coefficients of the control variables are represented by the vector δ. We follow the literature[41] to choose control variables in the vector $Controls_{i,t-1}$, which include the lagged values of the natural logarithm of the fund’s total net assets under management ($\ln TNA_{i,t-1}$), fund volatility ($Vol_{i,t-1}$), cross-sectional net alpha

[41] See, for example, Lynch and Musto (2003), Bollen (2007), Huang, Wei, and Yan (2007, 2012), Chen, Goldstein, and Jiang (2010), Spiegel and Zhang (2013), Brown and Wu (2016), Franzoni and Schmalz (2017), Harvey and Liu (2019), and Jiang, Starks, and Sun (2021).

dispersion ($Disp_{i,t-1}$), fund expense ratio ($Expense_{i,t-1}$), fund turnover ratio ($Turnover_{i,t-1}$), the weighted average flow of the fund class based on the Lipper fund classification, i.e., the style flow, ($StyleFlow_{i,t-1}$), fund flow ($Flow_{i,t-1}$), fund family net alpha ($FamAlpha_{i,t-1}$), the natural logarithm of fund family size ($\ln FamSize_{i,t-1}$), a dummy variable to indicate months with market risk premium between -5% and 5%, i.e., the moderate months, ($Mod_{i,t-1}$), the U.S. economic policy uncertainty index ($EPU_{i,t-1}$), and fund dummies and year dummies. The detailed definitions and constructions of these control variables are shown in the Data Appendix. When analyzing the flow–net alpha relations, we also include the interaction terms of $\ln TNA_{i,t-1}$ $Mod_{i,t-1}$, $Vol_{i,t-1}$, $Disp_{i,t-1}$, and $EPU_{i,t-1}$ with $\alpha_{i,t-1}$ because larger funds might experience less sensitive fund flows and the current literature shows that the flow–net alpha sensitivity is affected by fund volatility [Huang, Wei, and Yan (2012)], by market states [Franzoni and Schmalz (2017)], by cross-sectional net alpha dispersion [Harvey and Liu (2019)], and by economic policy uncertainty [Jiang, Starks, and Sun (2021)].

To analyze how the flow–net alpha convexity changes with fund age, we define a dummy variable $Pos_{i,t}$, where $Pos_{i,t} = 1$ if $\alpha_{i,t} \geq 0$, and $Pos_{i,t} = 0$ otherwise. We use the following model to do our analysis:

$$\begin{aligned} Flow_{i,t} = {} & \delta_0 + \delta_1 \alpha_{i,t-1} + \sum_j \beta_j \alpha_{i,t-1} \left(Age_{i,t-1}\right)^j + \delta_2 \alpha_{i,t-1} Pos_{i,t-1} \\ & + \sum_j \lambda_j \alpha_{i,t-1} Pos_{i,t-1} \left(Age_{i,t-1}\right)^j + \delta_3 Pos_{i,t-1} \\ & + \sum_j d_j \left(Age_{i,t-1}\right)^j + \delta * Controls_{i,t-1} + \varepsilon_{i,t}. \end{aligned} \tag{49}$$

The coefficients λ_j's show how the flow–net alpha convexity changes over fund age.

We also study how the probability of fund exit changes over fund age, especially for the old funds. We use the following model:

$$\text{Prob}(Exit_{i,t} = 1|\mathbf{X}_{i,t-1} = \mathbf{x}_{i,t-1}) \\ = F(\delta_0 + \delta_1 Age_{i,t-1} + \delta * Controls_{i,t-1} + \varepsilon_{i,t}), \tag{50}$$

where $Exit_{i,t}$ is a dummy variable equal to one if the fund exits the market and zero otherwise. A fund is defined as exited market if its last share class is delisted through liquidation or merge.[42] A fund can be delisted for other reasons. For example, it is changed to a closed-end fund, it closes to new investment and does not report fund information, or it is removed from the database with no reason reported. As these types of delisting might not indicate fund exits, we do not regard them as fund exits. The set of explanatory variables is represented by $\mathbf{X}_{i,t-1}$ and their realizations are represented by $\mathbf{x}_{i,t-1}$. We use the logistic cumulative distribution function for F to run a logit model, and for robustness check, we use the standard normal distribution function for F to run a probit model. Then, as we use monthly data, $\text{Prob}(Exit_{i,t} = 1|\mathbf{X}_{i,t-1} = \mathbf{x}_{i,t-1})$ represents funds' probability of exit in the next month, given the information in the current month.

5.2 Data

We collect our active fund data from the survivor-bias-free mutual fund database of the Center for Research in Security Prices (CRSP). Our sample period is from January 1990 to December 2020, and monthly data is used.[43] We first exclude index funds, variable annuity funds, and exchange-traded funds (ETFs). Then, we choose U.S. domestic equity mutual funds by using the Lipper fund classification.[44] This equity fund filter is similar to the one in Brown and Wu (2016), and the one in Feldman, Saxena, and Xu (2020), which is also close to the one

[42] In the CRSP database, in most of the cases where a fund's share class is merged, the acquirer's share class code is provided, showing that most of the "merges" in the database are actually acquisitions by other funds.

[43] Information on the Lipper fund classification and most of the information on the management company code to identify fund families begins in December of 1999. As we use a five-year rolling window to estimate fund net alpha, we start our sample from January 1990 so that our tests can include fund data starting from January 1995.

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

in Pastor, Stambaugh, and Taylor (2015).[45] Because we use a 5-year rolling window to estimate fund net alphas and because we require funds to have a long time-series of observations of all variables, i.e., at least 15 years, so that the variations of funds' ages are sufficiently large for analyzing how the flow–net alpha relation changes over fund age, we include funds that have at least 20 years of observations. We also require each of our equity funds to have fewer than 5 years of missing observations between the first observation and the last one, so that the style-matching model can perform well.

All fund returns are net of management expenses, 12b-fees, and front and rear load fees. We also obtain funds' net assets under management, the expense ratios, and turnover ratios from CRSP. While we analyze fund-level data, the CRSP data is offered at the fund share class level. We use the MFLINKS database to aggregate fund share class-level information to fund-level information. In particular, we calculate funds' total net assets under management by summing up its share classes' net assets under management, and calculate fund net returns, fund expense ratios, and fund turnover ratios as weighted averages of its share classes' net returns, fund expense ratios, and fund turnover ratios, respectively, using the lagged share class net assets under management as weights. Fund volatility is calculated as the standard deviation of the fund's net returns in the prior 12 months, and cross-sectional net alpha dispersion is calculated as the cross-sectional interquartile range of fund net alphas. Fund age is the time (in ten years) since the inception of the oldest share class. A fund's delisted time is the month after the last return observation of its last share class, and a fund is defined as exited market if its last share class is delisted due to liquidation, merge, or probably due to these, as indicated by CRSP. Fund family is identified by the management company code.[46] The fund family net alpha is calculated as the weighted average of the family members' net alphas, excluding the net alphas of the fund under consideration, where the lagged net asset under management is the

[45] See the discussion regarding the equity fund filter in Feldman, Saxena, and Xu (2020, Appendix).

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

weight.[47] Fund family size is calculated as the number of active equity funds in the family.

We also obtain data on index funds from Morningstar Direct and use the fund ID in the database to aggregate fund share class-level information to fund-level information for the index funds. These index funds, which we use as benchmark factors to estimate fund net alphas in the style-matching model in Equation (47), include a Large-Cap blend fund (Vanguard 500 Index), a Large-Cap equally-weighted fund (Invesco Equally-Weighted S&P 500), a Mid-Cap blend fund (Vanguard Extended Market Index), a Small-Cap blend fund (Vanguard Small-Cap Index Fund). We require index funds to have no missing observations in our sample period. The risk-free rate and excess return on the market (market risk premium) are collected from the Fama-French database in Wharton Research Data Services (WRDS). The economic policy uncertainty index is collected from the website Economic Policy Uncertainty[48].

Our *main sample* to analyze the flow–net alpha relations contains 769 funds, which are long-living active equity funds. To analyze the probability of fund exits, we extend our sample by including active equity funds with observations fewer than 15 years. This *extended sample* contains 3,030 funds.

5.3 Empirical Results

Table 1 reports the summary statistics. In our main sample, the fund net alpha, on average, is close to zero and its distribution tends to be symmetric. The fund flow is skewed to the right, as its mean is larger than its median. Also, the fund flow is large at the extremes. It is equal to 14% at the 99th percentile and -11% at the 1st percentile. Similar to Brown and Wu (2016), when analyzing the flow–net alpha relations, we winsorize the fund flow variable at the 1st and the 99th percentiles for each fund to mitigate the effects of extreme observations that are potentially due to fund mergers or data error.[49] Moreover, the R-squared of the style-

[47] In our sample, to be included in a family, a fund should be an active equity fund as defined above.
[48] The website address is http://www.policyuncertainty.com (accessed on April 1, 2021). Baker, Bloom, and Davis (2016) details the construction of this index.
[49] In unreported robustness tests, instead of winsorizing the flow observations, for each fund, we exclude the flow observations below the 1st or above the 99th percentile. We also winsorize or exclude the flow observations below the 1st or above the 99th percentile of the whole sample. We find very similar results in these robustness tests.

matching model is very high, with an average of around 90%, showing that the model fits well and it is unlikely that we have omitted relevant benchmark factors in estimating the fund net alphas. Furthermore, the average fund age in our sample is around twenty years, showing that they are long-living funds. In addition, a quarter of our funds belong to a small family with four or fewer member funds; there are also a few very big fund families in our main sample.

Flow–Net Alpha Relation and Fund Age

Table 2 illustrates the results of the model in Equation (48), which analyzes the flow–net alpha sensitivity. We find that, on average, the flow–net alpha sensitivity is positive, as shown in model specification (1). Then, we include the terms $\alpha_{i,t-1}\left(Age_{i,t-1}\right)^j$ and $\left(Age_{i,t-1}\right)^j$ up to $j = 4$ in the following model specifications.[50] We find that the interaction terms from $\alpha_{i,t-1}Age_{i,t-1}$ to $\alpha_{i,t-1}\left(Age_{i,t-1}\right)^4$ are all significant with negative, positive, negative, and positive signs, respectively, so, on average, the flow–net alpha sensitivity changes with fund age nonmonotonically.

Table 3 illustrates the results of the model in Equation (49), which analyzes the flow–net alpha convexity. We find that, on average, the flow–net alpha relation is convex, as shown in model specification (1), because fund flow is more sensitive to net alpha when the net alpha is positive than when it is negative. In particular, when the fund net alpha is positive (negative), a one percentage point higher in fund net alpha induces an increase in the fund flow by around 0.21% (0.11%). Corresponding to the results in Table 2, we include the interaction terms $\alpha_{i,t-1}Pos_{i,t-1}\left(Age_{i,t-1}\right)^j$ up to $j = 4$ in the following model specifications. We find that the interaction terms from $\alpha_{i,t-1}Pos_{i,t-1}Age_{i,t-1}$ to $\alpha_{i,t-1}Pos_{i,t-1}\left(Age_{i,t-1}\right)^4$ are all significant with negative, positive, negative, and positive signs, respectively, so on average, the flow–net alpha convexity changes with fund age nonmonotonically.

[50] The coefficients of these terms with $j = 5$ are insignificant.

We draw the flow–net alpha sensitivity and convexity over fund age, using the coefficient values of the model specification (5) in Table 2 and Table 3, respectively, and show the graphs in Figure 1. We plot the results for the average fund of ages of 10 to 80 years because around 90% of our observations correspond to a fund age within this range. When the average fund grows from 10 years old to 30 years old, the flow–net alpha sensitivity decreases continuously. When it is 30 years old, a one percentage point increase in net alpha induces an increase in fund flow by 0.1% lower than when it is 10 years old. The flow–net alpha sensitivity starts to increase after the fund age is more than 30 years, and then decreases again when fund age is around 45 years. After the average fund grows to around 65 years, the low–net alpha sensitivity increases again. The flow–net alpha convexity changes with fund age more volatilely. The convexity level decreases substantially when the average fund grows from 10 years old to 20 years old. After that, it gradually increases and then decreases again when the fund grows to around 45 years old. Then, the convexity level decreases again, and then increases again after the fund is 70 years old.

The decrease in the flow–net alpha sensitivity and convexity with fund age in the earliest years arises because investors have more precise estimates of manager ability during these years as more fund performances are revealed. The fluctuations of flow–net alpha sensitivity and convexity in the later years might be induced by multiple economic forces. For instance, Gaspar, Massa, and Matos (2006) and Eisele, Nefedova, Parise, and Peijnenburg (2020) show that fund families transfer returns from old funds to young funds when optimizing family profit. As shown in Section 4.1, with this subsidization, for the same level of manager ability, the fund's mean gross alpha is higher (lower) when it is young (old), i.e., $A(t, \xi_t)$ is large (small) when t is small (large). Also, Dangl, Wu, and Zechner (2008) show that a manager replacement should be preceded by a portfolio risk increase and followed by a portfolio risk decrease, and these patterns are also affected by the length of the manager's tenure. As discussed in Section 4.2, manager replacements would make gross alpha volatility,

$B(t, \xi_t)$, fluctuate over time. These forces, and other economic forces that create time-nonmonotonic gross alpha sensitivity to manager ability, fund return volatility, and sensitivity of inferred manager ability to fund performance all make the flow–net alpha sensitivities and convexities change nonmonotonically with fund age, resulting in the turning points in the graphs in the later years.

In short, our results show that the flow–net alpha sensitivity and convexity change with fund age nonmonotonically, even after controlling the factors, such as fund volatility, net alpha dispersion, market state, and economic policy uncertainty, that would affect the flow–net alpha relation shown in the current literature. These results support our nonlinear–dynamic ability framework but are inconsistent with linear–dynamic or constant ability frameworks. Therefore, we show that, besides factors found in the current literature, the dynamics of manager abilities and the nonlinear association of manager abilities and gross alphas induced by unexplored factors and latent factors are relevant forces driving the flow-performance relation.

We note that different funds can have different initial conditions and stochastic realizations for manager abilities and performances, and due to the dynamic nature of our model, we expect that different funds exhibit different patterns of flow–net alpha sensitivity and convexity. To study the flow–net alpha relations of individual funds, we re-run the model specification (5) in Table 2 and Table 3 for each fund without fund dummies and year dummies. We use the Newey-West estimator to estimate the standard errors, with the maximum lag of 12 to be considered in the autocorrelation structure of the regression error. In Table 5, we report the numbers of funds whose relevant coefficients are significant in Table 4.

Panel A shows the results of the flow–net alpha sensitivity. We find that, regarding each of the interaction terms from $\alpha_{i,t-1} Age_{i,t-1}$ to $\alpha_{i,t-1}(Age_{i,t-1})^4$, around 20% of funds have significant coefficients. Many funds have significantly positive coefficients of some of these interaction terms and significantly negative coefficients of other interaction terms, showing nonmonotonic flow–net alpha sensitivities over fund age. Panel B shows the results of the flow–net alpha convexity. We find that, regarding each of the interaction terms from $\alpha_{i,t-1}Pos_{i,t-1}Age_{i,t-1}$ to $\alpha_{i,t-1}Pos_{i,t-1}(Age_{i,t-1})^4$, around 22% of funds have significant coefficients. Many funds also exhibit nonmonotonic flow–net alpha convexities over fund age. Therefore, these results are also likely to support our nonlinear–dynamic ability framework, and these dynamic processes vary across funds.

Probability of Fund Exit and Fund Age

To analyze how the probability of fund exit changes with fund age, we use the extended sample in which we do not require funds to have at least 15 years of observations. First, we illustrate funds' survival rates at different age levels in Table 5. We can see that the survival rate does not monotonically increase with fund age but fluctuates with it. For example, the probability for a fund to survive in the next two years when it is 10 years old is 92.61%. This figure increases to 97.25% when it is 25 years old and decreases to 95.56% when it is 40 years old. Also, when funds reach 40 years old, the probability for them to exit the market at any time of the next five years is still close to 9%, showing that old funds exit the market with a probability that is not so low.

Table 6 reports the results of the logit model in Equation (50). We find that for the whole extended sample, holding other variables unchanged, a higher fund age significantly increases the probability of fund exit. This is also true in all the subsamples, in which we include observations such that the fund ages are above 10 years, above 15 years, above 20 years, and above 25 years, respectively. Also, the impact of fund age on the probability of fund exit is larger in the subsample containing funds older than 25 years, than in other subsamples and the whole extended sample. If we calculate the marginal effects, we find that at the average values of all the variables, a one-year increase in fund age induces a 0.0066% increase in the probability of fund exit in the next month for the subsample containing funds older than 25 years, whereas this figure is only 0.0019% for the whole extended sample. Thus, older funds are more likely to exit the market. This might be because, over time, funds face more intensive competition from new entrants who use similar portfolio strategies [see, for example, Wahal and Wand (2011)], and this type of competition has negative impacts on the tendency of incumbents' abilities, i.e., decreasing $a_0(t, \xi_t)$ and $a_1(t, \xi_t)$, as discussed in Section 4.3.

These results are inconsistent with a linear–dynamic ability framework or a linear–constant ability framework. Under the former framework, old funds, such as those with ages larger than 25 years, should have zero probability of exit, whereas under the latter framework, old funds' probability of exit should only change with fund size but not with fund age. However, we do not find these results. Instead, we find that old funds experience a nontrivial probability of exit (shown in Table 5), and their probability of exit increases with fund age after controlling for fund size. Thus, these results are consistent with our nonlinear–dynamic ability framework.

6 Insights into the Findings in the Literature

Current studies of the active fund management industry find interesting phenomena, and our model provides insights into these findings. Our model can even offer better explanations and insights to some particular phenomena than some of the current models do.

6.1 Insights on the Curvature of the Flow-Performance Relation

Several views can be found in the current literature on the curvature of the flow-performance relation. Some studies conclude that it is convex [see, for example, Lynch and Musto (2003), BG, and Brown and Wu (2016)], whereas other studies suggest that the relation is linear [see, for example, Spiegel and Zhang (2013)]. Our study complements this discussion by showing that the intercept, slope, and curvature of the function mapping flows to performance change over time nonmonotonically [see for example, Equations (17) and (32)], making empirical findings of these two types of curvature possible.

For illustration, in Figure 2, we show two situations of observations of fund flows and fund net alphas of a particular fund. For each situation, we plot four increasing and convex functions (the blue dashed curves) with the independent variable as the fund net alpha and the dependent variable as the fund flow, and the corresponding realized observation points (the red circles). These four functions show the flow–net alpha relation of the same fund in different time periods. In the first situation, shown on the left, the observations stay in an ellipse area. Consequently, the empirical fitted function (the black line) is an upward-sloping straight line. In the second situation, shown on the right, the observations stay in a crescent area, so the empirical fitted function (the black curve) is an increasing and convex curve. Thus, even though in theory, the flow-performance relation at each time is increasing and convex, empirically, we can observe that this relation is linear or convex.

If we put different funds' time-series observations of fund flows and fund performances together in a panel regression, the situation would be more complex. This is because in a panel regression, not only would the cross-sectional heterogeneity of functions mapping flows to performance affect the empirical findings of the curvature of this relation [see the discussion in the introduction of Spiegel and Zhang (2013)], but also the dynamics over time of these functions would affect the findings. Therefore, we need to incorporate these two types of effect in our analyses simultaneously. We leave this empirical issue for future studies.

6.2 Insights on the Flow-Performance Sensitivity and Market State

Franzoni and Schmalz (2017) find that the flow-performance sensitivity is steeper when the market excess return (i.e., the aggregate risk factor) is moderate than when it is extreme. They explain this "hump-shape" flow-performance sensitivity by developing a two-period learning model in which investors learn both the manager's skill and the loading of the fund's portfolio to the market factor. They show that in equilibrium, fund performance is less informative about manager skill when factor realizations are larger in absolute value, resulting in a "hump-shape" flow-performance sensitivity. However, as pointed out by Franzoni and Schmalz (2017), their model has a drawback: if the learning process has taken place for some more periods, the posterior estimates of manager skill and factor loading will be correlated, and this correlation can eliminate and even reverse the "hump shape" in the flow-performance sensitivity. Therefore, investors in their model learn and update the manager's skill and the fund's factor loading only once during their lifetime; then they change their investments to funds and pass on their fund holdings and beliefs to the next generation.[51] In other words, the Franzoni and Schmalz's (2017) model predicts the "hump-shape" flow-performance sensitivity for funds only in their earliest ages.

Our model can consistently explain and predict the "hump-shape" flow-performance sensitivities not only for young funds but also for old funds. In our model, if we assume that fund gross alphas are less sensitive to manager abilities under extreme market conditions, i.e., $A(t, \xi_t)$ is smaller when the absolute value of market excess return is large, then we predict a "hump-shape" equilibrium flow-performance sensitivity in a continuous-time framework for funds with different ages, as shown in Equations (17) and (32). This assumption is realistic because during periods with extreme market conditions, the market liquidity and volatility are less predictable. This makes funds' gross alphas less sensitive to manager abilities but more sensitive to luck. The lower sensitivity of gross alphas to manager abilities consequently reduces investors' reaction to fund returns, decreasing the flow-performance sensitivity.

6.3 Insights on Fund Marketing Activities

Current literature shows how funds' marketing activities affect the flow-performance relations. For example, Huang, Wei, and Yan (2007) find that funds with higher marketing expenses, in a fund family with star funds and in a large fund family, experience a less convex flow-performance relation. They theoretically show that given fund-level participation barriers, new investors can cover their participation costs only if the fund performance improves. Thus, a fund with high participation cost has fund flows increasing faster with fund performance (as part of the flows are from new investors), resulting in a more convex flow-performance relation. As marketing activities lower participation costs, funds with more marketing activities have a less convex flow-performance relation.

[51] See the discussions in Section 4.1 of Franzoni and Schmalz (2017).

We do not explicitly model the fixed up-front participation cost, as Huang, Wei, and Yan (2007) do, but our model can explain the less convex flow-performance relation due to more marketing activities. In our model, besides the variable fund costs, investors bear the other two types of cost over time: the management fee f and the estimation error of manager ability γ_t. If funds' marketing activities increase the former and/or decrease the latter, then the equilibrium flow-performance relation shown in our model [in Equations (17) and (32)] is less convex. This is plausible. For example, the fund manager is likely to charge a higher management fee f to cover additional marketing expenditures. Also, a fund's performance is likely to be correlated with other funds' performance in the same fund family,[52] so promoting the fund family and its star funds could offer additional information for investors to estimate the fund manager's ability, lowering the estimation error γ_t over time.

7 Conclusion

We introduce continuous-time rational models of the active fund management industry. We allow the dynamics of unobservable fund manager abilities and fund performances to follow a nonlinear framework, and allow economic factors to influence these dynamics, thus the market equilibrium. Our model predicts that in equilibrium, flow-performance sensitivity and convexity are nonmonotonic over time, and the probability of fund exit changes with time and with fund size at any fund age. In particular, we specialize our framework in three ways and demonstrate how each—cross-fund subsidization, manager replacement, and competition by new entrants—affects manager abilities and gross alpha productions and, consequently, induces nonmonotonicities in equilibrium. Our equilibrium results hold whether investors are risk neutral or mean-variance risk averse. On the other hand, if unobservable fund manager abilities and fund performances follow linear frameworks as in the current literature, equilibrium flow-performance sensitivity and convexity change over time only monotonically, and old funds' probability of exit changes with their size only. In our empirical study, we show

[52] Also see the discussions in Brown and Wu (2016) and Choi, Kahraman, and Mukherjee (2016).

nonmonotonicity in flow–net alpha sensitivity and convexity after controlling for the factors found in the current literature that would affect the flow–net alpha relations and show that the probabilities of funds' exit increase with fund age and decrease with fund size, even for old funds. Thus, our empirical results support our nonlinear framework.

Our framework enhances the explanatory and predictive power of relations and phenomena in the active fund management industry. We show that much of the empirical evidence in the current literature is consistent with our model. In particular, our theoretical results support, depending on parameter values, both linear and convex empirical flow-performance relations. We show that the flow-performance sensitivity is steeper when the market excess return is moderate at any fund age level, provided that fund gross alpha is less sensitive to manager ability under extreme market conditions. We also show that if marketing activities increase management fees and/or improve estimation precision of fund manager abilities, then the empirical flow-performance relation is less convex.

While this paper focuses on the flow-performance relation, our nonlinear dynamic unobservable manager abilities can be used to model dynamic unobservable human abilities in other areas of finance, economics, and other social sciences.

Data Appendix

This section details the definitions and constructions of the explanatory variables.

- $\alpha_{i,t}$ is the fund net alpha, calculated as the fund's net return minus that of the fund's style-matching benchmark portfolio estimated on a 5-year rolling basis. It is in decimal.
- $Pos_{i,t}$ is a dummy variable, with $Pos_{i,t} = 1$ if $\alpha_{i,t} \geq 0$, and $Pos_{i,t} = 0$ otherwise.
- $Age_{i,t}$ is the fund age. Fund age is calculated as the time since the inception of the fund's oldest share class. It is in ten years.
- $\ln Age_{i,t}$ is the natural logarithm fund age.
- $\ln TNA_{i,t}$ is the natural logarithm of the fund's total net assets under management. The total net assets under management is in million dollars.
- $Vol_{i,t}$ is the fund volatility, calculated as the standard deviation of the fund's net returns

in the prior 12 months. It is in decimal.

- $Disp_{i,t}$ is the cross-sectional net alpha dispersion, calculated as the cross-sectional interquartile range of net alphas, and it is in decimal.
- $Expense_{i,t}$ is fund expense ratio, the ratio of total investment that shareholders pay for the fund's operating expenses, including 12b-1 fees. It is in decimal.
- $TurnOver_{i,t}$ is fund turnover ratio, calculated as the minimum of aggregated sales and aggregated purchases of securities, divided by the average 12-month total net assets under management of the fund. It is in decimal.
- $StyleFlow_{i,t}$ is style flow, calculated as the weighted-average flow of the fund class based on Lipper fund classification, and is in decimal.
- $FamAlpha_{i,t}$ is fund family net alpha, calculated as the weighted average of the members' net alphas excluding the net alphas of fund i, where the lagged net asset under management is the weight. It is in decimal.
- $\ln FamSize_{i,t}$ is the natural logarithm of family size. Family size is the number of active equity funds that have net alpha observations in the family, and it is in integer.
- MRP_t is the market risk premium, calculated as the value-weight return on all NYSE, AMEX, and NASDAQ stocks (from CRSP) minus the one-month Treasury bill rate (from Ibbotson Associates). It is in decimal.
- $Mod_t = 1$ if $-5\% < MRP_{t-1} < 5\%$, and $Mod_{t-1} = 0$ otherwise.
- EPU_t is the three-component U.S. Economic Policy Uncertainty Index offered by the website http://www.policyuncertainty.com/. It is in decimal.

References

Alili, L, Patie, P., Pedersen, J. L., 2005. Representations of the first hitting time density of an Ornstein-Uhlenbeck process. Stochastic Models 21, 967–980.

Baker, S. R., Bloom, N., Davis, S. J., 2016. Measuring economic policy uncertainty. Quarterly Journal of Economics 131, 1593–1636.

Basak, S., Chabakauri, G., 2010. Dynamic mean-variance asset allocation. Review of Financial Studies 23, 2970–3016.

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

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

Berk, J. B., Stanton, R., 2007. Managerial ability, compensation, and the closed-end fund discount. The Journal of Finance 62, 529–556.

Berk, J. B., Tonks, I., 2007. Return persistence and fund flows in the worst performing mutual funds. Unpublished working paper. National Bureau of Economic Research.

Berk, J. B., van Binsbergen, J. H., 2015. Measuring skill in the mutual fund industry. Journal of Financial Economics 118, 1–20.

Bernoulli, D., 1954 (1738). Exposition of a new theory on the measurement of risk. Econometrica 22, 22–36.

Bollen, N. P. B., 2007. Mutual fund attributes and investor behavior. Journal of Financial and Quantitative Analysis 42, 683–708.

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

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

Carhart, M. M., 1997. On persistence in mutual fund performance. The Journal of Finance 52, 57–82.

Chen, J., Hong, H., Huang, M., Kubik, J., 2004. Does fund size erode mutual fund performance? The role of liquidity and organization. The American Economic Review 94, 1276–1302.

Chen, Q., Goldstein, I., Jiang, W., 2010. Payoff complementarities and financial fragility: Evidence from mutual fund outflows. Journal of Financial Economics 97, 239–262.

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

Cvitanic, J., Lazrak, A., Wang, T., 2008. Implications of the Sharpe ratio as a performance measure in multi-period settings. Journal of Economic Dynamics & Control 32, 1622–1649.

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

Detemple, J., 1986. Asset pricing in a production economy with incomplete information. The Journal of Finance 41, 383–391.

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

Evans, R. B., 2010. Mutual fund incubation. The Journal of Finance 65, 1581–1611.

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

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

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

Feldman, D., 2007. Incomplete information equilibrium: Separation theorems and other myths. Annals of Operations Research 151, 119–149.

Feldman, D., Leisen, D., 2021. Minimal dynamic equilibria. Unpublished working paper.

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

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

Feldman, D., Xu, J., 2021. Endogenous dynamic concentration of the active fund management industry. Unpublished working paper.

Ferreira, M. A., Keswani, A., Miguel, A. F., Ramos, S. B., 2012. The flow-performance relationship around the world. Journal of Banking and Finance 36, 1759–1780.

Ferreira, M. A., Keswani, A., Miguel, A. F., Ramos, S. B., 2013. The determinants of mutual fund performance: a cross-country study. Review of Finance 17, 483–525.

Franzoni, F., Schmalz, M.C., 2017. Fund flows and market states. The Review of Financial Studies 30, 2621–2673.

Gaspar, J., Massa, M., Matos, P., 2006. Favoritism in mutual fund families? Evidence on strategic cross-fund subsidization. The Journal of Finance 61, 73–104.

Harvey, C. R., Liu, Y., 2019. Cross-sectional alpha dispersion and performance evaluation. Journal of Financial Economics 134, 273–296.

Huang, J., Wei, K. D., Yan, H., 2007. Participation costs and the sensitivity of fund flows to past performance. The Journal of Finance 62, 1273–1311.

Huang, J., Wei, K. D., Yan, H., 2012. Investor learning and mutual fund flows. Unpublished working paper.

Ingersoll, J., Spiegel, M., Goetzmann, W., Welch, I., 2007. Portfolio performance manipulation and manipulation-proof performance measures. The Review of Financial Studies 20, 1503–1546.

Jiang, E., Starks, L.T., Sun, S., 2021. Economic policy uncertainty and learning: theory and evidence on mutual funds. Unpublished working paper.

Kacperczyk M., Nieuwerburgh S. V., Veldkamp L., 2014. Time-varying fund manager skill. The Journal of Finance 69, 1455–1484.

Kacperczyk M., Nieuwerburgh S. V., Veldkamp L., 2016. A rational theory of mutual funds' attention allocation. Econometrica 84, 571–626.

Kelly, J. L., 1956. A new interpretation of information rate. Bell System Technical Journal 35, 917–926.

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

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

Lynch, A. W., Musto, D. K., 2003. How investors interpret past fund returns. The Journal of Finance 58, 2033–2058.

Mamaysky, H., Spiegel, M., Zhang, H., 2007. Improved forecasting of mutual fund alphas and betas. Review of Finance 11, 359–400.

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

Pastor, L., Stambaugh, R. F., Taylor, L. A., 2015. Scale and skill in active management. Journal of Financial Economics 116, 23–45.

Rakowski, D., 2010. Fund flow volatility and performance. Journal of Financial and Quantitative Analysis 45, 223–237.

Roussanov, N., Ruan, H., Wei, Y., 2020. Mutual fund flows and performance in (imperfectly) rational markets? Unpublished working paper.

Spiegel, M., Zhang, H., 2013. Mutual fund risk and market share-adjusted fund flows. Journal of Financial Economics 108, 506–528.

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

Wang, Y., 2014. Mutual fund flows, performance persistence, and manager skill. Unpublished working paper. Chinese University of Hong Kong.

Xu, J., 2021. Hidden cross-fund subsidization, distortion of investors' belief, and flow-performance relation. Unpublished working paper.

Yan, X., 2008. Liquidity, investment style, and the relation between fund size and fund performance. Journal of Financial and Quantitative Analysis 43, 741–768.

Figure 1 Flow–Net Alpha Sensitivity and Convexity with Fund Age

Figure 1 illustrates the results of how flow–net alpha sensitivity and convexity change with fund age, in the upper plot and lower plot, respectively. The flow–net alpha sensitivity is expressed by $\delta_1 + \sum_{j=1}^{4} \beta_j \left(Age_{i,t-1}\right)^j$, and the parameter values δ_1 and β_1 to β_4 are from the estimated coefficient values of model specification (5) of Table 2. The flow–net alpha convexity is expressed by $\delta_2 + \sum_{j=1}^{4} \lambda_j \left(Age_{i,t-1}\right)^j$, and the parameter values δ_2 and λ_1 to λ_4 are from the estimated coefficient values of model specification (5) in Table 3. The vertical axes of the upper and lower plots are flow–net alpha sensitivity and convexity, respectively, and the horizontal axes of the two plots are $Age_{i,t-1}$ measured in ten years.

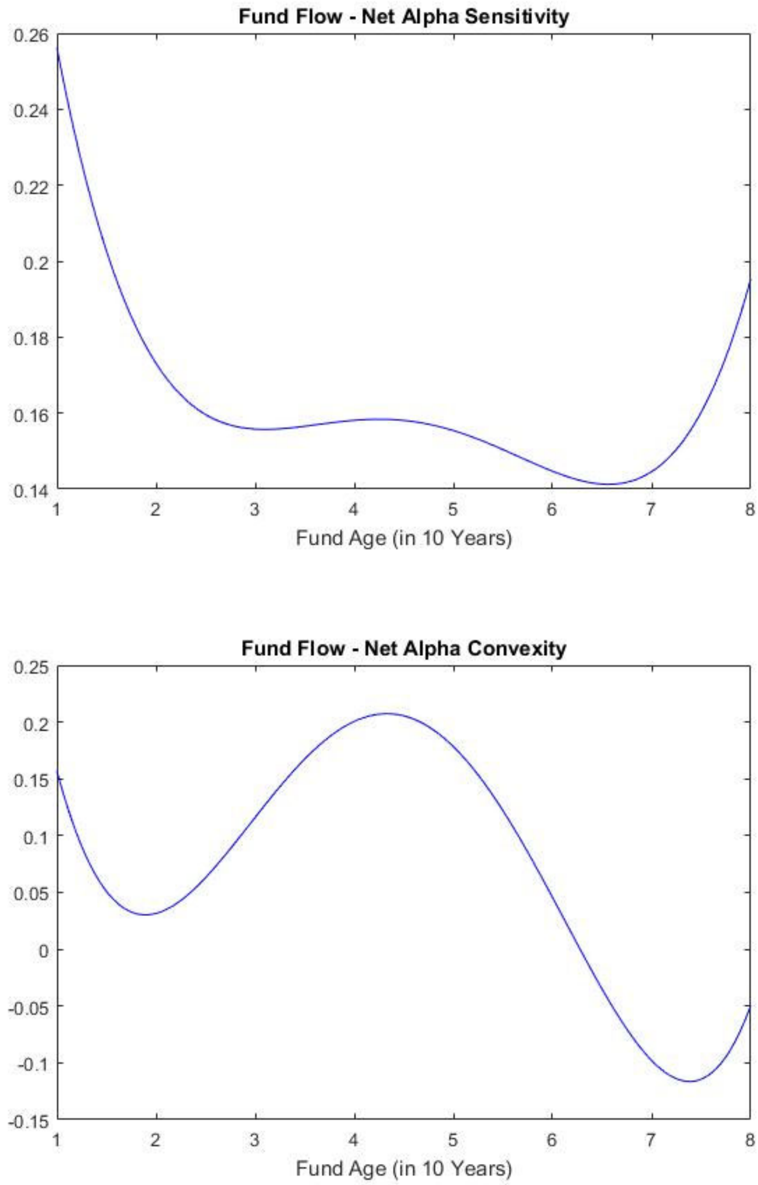

Figure 2. Examples of Fitting Observations of Fund Flows and Fund Net Alphas

Figure 2 illustrates two situations of fitting observations of fund flows and fund net alphas on a two-dimensional space. In each situation, each of the four blue dashed curves represents a function with fund net alpha (fund flow) as the independent variable (dependent variable) in a different time period for the same fund. Each of these functions is increasing and convex. The red circles represent the observations corresponding to these functions. Regarding the left (right) situation, the blue ellipse (blue crescent) indicates the area that the observations cover, and the black line (black curve) represents the empirical fitted function based on these observations.

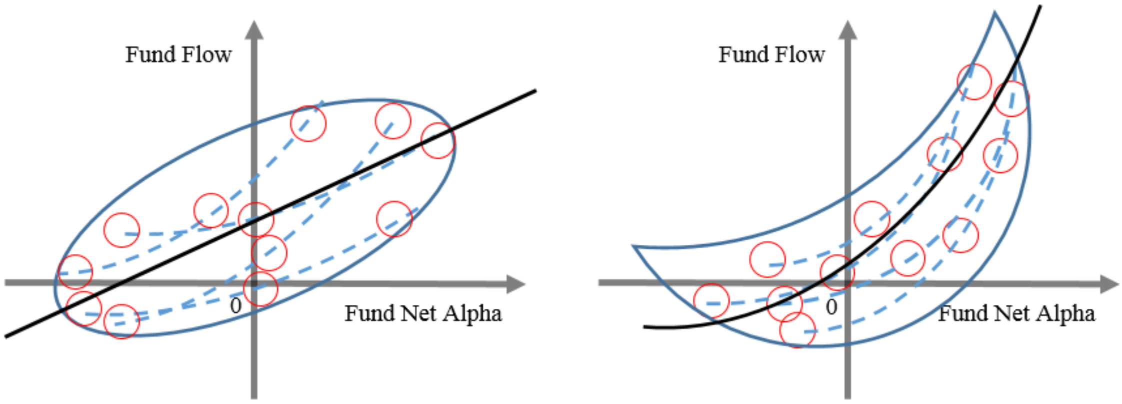

Table 1. Summary Statistics

Table 1 shows the summary statistics of our main sample, which contains 769 funds. The monthly observations from are January 1995 to December 2020. *Flow* is the fund percentage flow, calculated as the growth rate of total net asset under management minus fund net return, and it is in decimal. *Fund Net Return* is the fund return net of management expenses, 12b-fees, and front and rear load fees, and it is in decimal. *Alpha* is the fund net alpha $\alpha_{i,t}$ estimated by the style matching the model in Equation (47), and it is in decimal. The *Style-Matching Model R-squared* is the R^2 that we get by running the style matching model in Equation (47), and it is in decimal. *TNA* is the fund's total net asset under management measured in million dollars. *Expense* is the fund expense ratio as of the most recently completed fiscal year, including 12b-1 fees, and it is in decimal. *Turnover* is the fund turnover ratio, which is the minimum of aggregated sales and aggregated purchases of securities, divided by the average 12-month total net assets under management of the fund. It is in decimal. *Age* is the fund age, calculated as the time since the inception of the oldest share class, and it is in 10 years. *Vol* is the fund volatility, calculated as the standard deviation of the fund's net returns in the prior 12 months, and it is in decimal. *Disp* is the cross-sectional net alpha dispersion, calculated as the cross-sectional standard deviation of net alphas, and it is in decimal. *StyleFlow* is style flow, calculated as the weighted-average flow of the fund class based on Lipper fund classification, and it is in decimal. *FamAlpha* is the fund family's net alpha, calculated as the weighted average of the family members' net alphas, excluding the net alpha of the fund under consideration, where the lagged net asset under management is the weight, and it is in decimal. *FamSize* is the fund family size, calculated as the number of coexisting active equity funds in the family, and it is a number. *MRP* is market risk premium and it is in decimal. *Mod* is a dummy variable to indicate months with moderate MRP_t, and it is equal to one if $-5\% < MRP_t < 5\%$, and zero otherwise. *EPU* is the three-component U.S. Economic Policy Uncertainty Index.

Variable	Observation	Mean	Standard Deviation	Percentile 1st	25th	50th	75th	99th
Fund Variables								
Fund Flow (Decimal), *Flow*	190405	0.0006	0.3129	-0.1058	-0.0141	-0.0054	0.0048	0.1423
Fund Net Return (Decimal), *Ret*	190405	0.0085	0.0524	-0.1460	-0.0176	0.0127	0.0387	0.1346
Fund Net Alpha (Decimal), *Alpha*	190405	-0.0001	0.0189	-0.0504	-0.0089	-0.0003	0.0084	0.0510
Style-Matching Model R-Squared (Decimal)	190243	0.8845	0.0869	0.5870	0.8484	0.9054	0.9440	0.9884
Fund Total Net Asset (in 1 Million Dollar), *TNA*	190405	2946.12	9127.67	9.80	217.90	711.50	2156.60	44155.00
Fund Expense (Decimal), *Expense*	190405	0.0113	0.0037	0.0028	0.0091	0.0110	0.0132	0.0218
Fund Turn Over Ratio (Decimal), *TurnOver*	190405	0.7409	0.6073	0.0300	0.3261	0.5900	0.9700	2.9000
Fund Age (10 Years), *Age*	190405	2.2808	1.5728	0.5167	1.2417	1.8667	2.7167	7.9000
Fund Volatility (Decimal), *Vol*	190405	0.0466	0.0220	0.0148	0.0304	0.0423	0.0580	0.1139
Style Flow (Decimal), *StyleFlow*	190405	0.0007	0.0106	-0.0221	-0.0048	-0.0005	0.0052	0.0326
Family Net Alpha (Decimal), *FamAlpha*	190405	-0.0001	0.1336	-0.0377	-0.0060	-0.0003	0.0053	0.0369
Family Size (Number), *FamSize*	190405	12.3393	11.6432	2	4	9	16	59
Market Variable								
Net Alpha Dispersion (Decimal), *Disp*	311	0.0195	0.0085	0.0103	0.0138	0.0173	0.0226	0.0450
Market Risk Premium (Decimal), *MRP*	312	0.0076	0.0451	-0.1072	-0.0191	0.0134	0.0349	0.1018
Month with Moderate *MRP* (Dummy), *Mod*	312	0.7468	0.4355	0.0000	0.0000	1.0000	1.0000	1.0000
Economic Policy Uncertainty Index (Decimal), *EPU*	312	112.9226	44.0272	59.3240	81.0937	100.3744	133.2089	268.6164

Table 2. Flow–Net Alpha Sensitivity and Fund Age

Table 2 reports the results of the model in Equation (48). The dependent variable is *Flow*, the fund percentage flow. The explanatory variables are lagged by one month. The detailed definitions of the explanatory variables are in the Data Appendix. Standard errors are clustered by fund and presented in parentheses. The symbols ***, **, and * represent the 1%, 5%, and 10% significance levels, respectively, in a two-tail *t*-test.

	(1)	(2)	(3)	(4)	(5)
Lag_Alpha	0.1645***	0.2998***	0.3482***	0.4058***	0.4729***
	(0.0119)	(0.0599)	(0.0661)	(0.0777)	(0.0982)
*Lag_Alpha*Lag_Age*		-0.0312***	-0.0948***	-0.1840***	-0.3115***
		(0.0051)	(0.0191)	(0.0514)	(0.1116)
*Lag_Alpha*Lag_Age*^{2}			0.0095***	0.0398***	0.1106**
			(0.0024)	(0.0143)	(0.0503)
*Lag_Alpha*Lag_Age*^{3}				-0.0027**	-0.0167*
				(0.0011)	(0.0086)
*Lag_Alpha*Lag_Age*^{4}					0.0009*
					(0.0005)
Lag_Age	-0.0473***	-0.0486***	-0.0524***	-0.0602***	-0.0644***
	(0.0040)	(0.0040)	(0.0041)	(0.0044)	(0.0057)
Lag_Age^2			0.0007***	0.0035***	0.0057***
			(0.0001)	(0.0005)	(0.0018)
Lag_Age^3				-0.0002***	-0.0007**
				(0.0000)	(0.0003)
Lag_Age^4					0.0000
					(0.0000)
*Lag_Alpha*Lag_lnTNA*		-0.0066	-0.0049	-0.0044	-0.0042
		(0.0064)	(0.0062)	(0.0061)	(0.0061)
*Lag_Alpha*Lag_Vol*		-0.3578***	-0.4025***	-0.4306***	-0.4451***
		(0.0593)	(0.0633)	(0.0682)	(0.0712)
*Lag_Alpha*Lag_Disp*		-0.7393	-0.9607*	-1.0677*	-1.0867**
		(0.5498)	(0.5499)	(0.5496)	(0.5491)
*Lag_Alpha*Lag_Mod*		0.0632***	0.0675***	0.0692***	0.0684***
		(0.0144)	(0.0145)	(0.0145)	(0.0145)
*Lag_Alpha*Lag_EPU*		-0.0001	0.0000	0.0001	0.0001
		(0.0002)	(0.0002)	(0.0002)	(0.0002)
Lag_lnTNA	-0.0034***	-0.0033***	-0.0032***	-0.0030***	-0.0030***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Lag_Vol	0.0068	0.0088	0.0059	0.0078	0.0080
	(0.0151)	(0.0159)	(0.0159)	(0.0158)	(0.0158)
Lag_Disp	-0.0398**	-0.0378**	-0.0370**	-0.0363**	-0.0361**
	(0.0181)	(0.0181)	(0.0181)	(0.0181)	(0.0181)
Lag_Mod	0.0003	0.0004	0.0004	0.0004	0.0004
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Lag_EPU	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Lag_Expense	-0.5960***	-0.5910***	-0.6417***	-0.6828***	-0.6741***
	(0.1930)	(0.1925)	(0.1956)	(0.1976)	(0.1981)
Lag_TurnOver	-0.0020***	-0.0020***	-0.0019***	-0.0019***	-0.0019***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Lag_StlyeFlow	0.2646***	0.2616***	0.2579***	0.2576***	0.2572***
	(0.0189)	(0.0188)	(0.0188)	(0.0187)	(0.0187)
Lag_Flow	0.0123**	0.0122**	0.0121**	0.0120**	0.0120**
	(0.0059)	(0.0059)	(0.0058)	(0.0058)	(0.0058)
Lag_FamAlpha	0.0004***	0.0004***	0.0004***	0.0004***	0.0004***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Lag_lnFamSize	-0.0005	-0.0005	-0.0004	-0.0004	-0.0004
	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)
Constant	0.0789***	0.0796***	0.0829***	0.0876***	0.0894***
	(0.0061)	(0.0061)	(0.0062)	(0.0063)	(0.0066)
Year Dummies	Yes	Yes	Yes	Yes	Yes
Fund Dummies	Yes	Yes	Yes	Yes	Yes
Observations	190,405	190,405	190,405	190,405	190,405
R-squared	0.0294	0.0301	0.0310	0.0318	0.0319
Adjusted R-squared	0.0292	0.0299	0.0308	0.0316	0.0317

Table 3. Flow–Net Alpha Convexity and Fund Age

Table 3 reports the results of the model in Equation (49). The dependent variable is *Flow*, the fund percentage flow. The explanatory variables are lagged by one month. Other control variables of model specifications (1) to (5) are the same as those of model specifications (1) to (5) in Table 2, respectively. The detailed definitions of the explanatory variables are in the Data Appendix. Standard errors are clustered by fund and presented in parentheses. The symbols ***, **, and * represent the 1%, 5%, and 10% significance levels, respectively, in a two-tail *t*-test.

	(1)	(2)	(3)	(4)	(5)
Lag_Alpha	0.1029***	0.2138***	0.1914***	0.1735**	0.1070
	(0.0135)	(0.0577)	(0.0668)	(0.0824)	(0.1040)
*Lag_Alpha*Pos*	0.1081**	0.2459***	0.3815***	0.5286***	0.8058***
	(0.0452)	(0.0492)	(0.0842)	(0.1312)	(0.1982)
*Lag_Alpha*Pos*Lag_Age*		-0.0493***	-0.2038***	-0.4663***	-1.0405***
		(0.0165)	(0.0578)	(0.1413)	(0.3005)
*Lag_Alpha*Pos*Lag_Age* ^{2}			0.0278***	0.1342***	0.4655***
			(0.0078)	(0.0413)	(0.1407)
*Lag_Alpha*Pos*Lag_Age* ^{3}				-0.0106***	-0.0780***
				(0.0034)	(0.0248)
*Lag_Alpha*Pos*Lag_Age* ^{4}					0.0043***
					(0.0015)
*Lag_Alpha*Lag_Age*		-0.0059	0.0099	0.0591	0.2127*
		(0.0078)	(0.0273)	(0.0658)	(0.1289)
*Lag_Alpha*Lag_Age* ^{2}			-0.0047	-0.0291	-0.1199**
			(0.0036)	(0.0187)	(0.0585)
*Lag_Alpha*Lag_Age* ^{3}				0.0027*	0.0215**
				(0.0015)	(0.0101)
*Lag_Alpha*Lag_Age* ^{4}					-0.0012**
					(0.0006)
Pos	0.0003	-0.0008**	-0.0008**	-0.0008**	-0.0008**
	(0.0005)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
*Lag_Alpha*Lag_lnTNA*		-0.0070	-0.0049	-0.0047	-0.0045
		(0.0064)	(0.0063)	(0.0062)	(0.0062)
*Lag_Alpha*Lag_Vol*		-0.5706***	-0.6580***	-0.6972***	-0.7314***
		(0.0825)	(0.0983)	(0.1080)	(0.1124)
*Lag_Alpha*Lag_Disp*		-1.3410**	-1.5096**	-1.5702***	-1.5869***
		(0.5806)	(0.5847)	(0.5847)	(0.5830)
*Lag_Alpha*Lag_Mod*		0.0601***	0.0640***	0.0650***	0.0639***
		(0.0145)	(0.0146)	(0.0146)	(0.0145)
*Lag_Alpha*Lag_EPU*		-0.0001	0.0001	0.0001	0.0001
		(0.0002)	(0.0002)	(0.0002)	(0.0002)
Constant	0.0789***	0.0794***	0.0817***	0.0854***	0.0852***
	(0.0061)	(0.0061)	(0.0062)	(0.0064)	(0.0068)
Other Controls	Yes	Yes	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes
Fund Dummies	Yes	Yes	Yes	Yes	Yes
Observations	190,405	190,405	190,405	190,405	190,405
R-squared	0.0296	0.0306	0.0315	0.0323	0.0324
Adjusted R-squared	0.0294	0.0303	0.0313	0.0320	0.0322

Table 4. Flow–Net Alpha Sensitivity and Convexity: Results of Individual Funds

Table 4 reports the number of funds whose relevant coefficients in the models in Equations (48) and (49) are significant; these numbers are in Panel A and Panel B, respectively. The models are
$Flow_{i,t} = \delta_0 + \delta_1 \alpha_{i,t-1} + \sum_j \beta_j \alpha_{i,t-1} (Age_{i,t-1})^j + \sum_j d_j (Age_{i,t-1})^j + \delta * Controls_{i,t-1} + \varepsilon_{i,t}$, and
$Flow_{i,t} = \delta_0 + \delta_1 \alpha_{i,t-1} + \sum_j \beta_j \alpha_{i,t-1} (Age_{i,t-1})^j + \delta_2 \alpha_{i,t-1} Pos_{i,t-1} + \sum_j \lambda_j \alpha_{i,t-1} Pos_{i,t-1} (Age_{i,t-1})^j + \delta_3 Pos_{i,t-1} + \sum_j d_j (Age_{i,t-1})^j + \delta * Controls_{i,t-1} + \varepsilon_{i,t}$, respectively.
We set $j = 4$ for the above two models. The total number of funds in these tests is 769, and the models are run on each fund. The variables of these two models are the same as those in model specification (5) in Table 2 and Table 3, respectively, except that we do not include the fund dummies and year dummies. Newey-West estimator is used to estimate the standard errors, with the maximum lag of 12 to be considered in the autocorrelation structure of the regression error. The symbols ***, **, and * represent the 1%, 5%, and 10% significance levels, respectively, in a two-tail *t*-test. The last column shows the number of funds whose relevant coefficients are significant at least at the 10% significance level in a two-tail *t*-test.

Panel A: The Model in Equation (48) Flow–Net Alpha Sensitivity

Significance	*	**	***	**Total**
$\beta_1 > 0$	37	37	11	85
$\beta_1 < 0$	26	29	15	70
$\beta_2 > 0$	27	28	14	69
$\beta_2 < 0$	36	37	12	85
$\beta_3 > 0$	32	39	12	83
$\beta_3 < 0$	25	30	11	66
$\beta_4 > 0$	28	28	11	67
$\beta_4 < 0$	36	39	12	87

Panel B: The Model in Equation (49) Flow–Net Alpha Convexity

Significance	*	**	***	**Total**
$\lambda_1 > 0$	32	35	15	82
$\lambda_1 < 0$	24	36	25	85
$\lambda_2 > 0$	26	38	21	85
$\lambda_2 < 0$	30	34	17	81
$\lambda_3 > 0$	26	36	17	79
$\lambda_3 < 0$	31	38	17	86
$\lambda_4 > 0$	35	37	18	90
$\lambda_4 < 0$	27	32	19	78

Table 5. Fund Survival Rate and Fund Age

Table 5 reports the survival rates of funds at different fund age levels. The rows show the time that the funds can survive, and the columns show the ages of the funds, with the last column showing the average survival rate across all fund ages in our extended sample.

Fund Age	10 years	15 years	20 years	25 years	30 years	35 years	40 years	Sample Average
Survive in 1 year	96.18%	96.73%	97.20%	97.75%	98.80%	99.34%	98.89%	96.89%
Survive in 2 years	92.61%	94.96%	95.21%	97.25%	97.99%	96.71%	95.56%	94.15%
Survive in 3 years	89.99%	92.30%	92.54%	95.75%	97.59%	94.08%	95.56%	91.73%
Survive in 4 years	87.44%	90.18%	90.68%	95.00%	96.39%	93.42%	93.33%	89.60%
Survive in 5 years	84.77%	88.94%	89.48%	94.25%	93.98%	92.76%	91.11%	87.72%

Table 6. Probability of Fund Exit and Fund Age

Table 6 reports the results of the logit model in Equation (50). Column (1) reports the results of our whole extended sample, whereas columns (2), (3), (4), and (5) report the results of subsamples where funds' ages are above 10 years, above 15 years, above 20 years, and above 25 years, respectively. The dependent variable is *Exit*, which is one if the fund exits the market and zero otherwise. The explanatory variables are lagged by one month. The detailed definitions of the explanatory variables are in the Data Appendix. Fund random effects and year dummies are included in the logit model. Coefficients of the explanatory variables are reported. Standard errors are clustered by fund and presented in parentheses. The symbols ***, **, and * represent the 1%, 5%, and 10% significance levels, respectively, in a two-tail *t*-test.

	(1) All	(2) Age > 10 years	(3) Age > 15 years	(4) Age > 20 years	(5) Age > 25 years
Lag_Age	0.1051***	0.1311***	0.2774***	0.2294**	0.5091**
	(0.0340)	(0.0506)	(0.0879)	(0.0986)	(0.2160)
Lag_lnTNA	-0.4101***	-0.5430***	-0.7292***	-0.7308***	-1.0955***
	(0.0253)	(0.0433)	(0.0982)	(0.1098)	(0.2968)
Lag_Flow	-4.5150***	-4.8524***	-4.8603***	-5.5277***	-6.5042***
	(0.2352)	(0.3554)	(0.5472)	(0.7262)	(1.3238)
Lag_Alpha	-3.3538***	-7.4971**	-4.8451	-13.7987***	-11.0651
	(0.7408)	(3.5552)	(4.1532)	(3.7124)	(7.3056)
Lag_Vol	-3.8147	-7.4064**	-12.7267**	-16.0235**	-15.0987
	(2.4771)	(3.7360)	(5.8578)	(6.7914)	(10.7120)
Lag_Disp	-13.2655	-18.1927	-24.8829*	-30.1823	12.0848
	(8.2239)	(11.3132)	(14.4811)	(20.1245)	(25.0204)
Lag_Mod	0.1802*	0.1767	0.1978	0.1433	0.3176
	(0.1070)	(0.1393)	(0.1808)	(0.2488)	(0.3655)
Lag_EPU	0.0004	0.0016	0.0014	0.0010	-0.0044
	(0.0015)	(0.0020)	(0.0025)	(0.0033)	(0.0048)
Lag_Expense	-0.9580	-5.7381	28.0372**	7.5810	-29.8784
	(5.1402)	(6.2371)	(13.9906)	(26.5660)	(43.9043)
Lag_TurnOver	0.1743***	0.2303***	0.3525***	0.3187***	0.4058
	(0.0405)	(0.0598)	(0.0881)	(0.1112)	(0.2854)
Lag_StlyeFlow	0.0452***	0.0482***	-10.0412	-3.2099	-3.6440
	(0.0049)	(0.0047)	(6.8194)	(9.0784)	(13.4901)
Lag_FamAlpha	-6.4941	0.0043	-1.9275	-1.4659	4.2348
	(4.8169)	(0.0640)	(5.3084)	(7.0326)	(10.4950)
Lag_lnFamSize	0.3478***	0.4282***	0.6041***	0.7682***	1.2107***
	(0.0474)	(0.0633)	(0.1170)	(0.1512)	(0.2978)
Constant	-5.2954***	-4.7993***	-4.6575***	-4.7831***	-5.7979***
	(0.4050)	(0.5510)	(0.7801)	(1.0742)	(1.7506)
Year Dummies	Yes	Yes	Yes	Yes	Yes
Fund Random Effects	Yes	Yes	Yes	Yes	Yes
Observations	320,344	215,841	143,168	89,592	56,284
Number of Funds	3,030	2,010	1,382	944	565

Internet Appendix

This Internet Appendix provides the proofs of results of in Sections 2 and 3, and the simulation results of our theory.

Mathematical Proofs

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

Proof of Results in Section 2.3

The first-order condition with respect to q_t^a on the right-hand side of Equation (13), identifies $q_t^{a^*}$ as

$$q_t^{a^*} = \frac{A(t,\xi_t)m_t}{2c}. \tag{A1}$$

The second-order condition $-2c < 0$ shows that $q_t^{a^*}$ induces a maximum. Substituting Equation (A1) into Equation (13), the fund manager's optimal profit is

$$fq_t^* = \frac{[A(t,\xi_t)m_t]^2}{4c}. \tag{A2}$$

Rearranging, the optimal fund size is

$$q_t^* = \frac{[A(t,\xi_t)m_t]^2}{4cf}. \tag{A3}$$

Dividing Equation (A1) by Equation (A3) gives

$$\frac{q_t^{a^*}}{q_t^*} = \frac{2f}{A(t,\xi_t)m_t}. \tag{A4}$$

Here we assume that the manager sets f sufficiently low such that the constraint $0 \leq q_t^{a^*} \leq q_t^*$ is automatically satisfied and we do not incorporate this constraint in the optimization.

Also, substituting Equations (A1) and (A3) into Equation (9), we characterize the fund net alpha and gross alpha evolution relation as

$$\frac{dS_t}{S_t} = \frac{2f}{A(t,\xi_t)m_t}\frac{d\xi_t}{\xi_t} - 2f dt. \tag{A5}$$

Finally, substituting Equation (5) into Equation (A5), we have the fund net alpha evolution

$$\frac{dS_t}{S_t} = \frac{2fB(t,\xi_t)}{A(t,\xi_t)m_t} d\overline{W}_t. \tag{A6}$$

Thus, in equilibrium, the fund net alpha is normally distributed with mean zero and variance that decreases in inferred ability. That is, the higher the inferred ability, the lower is the noisy shocks' effect on net alpha.

Applying Itô's Lemma on q_t^* to Equation (A3) to derive dq_t^* and then dividing dq_t^* by q_t^* defined by Equation (A3), yields

$$\frac{dq_t^*}{q_t^*} = \frac{2A^2(t,\xi_t)m_t dm_t + A^2(t,\xi_t)(dm_t)^2}{[A(t,\xi_t)m_t]^2} = \frac{2m_t dm_t + (dm_t)^2}{m_t^2}. \tag{A7}$$

Substituting Equation (4) (for the dm_t terms) into Equation (A7) and then Equation (3) into the $d\overline{W}_t$ term yields

$$\begin{aligned}\frac{dq_t^*}{q_t^*} &= \frac{2\sigma_m(\gamma_t)}{m_t B(t,\xi_t)}\left(\frac{d\xi_t}{\xi_t}\right) + \frac{\sigma_m^2(\gamma_t)}{m_t^2 B^2(t,\xi_t)}\left(\frac{d\xi_t}{\xi_t}\right)^2 \\ &+ \frac{2}{m_t}\left[(a_0(t,\xi_t) + a_1(t,\xi_t)m_t) - \frac{A(t,\xi_t)\sigma_m(\gamma_t)m_t}{B(t,\xi_t)}\right]dt.\end{aligned} \tag{A8}$$

We substitute Equation (A5) into the flow-performance relation in Equation (A8) so that performance is measured by net alphas. We have

$$\begin{aligned}\frac{dq_t^*}{q_t^*} &= \frac{A(t,\xi_t)\sigma_m(\gamma_t)}{fB(t,\xi_t)}\left(\frac{dS_t}{S_t}\right) + \frac{A^2(t,\xi_t)\sigma_m^2(\gamma_t)}{4f^2B^2(t,\xi_t)}\left(\frac{dS_t}{S_t}\right)^2 \\ &+2\left[\frac{a_0(t,\xi_t)}{m_t} + a_1(t,\xi_t)\right]dt.\end{aligned} \tag{A9}$$

Q.E.D.

Proof of Results in Section 2.4

Going through the same process as the previous proof, we have a similar equilibrium relation between fund flows and expected manager abilities:

$$\frac{dq_t^*}{q_t^*} = \frac{2m_t dm_t + (dm_t)^2}{m_t^2}. \tag{A10}$$

Then, we directly substitute Equation (19) into (A10) and have

$$\frac{dq_t^*}{q_t^*} = \frac{2\gamma_t}{B^2 m_t}\left(\frac{d\xi_t}{\xi_t}\right) + \frac{\gamma_t^2}{B^4 m_t^2}\left(\frac{d\xi_t}{\xi_t}\right)^2 - \frac{2\gamma_t}{B^2}dt. \tag{A11}$$

Substituting Equations (A5) and (20) into Equation (A11) (with $A = 1$), we have

$$\frac{dq_t^*}{q_t^*} = \frac{1}{f}\left(\frac{\gamma_0}{B^2 + \gamma_0 t}\right)\left(\frac{dS_t}{S_t}\right) + \frac{1}{4f^2}\left(\frac{\gamma_0}{B^2 + \gamma_0 t}\right)^2\left(\frac{dS_t}{S_t}\right)^2. \quad \text{(A12)}$$

Q.E.D.

Proof of Results in Section 2.5

Substituting Equation (9) and then Equation (5) into Equation (26), and regarding q_t^a, q_t, and f as exogenous to the investor, we calculate $\mathrm{E}\left[\frac{dp_t}{p_t}\middle|\mathcal{F}_t^{\xi}\right]$ and $\mathrm{Var}\left[\frac{dp_t}{p_t}\middle|\mathcal{F}_t^{\xi}\right]$. Then, the investor's problem becomes

$$\max_{w_t's} \frac{\left[w_t\left(\frac{q_t^a}{q_t}A(t,\xi_t)m_t - \frac{c{q_t^a}^2}{q_t} - f\right) + \mu_p\right]dt}{\sqrt{\left[w_t^2\left(\frac{q_t^a}{q_t}\right)^2 B^2(t,\xi_t) + \sigma_p^2\right]dt}}, \quad \text{(A13)}$$

subject to

$$0 \leq w_t \leq 1. \quad \text{(A14)}$$

At each time t, the first-order condition with respect to w_t generates the optimal weight w_t^*:

$$w_t^* = \frac{\left(\frac{q_t^a}{q_t}A(t,\xi_t)m_t - \frac{c{q_t^a}^2}{q_t} - f\right)\sigma_p^2}{\left(\frac{q_t^a}{q_t}\right)^2 B^2(t,\xi_t)\mu_p}. \quad \text{(A15)}$$

The second-order condition is satisfied (the proof is omitted for brevity), so w_t^* is the maximizer.

As investors face the same risk-return tradeoff and have the same objective function, they all make the same optimal decision of w_t^*. Here, the amount of wealth allocated to the fund, i.e., the fund's size, is

$$q_t = w_t^* V = V\frac{\left(\frac{q_t^a}{q_t}A(t,\xi_t)m_t - \frac{c{q_t^a}^2}{q_t} - f\right)\sigma_p^2}{\left(\frac{q_t^a}{q_t}\right)^2 B^2(t,\xi_t)\mu_p}. \quad \text{(A16)}$$

By rearranging Equation (A16), we can express the fund manager's profit as

$$f q_t = -\frac{{q_t^a}^2 B^2(t,\xi_t)\mu_p}{V\sigma_p^2} - c{q_t^a}^2 + q_t^a A(t,\xi_t) m_t. \quad \text{(A17)}$$

The fund manager's objective is to maximize the fund's profit, $f q_t$, and to do so, the manager has to choose q_t^a to maximize the right-hand side of Equation (A17). Thus, the manager's problem can be written as

$$\max_{q_t^a} -\frac{{q_t^a}^2 B^2(t,\xi_t)\mu_p}{V\sigma_p^2} - c{q_t^a}^2 + q_t^a A(t,\xi_t) m_t, \quad \text{(A18)}$$

subject to

$$0 \le q_t^a \le q_t. \quad \text{(A19)}$$

Here, q_t and V are exogenous to the manager so are unaffected by his/her choice of q_t^a. Then, the first-order condition with respect to q_t^a generates the optimal weight ${q_t^a}^*$:

$$ {q_t^a}^* = \frac{A(t,\xi_t) m_t V\sigma_p^2}{2\left[B^2(t,\xi_t)\mu_p + cV\sigma_p^2\right]}. \quad \text{(A20)}$$

The second-order condition is $-\frac{2B^2(t,\xi_t)\mu_p}{V\sigma_p^2} - 2c < 0$, showing that ${q_t^a}^*$ is a maximizer.

Then, after substituting Equation (A20) into Equation (A17) and rearranging, we have the optimal fund size:

$$q_t^* = \frac{[A(t,\xi_t) m_t]^2 V\sigma_p^2}{4f\left[B^2(t,\xi_t)\mu_p + cV\sigma_p^2\right]}. \quad \text{(A21)}$$

We can see that

$$\frac{{q_t^a}^*}{q_t^*} = \frac{2f}{A(t,\xi_t) m_t}. \quad \text{(A22)}$$

We assume that manager i sets f sufficiently low such that the condition $0 \le q_t^{a*} \le q_t^*$ is automatically satisfied, and we do not incorporate this constraint in the optimization problem in Equation (A18).

The fund manager's optimal profit is

$$f q_t^* = \frac{[A(t,\xi_t) m_t]^2 V\sigma_p^2}{4\left[B^2(t,\xi_t)\mu_p + cV\sigma_p^2\right]}. \quad \text{(A23)}$$

A fund manager's higher expected ability and a higher benchmark volatility induce a higher optimal profit. On the other hand, a higher fund gross alpha volatility, a higher benchmark mean return, and higher fund cost sensitivity to size, induce a lower optimal profit.

Then, substituting Equations (A20) and (A21) into Equation (9), we get a relation between net alpha and gross alpha as follows:

$$\begin{aligned}\frac{dS_t}{S_t} &= \frac{2f}{A(t,\xi_t)m_t}\frac{d\xi_t}{\xi_t} - \frac{fcV\sigma_p^2}{B^2(t,\xi_t)\mu_p + cV\sigma_p^2}dt - fdt \\ &= f\left(\frac{2}{A(t,\xi_t)m_t}\frac{d\xi_t}{\xi_t} - \frac{B^2(t,\xi_t)\mu_p + 2cV\sigma_p^2}{B^2(t,\xi_t)\mu_p + cV\sigma_p^2}dt\right).\end{aligned} \tag{A24}$$

Then, substituting Equation (5) into Equation (A24), we have the fund net alpha:

$$\frac{dS_t}{S_t} = \frac{fB^2(t,\xi_t)\mu_p}{B^2(t,\xi_t)\mu_p + cV\sigma_p^2}dt + \frac{2fB(t,\xi_t)}{A(t,\xi_t)m_t}d\overline{W}_t. \tag{A25}$$

Substituting Equations (A20) and (A21) into Equation (A15), we have the optimal weight allocated to the active fund as

$$w_t^* = \frac{[A(t,\xi_t)m_t]^2 V\sigma_p^2}{4f\left[B^2(t,\xi_t)\mu_p + cV\sigma_p^2\right]}. \tag{A26}$$

As $m_t \geq \underline{m_t} \geq 0$ and all other parameters on the right-hand side of Equation (A26) are positive, we have $w_t^* \geq 0$, i.e., investors do not short sell the active fund. Also, with a sufficiently large μ_p or a sufficiently small σ_p^2, we have $w_t^* \leq 1$. The intuition is that as long as the passive benchmark portfolio provides sufficiently high expected return or sufficiently low risk, investors do not short sell it. These results are realistic because, in reality, we observe that investors invest part of their wealth in active funds and another part in passive benchmark portfolios. Then, the condition $0 \leq w_t \leq 1$ is automatically satisfied and we do not incorporate this constraint in solving the investors' optimization problems.

Then, substituting Equations (A22), (31), and (A26) into Equation (A13), we have the investor's optimal instantaneous Sharpe ratio at time t:

$$\frac{\left[\frac{[A(t,\xi_t)m_t]^2\sigma_p^2}{4f\left[B^2(t,\xi_t)\mu_p+cV\sigma_p^2\right]}\times\frac{fB^2(t,\xi_t)\mu_p}{B^2(t,\xi_t)\mu_p+cV\sigma_p^2}+\mu_p\right]dt}{\sqrt{\left[\left(\frac{[A(t,\xi_t)m_t]^2\sigma_p^2}{4f\left[B^2(t,\xi_t)\mu_p+cV\sigma_p^2\right]}\right)^2\left(\frac{2f}{A(t,\xi_t)m_t}\right)^2B^2(t,\xi_t)+\sigma_p^2\right]dt}}$$

$$=\frac{\left[\frac{[A(t,\xi_t)m_t]^2\sigma_p^2B^2(t,\xi_t)\mu_p}{4\left[B^2(t,\xi_t)\mu_p+cV\sigma_p^2\right]^2}+\mu_p\right]dt}{\sqrt{\left[\frac{[A(t,\xi_t)m_t]^2\sigma_p^4B^2(t,\xi_t)}{4\left[B^2(t,\xi_t)\mu_p+cV\sigma_p^2\right]^2}+\sigma_p^2\right]dt}}. \tag{A27}$$

Now we are ready to derive the flow-performance relation. Applying Itô's Lemma to Equation (A21) to derive dq_t^*, then dividing by q_t^* from Equation (A21), we have

$$\frac{dq_t^*}{q_t^*}=\frac{2m_tdm_t+(dm_t)^2}{m_t^2}. \tag{A28}$$

We note that the above result is valid only if $m_t>\underline{m_t}$. If $m_t\leq\underline{m_t}$, then $dq_t^*/q_t^*=0$.

Given $m_t>\underline{m_t}$, we substitute Equation (4) (for the dm_t terms) into Equation (A28), and then replace the $d\bar{W}_t$ term by its definition in Equation (3). We have the flow-performance relation using gross alpha as the performance measure:

$$\frac{dq_t^*}{q_t^*}=\frac{2\sigma_m(\gamma_t)}{m_tB(t,\xi_t)}\left(\frac{d\xi_t}{\xi_t}\right)+\frac{\sigma_m^2(\gamma_t)}{m_t^2B^2(t,\xi_t)}\left(\frac{d\xi_t}{\xi_t}\right)^2+\frac{2}{m_t}\left[(a_0(t,\xi_t)+a_1(t,\xi_t)m_t)-\frac{A(t,\xi_t)\sigma_m(\gamma_t)m_t}{B(t,\xi_t)}\right]dt. \tag{A29}$$

We then substitute Equation (A24) into Equation (A29) and get the flow-performance relation using net alpha as the performance measure:

$$\frac{dq_t^*}{q_t^*}=\frac{A(t,\xi_t)\sigma_m(\gamma_t)}{fB(t,\xi_t)}\left(\frac{dS_t}{S_t}\right)+\frac{A^2(t,\xi_t)\sigma_m^2(\gamma_t)}{4f^2B^2(t,\xi_t)}\left(\frac{dS_t}{S_t}\right)^2+Y_tdt, \tag{A30}$$

where

$$Y_t=\frac{2}{m_t}\left[(a_0(t,\xi_t)+a_1(t,\xi_t)m_t)-\frac{A(t,\xi_t)\sigma_m(\gamma_t)m_t}{B(t,\xi_t)}\right]+\frac{A(t,\xi_t)\sigma_m(\gamma_t)\left(B^2(t,\xi_t)\mu_p+2cV\sigma_p^2\right)}{B(t,\xi_t)\left(B^2(t,\xi_t)\mu_p+cV\sigma_p^2\right)} \tag{A31}$$

$$= \frac{2a_0(t,\xi_t)}{m_t} + 2a_1(t,\xi_t) - \frac{A(t,\xi_t)\sigma_m(\gamma_t)B(t,\xi_t)\mu_p}{B^2(t,\xi_t)\mu_p + cV\sigma_p^2}.$$

Here Y_t is independent of either $d\xi_t/\xi_t$ or dS_t/S_t.

Q.E.D.

Simulation Results

We use simulation to illustrate our equilibrium flow-performance relation under linear–constant ability framework [as in BG, Case One)], linear–dynamic ability framework (Case Two), and nonlinear–dynamic ability framework (Case Three). We assume risk-neutral investors in this illustration. We discretize our continuous-time processes into discrete-time processes, setting $dt = \Delta t$ to be one month and $d\bar{W}_t = \Delta\bar{W}_t$ to follow a normal distribution of mean zero and variance Δt.

In our simulation, we use some statistics from our sample of the active equity mutual funds in the U.S. market. A detailed description of the sample is in the Data section. In the sample, the average annual fund expense, including the 12b-1 fee, is 1.13%; the average monthly net alpha is -0.01%; and the standard deviation of the monthly net return is 5.24%. Then, we set our model's parameters as follows. Our monthly fee f is $1.13\%/12 = 0.094\%$ (which is approximated by the average monthly fund expense); our initial expected management ability m_0 is $1.13\% - 0.01\% = 1.12\%$ (which is approximated by the average fund gross alpha in our sample); and our gross fund share price $\xi_0 = 1$. We set other parameters under three cases:

- Case One: $a_0 = 0,\ a_1 = 0,\ b_1 = 0,\ b_2 = 0,\ A = 1,\ B = 0.0524,\ \gamma_0 = 0.0008$.
- Case Two: $a_0 = 0,\ a_1 = 0,\ b_1 = 0.04,\ b_2 = 0.001,\ A = 0.01,\ B = 0.0524,\ \gamma_0 = 0.0001$.
- Case Three: $a_0 = 0$, $a_1 = 0$, $b_1 = 0.0005 + 0.000012t + 0.000004\ln(1 + \xi_t)$, $b_2 = 0.0001,\ A = 0.55 + 0.001\ln(1 + \xi_t),\ B = 0.1 - 0.01\ln(t) + 0.01\ln(1 + \xi_t)$, $\gamma_0 = 0.001$.

We then simulate m_t, γ_t, $\sigma_m(\gamma_t)$, fund net alphas $\Delta S_t/S_t$, and fund flows $\Delta q_t^*/q_t^*$. We plot the results of $\sigma_m(\gamma_t)$ in the three cases, from Month 1 to Month 300, in Figure A1. Also, in Figure A1, we use blue circles to plot the values of fund flows and fund net alphas from Month 13 to Month 36, green stars to plot these values from Month 61 to Month 84, and red plus signs to plot these values from Month 241 to Month 264.

Case One is the case in BG, where the estimation of constant manager ability become more and more precise over time. In this case, $\sigma_m(\gamma_t)$ is deterministic and decreasing over time. With more precise ability estimates, investors rely less and less on realized performance to infer ability; consequently, the conditional expected manager ability is less and less sensitive to shocks to gross alphas. As a result, the flow–net alpha sensitivity decreases over time.

Case Two is the case of a linear–dynamic ability framework. We choose the value of γ_0 to be below the steady state value of γ_t, and over time, γ_t increases towards its steady state value. In this case, $\sigma_m(\gamma_t)$ is deterministic and increasing over time, i.e., the inferred ability is more and more sensitive to new shocks to gross alphas. Consequently, the flow–net alpha sensitivity increases over time. If we set the value of γ_0 to be above the steady state value of γ_t, then γ_t decreases over time toward its steady state value. Consequently, $\sigma_m(\gamma_t)$ decreases over time and the flow–net alpha sensitivity decreases over time.

Case Three is the case of a nonlinear–dynamic ability framework. $\sigma_m(\gamma_t)$ is stochastic, and it first decreases over time and then increases, i.e., the inferred ability is less and less sensitive to new shocks to gross alphas over the early months and then becomes more and more sensitive over the later months. Also, as some parameters are functions of ξ_t, the randomness of ξ_t affects $\sigma_m(\gamma_t)$'s value, making it fluctuate slightly over time. Eventually, the flow–net alpha sensitivity first decreases and then increases.

Different from the results in BG that the flow–net alpha sensitivity decreases monotonically over time, our results show that it can change with different patterns over time. In reality, we expect that the pattern of $\sigma_m(\gamma_t)$ over time and that of the flow–net alpha sensitivity may be complex.

Next, we impose the survival ability level $\underline{m} = 0$, such that if $m_t < 0$, the fund exits the market. We show the results of m_t and funds' exit density in Figure A2. In Case One, as the manager ability is constant, its estimate m_t converges to the true ability level very quickly. As this level is positive, the fund never exits the market, and the exit density is zero over time. In Case Two, the manager ability is dynamic, resulting in a more volatile m_t. Over time, when m_t decreases and close to zero, the exit density surges. The fund exits the market in Month 121, and the exit density is very high in the previous few months. In Case Three, the manager ability is dynamic and associated with gross alpha in a nonlinear framework. Over time, m_t is volatile and starts to decrease after around 280 months. The fund exits the market even when it is very old (in Month 343). Thus, these results show that under our nonlinear–dynamic ability framework, old funds can still exit the market, which is different from the prediction of BG that old funds are very unlikely to exit the market.

Figure A1. Simulation Results of Fund Flows and Performances

Figure A1 illustrates the simulation results using parameters defined in Case One, Case Two, and Case Three, in the two subplots on the top, two subplots in the middle, and two subplots at the bottom, respectively. For each case, on the left-hand side, we illustrate the sensitivity of expected manager ability to shocks in gross alphas, $\sigma_m(\gamma_t)$, from Month 1 to Month 300; and on the right-hand side, we illustrate the fund flows (vertical axis) and fund net alpha (horizontal axis) from Month 13 to Month 36 in blue circles, from Month 61 to Month 84 in green stars, and from Month 241 to Month 264 in red plus signs.

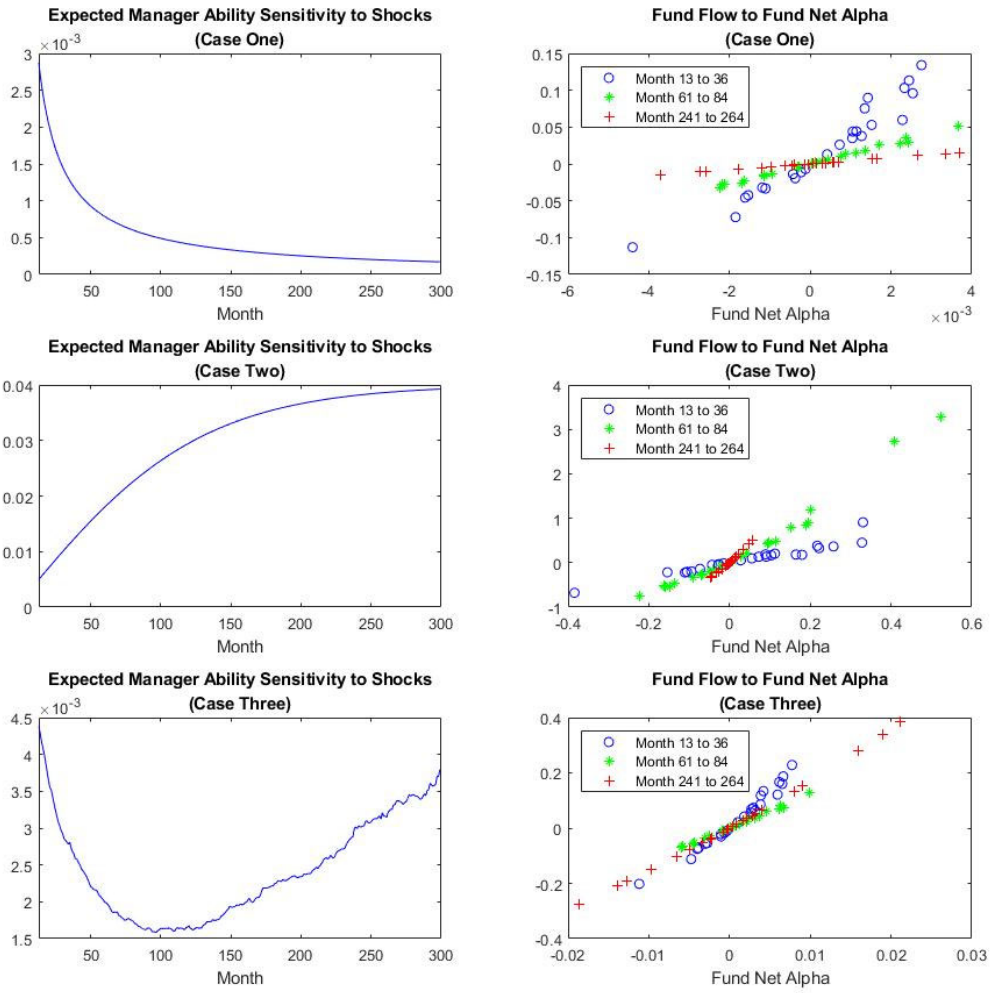

Figure A2. Simulation Results of Fund Exit Densities

Figure A2 illustrates the simulation results using parameters defined in Case One, Case Two, and Case Three, in the two subplots on the top, two subplots in the middle, and two subplots at the bottom, respectively. For each case, on the left-hand side, we illustrate the expected manager ability, m_t; and on the right-hand side, we illustrate the density that the fund will exit in the next month. In Case One, the fund never exits, and we plot the results from Month 1 to Month 300. In Case Two (Case Three), the fund exits in Month 121 (Month 343), and we plot the results before that time.