FUND FLOWS AND PERFORMANCE UNDER DYNAMIC UNOBSERVABLE MANAGING ABILITY

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

Latest revision, January 1, 2021

Abstract. We introduce continuous-time rational models of dynamic unobservable fund manager abilities with risk-neutral or risk-averse investors. Investors forever face time nonmonotonic ability-tracking problems. Consequently, precision of inferred abilities and inferred abilities' sensitivities to fund returns' innovation shocks are time nonmonotonic. These, in turn, induce fund flow-performance sensitivities that are time nonmonotonic. Our empirical evidence of nonmonotonic flow-performance sensitivities supports our theoretical framework, showing that our model explains real-world flow-performance relations better than current models, which predict only monotonic flow-performance sensitivities. We also offer insights to help resolve an ongoing dispute of whether empirical flow-performance relations are linear or convex.

JEL Codes: G11, G14, G23
Keywords: Active fund management, Fund flows, Fund performance, Alpha, 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, David Colwell, Chang-Mo Kang, Abhiroop Mukherjee, Linda Pesante, Wing Wah Tham, Nhu Tran, and Youchang Wu 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; Southern Finance Association Annual Meeting, and Greater China Area Finance Conference. Jingrui Xu's research was supported by the Fundamental Research Funds for the Central Universities, Grant Number 20720181061.

1 Introduction

Current literature demonstrates that fund return and size together reflect manager abilities, whereas fund flows reacting to fund performance reflect investors' rationality and competitiveness [see, for example, the innovative Berk and Green (2004), Berk (2005), and Berk and van Binsbergen (2015)]. As fund managers' abilities to outperform passive benchmarks are unobservable, to make investment decisions, investors use fund performance to form and update estimates of these unobservable abilities. Most current models studying this flow-performance relation assume that managers' abilities are unobservable constants.[1] We wonder, by allowing dynamic manager abilities, whether we can better explain and predict the flow-performance relation, and offer more insights to investors' rationality and competitiveness.

Actually, current empirical evidence is likely to support dynamic nonlinear manager abilities evolutions, which induce nonmonotonic manager abilities' precisions. Researchers find that funds' 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 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 fund managers' abilities to outperform passive benchmarks are affected by fund family activities [see, for example, Gaspar, Massa, Matos (2006) and Evans (2010), and Brown and Wu (2016)], by changing attention allocation [Kacperczyk, Nieuwerburgh, and Veldkamp (2016)], by managers' replacements [Dangl, Wu, and Zechner (2008)], and by macroeconomic conditions [see, for example, Ferreira, Keswani, Miguel, and Ramos (2012, 2013), Kacperczyk, Nieuwerburgh, and Veldkamp (2014), and Feldman, Saxena, and Xu (2020, 2021)]. Because these fund family activities, managers' replacements, changing attention allocation, and macroeconomic factors are dynamic, they might drive fund managers' abilities to change over time. Thus, in this study, we investigate,

[1] See, for example, Berk and Green (2004), Brown and Wu (2016), Huang, Wei, and Yan (2007), and Lynch and Musto (2003). Brown and Wu (2013), a working paper version of Brown, and Wu (2016) investigating fund families, and Dangl, Wu and Zechner (2008), studying managers' optimal replacement, model dynamic unobservable managing ability under a linear framework, allowing only monotonic precisions of abilities.

[2] Theoretical models, such as those of Berk and Green (2004) and Pastor and Stambaugh (2012), show that investors invest more (less) in the funds that perform better (worse), and this larger (smaller) amount of investments 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.

theoretically and empirically, the impact of dynamic unobservable fund manager abilities, on explanatory and predictive powers of fund flow-performance relations.

The first question we address (within a linear framework)

The first question we pose and answer is how investors and managers learn fund managers' dynamic unobservable abilities. To model this learning process, we apply optimal nonlinear filtering techniques used in the asset pricing and fund management literature [see, for example, in funds context, Berk and Stanton (2007), Dangl, Wu, and Zechner (2008), Brown and Wu (2013, 2016), Choi, Kahraman, and Mukherjee (2016), and in asset pricing context, Dothan and Feldman (1986), Detemple (1986), and Feldman (1989, 2007)[3]]. 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, where instantaneous expected returns (the drift terms) depend on the dynamic unobservable manager ability levels. These ability levels also follow Itô processes. Their instantaneous expected returns (drift terms) are affine functions of current managers' ability levels, and their diffusions are (locally, imperfectly) correlated with those of funds' gross alpha processes. With their prior beliefs of managers' ability levels and fund gross alphas, and conditional on observable realized fund gross alphas, investors generate posterior beliefs, or estimates, of managers' ability levels and of forthcoming fund expected gross alphas. As manager abilities are dynamic, estimation errors (or precision) of conditional expected manager abilities are dynamic as well. Depending on parameter values and realizations, the time pattern of these estimation errors can be increasing, decreasing, or constant.[4] Consequently, sensitivities of conditional expected manager abilities to new observations of fund gross alphas have time patterns that are increasing, decreasing, or constant, respectively.

We maintain many of the features of the model in Berk and Green (2004) and Brown and Wu (2016). In particular, we assume decreasing returns to scale (i.e., funds' total costs are

[3] Assuming constant unobservable manager abilities, Berk and Stanton (2007) study closed-end fund discounts, Brown and Wu (2016) study cross-fund learning within fund families, and Choi, Kahraman, and Mukherjee (2016) study investors' learning on managers' abilities from their performance in other funds that they manage. They both use filtering techniques to analyze the learning of constant abilities by observing fund returns.

[4] Our linear framework does not allow nonmonotonicity of the estimation error (or precision). However, the nonlinear framework that we introduce later and describe below, allow nonmonotonicity, and gives rise to some of our most substantial results.

increasing and convex in the size of assets under active management); 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 (i.e., investors supply capital with infinite elasticity to funds that have positive excess expected returns), and then we model the equilibrium for mean-variance risk-averse investors.

The second question we address (within a linear framework)

The second question we pose and answer is how do dynamics of fund manager abilities affect the equilibrium flow-performance relations? We show that in equilibrium, consistent with previous theoretical models, fund flows are increasing and convex in fund performance realizations (measured by fund net alphas), and that higher management fees and higher return volatilities each reduces fund flows' sensitivity to fund performance (flow-performance sensitivity). New to the literature, we find that higher (systematic) sensitivity of funds' gross alphas to inferred manager abilities and higher (local) sensitivity of inferred manager abilities to innovation shocks in fund gross alphas increase the flow-performance sensitivity. This is because a higher level of the former sensitivity increases fund performance, given the same level of manager abilities, and a higher level of the latter sensitivity makes performance shocks more informative on manager abilities.[5]

In our model, the increasing, decreasing, or constant time patterns of sensitivities of inferred manager abilities to innovation shocks in fund gross alphas, drive the flow-performance sensitivities to be increasing, decreasing, or constant over time, respectively. This feature is different from the previous models such as Berk and Green (2004) and Brown and Wu (2016), in which the decreasing sensitivity of inferred manager abilities to new realizations of fund returns drives the flow-performance sensitivity to decrease monotonically over time. In addition, we show that, for certain parameter values, our model degenerates to a continuous-time analog of the Berk and Green (2004) model, recreating the insights of their model in a dynamic context. Also, for certain parameter values, the relation of fund flows and manager abilities in our model degenerates to a single fund version of Brown and Wu (2016) model.

[5] Please see discussions of equilibrium flow-performance relations in Section 2.3.

Risk-averse investors

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 those of investors with Bernoulli logarithmic preferences [see, for example, Feldman (1992)]. Moreover, these portfolios are "growth optimal," as independently discovered by Bernoulli [Bernoulli (1954)] and implied by the "Kelly Criterion" [Kelly (1956)].[6] We are able to demonstrate that, as explained below, various effects offset each other such that investors' risk aversion does not affect the equilibrium flow-performance sensitivity. The reason is that investors' risk aversion affects only the sensitivity of dollar amounts of fund flows to performance; when fund flows are calculated as flow percentages as we do here, effects of investors' risk aversion cancel out.[7] Thus, the equilibrium flow-performance sensitivity where investors are mean-variance risk-averse is similar to the one where investors are risk neutral.

Nonlinear framework of manager ability levels and fund gross alphas

We further allow a nonlinear structure of active funds' observable gross alphas and the dynamic unobservable manager ability levels. Instantaneous expected returns (drift terms) and diffusion terms of the processes of manager ability levels and gross alphas are functions of time and fund's gross share price.[8] Under this nonlinear framework, the equilibrium flow-performance sensitivity can have more complex patterns over time, including nonmonotonic ones; thus, these theoretical results offer additional insights into complex empirical flow-performance relations. Specifically, nonmonotonic (over time) precisions of inferred abilities, induce nonmonotonic (over time) flow-performance sensitivities. This feature is different from the one in Berk and Green (2004) and subsequent models, in which estimation errors of the conditional expected manager abilities decrease over time, asymptotically to zero, inducing decreasing flow-performance sensitivities, or the estimation errors change monotonically over

[6] 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 (2019).

[7] 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 in fund sizes divided by fund sizes, and the effects of investors' risk aversion on the numerator and on the denominator cancel out.

[8] We use gross share price, i.e., share price before fund costs and fees, to calculate fund gross alphas.

time, inducing monotonic flow-performance sensitivities.

Empirical study

Using the U.S. active equity mutual fund data in the Center for Research in Security Prices (CRSP), we empirically test the flow-performance sensitivity over time. We find that in our sample period, on average, fund flow sensitivity to fund net alphas (flow–net alpha sensitivity), over time, first decreases, then increases, and then decreases again. We stress that as we model dynamic unobservable abilities, investors face continuous "tracking problems." Thus, patterns of inferred abilities are dynamic as well. Patterns depend on initial conditions, parameter values, and stochastic realizations. Thus, the pattern that we identified, of decreasing-increasing-decreasing flow–net alpha sensitivities, is specific to our sample period. Nonetheless, this nonmonotonic pattern is sufficient to support dynamic abilities with nonmonotonic precisions rather than with monotonic ones. Also, our empirical findings on individual funds show[9] that the majority of funds experience no significant changes in flow–net alpha sensitivity over time, and some funds even experience a significant increase in flow–net alpha sensitivity from one period to another. We also test how the flow–net alpha sensitivity changes with fund age. We find that, on average, the flow–net alpha sensitivity first decreases with fund age, then increases with it, then decreases with it again, and finally increases with it.

These findings are consistent with dynamic abilities with nonmonotonic precisions and are inconsistent with monotonic precisions. It is plausible that, in our sample, the decrease in the flow–net alpha sensitivity in the earliest period or in the earliest age years for an average fund, is due to the fact that investors have more and more precise estimates of manager abilities as the number of observations increases. However, manger abilities are dynamic over time as they are driven by dynamic factors such as fund family activities, managers' replacements, changing attention allocation, and macroeconomic factors. The dynamic manager abilities induce lower or higher estimation precisions in the later periods or age years. Consequently, the flow–net alpha sensitivity changes with time period or fund age nonmonotonically. Also, in the real world, it is likely that dynamic unobservable manager abilities, which possibly vary across funds, generate different patterns of time-varying flow-performance sensitivity for

---

[9] See Table 5.

different funds.

More on the literature and our findings

In the literature of the flow-performance relation, some studies suggest that this relation is convex [see, for example, Berk and Green (2004), Brown and Wu (2016), and Lynch and Musto (2003)], and other studies suggest that it is linear [see, for example, Spiegel and Zhang (2013)]. Our study complements this discussion by showing that the flow-performance relation is dynamic (specifically, the intercept, slope, and curvature of this function change over time), making empirical identification of the relation complex. In the real world, both cross-sectional heterogeneity and time dynamics of flow-performance relations affect empirical results; thus, both should be considered.

There is also a discussion 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 show that as funds with marketing activities reduce investors' participation costs, i.e., costs to obtain fund information, the lower participation costs also lower new investors' requirements for fund performance, making fund flows increase more slowly with fund performance. Different from their insights, we show that if fund marketing expenditures are eventually transferred to investors through management fees, then the higher fees lead to a less convex flow-performance relation. More importantly, if funds marketing activities help investors estimate managers' abilities, then investors have more precise estimates of these abilities over time. The improvements in estimation precision lower the sensitivity of these inferred abilities to fund performance innovation shocks and, consequently, decrease the convexity of flow-performance relations.

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 providing, a model of dynamic unobservable manager abilities under a nonlinear framework, which better explains the real-world nonmonotonic time-varying flow-performance sensitivity.[10]

Second, we provide empirical evidence that supports the case of dynamic abilities with nonmonotonic precisions rather than with monotonic ones, and by developing an empirical framework for testing this issue. Third, our model offers new insights into empirical findings in the current literature, including the complex curvature of flow-performance relations, by using panel data and the findings that marketing activities induce smaller such curvature.

Finally, our findings imply that the innovative insights of Berk and Green (2004) regarding myth of active portfolio management [see also Berk (2005)], which were demonstrated within a parsimonious model, hold in a wider class of equilibria in terms of production structure (dynamic abilities with nonmonotonic precisions rather than with monotonic precisions), information structure (unobservable processes, including nonlinear ones, rather than unobservable constant, which gradually becomes observable or linear structures with monotonic precisions of ability estimates), and preferences structure (risk-neutral or risk-averse investors, rather than risk-neutral investors only).

Section 2 introduces our model. Section 3 provides simulation results of our equilibria. Section 4 illustrates our empirical analyses. Section 5 discusses our model's insights into current empirical phenomena. Section 6 concludes.

2 The Model

We introduce a rational equilibrium framework to study how the dynamics of unobservable manager abilities affect equilibrium flow-performance relations. We first show how managers and investors infer dynamic unobservable manager abilities and form equilibrium flow-performance relations by solving, respectively, managers' profit-maximizing problems and investors' portfolio optimization problems. Our baseline model assumes risk-neutral investors and a linear filtering framework of dynamic unobservable manager abilities, as inferred from share prices. Then, we study mean-variance risk-averse investors, and develop a nonlinear filtering framework of the dynamic unobservable manager abilities, as inferred from share prices. Some of the mathematical proofs of our results are in the Appendix.

[10] 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.

Some of our model settings are similar to those of Berk and Green (2004) and Brown and Wu (2016).[11] We use a two-fund setting, i.e., investors can invest in a representative active fund, with one manager, and in a passive benchmark portfolio.[12] 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 active fund manager and investors over a finite time interval/period, at times t, $t \in [0, T]$, where $T, T > 0$, is a constant.

2.1 Observable Fund Returns and Unobservable Manager Ability: Linear Filtering

Let ξ_t, $0 \leq t \leq T$ be the active fund's gross share price, i.e., share price before fund costs and fees,[13] so $d\xi_t/\xi_t$ is the instantaneous fund gross return. For simplification, we also assume that this active fund has a beta loading of one on the passive benchmark portfolio. To focus on the active fund's return, similar to previous models,[14] we normalize the passive benchmark portfolio's return to zero so that the fund instantaneous gross return in excess of the passive benchmark is $d\xi_t/\xi_t - 0 = d\xi_t/\xi_t$. Hereafter, we call the active fund's instantaneous gross alpha $d\xi_t/\xi_t$, briefly called gross alpha.

The active fund's gross alpha depends on the active fund manager's instantaneous ability, θ_t, $0 \leq t \leq T$, to beat the benchmark. We will call it briefly, ability. This ability is unobservable to both the fund manager and investors. The fund manager and investors learn about θ_t by observing the history of the fund gross alpha $d\xi_s/\xi_s$, $0 \leq s \leq t$ (or equivalently by observing gross fund share price ξ_s, $0 \leq s \leq t$). We assume a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $\{\mathcal{F}_t\}_{0 \leq t \leq T}$, a right-continuous, nondecreasing family of sub-σ-algebras. Two independent Wiener processes, $W_{1,t}$ and $W_{2,t}$, $0 \leq t \leq T$, are adapted to this filtration. The unobservable θ_t and the observable ξ_t evolve as follows.

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

[11] Similar to Berk and Green (2004) and Brown and Wu (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.

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

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

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

$$\frac{d\xi_t}{\xi_t} = A_1\theta_t dt + BdW_{2,t}, \qquad (2)$$

with initial conditions θ_0 and ξ_0, respectively. The parameters a_0, a_1, b_1, b_2, A_1, and B are known constants. To make economic sense, we assume that $A_1 > 0$ (otherwise the ability becomes a "disability"). For simplicity and without loss of generality, we assume $B > 0$.

While ability is unobservable to managers and investors, the evolution processes ("laws of motion") and all parameter values are common knowledge.

This setting implies the following. First, the ability, θ_t, to beat the benchmark follows a dynamic process. Second, the fund gross alpha, $d\xi_t/\xi_t$, depends on the manager's ability and on random shocks. As $A_1 > 0$, a manager with positive (negative) ability tends to create positive (negative) fund gross alpha; and the larger A_1 is, the higher the sensitivity of gross alpha to ability is. Also, B is the diffusion coefficient of fund gross alpha, which positively corresponds to its volatility. Third, where $b_2 > 0$ ($b_2 < 0$), that is, b_2 is strictly positive (negative), the shock $W_{2,t}$ affects both ability and fund gross alpha, which, consequently, are instantaneously positively (negatively) correlated, as $b_2B > 0$ ($b_2B < 0$). Where $b_2 = 0$, and $b_1 > 0$ ability and gross alpha are affected by independent shocks, thus, instantaneously uncorrelated. A larger b_2 relative to b_1 implies a higher instantaneous correlation between gross alpha and ability.

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.

The manager and investors update their estimates of θ_t using their observations ξ_t, in a Bayesian fashion. This type of model is presented in Liptser and Shiryaev (2001a, Ch. 8;

2001b, Ch. 12).[15] The techniques are called optimal nonlinear filtering and are used in numerous previous studies [see, for example, Dothan and Feldman (1986), Detemple (1986), Feldman (1989, 2007), and Berk and Stanton (2007)]. The following proposition describes how the managers and investors form and update their estimates of θ_t.

Proposition 1.

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

$$\bar{W}_t = \int_0^t \frac{d\xi_s/\xi_s - A_1 m_s ds}{B} \tag{3}$$

is a Wiener process with respect the filtration $\left\{\mathcal{F}_t^{\xi}\right\}_{0\le t\le T}$, with $\bar{W}_0 = 0$; and the σ-algebras $\mathcal{F}_t^{\xi}$ and $\mathcal{F}_t^{\xi_0,\bar{W}}$ are equivalent.

b. $\bar{W}_t$ innovates the (observable) conditional mean, m_t, of the unobservable fund manager ability, θ_t, to beat the benchmark. The variables m_t, ξ_t, and γ_t are the unique, continuous, $\mathcal{F}_t^{\xi}$-measurable solutions of the system of equations

$$dm_t = (a_0 + a_1 m_t)dt + \sigma_m(\gamma_t)d\bar{W}_t, \tag{4}$$

$$\frac{d\xi_t}{\xi_t} = A_1 m_t dt + B d\bar{W}_t, \tag{5}$$

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

where

$$\sigma_m(\gamma_t) \triangleq \frac{b_2 B + A_1 \gamma_t}{B}, \tag{7}$$

with initial conditions ξ_0, m_0, and γ_0.

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

Proof. Theorem 8.1 of Liptser and Shiryaev (2001a) and Theorem 12.5 of Liptser and Shiryaev (2001b) jointly provide the proof of Proposition 1a. Theorem 12.5 of Liptser and Shiryaev (2001b) provides the proof of Proposition 1b. Theorem 11.1 of Liptser and Shiryaev (2001b) provides the proof of Proposition 1c. The technical requirements to prove the theorems are regular conditions over the period $0 \le t \le T$, such as bounded parameter values, integrality

[15] The models presented by Liptser and Shiryaev (2001a,b) allow all model parameters to be stochastic, functions of the stochastic gross alpha. For simplicity, we first introduce a linear framework with constant parameters. We analyze the nonlinear framework with stochastic parameters in Section 2.6.

of variables, and finite moments of variables.[16] The intuition of these requirements is that, over a finite time period, almost surely the manager ability, fund gross alphas, and their variations should be finite so that the learning process is well defined. These requirements are satisfied, due to our finite parameter values, finite initial values, and the finite horizon within which we study our model. In the real world, abilities that keep improving or deteriorating over a short period, or abilities that revert to a finite mean over a long period, would satisfy the technical requirements and follow our learning process.

The Wiener process $\bar{W}_t$ represents the innovation shocks to estimates of manager unobservable abilities. By Proposition 1a, the process ξ_t and the innovation process $\bar{W}_t$ with ξ_0 generate the same information.

Proposition 1b implies that investors make their optimal decisions in two steps. First, by observing the history of the fund's share price ξ_t and restructuring the state space to consists of observable processes only while maintaining informational equivalence,[17] they generate a posterior estimate of the fund manager ability m_t. Second, they use their posterior estimate m_t to predict the fund gross alpha in the forthcoming future, as shown by Equation (5). They use this prediction in solving their investor problems, as shown in the next sections. Notice that in these optimization processes, the unobservable manager ability θ_t is replaced by its observable conditional mean m_t, updated by a new Wiener process $\bar{W}_t$, and that m_t is continuously updated as a function of the dynamic conditional variance γ_t, representing the imprecision of the estimate. Hence, investors' problems become Markovian, which makes the problems tractable.

To make economic sense, we assume a nonnegative b_2, which induces a positive correlation $(b_2B + A_1\gamma_t)$ between inferred ability and performance shocks (because B and A_1 are positive).[18] Then, the sensitivity of expected manager ability to innovation shocks in

[16] See the requirements of the corresponding theorems in Liptser and Shiryaev (2001a, 2001b).
[17] See Feldman (1992).
[18] This is because a negative b_2 induces a negative instantaneous/idiosyncratic correlation, which can give rise to negative total correlation. If γ_t weighs the positive systematic source of correlation, A_1, insufficiently high; then, the negative instantaneous/idiosyncratic source of correlation (b_2B) dominates. Thus, under these special parameter values, which we do not allow here, the dynamics γ_t may induce correlation, between inferred ability and performance shocks, that changes sign over time, resulting in a transient nonmonotonic relation between performance shocks and inferred ability even under the linear structure that we analyze in this section. For detailed analysis of this nonmonotonicity, see Feldman (1989, Proposition 4).

fund gross alpha, $\sigma_m(\gamma_t)$ is always positive. In other words, under this setting, a positive (negative) shock in fund gross alpha induces an increase (a decrease) in inferred manager ability.

By Proposition 1c, the conditional distribution of θ_t is Gaussian. Then, conditional distribution of θ_t is determined by the first two moments, m_t and γ_t. As the parameters a_1, b_1, b_2, A_1, and B are constants and γ_0 is given, γ_t is deterministic, as shown in Proposition 1b. Consequently, $\sigma_m(\gamma_t)$, the sensitivity of expected manager ability to innovation shocks in fund gross alpha, is also deterministic but dynamic. However, m_t is stochastic and its future values are unknown. Therefore, investors know the precision of their future estimates of manager ability in advance, although they do not know the future estimates of this ability. 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.

Taking a closer look at $d\gamma_t$, we find that depending on the parameter values, it can be positive, negative, or zero; that is, the precision of the future estimates of the manager ability level can increase, decrease, or be unchanged for the next small time period. In particular, where b_2B, the instantaneous covariance between $d\xi_t/\xi_t$ and θ_t,[19] and/or A_1, the sensitivity of the drift of $d\xi_t/\xi_t$ to θ_t, are sufficiently large (small), $d\gamma_t$ would be negative (positive). In other words, if the gross alpha and the manager ability are more (less) correlated instantaneously and/or the change in the gross alpha is more (less) sensitive to the manager ability, then the precision of investors posterior estimates of the manager ability increases (decreases). Also, where B^2, the instantaneous variance of $d\xi_t/\xi_t$,[20] is sufficiently small (large), $d\gamma_t$ also is negative (positive). In other words, if the gross alpha process is less (more) volatile, then the precision of investors posterior estimates of the manager ability increases (decreases).

Then, depending on parameter values, the dynamics of $d\gamma_t$, induces a γ_t that monotonically increases, decreases, or stays unchanged over time. Consequently, $\sigma_m(\gamma_t)$, monotonically increases, decreases, or stays unchanged, respectively, over time.

The dynamics of γ_t is one of the key differences between our model and Berk and

---

[19] This is because $\mathrm{Cov}\left(d\theta_t, \frac{d\xi_t}{\xi_t}\right) = b_2 B dt$.

[20] This is because $\mathrm{Cov}\left(\frac{d\xi_t}{\xi_t}, \frac{d\xi_t}{\xi_t}\right) = B^2 dt$.

Green's (2004) model 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 to the innovation shocks in fund returns. In contrast to those studies, our study allows the fund manager ability to follow a dynamic process. Within our more general structure, the precision of investors' posterior estimates of manager ability can increase, decrease, or stay unchanged over time. Then, in turn, over time, the sensitivity of inferred manager ability to the innovation shocks in fund returns can increase, decrease, or stay unchanged. These features generate a framework that has stronger theoretical and in empirical explanatory and prediction powers, studying the flow-performance relation.[21]

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 $\{\theta_t, m_t, \gamma_t\}$, we can solve investors' and the fund manager's optimization problems.

We assume that there are infinitely many small risk-neutral investors in the market and that each investor's investment decision does not affect the fund's return and size, although all investors together do affect the fund's return and size. An investor's portfolio return depends on three components: gross alpha, management fee, and fund costs. Berk and Green (2004) show that the case where the fund manager actively manages the whole fund and chooses his/her management fee f_t at each time t, is equivalent to the case where the fund manager chooses the amount of the fund to actively manage at each time t under a fixed management fee f. As the latter case is more realistic, we focus on it to conduct our analysis.

At time t, fund costs variable $C(q_t^a)$ is a function of the fund amount that is under active management q_t^a. Out of the q_t, the total asset managed by the fund, 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, inducing no fund costs. There are decreasing returns to scale at the fund level, similar

[21] We note that, for special parameter values that we do not allow here, the dynamics of γ_t might induce, for this linear version of the model, a transient nonmonotonic time pattern of the diffusion coefficient of ability $\sigma_m(\gamma_t)$ [as can be seen from Equation (7)]. (See also Footnote 18.)

to Berk and Green (2004) and Feldman, Saxena, and Xu (2020, 2021). Thus, $C(q_t^a)$ is increasing and convex in q_t^a, and we assume that

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

where c, a positive known constant, 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 call dS_t/S_t the active fund's (instantaneous) net alpha, or, briefly, net alpha. Based on the above discussion, we have,

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

Similar to Berk and Green (2004), 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, we have the following condition:

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

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

$$\frac{q_t^a}{q_t}A_1 m_t - \frac{c{q_t^a}^2}{q_t} - f = 0. \tag{11}$$

Rearranging,

$$fq_t = A_1 m_t q_t^a - c{q_t^a}^2. \tag{12}$$

The fund manager wants to maximize fund profit fq_t by choosing q_t^a. Then, the manager's problem is

$$\max_{q_t^a} fq_t = \max_{q_t^a} A_1 m_t q_t^a - c{q_t^a}^2 \tag{13}$$

subject to

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

By solving the investors' problem and the manager's problem, we can obtain the equilibrium flow-performance relation.

2.3 The Flow-performance Relation

As in Berk and Green (2004), we define the lowest level of conditional expected manager ability that makes the fund survive, $\underline{m_t}$. If $m_t \leq \underline{m_t}$, the fund receives no investments from investors and exits the market. Hereafter, we call $\underline{m_t}$ the survival level. Here we assume $\underline{m_t} \geq 0$. The reason is that given updated information, the expected instantaneous gross alpha accumulated in dt is $\mathrm{E}\left(d\xi_t/\xi_t|\mathcal{F}_t^{\xi}\right) = A_1 m_t dt$, with $A_1 > 0$. 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 they would switch their investments to the passive benchmark portfolio. The optimal amount under active management and the optimal total asset under management, q_t^{a*} and q_t^*, are not trivial where $m_t > \underline{m_t} \geq 0$; otherwise, they are both zero. Also, we assume that the manager would set the fee f so that $q_t^{a*} \leq q_t^*$, as Berk and Green (2004) assume.

To characterize the flow-performance relation using gross alpha as the performance measure, we apply Itô's Lemma to calculate dq_t^*, and divide it by q_t^* to get the equilibrium percentage fund flow. We then substitute fund gross alpha into the expression, getting the desired characterization in the following equation:

$$\frac{dq_t^*}{q_t^*} = \frac{2\sigma_m(\gamma_t)}{m_t B}\left(\frac{d\xi_t}{\xi_t}\right) + \frac{\sigma_m^2(\gamma_t)}{m_t^2 B^2}\left(\frac{d\xi_t}{\xi_t}\right)^2 + \frac{2}{m_t}\left[(a_0 + a_1 m_t) - \frac{A_1 \sigma_m(\gamma_t) m_t}{B}\right] dt. \tag{15}$$

To characterize the flow-performance relation using net alpha as the performance measure, we repeat the previous procedure but substitute fund net alpha, getting the desired characterization in the following equation:

$$\frac{dq_t^*}{q_t^*} = \frac{A_1\sigma_m(\gamma_t)}{fB}\left(\frac{dS_t}{S_t}\right) + \frac{A_1^2\sigma_m^2(\gamma_t)}{4f^2B^2}\left(\frac{dS_t}{S_t}\right)^2 + 2\left(\frac{a_0}{m_t} + a_1\right)dt.^{22} \quad (16)$$

The following proposition summarizes the flow-performance relations.

Proposition RN. If $m_t \leq \underline{m_t}$, then the fund receives zero investments from investors. If $m_t > \underline{m_t}$, then the flow-performance relation has the following characteristics.

a. Fund flow increases with and is convex[23] in fund gross alpha. A higher conditional expected manager ability m_t, and a higher volatility of fund gross alphas B both induce lower sensitivity of fund flow to fund gross alpha. Higher sensitivity of expected manager ability to innovation shocks in fund gross alpha, $\sigma_m(\gamma_t)$, induces higher sensitivity of fund flow to fund gross alpha.

b. When we use the fund net alpha as the measure of fund performance, fund flow still increases with and is convex in fund performance, decreases with B, and increases with $\sigma_m(\gamma_t)$. Also, a change in m_t does not affect the flow-performance sensitivity. In addition, a higher management fee f induces lower flow-performance sensitivity, whereas higher sensitivity of fund returns to manager ability, A_1, induces higher flow-performance sensitivity.

Proof. See the Appendix.

The intuition of Proposition RNa is as follows. When the fund manager ability is expected to be sufficiently high, the fund receives investments from investors. A higher fund gross alpha reflects higher manager ability to generate returns for investors, so fund flows are positively related to fund gross alphas. Under decreasing returns to scale, it is more difficult to improve gross alpha when it is already high; if a manager can do so, this is a signal of extremely high ability. Thus, fund flows are more sensitive to fund gross alpha where it is higher, resulting in a convex flow-performance relation. Higher sensitivity of expected manager ability to

---

22 The term $\left(\frac{dS_t}{S_t}\right)^2$ in Equation (16), in its continuous-time limit (the quadratic variation), is, in equilibrium, $\left(\frac{2fB}{A_1 m_t}\right)^2 dt$ [see Equation (A6)], suggesting an instantaneous linear flow-performance sensitivity. However, as investors allocate wealth to funds discretely, Equation (16) implies that flow-performance sensitivities are convex.

23 Please see Footnote 22.

innovation shocks in fund gross alpha, $\sigma_m(\gamma_t)$, 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. As $\sigma_m(\gamma_t)$ may monotonically increase, decrease, or stay unchanged over time, so may the flow-performance sensitivity. On the other hand, higher fund gross alpha volatility B implies that the fund gross alpha contains less information about manager ability, so fund flows are less sensitive to fund gross alphas. If investors have already high expectations of the manager's ability (i.e., m_t is high), investments to the fund are already large. Given this larger fund size, it is more difficult for fund managers to induce a particular percentage fund inflow, so fund flows are less sensitive to fund gross alphas.

The intuition of Proposition RNb is as follows. Both funds net alphas and gross alphas reflect information on managers' abilities, so most of the results in RNa, where performance is measured by gross alphas, are maintained in RNb, where performance is measured by net alphas. We have that dS_t/S_t is a function of f and A_1; if any of these parameters changes, there is an impact on the coefficient of dS_t/S_t shown in Equation (16), and a reverse impact on dS_t/S_t.[24] These two types of impact might cancel each other out; but if the fund size adjusts to the changes in f and A_1 so quickly that dS_t/S_t is driven back toward its original level, then the changes in f and A_1 would still affect the flow–net alpha sensitivity.

If a fund manager charges a higher fixed fee f, this higher fee induces a smaller fund size, thus a smaller fund flow. As the flow-performance relation is convex, fund flows are less sensitive to fund net alphas at this smaller fund flow level. Moreover, a higher sensitivity of fund gross alpha to manager ability A_1 induces a larger fund size, thus a large fund flow. As the flow-performance relation is convex, fund flows are more sensitive to fund net alphas at this larger fund flow level. In addition, the expected manager ability determines the optimal amount of fund that is under active management, which consequently determines fund net alphas. Therefore, fund net alphas capture investors' expectation of manager ability, so the conditional expected manager ability m_t does not explicitly affect the flow-performance sensitivity when using net alphas to measure performance.

[24] See the expression of $\frac{dS_t}{S_t}$ in (A6) in the Appendix.

2.4 Relation to Berk and Green (2004) and Brown and Wu (2016)

Berk and Green (2004) provide one of the earliest discrete-time models that studies the flow-performance relation and offers relevant insights. In their model, manager ability is an unknown constant that investors and the fund manager learn by observing fund returns. Our model nests Berk and Green (2004) in the sense that we can degenerate it to a continuous-time analog of it.

To make the manager ability $\theta_t,\ t \geq 0$, 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 $A_1 = 1$, i.e., the sensitivity of fund gross alpha to manager ability is one, to further simplify our model and match it with Berk and Green (2004)'s. Then, Equations (4), (5), and (6) become, with parameter values we set above,

$$\sigma_m(\gamma_t) = \frac{\gamma_t}{B} \tag{17}$$

$$dm_t = \frac{\gamma_t}{B^2}\left(\frac{d\xi_t}{\xi_t} - m_t dt\right) \tag{18}$$

$$\gamma_t = \frac{\gamma_0 B^2}{B^2 + \gamma_0 t}. \tag{19}$$

The equilibrium flow-performance relation using net alpha as the performance measure becomes

$$\frac{dq_t^*}{q_t^*} = \frac{1}{f}\left(\frac{\gamma_0}{B^2 + \gamma_0 t}\right)\left(\frac{dS_t}{S_t}\right) + \frac{1}{4f^2}\left(\frac{\gamma_0}{B^2 + \gamma_0 t}\right)^2\left(\frac{dS_t}{S_t}\right)^2. \tag{20}$$

This result is valid only if $m_t > \underline{m_t}$. Otherwise, the fund receives zero investments and $dq_t^*/q_t^* = 0$. The flow-performance relation in Equation (20) is a continuous-time analog of Equation (30) in Berk and Green (2004). We summarize our results of Equation (20) in the following proposition, which provides the same insights as those offered by Equation (30) of Berk and Green (2004).[25]

Proposition BG. Where the manager ability is an unknown constant θ, if $m_t \leq \underline{m_t}$, then the fund receives zero investments from investors; if $m_t > \underline{m_t}$, then the flow-performance

[25] See the discussion below Equation (30) of Berk and Green (2004).

relation has these characteristics.

a. If the noise in the observed fund returns increases relative to the noise of the prior estimate of the manager ability, i.e., B^2 increases relative to γ_0, investors learn less from the fund returns. Thus, fund flows are less sensitive to net alphas.

b. As the age of the fund increases, i.e., t increases, investors have more observations of the fund's performance and develop a more precise estimate of the manager's ability. Thus, fund flows are less sensitive to (current and future) net alphas.

c. As the management fee f increases, the fund size and the fund flow are smaller. Thus, fund flows are less sensitive to net alphas due to the convexity of the flow-performance relation.

Proof. See the Appendix.

This case of constant unobservable manager abilities is also related to the one in Brown and Wu (2016). In particular, if there is only one fund in Brown and Wu (2016)'s model, i.e., there is no cross-fund learning, Equations (5) and (6) in Brown and Wu (2016) become our Equations (18) and (19). With cross-fund learning within fund families, Brown and Wu (2016) also find that sensitivity of fund flows to fund performance decrease monotonically, similar to our Proposition BG.

As the above results show, the constant unobservable manager ability framework generates an equilibrium flow-performance relation that is transient, that is, as time passes, flow–net alpha sensitivity decreases to zero monotonically. This result does not match with empirical data because empirically, even for very old funds, fund flows are still sensitive to performance and, as we show later in the empirical section, some funds' flows are more sensitive to fund performance over some periods. Our dynamic unobservable manager ability framework does not have this restriction because it allows the equilibrium flow-performance sensitivity to increase, decrease, or stay unchanged over time.

Lynch and Musto (2003) offer insights into the convex flow-performance relation by modeling managers' choice of strategies. They show that past performance tells less about future performance if funds discard previous strategies that generate bad results. Thus, flows are less sensitive to past performance when it is poor. Similar to Berk and Green (2004), we do not introduce the manager strategy choice structure in our model; instead, the convex flow-performance relation in Berk and Green (2004) and in our model results from the decreasing returns to scale, i.e., the convexity of fund costs in fund size.

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 [see Bernoulli (1954)] and the "Kelly Criterion" [see Kelly (1956)]. This setting is also similar to the one in Pastor and Stambaugh (2012), and Feldman, Saxena, and Xu (2020, 2021).

Sharpe ratio maximization is a common assumption in 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 incentives for manipulation. To 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. In our model, as investors act on their own interests, they do not 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).

Because risk-averse investors, trading off risk and return, arrive at an endogenous equilibrium, we need to redefine the model. First, we cannot normalize the passive benchmark portfolio return to be zero. Instead, we define the share price of the passive benchmark portfolio at time t, $0 \leq t \leq T$, as η_t. This, in turn, includes redefinitions of net and gross alphas. We assume that the passive benchmark portfolio return $d\eta_t/\eta_t$ follows

$$\frac{d\eta_t}{\eta_t} = \mu_p dt + \sigma_p dW_{p,t}, \tag{21}$$

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

$$dW_{p,t}d\overline{W}_t = 0. \tag{22}$$

Third, to simplify our discussion, we normalize the risk-free rate to zero.[26] All other terms are the same as before.

An investor's problem, then, is to maximize the portfolio's instantaneous Sharpe ratio:

$$\max_{w_t} \frac{\mathrm{E}\left[\frac{dp_t}{p_t}|\mathcal{F}_t^{\xi}\right]}{\sqrt{\mathrm{Var}\left[\frac{dp_t}{p_t}|\mathcal{F}_t^{\xi}\right]}}, \tag{23}$$

subject to

$$0 \le w_t \le 1, \tag{24}$$

where w_t is the weight allocation to the active fund,[27] 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

$$\frac{dp_t}{p_t} = w_t\left(\frac{dS_t}{S_t} + \frac{d\eta_t}{\eta_t}\right) + (1 - w_t)\frac{d\eta_t}{\eta_t} = w_t\frac{dS_t}{S_t} + \frac{d\eta_t}{\eta_t}. \tag{25}$$

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_t, $V_t \in (0, +\infty)$, $0 \le t \le T$. In reality, V_t 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. Thus, to simplify our analysis, we assume that V_t is exogenous to both investors and managers. Here, the amount of wealth that is allocated to the fund or fund size is $q_t = w_t^*V_t$. 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

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

[27] 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.

is some function.

The manager's problem is, then,

$$\max_{q_t^a} f q_t = \max_{q_t^a} g(q_t^a). \tag{26}$$

subject to

$$0 \le q_t^a \le q_t. \tag{27}$$

Therefore, as in Berk and Green (2004), in our equilibrium the fund manager's optimal profit does not depend on the fee set by the manager.

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

$$\frac{dS_t}{S_t} = \frac{fB^2\mu_p}{B^2\mu_p + cV_t\sigma_p^2}dt + \frac{2fB}{A_1 m_t}d\bar{W}_t. \tag{28}$$

We can see that on average, the fund net alpha increases with the management fee, fund gross alpha volatility, and benchmark mean return because the drift term on the right-hand side of Equation (28) increases with f, B, and μ_p. On the other hand, the fund net alpha, on average, decreases with the fund's cost sensitivity to size and benchmark volatility because the drift term decreases with c and σ_p^2. More importantly, the result in Equation (28) shows that expected fund net alpha (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, so it has to provide a higher return to induce 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 using gross alpha as the performance measure as

$$\begin{aligned} \frac{dq_t^*}{q_t^*} &= \frac{2\sigma_m(\gamma_t)}{m_t B}\left(\frac{d\xi_t}{\xi_t}\right) + \frac{\sigma_m^2(\gamma_t)}{m_t^2 B^2}\left(\frac{d\xi_t}{\xi_t}\right)^2 \\ &+ \frac{2}{m_t}\left[(a_0 + a_1 m_t) - \frac{A_1\sigma_m(\gamma_t)m_t}{B}\right]dt + X_{t,} \end{aligned} \tag{29}$$

where X_t contains all the terms related to dV_t and $(dV_t)^2$. As we assume that V_t is exogenous to the market, thus independent of m_t at each time t, we assume $dV_t dm_t = 0$.

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

Analysis similar to the one in the risk-neutral case, yields the equilibrium flow-performance relation using net alpha as the performance measure:

$$\frac{dq_t^*}{q_t^*} = \frac{A_1\sigma_m(\gamma_t)}{fB}\left(\frac{dS_t}{S_t}\right) + \frac{A_1^2\sigma_m^2(\gamma_t)}{4f^2B^2}\left(\frac{dS_t}{S_t}\right)^2 + Y_t dt + X_t, \tag{30}$$

where

$$Y_t = \frac{2a_0}{m_t} + 2a_1 - \frac{A_1\sigma_m(\gamma_t)B^2\mu_p}{B\left(B^2\mu_p + cV_t\sigma_p^2\right)}. \tag{31}$$

Here both X_t and Y_t are independent of either $d\xi_t/\xi_t$ or dS_t/S_t.

From Equation (30), we can see that investors' risk aversion does not affect the flow-performance sensitivity whether the performance is measured by gross alphas or net alphas. Investors' risk aversion affects only the components of the fund flow that is unrelated to fund performance, as shown in Equations (30) and (31). The intuition is that investors' risk aversion affects the amount of the investment allocated to the risky active fund q_t^*, so it affects the sensitivity of the dollar amount of the fund flow dq_t^* to performance. However, when the fund flow is calculated as percentage flow dq_t^*/q_t^*, the effects of risk-aversion cancel out. Therefore, the flow-performance sensitivity where investors are mean-variance risk-averse is similar to the one where investors are risk-neutral. The following proposition summarizes the results in this section.

Proposition RA. Mean-variance risk-averse investors, who maximize their portfolios' instantaneous Sharpe ratios, induce equilibrium fund flow-performance sensitivity of the same characteristics as risk-neutral investors do.

Proof. See the Appendix.

2.6 Nonlinear Filtering Framework and the Flow-performance Relation

Suppose that the ability and gross alpha processes are more general in the sense that their parameters, rather than being constants, change as functions of time and fund (gross) share price levels. The state equations, then, follow the nonlinear system,

$$d\theta_t = [a_0(t,\xi_t) + a_1(t,\xi_t)\theta_t]dt + b_1(t,\xi_t)dW_{1,t} + b_2(t,\xi_t)dW_{2,t}, \tag{32}$$

$$\frac{d\xi_t}{\xi_t} = A_1(t,\xi_t)\theta_t dt + B(t,\xi_t)dW_{2,t}, \tag{33}$$

with initial conditions θ_0 and ξ_0, respectively. The parameters $a_0(t,\xi_t)$, $a_1(t,\xi_t)$, $b_1(t,\xi_t)$, $b_2(t,\xi_t)$, $A_1(t,\xi_t)$, and $B(t,\xi_t)$ are functions of t and ξ_t. Similar to the linear case, in order to make economic sense, we assume that $A_1(t,\xi_t) > 0$ (otherwise the ability becomes a "disability"), and for simplicity, and without loss of generality, we assume $B(t,\xi_t) > 0$. The definitions of other variables are the same as before. Then, the following proposition describes how the manager and investors form and update their estimates of θ_t within this nonlinear filtering framework.

The intuition regarding our ability to solve this nonlinear system and the nature of its equilibrium is as follows. Because time and price levels are observable, at each time point the parameters of the ability distribution conditional on realization are known and the distribution is Gaussian. In the next instant, however, realizations stochastically change. Consequently, the conditional distribution parameters stochastically change as well, and the conditional distribution is a "new" Gaussian distribution with different moments. Thus, the moments of the ability-conditional distributions evolve stochastically. We can think of the equilibrium here, relative to one in the linear case, as an equilibrium that stochastically travels among the stochastic equilibria of the type that we establish above. We can now state the following propositions.

Proposition 2.

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

$$\bar{W}_t = \int_0^t \frac{d\xi_s/\xi_s - A_1(t,\xi_t)m_s ds}{B(t,\xi_t)} \tag{34}$$

is a Wiener process with respect the filtration $\left\{\mathcal{F}_t^{\xi}\right\}_{0\le t\le T}$, with $\bar{W}_0 = 0$; and the σ-algebras $\mathcal{F}_t^{\xi}$ and $\mathcal{F}_t^{\xi_0,\bar{W}}$ are equivalent.

b. $\bar{W}_t$ innovates the (observable) conditional mean m_t of the unobservable fund manager ability θ_t to beat the benchmark. The variables m_t, ξ_t, and γ_t are the unique, continuous, $\mathcal{F}_t^{\xi}$-measurable solutions of the system of equations

$$dm_t = [a_0(t,\xi_t) + a_1(t,\xi_t)m_t]dt + \sigma_m(\gamma_t)d\bar{W}_t, \tag{35}$$

$$\frac{d\xi_t}{\xi_t} = A_1(t,\xi_t)m_t dt + B(t,\xi_t)d\bar{W}_t, \tag{36}$$

$$d\gamma_t = [b_1^2(t,\xi_t) + b_2^2(t,\xi_t) + 2a_1(t,\xi_t)\gamma_t - \sigma_m^2(\gamma_t)]dt, \tag{37}$$

where

$$\sigma_m(\gamma_t) \triangleq \frac{b_2(t,\xi_t)B(t,\xi_t) + A_1(t,\xi_t)\gamma_t}{B(t,\xi_t)}, \tag{38}$$

with initial conditions ξ_0, m_0, and γ_0.

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

Proof. Theorem 8.1 of Liptser and Shiryaev (2001a) and Theorem 12.5 of Liptser and Shiryaev (2001b) jointly provide the proof of Proposition 2a. Theorem 12.5 of Liptser and Shiryaev (2001b) provides the proof of Proposition 2b. Theorem 11.1 of Liptser and Shiryaev (2001b) provides the proof of Proposition 2c. We also assume that the technical requirements to prove the theorems over the period $0 \le t \le T$ are satisfied.

Again, to make economic sense, we assume a nonnegative $b_2(t,\xi_t)$, which ensures a positive $\sigma_m(\gamma_t)$ in Equation (38). That is, a positive (negative) shock in fund gross alpha induces an increase (a decrease) in inferred manager ability.

Then, we have the equilibrium results stated in the following proposition.

Proposition NL. If the fund manager's ability and fund gross alpha follow the nonlinear system represented by Equations (32) and (33), the equilibrium flow-performance relations are as follows.

a. If investors are risk-neutral, the equilibrium flow-performance relation using gross alpha as the performance measure is

$$\frac{dq_t^*}{q_t^*} = \frac{2\sigma_m(\gamma_t)}{m_t B(t,\xi_t)}\left(\frac{d\xi_t}{\xi_t}\right) + \frac{\sigma_m^2(\gamma_t)}{m_t^2 B^2(t,\xi_t)}\left(\frac{d\xi_t}{\xi_t}\right)^2 + \frac{2}{m_t} \times \left[a_0(t,\xi_t) + a_1(t,\xi_t)m_t - \frac{A_1(t,\xi_t)\sigma_m(\gamma_t)m_t}{B(t,\xi_t)}\right]dt, \tag{39}$$

and the equilibrium flow-performance relation using net alpha as the performance measure is

$$\frac{dq_t^*}{q_t^*} = \frac{A_1(t,\xi_t)\sigma_m(\gamma_t)}{fB(t,\xi_t)}\left(\frac{dS_t}{S_t}\right) \tag{40}$$

$$+\frac{A_1^2(t,\xi_t)\sigma_m^2(\gamma_t)}{4f^2B^2(t,\xi_t)}\left(\frac{dS_t}{S_t}\right)^2+2\left[\frac{a_0(t,\xi_t)}{m_t}+a_1(t,\xi_t)\right]dt.$$

b. If investors are risk-averse, the equilibrium flow-performance relation using gross alpha as the performance measure is

$$\frac{dq_t^*}{q_t^*}=\frac{2\sigma_m(\gamma_t)}{m_tB(t,\xi_t)}\left(\frac{d\xi_t}{\xi_t}\right)+\frac{\sigma_m^2(\gamma_t)}{m_t^2B^2(t,\xi_t)}\left(\frac{d\xi_t}{\xi_t}\right)^2+\frac{2}{m_t}$$
$$\times\left[a_0(t,\xi_t)+a_1(t,\xi_t)m_t-\frac{A_1(t,\xi_t)\sigma_m(\gamma_t)m_t}{B(t,\xi_t)}\right]dt+X_t \tag{41}$$

where X_t contains all the terms related to dV_t and $(dV_t)^2$, and the equilibrium flow-performance relation using net alpha as the performance measure is

$$\frac{dq_t^*}{q_t^*}=\frac{A_1(t,\xi_t)\sigma_m(\gamma_t)}{fB(t,\xi_t)}\left(\frac{dS_t}{S_t}\right)+\frac{A_1^2(t,\xi_t)\sigma_m^2(\gamma_t)}{4f^2B^2(t,\xi_t)}\left(\frac{dS_t}{S_t}\right)^2+Y_tdt$$
$$+X_t \tag{42}$$

where

$$Y_t=\frac{2a_0(t,\xi_t)}{m_t}+2a_1(t,\xi_t)-\frac{A_1(t,\xi_t)\sigma_m(\gamma_t)B^2(t,\xi_t)\mu_p}{B(t,\xi_t)\left(B^2(t,\xi_t)\mu_p+cV_t\sigma_p^2\right)}. \tag{43}$$

Proof. Similar to those in the previous sections.

The results of this nonlinear filtering framework are different from those of the linear filtering framework in the following ways. First, the estimation precision of manager ability, γ_t, is stochastic. Second, the sensitivity of expected manager ability to innovation shocks in fund gross alpha, $\sigma_m(\gamma_t)$, is stochastic. Third, these stochastic parameters yield more complex patterns of equilibrium flow-performance sensitivity over time. In this framework, γ_t, $\sigma_m(\gamma_t)$, and the equilibrium flow-performance sensitivity can change nonmonotonically over time.[28] This is one of the key differences between our results and those of Berk and Green (2004) and subsequent models, where flow-performance sensitivity decreases, or changes monotonically, over time. Consequentially, this framework, which has no added parameters but allows

[28] Nonmonotonic such patterns arise naturally, whereas within the linear setup that we studied earlier, transient such patterns may arise only under special parameters values. These special parameter values induce negative correlation between performance and inferred abilities. To make economic sense, and for brevity, we do not allow these parameter values. See Footnotes 18 and 21.

nonlinear structures, better explains empirical nonmonotonic flow-performance relations.

3 Simulation Results

We use simulation to illustrate our equilibrium flow-performance relation under a case of constant unobservable manager ability [as in Berk and Green (2004), Case One)], a case of dynamic unobservable manager ability with linear filtering (Case Two), and a case of dynamic unobservable manager ability with nonlinear filtering (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 $d\overline{W}_t = \Delta\overline{W}_t$ to follow a normal distribution of mean zero and variance Δt.

In our simulation, we use some statistics from our sample of the active managed equity mutual funds in the U.S. market. A detailed description of the sample is in the data section of our empirical study. In the sample, the average annual fund expense, including the 12b-1 fee, is 1.31%; the average monthly net alpha is -0.04%; and the standard deviation of the monthly net return is 5.43%. Then, we set our model's parameters as follows. Our monthly fee f is $1.31\%/12 = 0.11\%$ (which is approximated by the average monthly fund expense); our initial expected management ability m_0 is $1.31\% - 0.04\% = 1.27\%$ (which is approximated by the average fund gross alpha in our sample); our gross fund share price $\xi_0 = 1$; and our instantaneous volatility of gross alpha B is 5.43% (which is the same as the sample standard deviation of the monthly net return). We set other parameters under three cases:

- Case One: $a_0 = 0,\ a_1 = 0,\ A_1 = 1,\ b_1 = 0,\ b_2 = 0,\ \gamma_0 = 0.0006$;
- Case Two: $a_0 = 0.005,\ a_1 = -0.001,\ A_1 = 0.01,\ b_1 = 0.06,\ b_2 = 0.01,\ \gamma_0 = 0.0001$;
- Case Three: $a_0 = 0.005,\ a_1 = -0.001,\ A_1 = 0.55 + 0.001\ln(1+\xi_t),\ b_1 = 0.0005 + 0.000012t + 0.000004\ln(1+\xi_t),\ b_2 = 0.0001,\ \gamma_0 = 0.0006$.

We then simulate m_t, γ_t, $\sigma_m(\gamma_t)$, fund net alphas $\Delta S_t/S_t$, and fund flow $\Delta q_t^*/q_t^*$, for 240 months (20 years). We plot the results of $\sigma_m(\gamma_t)$ in the three cases, from Month 13 to Month 240, in Figure 1. Also, in Figure 1, we use blue circles to plot the values of fund flows and fund net alphas from Month 25 to Month 48, green stars to plot these values from Month 85 to Month 108, and red pluses to plot these values from Month 217 to Month 240.

In Case One, the manager ability is an unobservable constant. This is the case in Berk and Green (2004), where the estimation of manager ability becomes more and more precise over time, i.e., γ_t decreases over time towards its static state value zero. 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.

In Case Two, the unobservable manager ability is dynamic under a linear filtering 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 conditional expected manager 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 towards its steady state value. Consequently, $\sigma_m(\gamma_t)$ decreases over time and the flow–net alpha sensitivity decreases over time.

In Case Three, the unobservable manager ability is dynamic under a nonlinear filtering framework. In this case, the static state value of γ_t moves over time. Due to parameter values, γ_t is above its steady state value in the earlier months and below it in the later months. Consequently, γ_t decreases toward steady state values in the earlier months and increases towards new steady state values in the later months. As a result, $\sigma_m(\gamma_t)$ is stochastic. It first decreases over time and then increases, i.e., the conditional expected manager 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. Eventually, the flow–net alpha sensitivity first decreases and then increases.

Different from the results in Berk and Green (2004) that the flow–net alpha sensitivity decreases monotonically over time, our results show that the flow–net alpha sensitivity can change with different patterns over time, and these new patterns result from the dynamics of manager abilities. 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.

4 Empirical Study

Numerous papers have studied empirically flow-performance relations under different contexts [see, for example, Bollen (2007), Brown and Wu (2016), Chen, Goldstein, and Jiang (2010), Huang, Wei, and Yan (2007), Lynch and Musto (2003), and Spiegel and Zhang (2013)]. The purpose of our empirical analysis is to examine whether the empirical findings support a framework of dynamic unobservable manager abilities with nonmonotonic precisions (a framework which must be nonlinear) or a framework with monotonic precisions (a framework which must be linear).

Based on our theoretical results, if manager abilities' precisions are constant or monotonic, flow-performance sensitivities change monotonically over time. If managers' abilities are dynamic, flow-performance sensitivities can have different patterns over time. How the flow-performance sensitivity changes over time implies whether the fund manager abilities' precisions are constant, monotonic, or nonmonotonic. We test this sensitivity below.

4.1 Methodology

We first analyze the flow-performance relation using methods common in the literature, specifically cited below. The model we use is

$$Flow_{i,t} = \beta_0 + \beta_1 \alpha_{i,t-1} + \beta_2 \alpha_{i,t-1} * D_{i,t-1} + \gamma Controls_{i,t} + \varepsilon_{i,t}, \tag{44}$$

where

$$Flow_{i,t} = \frac{TNA_t - TNA_{t-1}(1 + Ret_t)}{TNA_{t-1}}. \tag{45}$$

Here, i is the fund index and t is the time (month) index. The dependent variable $Flow_{i,t}$ is the (percentage) fund flow, TNA_t is the fund's total net asset under management, Ret_t is the fund net return, and $\alpha_{i,t}$ is the fund net alpha to measure fund performance. Following Feldman, Saxena, and Xu (2020, 2021), we estimate the following style-matching model:

$$Ret_t = \alpha_{i,t} + b_{i,t}^1 F_t^1 + b_{i,t}^2 F_t^2 + \cdots + b_{i,t}^n F_t^n, \tag{46}$$

where F_t^1 through F_t^n are the net returns of tradable index funds of different asset classes. We use tradable index funds as factors in this model because index funds are intended to represent the next-best investment opportunity available to investors as a tradable asset [Berk and van Binsbergen (2015)]. 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 $\alpha_{i,t}$. $D_{i,t}$ is a dummy variable; and $D_{i,t} = 1$ if $\alpha_{i,t} \geq 0$, and 0 otherwise. A positive β_2 implies that the flow–net alpha sensitivity, where the net alphas are positive, is higher than the one where the net alphas are negative, so a positive β_2 implies convexity in the flow–net alpha relation.

The coefficients of the control variables are represented by the vector γ. We follow the literature [e.g., Brown and Wu (2016), Chen, Goldstein, and Jiang (2010) and Spiegel and Zhang (2013)] to choose control variables in the vector $Controls_{i,t}$, listed below.

- $Expense_{i,t}$: fund expense ratio as of the most recently completed fiscal year, which includes 12b-1 fees;
- $Flow_{i,t-1}$: the lagged fund flow;
- $Vol_{i,t}$: fund volatility, calculated as the standard deviation of the fund's net returns in the prior 12 months;
- $\ln Age_{i,t-1}$: the lagged natural logarithm of fund age, which is calculated as the number of months since the inception of the oldest share class;
- $\ln TNA_{i,t-1}$: the lagged natural logarithm of the fund's total net assets under management;
- $\alpha_{i,t-1} * \ln TNA_{i,t-1}$: the interaction term of lagged fund net alpha and the lagged natural logarithm of the fund's total net assets under management, as larger funds might experience less volatile fund flows, holding other factors unchanged;
- $FamAlpha_{i,t-1}$: lagged fund family net alpha, calculated as the weighted average of the family members' net alphas excluding the net alphas of fund i, where the lagged

net asset under management is the weight;

- $\ln FamSize_{i,t-1}$: the lagged natural logarithm of the family size, which is the number of coexisting active equity-only funds in the family;
- fund dummies and year dummies.

Our main purpose is to find out whether flow-performance relations change over time. Thus, we require the funds in our sample to have sufficiently long lives, i.e., at least 15 years of fund flow and net alpha observations. We divide a fund's time-series observations of fund flows and net alphas into four periods: Period 0: the first 5 years of observations in the sample, Period 1: the second 5 years, Period 2: the third 5 years, and Period 3: the remaining years of observations. To analyze how the flow–net alpha sensitivity changes over time, we use the model

$$\begin{aligned} Flow_{i,t} = \beta_0 &+ \beta_1 \alpha_{i,t-1} + \beta_2 \alpha_{i,t-1} * M1_{i,t} + \beta_3 \alpha_{i,t-1} * M2_{i,t} \\ &+ \beta_4 \alpha_{i,t-1} * M3_{i,t} + \beta_5 M1_{i,t} + \beta_6 M2_{i,t} + \beta_7 M3_{i,t} \\ &+ \gamma Controls_{i,t} + \varepsilon_{i,t}, \end{aligned} \tag{47}$$

where $M1_{i,t}$, $M2_{i,t}$, and $M3_{i,t}$ are 1 if the time is in Period 1, Period 2, and Period 3, respectively, and 0 otherwise. In this model, Period 0 is the base group. We focus on β_2, β_3, β_4, $\beta_3 - \beta_2$, $\beta_4 - \beta_3$, and $\beta_4 - \beta_2$, which measure the differences in the flow–net alpha sensitivity between Periods 0 and 1, between Periods 0 and 2, between Periods 0 and 3, between Periods 1 and 2, between Periods 2 and 3, and between Periods 1 and 3, respectively.

To analyze how the convexity of the flow-performance relation changes over time, we use the model

$$\begin{aligned} Flow_{i,t} = \beta_0 &+ \beta_1 \alpha_{i,t-1} + \beta_2 \alpha_{i,t-1} * D_{i,t-1} + \beta_3 \alpha_{i,t-1} * D_{i,t-1} * M1_{i,t} \\ &+ \beta_4 \alpha_{i,t-1} * D_{i,t-1} * M2_{i,t} + \beta_5 \alpha_{i,t-1} * D_{i,t-1} * M3_{i,t} \\ &+ \beta_6 \alpha_{i,t-1} * M1_{i,t} + \beta_7 \alpha_{i,t-1} * M2_{i,t} + \beta_8 \alpha_{i,t-1} * M3_{i,t} \\ &+ \beta_9 M1_{i,t} + \beta_{10} M2_{i,t} + \beta_{11} M3_{i,t} + \gamma Controls_{i,t} + \varepsilon_{i,t}. \end{aligned} \tag{48}$$

We focus on β_3, β_4, β_5 $\beta_4 - \beta_3$, $\beta_5 - \beta_4$, and $\beta_5 - \beta_3$, which measure the differences in the convexity of the flow–net alpha relation, between Periods 0 and 1, between Periods 0 and 2, between Periods 0 and 3, between Periods 1 and 2, between Periods 2 and 3, and between

Periods 1 and 3, respectively.

4.2 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 1995 to December 2019, and monthly data is used.[29] We first exclude index funds, variable annuity funds, and exchange-traded funds (ETFs). Then, we choose U.S. domestic equity-only mutual funds by using the Lipper fund classification.[30] 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 in Pastor, Stambaugh, and Taylor (2015).[31] 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 fund flows and net alphas, i.e., at least 15 years, to analyze how the flow–net alpha sensitivity changes over time, we include the funds that have at least 20 years of observations. We also require each of our equity funds to have fewer than 5 years' missing observations between the first observation and the last one, so that the style-matching model can perform well. Briefly, our sample includes long-living active equity-only funds.

All fund returns are net of management expenses, 12b-fees, and front and rear load fees. We also obtain funds' net assets under management and the expense ratio 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 and fund expenses as weighted averages of its share classes' net returns and fund expenses, respectively, using the lagged net assets under management as weights. Fund age is the number of months since the inception of the oldest share class. Fund family is identified by the management company

[29] 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 widow to estimate fund net alpha, we start our sample from January 1995 so that our tests can include fund data starting from January 2000.
[30] 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.
[31] See the discussion in the appendix regarding the equity fund filter in Feldman, Saxena, and Xu (2020a).

code.[32] 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 weight.[33] Fund family size is calculated as the number of active equity-only funds in the family.

We also obtain data on index funds from CRSP and use the MFLINKS database to aggregate fund share class-level information to fund-level information for the index funds similar to what we do for the active equity-only funds. These index funds, which we use as benchmark factors to estimate fund net alphas in the style-matching model in (46), include a Large-Cap Core fund (Schwab 1000 Index Fund), a Large-Cap Growth fund (Vanguard Growth Index Fund), a Large-Cap Value fund (Vanguard Value Index Fund), a Small-Cap Core fund (Vanguard Small-Cap Index Fund), and a Multi-Cap Core fund (Vanguard Total Stock Market Index Fund). We require index funds to have no missing observations in our sample period. The CRSP Fama-French risk-free rate is also used as one of the benchmark factors to estimate fund net alphas.

Our sample contains 1,555 actively managed U.S. domestic equity-only mutual funds, and 1,528 of them belong to a fund family with more than one funds. This shows that the number of funds in our sample is not far away from the one of Brown and Wu (2016)[34] although we include only long-living funds. Also, the numbers of observations in our tests are close to triple of theirs, as we have a longer sample period.

4.3 Empirical Results

Table 1 reports the summary statistics. In our 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 23% at the 99^{th} percentile and -15% at the 1^{st} percentile. Similar to Brown and Wu (2016), to mitigate the effects of extreme observations that are potentially due to fund mergers or data error, for each fund, we winsorize the fund flow variable at the 1^{st} and the 99^{th} percentiles.[35]

[32] 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.
[33] In our sample, to be included in a family, a fund should be an active equity-only fund as defined above.
[34] The sample of Brown and Wu (2016) contains 2,053 funds.
[35] Instead of winsorizing the flow observations, in unreported robustness tests, we winsorize all variables at the

Moreover, the R-squared of the style-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. In addition, 99% of our funds belong to a family with two or more funds, showing that most of the funds in our sample are managed by fund families.

Table 2 illustrates the results of the model in Equation (44). We find that, on average, fund flows increase with lagged fund net alphas, as the coefficient of $\alpha_{i,t-1}$ is highly significantly positive after we include all controls. By the results in (3), if the fund net alpha increases by 1 percentage point, a fund with the average size of \$1170.67 million would experience an increase in fund flow by around 0.14 (= -0.0214ln(1170.67) + 0.2865) percentage points. Results of model specification (4) in Table 2 show that the coefficient of the interaction term $\alpha_{i,t-1} * D_{i,t-1}$ is positive and significant at the 1% significance level. Then, fund flows are more sensitive to fund net alphas when fund net alphas are positive than when they are negative. If the fund net alpha increases by 1 percentage point, the fund flow increases by 0.13 percentage points more if the fund net alpha is positive than if it is negative. Thus, the flow–net alpha relation in our sample exhibits convexity. These results are consistent with the findings in the previous literature, such as Brown and Wu (2016), Lynch and Musto (2003), and Spiegel and Zhang (2013).

Table 3 shows the results of the model in Equation (47). We can see that the coefficient of $\alpha_{i,t-1}$ is positive and highly significant, showing that in the base group, i.e., in Period 0, fund flows increase with fund net alphas. Also, the coefficient of $\alpha_{i,t-1} * M1_{i,t}$ is negative and significant, showing that from Period 0 to Period 1, the flow–net alpha sensitivity decreases [if the fund net alpha increases by 1 percentage point, the fund flow decreases by 0.08 percentage points fewer in Period 1 than in Period 0, by model specification (3)]. Moreover, shown by the results in Panel B, the coefficient of $\alpha_{i,t-1} * M2_{i,t}$ is significantly higher than that of $\alpha_{i,t-1} * M1_{i,t}$, so from Period 1 to Period 2, the flow–net alpha sensitivity increases [if the fund net alpha increases by 1 percentage point, the fund flow increases by 0.09 percentage points more in Period 2 than in Period 1, by model specification (3)]. In addition, the coefficient

1st and 99th percentiles or exclude the flow observations below the 1st or above the 99th percentile following Brown and Wu (2016). We find very similar results in these robustness tests.

of $\alpha_{i,t-1} * M3_{i,t}$ is significantly higher than that of $\alpha_{i,t-1} * M2_{i,t}$, so from Period 2 to Period 3, the flow–net alpha sensitivity decreases [if the fund net alpha increases by 1 percentage point, the fund flow decreases by 0.06 percentage points fewer in Period 3 than in Period 2, by model specification (3)]. To conclude, the flow–net alpha sensitivity decreases from Period 0 to Period 1, increases from Period 1 to Period 2, and then decreases from Period 2 to Period 3; and, eventually, the flow–net alpha sensitivity in Period 3 is insignificantly different from the one in Period 0, as the coefficient of $\alpha_{i,t-1} * M3_{i,t}$ is insignificant.

We stress again that because of the dynamic nature of our model, the decreasing-increasing-decreasing pattern of flow–net alpha sensitivity we detect depends on our sample period. No pattern, in this dynamic context, is expected to replicate itself due to different initial condition and stochastic realizations.

If manager abilities' precisions are constant or monotonic, the flow–net alpha sensitivity should change with time monotonically. The nonmonotonic flow–net alpha sensitivities that we identify above and report in Table 3 do not support the framework where fund manager abilities' precisions are constant or monotonic. Instead, these findings support the case of dynamic unobservable manager abilities that drive the flow–net alpha sensitivities to increase during some periods and to decrease over others.

Table 4 reports the results of the model in Equation (48). We do not find strong evidence that the flow–net alpha convexity changes over time. The difference in the coefficients of $\alpha_{i,t-1} * D_{i,t-1} * M3_{i,t}$ and $\alpha_{i,t-1} * D_{i,t-1} * M1_{i,t}$ are negative but only marginally significant [shown in Panel B model specification (3)]. The coefficients and other differences in the coefficients of $\alpha_{i,t-1} * D_{i,t-1} * M1_{i,t}$, $\alpha_{i,t-1} * D_{i,t-1} * M2_{i,t}$, and $\alpha_{i,t-1} * D_{i,t-1} * M3_{i,t}$ are insignificant either in all or in model specifications (2) and (3). If manager abilities' precisions are constant or monotonic, the flow–net alpha convexity should change with time monotonically. We do not find this result here, and our results in Table 4 are unlikely to support a framework of unobservable abilities with monotonic precisions.

We also re-run our models in Equations (47) and (48) for each individual fund and report the numbers of funds whose relevant coefficients are significant in Table 5.[36] We further

[36] In the regressions for each fund, we do not include the fund dummies and year dummies, and we use the Newey-West estimator to estimate the standard errors, with the maximum lag of 12 to be considered in the

require the funds to have no missing observations between the first observation and the last one so that we can use the Newey-West estimator to estimate the standard errors, and we require the funds to have at least two years' observations in Period 3. There is a total of 1029 funds in these tests on individual funds.

Panel A shows the results of the model in Equation (47) regarding flow–net alpha sensitivity. We find that out of 1029 funds, 79 (139) of them experience an increase (decrease) in the flow–net alpha sensitivity from Period 0 to Period 1; 140 (90) of them experience an increase (decrease) in such sensitivity from Period 1 to Period 2; and 95 (117) of them experience an increase (decrease) in such sensitivity from Period 2 to Period 3. Also, 108 (128) funds experience an increase (decrease) in flow–net alpha sensitivity from Period 0 to Period 2; 104 (135) funds experience an increase (decrease) in such sensitivity from Period 0 to Period 3; and 124 (116) funds experience an increase (decrease) in such sensitivity from Period 1 to Period 3. Also, we can see that the majority of the funds have insignificant changes in flow–net alpha sensitivity over time.

Panel B shows the results of the model in Equation (48) regarding flow–net alpha convexity. We find that of 1029 funds, 121 (93) funds experience an increase (decrease) in the flow–net alpha convexity from Period 0 to Period 1; 94 (101) funds experience an increase (decrease) in such convexity from Period 1 to Period 2; and 97 (83) funds experience an increase (decrease) in such convexity from Period 2 to Period 3. Also, 95 (91) funds experience an increase (decrease) in flow–net alpha convexity from Period 0 to Period 2; 101 (94) funds experience an increase (decrease) in such convexity from Period 0 to Period 3; and 116 (111) funds experience an increase (decrease) in such convexity from Period 1 to Period 3. Also, we can see that the majority of the funds have insignificant changes in flow–net alpha convexity over time.

The results in this table are unlikely to support the framework of constant or monotonic unobservable manager abilities' precisions. If managers' abilities' precisions are constant or monotonic, we should find that a large portion of the funds experience a monotonic change in the flow–net alpha sensitivity and convexity from one period to another. However, we find that

---

autocorrelation structure of the regression error.

the majority of funds experience no change in flow–net alpha sensitivity or convexity, and some funds even experience a nonmonotonic change in such sensitivity and convexity from one period to another. Therefore, the results in this table are likely to support the framework where managers' abilities' precisions are dynamic nonmonotonic and these dynamic processes vary across funds.

Robustness Check

For robustness checking, we redefine the time periods using calendar time, such that Period 0, Period 1, Period 2, and Period 3 are from January 2000 to December 2004, from January 2005 to December 2009, from January 2010 to December 2014, and from January 2015 to December 2019, respectively. We then rerun our models and find quite consistent results. We skip the results for brevity.

For our second robustness check, we analyze how the flow–net alpha sensitivity changes with funds' age. If the flow–net alpha sensitivity changes monotonically with funds' age, then it is consistent with a framework of unobservable abilities with monotonic precisions. On the other hand, if the flow–net alpha sensitivity changes with funds' age nonmonotonically, then it is consistent with a nonlinear dynamic unobservable ability framework with nonmonotonic precisions. Our model is

$$\begin{aligned} Flow_{i,t} = \beta_0 + \beta_1 \alpha_{i,t-1} + \sum_j c_j \alpha_{i,t-1} * \left(Age_{i,t-1}\right)^j \\ + \sum_j d_j \left(Age_{i,t-1}\right)^j + \gamma Controls_{i,t} + \varepsilon_{i,t}. \end{aligned} \tag{49}$$

Then, the flow–net alpha sensitivity from this model is $\beta_1 + \sum_{j=1}^{n} c_j \left(Age_{i,t-1}\right)^j$.

Table 6 shows the results of this model. 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 $j = 6$ because the coefficients of these terms with $j = 7$ are insignificant. As the interaction terms from $\alpha_{i,t-1} * Age_{i,t-1}$ to $\alpha_{i,t-1} * \left(Age_{i,t-1}\right)^6$ are significant, then on average, the flow–net alpha sensitivity changes with fund age nonmonotonically. We also notice that the interaction term $\alpha_{i,t-1} * \ln TNA_{i,t-1}$ becomes insignificant when we include the terms $\alpha_{i,t-1} * \left(Age_{i,t-1}\right)^j$ into the model. Thus, the effect of fund age on the flow–net alpha sensitivity overwhelms that of fund size on the flow–net alpha sensitivity.

We draw the flow–net alpha sensitivity in Table 6 using the coefficient values of the model specification in Equation (6) and show that sensitivity in Figure 2. We plot the results for average funds of ages of 5 to 40 years because around 90% of our observations have a fund age within this range. When the average fund is 5 years old, if the fund net alpha increases by 1 percentage point, the fund flow increases by around 0.3 percentage points. As the average fund grows from 5 years old to 10 years old, the flow–net alpha sensitivity decreases. When the average fund is 10 years old, a percentage point increase in fund net alpha stimulates an increase in fund flow by only 0.2 percentage points. After the average fund's age reaches 10 years, the flow–net alpha sensitivity gradually increases with fund age. After the average fund's age is higher than around 25 years, the flow–net alpha sensitivity decreases again, and then increases after the average fund's age is higher than around 35 years. When the average fund's age is higher than 40 years, the flow–net alpha sensitivity increases to a level similar to the one when the fund is 5 years old.

In general, the result shows that the flow–net alpha sensitivity changes with fund age nonmonotonically, supporting a nonlinear framework of dynamic unobservable ability. The decrease, in our sample, in flow–net alpha sensitivity with fund age in the earliest years might imply that investors have more and more precise estimates of manager abilities during these early years. However, manger abilities are dynamic over time, inducing lower or higher estimation precisions in the later years. Consequently, flow–net alpha sensitivity changes with fund age nonmonotonically, and there are turning points in the graph in the later years.

5 Insights into the Findings in the Literature

Current studies of the active fund management industry find interesting phenomena, and we show how our model provides insights into these findings.

5.1 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 the flow-performance relation is convex [see, for example, Berk and Green (2004), Brown and Wu (2016), and Lynch and Musto (2003)], whereas other studies suggest that this relation is linear [see, for example, Spiegel and Zhang

(2013)]. This study complements this discussion by showing that the flow-performance sensitivity can increase, decrease, be nonmonotonic, or stay unchanged over time, making empirical findings of these two types of curvature possible.

Whereas the cross-sectional heterogeneity of functions mapping flows to performance would affect the empirical findings of the curvature of this relation using panel data [see the discussion in the introduction of Spiegel and Zhang (2013)], the dynamics over time of these functions would also affect the empirical findings. For a particular fund, our theoretical results suggest that the intercept, slope, and curvature of the function mapping flows to performance change over time, and these changes might be nonmonotonic [see for example, Equations (16) and (40)]. Thus, if we plot the time-series observations of its fund flows and fund net alphas on a two-dimensional space, the observation points could fill in an area such that the empirical fitted function shows linearity or convexity.

For illustration, in Figure 3, 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 (i.e., the blue dashed curves) with the independent variable as the fund net alpha and the dependent variable as the fund flow, and the corresponding observation points (i.e., the red circles). These four functions show the flow–net alpha relation of different time periods of the same fund. In the first situation shown on the left, we can see that the observations stay in an ellipse area. Consequently, the empirical fitted function (i.e., the black line) based on these observations is an upward-sloping line. In the second situation shown on the right, the observations stay in a crescent area, so the empirical fitted function (i.e., the black curve) based on these observations is an increasing and convex curve. In the real data, observations of fund flows and fund net alphas can stay in an area with more complex shapes. Thus, even though by theory, the flow-performance relation at each time period is increasing and convex, empirically, we can observe that this relation is linear or convex, depending on the situation.

If we put different funds' observations of fund flows and fund performances together in a panel regression, the situation would be more complex, because both the cross-sectional heterogeneity and the dynamics over time of the flow-performance relation exert impacts on the empirical results. Therefore, when analyzing the curvature of the flow-performance relation, we need to incorporate these two types of effects simultaneously. As this paper focuses on modelling dynamic unobservable fund manager ability, we leave this empirical issue for future studies.

5.2 Insights on Fund Marketing Activities

There is interesting discussion in the literature of how funds' marketing activities affect the flow-performance relation. 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, have a higher flow sensitivity to medium performance and a lower flow sensitivity to high performance, i.e., a less convex flow-performance relation. They develop a model and show that given fund-level participation barriers, new investors can cover their participation costs (cost to obtain information about a fund) only if the fund performance improves. Hence, as the fund performance increases, the investments from both existing investors and new investors increase. Thus, compared to a fund with low participation cost, a fund with high participation cost would have fund flows increasing faster with fund performance, resulting in a more convex flow-performance relation. As funds with more marketing activities tend to have lower participation costs, they have a less convex flow-performance relation.

Our model also offers insights into this phenomenon. First, the marketing expenditures are not directly related to fund size, so they are unlikely to be captured by the fund costs variable $C(q_t^a)$; instead, the fund manager charges management fee f to cover these expenditures. Therefore, if more marketing expenditures imply a higher management fee, then they would induce a less convex flow-performance relation based on our theoretical results.

Secondly and more importantly, a fund's performance is likely to be (highly) correlated with other funds in the same fund family because the fund manager can obtain assistance from other managers in the family.[37] Thus, marketing activities, such as promoting the fund family and its star funds, could offer additional information for investors to estimate the fund manager's ability. If this is the case, then investors would have a more precise prior estimate of the manager's ability (i.e., γ_0 is smaller) and, consequently, more precise estimates of the manager's ability over time, making the inferred manager ability less sensitive to innovation shocks in fund performance (i.e., $\sigma_m(\gamma_t)$ is smaller). Then, the flow-performance relation is

[37] Also see the discussions in Brown and Wu (2016).

less convex.

We do not explicitly model the fixed up-front participation cost, as Huang, Wei, and Yan (2007) do. Also, in reality, whether funds' marketing activities would reduce investors' up-front participation costs is difficult to observe and confirm. 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 we show that the flow-performance relation is less convex.

6 Conclusion

We introduce a continuous-time rational model of the active fund management industry where unobservable fund manager abilities are dynamic. Our model predicts that flow-performance relations change over time in a variety of patterns. Depending on parameter values and realizations, dynamics of managers' abilities result in increasing, decreasing, nonmonotonic, or constant sensitivity of inferred manager abilities to innovation shocks in fund returns over time, consequently leading to increasing, decreasing, nonmonotonic, or constant flow-performance sensitivity, respectively. We show that these equilibrium results hold whether investors are risk-neutral or mean-variance risk-averse. We also show that if unobservable fund manager abilities and fund performance follow a nonlinear framework, the equilibrium flow-performance relation can have a more complex pattern over time. In addition, we offer empirical evidence of the dynamics of the flow–net alpha sensitivity, which supports a nonlinear framework of dynamic manager abilities.

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 also show that if marketing activities increase management fees and/or improve estimates of fund manager abilities, then the empirical flow-performance relation is less convex.

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

Appendix

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

$$q_t^{a^*} = \frac{A_1 m_t}{2c}. \tag{A1}$$

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

$$f q_t^* = \frac{(A_1 m_t)^2}{4c}. \tag{A2}$$

Rearranging, the optimal fund size is

$$q_t^* = \frac{(A_1 m_t)^2}{4cf}. \tag{A3}$$

Dividing Equation (A1) by Equation (A3) gives

$$\frac{q_t^{a^*}}{q_t^*} = \frac{2f}{A_1 m_t}. \tag{A4}$$

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

$$\frac{dS_t}{S_t} = \frac{2f}{A_1 m_t}\frac{d\xi_t}{\xi_t} - 2f dt. \tag{A5}$$

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

$$\frac{dS_t}{S_t} = \frac{2fB}{A_1 m_t} d\overline{W}_t. \tag{A6}$$

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

$$\frac{dq_t^*}{q_t^*} = \frac{2A_1^2 m_t dm_t + A_1^2 (dm_t)^2}{(A_1 m_t)^2} = \frac{2m_t dm_t + (dm_t)^2}{m_t^2}. \tag{A7}$$

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

$$\frac{dq_t^*}{q_t^*} = \frac{2\sigma_m(\gamma_t)}{m_t B}\left(\frac{d\xi_t}{\xi_t}\right) + \frac{\sigma_m^2(\gamma_t)}{m_t^2 B^2}\left(\frac{d\xi_t}{\xi_t}\right)^2 + \frac{2}{m_t}\left[(a_0 + a_1 m_t) - \frac{A_1 \sigma_m(\gamma_t) m_t}{B}\right] dt. \tag{A8}$$

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

$$\frac{dq_t^*}{q_t^*} = \frac{A_1 \sigma_m(\gamma_t)}{fB}\left(\frac{dS_t}{S_t}\right) + \frac{A_1^2 \sigma_m^2(\gamma_t)}{4f^2 B^2}\left(\frac{dS_t}{S_t}\right)^2 + 2\left(\frac{a_0}{m_t} + a_1\right) dt. \tag{A9}$$

Q.E.D.

Proof of Results in Section 2.4

Going through the process that is the same 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 (18) 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 (A5) and (19) into (A11) (with $A_1 = 1$), we have

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

Q.E.D.

Proof of Results in Section 2.5

Substituting Equation (9) and then Equation (5) into Equation (25), and regarding q_t^a, q_t, and f as exogenous to the investor, we calculate $\mathrm{E}\left[\frac{dp_t}{p_t}|\mathcal{F}_t^{\xi}\right]$ and $\mathrm{Var}\left[\frac{dp_t}{p_t}|\mathcal{F}_t^{\xi}\right]$. Then, the investor's problem becomes

$$\max_{w_t's} \frac{\left[w_t\left(\frac{q_t^a}{q_t}A_1 m_t - \frac{c{q_t^a}^2}{q_t} - f\right) + \mu_p\right]dt}{\sqrt{\left[w_t^2\left(\frac{q_t^a}{q_t}\right)^2 B^2 + \sigma_p^2\right]dt}}, \tag{A13}$$

subject to

$$0 \le w_t \le 1. \tag{A14}$$

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

$$w_t^* = \frac{\left(\frac{q_t^a}{q_t}A_1 m_t - \frac{c{q_t^a}^2}{q_t} - f\right)\sigma_p^2}{\left(\frac{q_t^a}{q_t}\right)^2 B^2 \mu_p}. \tag{A15}$$

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^*. 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_t, $V_t \in (0, +\infty)$, $0 \le t \le T$. In reality, V_t not only depends on the returns from financial assets, but also depends on production activities, research and development expenditures, consumption, taxes, and many other aspects of the economy that we do not model here. Thus, to simplify our analysis, we assume that V_t is exogenous to both investors and managers. Here, the amount of wealth allocated to the fund, i.e., the fund's size, is

$$q_t = w_t^* V_t = V_t \frac{\left(\frac{q_t^a}{q_t}A_1 m_t - \frac{c{q_t^a}^2}{q_t} - f\right)\sigma_p^2}{\left(\frac{q_t^a}{q_t}\right)^2 B^2 \mu_p}. \tag{A16}$$

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

$$fq_t = -\frac{{q_t^a}^2 B^2 \mu_p}{V_t \sigma_p^2} - c{q_t^a}^2 + q_t^a A_1 m_t. \tag{A17}$$

The fund manager's objective is to maximize the fund's profit, fq_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

$$\max_{q_t^a} -\frac{{q_t^a}^2 B^2 \mu_p}{V_t \sigma_p^2} - c{q_t^a}^2 + q_t^a A_1 m_t, \tag{A18}$$

subject to

$$0 \le q_t^a \le q_t. \tag{A19}$$

Here, q_t and V_t 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 ${q_t^a}^*$:

$$ {q_t^a}^* = \frac{A_1 m_t V_t \sigma_p^2}{2\left(B^2 \mu_p + cV_t \sigma_p^2\right)}. \tag{A20}$$

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

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

$$q_t^* = \frac{(A_1 m_t)^2 V_t \sigma_p^2}{4f\left(B^2 \mu_p + cV_t \sigma_p^2\right)}. \tag{A21}$$

We can see that

$$\frac{{q_t^a}^*}{q_t^*} = \frac{2f}{A_1 m_t}. \tag{A22}$$

The fund manager's optimal profit is

$$fq_t^* = \frac{(A_1 m_t)^2 V_t \sigma_p^2}{4\left(B^2 \mu_p + cV_t \sigma_p^2\right)}. \tag{A23}$$

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:

$$\frac{dS_t}{S_t} = \frac{2f}{A_1 m_t}\frac{d\xi_t}{\xi_t} - \frac{fcV_t\sigma_p^2}{B^2\mu_p + cV_t\sigma_p^2}dt - fdt$$
$$= f\left(\frac{2}{A_1 m_t}\frac{d\xi_t}{\xi_t} - \frac{B^2\mu_p + 2cV_t\sigma_p^2}{B^2\mu_p + cV_t\sigma_p^2}dt\right). \tag{A24}$$

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

$$\frac{dS_t}{S_t} = \frac{fB^2\mu_p}{B^2\mu_p + cV_t\sigma_p^2}dt + \frac{2fB}{A_1 m_t}d\overline{W}_t. \tag{A25}$$

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

$$w_t^* = \frac{(A_1 m_t)^2\sigma_p^2}{4f\left(B^2\mu_p + cV_t\sigma_p^2\right)}. \tag{A26}$$

Then, substituting Equations (A22), (28), and (A26) into Equation (A13), we have the investor's optimal instantaneous Sharpe ratio at time t,

$$\frac{\left[\frac{(A_1 m_t)^2\sigma_p^2}{4f\left(B^2\mu_p + cV_t\sigma_p^2\right)} \times \frac{fB^2\mu_p}{B^2\mu_p + cV_t\sigma_p^2} + \mu_p\right]dt}{\sqrt{\left[\left(\frac{(A_1 m_t)^2\sigma_p^2}{4f\left(B^2\mu_p + cV_t\sigma_p^2\right)}\right)^2\left(\frac{2f}{A_1 m_t}\right)^2 B^2 + \sigma_p^2\right]dt}}$$
$$= \frac{\left[\frac{(A_1 m_t)^2\sigma_p^2 B^2\mu_p}{4\left(B^2\mu_p + cV_t\sigma_p^2\right)^2} + \mu_p\right]dt}{\sqrt{\left[\frac{(A_1 m_t)^2\sigma_p^4 B^2}{4\left(B^2\mu_p + cV_t\sigma_p^2\right)^2} + \sigma_p^2\right]dt}}. \tag{A27}$$

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

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

where X_t contains all the terms related to dV_t and $(dV_t)^2$. As we assume that V_t is exogenous to the market, thus independent of m_t at each time t, we assume $dV_t dm_t = 0$. 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:

$$\begin{aligned}\frac{dq_t^*}{q_t^*} &= \frac{2\sigma_m(\gamma_t)}{m_t B}\left(\frac{d\xi_t}{\xi_t}\right) + \frac{\sigma_m^2(\gamma_t)}{m_t^2 B^2}\left(\frac{d\xi_t}{\xi_t}\right)^2 \\ &+ \frac{2}{m_t}\left[(a_0 + a_1 m_t) - \frac{A_1\sigma_m(\gamma_t)m_t}{B}\right]dt + X_t.\end{aligned} \tag{A29}$$

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

$$\frac{dq_t^*}{q_t^*} = \frac{A_1\sigma_m(\gamma_t)}{fB}\left(\frac{dS_t}{S_t}\right) + \frac{A_1^2\sigma_m^2(\gamma_t)}{4f^2B^2}\left(\frac{dS_t}{S_t}\right)^2 + Y_t dt + X_t, \tag{A30}$$

where

$$\begin{aligned}Y_t &= \frac{2}{m_t}\left[(a_0 + a_1 m_t) - \frac{A_1\sigma_m(\gamma_t)m_t}{B}\right] + \frac{A_1\sigma_m(\gamma_t)\left(B^2\mu_p + 2cV_t\sigma_p^2\right)}{B\left(B^2\mu_p + cV_t\sigma_p^2\right)} \\ &= \frac{2a_0}{m_t} + 2a_1 - \frac{A_1\sigma_m(\gamma_t)B^2\mu_p}{B\left(B^2\mu_p + cV_t\sigma_p^2\right)}.\end{aligned} \tag{A31}$$

Here both X_t and Y_t are independent of either $d\xi_t/\xi_t$ or dS_t/S_t.

Q.E.D.

References

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.

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., 2019. 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.

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.

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.

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.

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.

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.

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

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

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. Simulation Results

Figure 1 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 13 to Month 240, and on the right-hand side, we illustrate the fund flow (vertical axis) and fund net alpha (horizontal axis) from Month 25 to Month 48 in blue circles, from Month 85 to Month 108 in green stars, and from Month 217 to Month 240 in red pluses.

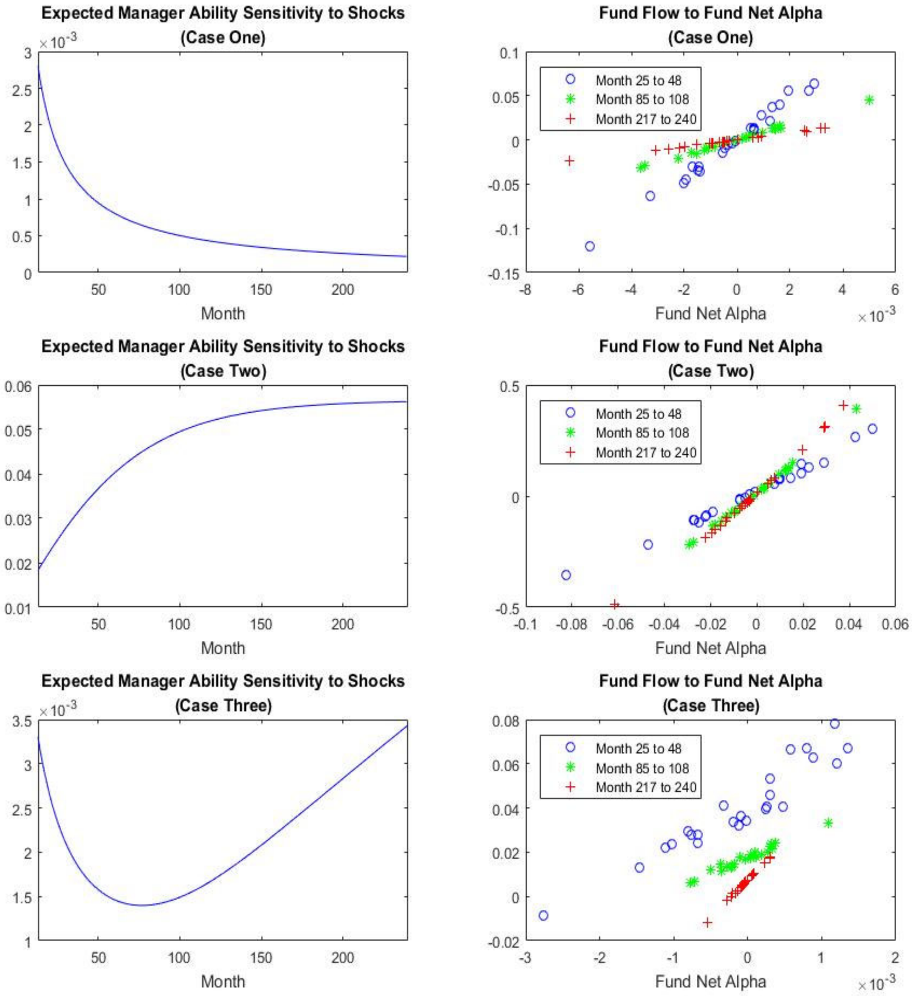

Figure 2 Flow–net alpha Sensitivity and Fund Age

Figure 2 illustrates the result of how flow–net alpha sensitivity changes with fund age. The flow–net alpha sensitivity is expressed by $\beta_1 + \sum_{j=1}^{6} c_j \left(Age_{i,t-1}\right)^j$, and we plot the values of flow–net alpha sensitivity (vertical axis) over the values of $Age_{i,t-1}$ (horizontal axis). The parameter values β_1 and c_1 to c_6 are from the estimated coefficient values of model specification (6) in Table 6.

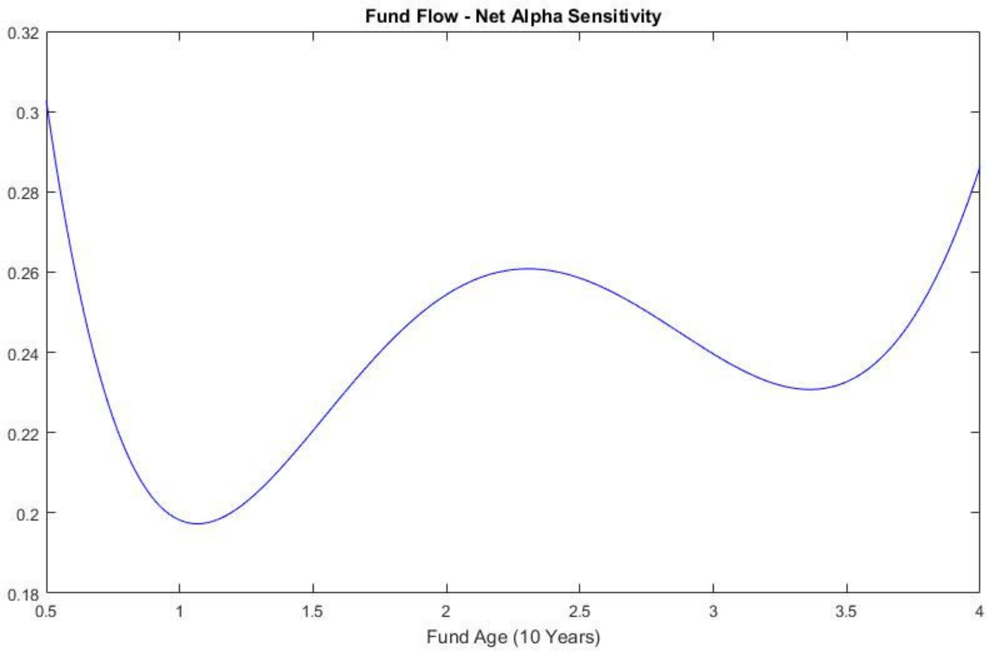

Figure 3. Examples of Fitting Observations of Fund Flows and Fund Net Alphas

Figure 3 illustrates two situations of fitting observations of fund flows and fund net alphas on a two-dimension space. In each of the 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.

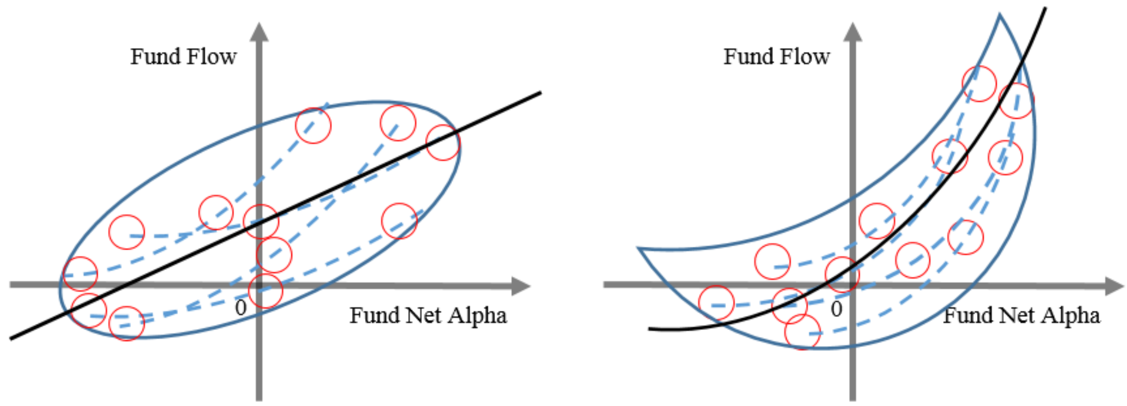

Table 1. Summary Statistics

Table 1 shows the summary statistics of our monthly observations from January 2000 to December 2019. *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 (46), 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 (46), 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. *Age* is the fund age, calculated as the number of 10 years since the inception of the oldest share class. *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. *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-only funds in the family, and it is a number.

				Percentile				
Variable	Observatior	Mean	Standard Deviation	1st	25th	50th	75th	99th
Fund Flow (Decimal), *Flow*	338764	0.0163	2.5522	-0.1535	-0.0174	-0.0066	0.0051	0.2269
Fund Net Return (Decimal), *Ret*	338764	0.0070	0.0543	-0.1296	-0.0174	0.0116	0.0357	0.1196
Fund Net Alpha (Decimal), *Alpha*	338764	-0.0004	0.0291	-0.0471	-0.0078	-0.0004	0.0071	0.0450
Style-Matching Model R-Squared (Decimal)	338764	0.8977	0.0909	0.5867	0.8681	0.9241	0.9570	0.9899
Fund Total Net Asset (in 1 Million Dollar), *TNA*	338764	1170.67	4505.53	0.60	42.70	185.10	742.80	17795.30
Fund Expense (Decimal), *Expense*	338764	0.0131	0.0054	0.0038	0.0095	0.0122	0.0161	0.0259
Fund Age (10 Years), *Age*	338764	1.7813	1.2750	0.1417	1.0083	1.5333	2.0917	7.3250
Fund Volatility (Decimal), *Vol*	338756	0.0448	0.0313	0.0153	0.0293	0.0406	0.0553	0.1117
Family Net Alpha (Decimal), *FamAlpha*	324944	-0.0004	0.1074	-0.0337	-0.0052	-0.0003	0.0043	0.0304
Family Size (Number), *FamSize*	324944	47.4851	39.1923	2	14	39	70	165

Table 2. Flow–Net Alpha Sensitivity and Convexity

Table 2 reports the results of the model in Equation (44). The dependent variable is *Flow*, the fund percentage flow. *Lag_Alpha* is the fund net alpha lagged by one month, $\alpha_{i,t-1}$. *D* is a dummy variable, which is one if *Lag_Alpha* is positive or zero, and is zero otherwise. *Expense* is the fund expense ratio as of the most recently completed fiscal year, including 12b-1 fees. *Lag_LnTNA* is the natural logarithm of the fund's total net assets under management lagged by one month. *Lag_Flow* is the *Flow* lagged by one month. *Lag_LnAge* is the natural logarithm fund age lagged by one month. *Vol* is the fund volatility. *Lag_FamAlpha* is the fund family's net alpha lagged by one month. *Lag_LnFamSize* is the natural logarithm fund family size lagged by one month. Standard errors are clustered by fund and presented in parentheses. 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)
Lag_Alpha	0.0498	0.3175***	0.2865***	0.2290***
	(0.0847)	(0.0597)	(0.0562)	(0.0562)
*Lag_Alpha*D*				0.1278***
				(0.0421)
Expense	-1.5423***	-1.6199***	-2.4671***	-2.4649***
	(0.1906)	(0.1971)	(0.4181)	(0.4192)
Lag_Flow	0.0007	0.0007	0.0005	0.0005
	(0.0006)	(0.0006)	(0.0004)	(0.0004)
Vol	-0.0123*	-0.0132	-0.0625***	-0.0772***
	(0.0071)	(0.0131)	(0.0241)	(0.0254)
Lag_LnAge	-0.0256***	-0.0260***	-0.0681***	-0.0681***
	(0.0011)	(0.0011)	(0.0035)	(0.0035)
Lag_LnTNA	-0.0022***	-0.0022***	-0.0054***	-0.0054***
	(0.0005)	(0.0005)	(0.0008)	(0.0008)
*Lag_Alpha*Lag_LnTNA*	0.0049	-0.0238**	-0.0214**	-0.0231**
	(0.0106)	(0.0099)	(0.0092)	(0.0095)
Lag_FamAlpha		0.0001	0.0004**	0.0004**
		(0.0002)	(0.0002)	(0.0002)
Lag_LnFamSize		-0.0013***	-0.0022**	-0.0022**
		(0.0003)	(0.0009)	(0.0009)
Constant	0.0401***	0.0458***	0.0597***	0.0588***
	(0.0050)	(0.0051)	(0.0087)	(0.0087)
Year Dummies	No	No	Yes	Yes
Fund Dummies	No	No	Yes	Yes
Observations	337,197	323,330	323,330	323,330
R-squared	0.0319	0.0327	0.0436	0.0436
Adjusted R-squared	0.0318	0.0327	0.0435	0.0435

Table 3. Time Pattern of Flow–Net Alpha Sensitivity

Table 3 reports the results of the model in Equation (47). The dependent variable is *Flow*, the fund percentage flow and *Lag_Alpha* is the fund net alpha lagged by one month, $\alpha_{i,t-1}$. We divide a fund's time-series observations into four parts: the first five years of observations in the sample (i.e., Period 0), the second five years (i.e., Period 1), the third five years (i.e., Period 2), and the observations in the remaining period (i.e., Period 3). *M1*, *M2*, and *M3* is one if the time is in Period 1, Period 2, and Period 3, respectively, and zero otherwise. *Expense* is the fund expense ratio as of the most recently completed fiscal year, including 12b-1 fees. *Lag_LnTNA* is the natural logarithm of the fund's total net assets under management lagged by one month. *Lag_Flow* is the *Flow* lagged by one month. *Lag_LnAge* is the natural logarithm fund age lagged by one month. *Vol* is the fund volatility. *Lag_FamAlpha* is the fund family's net alpha lagged by one month. *Lag_LnFamSize* is the natural logarithm fund family size lagged by one month. Standard errors are clustered by fund and presented in parentheses. The symbols ***, **, and * represent the 1%, 5%, and 10% significance level, respectively, in a two-tail *t*-test. Panel A reports the regression results. Panel B reports the results of two-tailed *t*-tests on the differences in coefficients.

Panel A	(1)	(2)	(3)
Lag_Alpha	0.2887***	0.3533***	0.3148***
	(0.0309)	(0.0693)	(0.0639)
*Lag_Alpha*M1*	-0.2076***	-0.0836***	-0.0764***
	(0.0181)	(0.0309)	(0.0289)
*Lag_Alpha*M2*	0.0050	-0.0008	0.0136
	(0.0326)	(0.0349)	(0.0320)
*Lag_Alpha*M3*	-0.0447	-0.0542*	-0.0454
	(0.0354)	(0.0309)	(0.0280)
M1	-0.0039***	-0.0029***	0.0129***
	(0.0005)	(0.0011)	(0.0020)
M2	-0.0005	0.0011	0.0108***
	(0.0006)	(0.0014)	(0.0027)
M3	-0.0000	0.0016	0.0011
	(0.0007)	(0.0018)	(0.0029)
Expense	-1.5246***	-1.6052***	-2.4256***
	(0.0397)	(0.1974)	(0.4134)
Lag_Flow	0.0007***	0.0007	0.0005
	(0.0001)	(0.0006)	(0.0004)
Vol	0.0017	0.0011	-0.0626***
	(0.0062)	(0.0142)	(0.0241)
Lag_LnAge	-0.0255***	-0.0266***	-0.0712***
	(0.0004)	(0.0016)	(0.0036)
Lag_LnTNA	-0.0022***	-0.0021***	-0.0055***
	(0.0001)	(0.0005)	(0.0008)
*Lag_Alpha*Lag_LnTNA*	-0.0141***	-0.0253**	-0.0227**
	(0.0050)	(0.0103)	(0.0095)
Lag_FamAlpha		0.0002	0.0004**
		(0.0002)	(0.0002)
Lag_LnFamSize		-0.0013***	-0.0021**
		(0.0003)	(0.0009)
Constant	0.0404***	0.0450***	0.0581***
	(0.0010)	(0.0052)	(0.0086)
Year Dummies	No	No	Yes
Fund Dummies	No	No	Yes
Observations	337,197	323,330	323,330
R-squared	0.0326	0.0330	0.0446
Adjusted R-squared	0.0325	0.0329	0.0445

Panel B	(1)	(2)	(3)
	*Lag_Alpha*M2 - Lag_Alpha*M1*		
Coefficient	0.2130***	0.0829***	0.0900***
Standard Error	0.0296	0.0291	0.0286
Student-t Statistics	51.5500	8.1330	9.9020
Two-Tailed P-Value	0.0000	0.0044	0.0017
	*Lag_Alpha*M3 - Lag_Alpha*M2*		
Coefficient	-0.0496	-0.0534*	-0.0590**
Standard Error	0.0427	0.0286	0.0286
Student-t Statistics	1.3540	3.4890	4.2520
Two-Tailed P-Value	0.2450	0.0620	0.0394
	*Lag_Alpha*M3 - Lag_Alpha*M1*		
Coefficient	0.1630***	0.0295	0.0310
Standard Error	0.0328	0.0256	0.0268
Student-t Statistics	24.6000	1.3230	1.3370
Two-Tailed P-Value	0.0000	0.2500	0.2480

Table 4. Time Pattern of Flow–Net Alpha Convexity

Table 4 reports the results of the model in Equation (48). The dependent variable is *Flow*, the fund percentage flow and *Lag_Alpha* is the fund net alpha lagged by one month, $\alpha_{i,t-1}$. *D* is a dummy variable, which is one if *Lag_Alpha* is positive or zero, and is zero otherwise. *M1*, *M2*, and *M3* is one if the time is in Period 1, Period 2, and Period 3, respectively, and zero otherwise. *Expense* is the fund expense ratio as of the most recently completed fiscal year, including 12b-1 fees. *Lag_LnTNA* is the natural logarithm of the fund's total net asset under management lagged by one month. *Lag_Flow* is the *Flow* lagged by one month. *Lag_LnAge* is the natural logarithm fund age lagged by one month. *Vol* is the fund volatility. *Lag_FamAlpha* is the fund family's net alpha lagged by one month. *Lag_LnFamSize* is the natural logarithm fund family size lagged by one month. Standard errors are clustered by fund and presented in parentheses. The symbols ***, **, and * represent the 1%, 5%, and 10% significance level, respectively, in a two-tail *t*-test. Panel A reports the regression results. Panel B reports the results of two-tailed *t*-tests on the differences in coefficients.

Panel A	(1)	(2)	(3)
Lag_Alpha	0.2320***	0.2846***	0.2548***
	(0.0374)	(0.0729)	(0.0653)
*Lag_Alpha*D*	0.1306***	0.1341*	0.1179*
	(0.0407)	(0.0734)	(0.0687)
*Lag_Alpha*M1*D*	-0.2363***	0.1167	0.0784
	(0.0549)	(0.0900)	(0.0887)
*Lag_Alpha*M2*D*	0.0022	-0.0595	-0.0248
	(0.0907)	(0.1076)	(0.1065)
*Lag_Alpha*M3*D*	-0.1013	-0.1536	-0.0737
	(0.0962)	(0.0937)	(0.0897)
*Lag_Alpha*M1*	-0.0376	-0.1309***	-0.1069**
	(0.0447)	(0.0430)	(0.0475)
*Lag_Alpha*M2*	0.0148	0.0389	0.0360
	(0.0545)	(0.0477)	(0.0475)
*Lag_Alpha*M3*	0.0143	0.0291	0.0004
	(0.0572)	(0.0445)	(0.0433)
M1	-0.0024***	-0.0033***	0.0126***
	(0.0006)	(0.0012)	(0.0020)
M2	-0.0001	0.0016	0.0109***
	(0.0008)	(0.0015)	(0.0028)
M3	0.0008	0.0025	0.0014
	(0.0009)	(0.0019)	(0.0030)
Expense	-1.5300***	-1.6155***	-2.4220***
	(0.0398)	(0.1977)	(0.4147)
Lag_Flow	0.0007***	0.0007	0.0005
	(0.0001)	(0.0006)	(0.0004)
Vol	-0.0004	-0.0122	-0.0773***
	(0.0063)	(0.0149)	(0.0260)
Lag_LnAge	-0.0255***	-0.0266***	-0.0713***
	(0.0004)	(0.0016)	(0.0036)
Lag_LnTNA	-0.0022***	-0.0022***	-0.0055***
	(0.0001)	(0.0005)	(0.0008)
*Lag_Alpha*Lag_LnTNA*	-0.0172***	-0.0266**	-0.0239**
	(0.0050)	(0.0105)	(0.0097)
Lag_FamAlpha		0.0002	0.0004**
		(0.0002)	(0.0002)
Lag_LnFamSize		-0.0013***	-0.0021**
		(0.0003)	(0.0009)
Constant	0.0397***	0.0448***	0.0573***
	(0.0011)	(0.0052)	(0.0086)
Year Dummies	No	No	Yes
Fund Dummies	No	No	Yes
Observations	337,197	323,330	323,330
R-squared	0.0326	0.0330	0.0446
Adjusted R-squared	0.0326	0.0330	0.0445

Panel B	(1)	(2)	(3)
	*Lag_Alpha*M2*D - Lag_Alpha*M1*D*		
Coefficient	0.2390***	-0.1760*	-0.1030
Standard Error	0.0894	0.1020	0.1060
Student-t Statistics	7.1210	2.9810	0.9530
Two-Tailed P-Value	0.0076	0.0844	0.3290
	*Lag_Alpha*M3*D - Lag_Alpha*M2*D*		
Coefficient	-0.1030	-0.0941	-0.0489
Standard Error	0.1190	0.0919	0.0938
Student-t Statistics	0.7520	1.0490	0.2710
Two-Tailed P-Value	0.3860	0.3060	0.6030
	*Lag_Alpha*M3*D - Lag_Alpha*M1*D*		
Coefficient	0.1350	-0.2700***	-0.1520*
Standard Error	0.0950	0.0829	0.0907
Student-t Statistics	2.0220	10.6300	2.8120
Two-Tailed P-Value	0.1550	0.0011	0.0938

Table 5. Flow–Net Alpha Sensitivity and Convexity: Results of Individual Funds

Table 5 reports the number of funds whose relevant coefficients or differences in coefficients in the model, in Equations (47) and (48), are significant; these numbers are in Panel A and Panel B, respectively. The models are
$Flow_{i,t} = \beta_0 + \beta_1\alpha_{i,t-1} + \beta_2\alpha_{i,t-1} * M1_{i,t} + \beta_3\alpha_{i,t-1} * M2_{i,t} + \beta_4\alpha_{i,t-1} * M3_{i,t} + \beta_5 M1_{i,t} + \beta_6 M2_{i,t} + \beta_7 M3_{i,t} + \gamma Controls_{i,t} + \varepsilon_{i,t}$, and
$Flow_{i,t} = \beta_0 + \beta_1\alpha_{i,t-1} + \beta_2\alpha_{i,t-1} * D_{i,t-1} + \beta_3\alpha_{i,t-1} * D_{i,t-1} * M1_{i,t} + \beta_4\alpha_{i,t-1} * D_{i,t-1} * M2_{i,t} + \beta_5\alpha_{i,t-1} * D_{i,t-1} * M3_{i,t} + \beta_6\alpha_{i,t-1} * M1_{i,t} + \beta_7\alpha_{i,t-1} * M2_{i,t} + \beta_8\alpha_{i,t-1} * M3_{i,t} + \beta_9 M1_{i,t} + \beta_{10} M2_{i,t} + \beta_{11} M3_{i,t} + \gamma Controls_{i,t} + \varepsilon_{i,t}$, respectively.
The total number of funds in these tests is 1029. The models are regressed on each fund, with all the control variables except 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 level, respectively, in a two-tail *t*-test. The last column shows the number of funds whose relevant coefficients or differences in coefficients are significant at least at 10% significance level in a two-tail *t*-test.

Panel A: The Model in Equation (47) Flow–Net Alpha Sensitivity

Significance	*	**	***	**Total**
$\beta_2 > 0$	34	29	16	79
$\beta_2 < 0$	53	54	32	139
$\beta_3 > 0$	34	45	29	108
$\beta_3 < 0$	43	55	30	128
$\beta_4 > 0$	23	53	28	104
$\beta_4 < 0$	41	56	38	135
$\beta_3 - \beta_2 > 0$	59	57	24	140
$\beta_3 - \beta_2 < 0$	30	40	20	90
$\beta_4 - \beta_3 > 0$	41	36	18	95
$\beta_4 - \beta_3 < 0$	48	54	15	117
$\beta_4 - \beta_2 > 0$	39	53	32	124
$\beta_4 - \beta_2 < 0$	29	48	39	116

Panel B: The Model in Equation (48) Flow–Net Alpha Convexity

Significance	*	**	***	**Total**
$\beta_3 > 0$	40	46	35	121
$\beta_3 < 0$	29	43	21	93
$\beta_4 > 0$	36	37	22	95
$\beta_4 < 0$	26	36	29	91
$\beta_5 > 0$	34	40	27	101
$\beta_5 < 0$	35	37	22	94
$\beta_4 - \beta_3 > 0$	31	44	19	94
$\beta_4 - \beta_3 < 0$	40	42	19	101
$\beta_5 - \beta_4 > 0$	51	28	18	97
$\beta_5 - \beta_4 < 0$	35	30	18	83
$\beta_5 - \beta_3 > 0$	50	38	28	116
$\beta_5 - \beta_3 < 0$	46	48	17	111

Table 6. Flow–Net Alpha Sensitivity and Fund Age

Table 6 reports the results of the model in Equation (49). The dependent variable is *Flow*, the fund percentage flow and *Lag_Alpha* is the fund net alpha lagged by one month, $\alpha_{i,t-1}$. *Lag_Age* is the fund age in 10 years lagged by one month. *Expense* is the fund expense ratio as of the most recently completed fiscal year, including 12b-1 fees. *Lag_LnTNA* is the natural logarithm of the fund's total net asset under management lagged by one month. *Lag_Flow* is the *Flow* lagged by one month. *Vol* is the fund volatility. *Lag_FamAlpha* is the fund family's net alpha lagged by one month. *Lag_LnFamSize* is the natural logarithm fund family size lagged by one month. Standard errors are clustered by fund and presented in parentheses. The symbols ***, **, and * represent the 1%, 5%, and 10% significance level, respectively, in a two-tail *t*-test.

Panel A	(1)	(2)	(3)	(4)	(5)	(6)
Lag_Alpha	0.3043***	0.3400***	0.4031***	0.5080***	0.6472***	0.7544***
	(0.0628)	(0.0778)	(0.1055)	(0.1524)	(0.2125)	(0.2685)
*Lag_Alpha*Lag_Age*	-0.0187**	-0.0678**	-0.1852**	-0.4547**	-0.9396**	-1.4646**
	(0.0084)	(0.0341)	(0.0938)	(0.2245)	(0.4392)	(0.7071)
*Lag_Alpha*Lag_Age*^{2}		0.0084*	0.0554*	0.2322**	0.6838**	1.3889**
		(0.0048)	(0.0293)	(0.1161)	(0.3165)	(0.6666)
*Lag_Alpha*Lag_Age*^{3}			-0.0046*	-0.0434**	-0.2064**	-0.5914**
			(0.0025)	(0.0217)	(0.0942)	(0.2827)
*Lag_Alpha*Lag_Age*^{4}				0.0026**	0.0268**	0.1225**
				(0.0013)	(0.0121)	(0.0588)
*Lag_Alpha*Lag_Age*^{5}					-0.0012**	-0.0121**
					(0.0006)	(0.0058)
*Lag_Alpha*Lag_Age*^{6}						0.0005**
						(0.0002)
Lag_Age	-0.0531***	-0.0650***	-0.1009***	-0.1658***	-0.2579***	-0.3949***
	(0.0048)	(0.0049)	(0.0057)	(0.0082)	(0.0133)	(0.0201)
Lag_Age^2		0.0031***	0.0184***	0.0599***	0.1454***	0.3141***
		(0.0003)	(0.0014)	(0.0039)	(0.0103)	(0.0202)
Lag_Age^3			-0.0014***	-0.0103***	-0.0413***	-0.1277***
			(0.0001)	(0.0008)	(0.0035)	(0.0093)
Lag_Age^4				0.0006***	0.0051***	0.0254***
				(0.0000)	(0.0005)	(0.0021)
Lag_Age^5					-0.0002***	-0.0024***
					(0.0000)	(0.0002)
Lag_Age^6						0.0001***
						(0.0000)
Expense	-2.8829***	-2.9761***	-3.0003***	-2.6938***	-2.5145***	-2.3891***
	(0.4735)	(0.4780)	(0.4757)	(0.4529)	(0.4260)	(0.3947)
Lag_Flow	0.0010	0.0010	0.0010	0.0010	0.0009	0.0009
	(0.0008)	(0.0008)	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Vol	-0.0738***	-0.0784***	-0.0643**	-0.0660**	-0.0731***	-0.0670***
	(0.0272)	(0.0268)	(0.0262)	(0.0259)	(0.0255)	(0.0250)
Lag_LnTNA	-0.0112***	-0.0107***	-0.0097***	-0.0086***	-0.0078***	-0.0072***
	(0.0010)	(0.0010)	(0.0009)	(0.0009)	(0.0009)	(0.0008)
*Lag_Alpha*Lag_LnTNA*	-0.0202**	-0.0186**	-0.0172*	-0.0159*	-0.0135*	-0.0122
	(0.0095)	(0.0092)	(0.0089)	(0.0085)	(0.0081)	(0.0080)
Lag_FamAlpha	0.0004*	0.0004*	0.0005**	0.0005***	0.0005**	0.0004**
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Lag_LnFamSize	-0.0000	-0.0003	-0.0011	-0.0017*	-0.0023**	-0.0025***
	(0.0011)	(0.0011)	(0.0010)	(0.0010)	(0.0010)	(0.0009)
Constant	0.1636***	0.1694***	0.1822***	0.2009***	0.2258***	0.2542***
	(0.0108)	(0.0109)	(0.0110)	(0.0110)	(0.0113)	(0.0116)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Fund Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	323,410	323,410	323,410	323,410	323,410	323,410
R-squared	0.0242	0.0255	0.0281	0.0315	0.0350	0.0389
Adjusted R-squared	0.0241	0.0254	0.0280	0.0314	0.0349	0.0388