Investor Learning and the Aggregate Allocation of Capital to Active Management*

Abstract

We estimate a model in which Bayesian investors learn not only about differential ability across funds but also about the nature of decreasing returns to scale (DRS)—how a fund's performance depends on its size versus the size of its competition— in real time and competitively allocate capital to funds, conditional on their current beliefs. We find that prior beliefs in the early 1990s feature fund-level DRS that are much steeper than indicated by their unbiased estimates, and that distorted beliefs about fund-level DRS alone account for about 5% of the variation in the aggregate allocation of capital to the mutual fund industry in the data. Aggregate allocation to active management can be explained without appealing to behavioral arguments.

*This paper is a substantially revised version of Chapter 1 of my PhD thesis at Princeton University. An earlier draft was circulated as "Why Has Active Asset Management Grown?" I am grateful for advice and encouragement from my thesis committee: Chris Sims (chair), Wei Xiong, and Motohiro Yogo. I have also benefited from comments by Jules van Binsbergen, John Cochrane, Diego Garcia, Valentin Haddad, Jakub Kastl, Stefan Nagel, Ľubos Pástor, Robert Stambaugh, Youchang Wu, Tao Zha, and seminar participants at AQR Capital Management, Atlanta Fed, Emory University, Hitotsubashi University, Princeton University, Richmond Fed, Rutgers University, and University of Alberta.

1 Introduction

Recently, there has been an increased interest in capital allocation within the market for mutual funds. For a long time, findings on capital allocation have generally been interpreted as irrational return chasing by naive investors: fund flows respond to past performance despite the fact that performance tends to be not persistent.[1] But there is growing evidence that the flow-performance relation can be consistent with rational response by learning investors.[2] When a mutual fund outperforms, Bayesian updating leads investors to learn that the fund will deliver positive average risk-adjusted returns (net alphas) at its current size. In turn, there will be flows into that fund, which will stop when the fund is no longer expected to deliver positive net alphas.[3] Key to this equilibration is *investors anticipating* decreasing returns to scale (DRS): as the fund grows, its trades are associated with a larger price impact, eroding its performance. Numerous studies provide ample panel evidence that confirms such diseconomies of scale.[4] While investors would learn the true degree of DRS in the limit as the sample size approaches infinity, their perceptions of scalability at the time of capital allocation decisions generally differ from the true scalability as it is unobservable. Therefore, to quantify the capital response driven by investor learning, we must understand the dynamics of their perceptions of not only skill but also scalability.

This paper investigates these issues by using the aggregate dynamics of capital allocation to the mutual fund industry to reveal the beliefs of investors and therefore reveal which alpha generating technology investors are using to process information. Specifically, we consider a model in which investors learn about and allocate capital to active funds, characterized by the two attributes mentioned above: (i) they have differential skills to generate alpha but (ii) they face decreasing returns to scale (DRS) in deploying these skills, both at the fund level

[1] See Chevalier and Ellison (1997) and Sirri and Tufano (1998), among others.

[2] See Berk and van Binsbergen (2017) for a review of recent advances in this strand of the mutual fund literature.

[3] Berk and van Binsbergen (2016) and Barber, Huang, and Odean (2016) conclude that mutual fund flows are most responsive to capital asset pricing model (CAPM) alpha, whereas Ben-David et al. (2022) argue that investors do not adjust fund performance using asset pricing models.

[4] See, e.g., Chen et al. (2004), Yan (2008), Pástor, Stambaugh, and Taylor (2015), and Zhu (2018).

and at the strategy level.[5] The Bayesian investors in our model learn about these parameters governing fund alphas in real time and competitively allocate capital to funds, conditional on their current beliefs. Consistent with the inherent difficulty in estimating returns to scale, we find that prior beliefs in the early 1990s feature fund-level DRS that are much steeper than indicated by their unbiased estimates. We also find that such beliefs persist for two reasons. First, our estimates indicate that the priors are relatively tighter for the effects of fund scale than for skill. Second, fund sizes are determined in equilibrium not only by fund-level DRS, but also by strategy-level DRS, both of which shrink fund sizes, making realized returns less informative about the effects of fund scale than about skill. Overall, distorted beliefs about fund-level DRS alone account for about 5% of the variation in the aggregate allocation of capital to the mutual fund industry in the data.

To derive the model implications for the evolution of the size of the mutual fund industry, we need to specify investors' prior beliefs, which we would update with the data to obtain, at any point in time, their posterior beliefs and in turn their equilibrium allocation. In our model, the size of a strategy segment is given by the ratio of the average incumbent fund's perceived skill to the perceived *total* strategy-level decreasing returns to scale the average fund faces, where the total strategy-level DRS is equal to the direct strategy-level DRS plus the product of fund-level DRS and the reciprocal of M, the number of incumbent funds. Intuitively, if the size of a strategy segment increases by \$1, so does the size of a fund in the strategy segment by $\$1/M$ on average. Thus, the decrease in the average fund's performance associated with such an increase in strategy size is effectively the weighted sum of the direct strategy-level DRS (with a weight of one) and fund-level DRS (with a weight of $1/M$).

We estimate the parameters governing investors' prior beliefs as of the beginning of 1991 by comparing the aggregate allocation of capital across different strategy segments of the

[5] In Appendix A, we validate the particular regression model that investors in our model believe funds' net alphas obey by empirically analyzing the nature of returns to scale for active mutual funds. Our evidence suggests that our proxies for fund size and for strategy size are reasonably accurate in measuring the limitations on a fund due to its size and due to the size of its competition, respectively.

mutual fund industry from the model against the data for each month over the next 25 years, during which the industry grew dramatically.[6] We do not use the first few years for comparing the model against the data to avoid the initial condition from having an unduly large impact on the estimation results.

Our estimates suggest that the investors' initial beliefs are strongly at odds with the true values of the parameters governing alphas. Specifically, investors believed at the beginning of 1991 in average levels of skills considerably higher than the average estimated fund fixed effects (which would lead to an overallocation) and in returns to scale substantially more negative than the estimates of the size-performance relation (which would lead to an underallocation). Overall, we find that the latter bias is dominant, which induces investors to initially underallocate to active mutual funds. Interestingly, while the priors are strongly at odds with the true values of the parameters, they are also considerably disperse, indicating substantial prior uncertainty about the key parameters.

Having actively learned about actively managed funds in their early years, investors as of March 1993 already held beliefs quite different from their beliefs at the beginning of 1991. First, expectations about skills are broadly in line with the data. Second, expectations about strategy-level decreasing returns to scale are no longer exaggerated; if anything, they are somewhat understated. Only the expectations about fund-level decreasing returns to scale continue to be exaggerated, although not as dramatic as under the priors. Importantly, examining the evolution of these expectations, we find that the investors' beliefs as of March 1993 persist for a long time, and their beliefs, even toward the end of our sample period, are quite similar.

Our results highlight that learning about the nature of returns to scale is generally slower than learning about skill. Investors persist in their beliefs that decreasing returns to scale are driven mainly by fund size. This bias, combined with the observed steady entry of new

[6] We use the nine categories corresponding to Morningstar's 3×3 stylebox to assign each fund its strategy segment. Funds in the same Morningstar category presumably follow similar investment strategies and thus compete against each other.

funds, explains why the mutual fund industry grew dramatically during our sample period. To see this, recall that the total strategy-level DRS the average fund faces is affected by the product of fund-level DRS and the reciprocal of the number of incumbent funds. Since there are few funds initially, investors initially exaggerate the total impact of strategy scale based on their perceptions of steep fund-level DRS. In addition, investors interpret the sustained fund entry as signaling a decline in the total impact of strategy scale much sharper than warranted because their beliefs about steep fund-level DRS persist. This decline in the perceived total impact of strategy scale, which coincides with a relatively stable perceived skill of the average incumbent fund (because it is broadly unbiased to begin with), translates into industry growth. Quantitatively, we find that around 37% of the variance of aggregate sizes (across strategy segments and over time) can be related to fluctuations in the perceived total impact of strategy scale, which is economically significant.

The key role played by steeper-than-warranted perceived fund-level decreasing returns in accounting for the aggregate allocation of capital over time raises a natural question: why does this bias in investors' beliefs endure? Bayesian learning implies the % change in investors' perception of each parameter upon observing fund returns is proportional to the product of investors' uncertainty about the parameter—its perceived variance over its perceived value—and returns' weight on the parameter.[7] Now, for any given fund, its returns have a weight of one on its skill, whereas its returns have weights equal to its size and the size of its competition on fund-level DRS and strategy-level DRS, respectively. We find that, in equilibrium, both fund size and strategy size are generally not large enough for fund returns to be as informative about parameters governing DRS as they are about fund skill. Heuristically, learning about the nature of returns to scale is generally slower than learning about managerial skill because investors first decompose performance into skill and the total effects of scale, which they then decompose to analyze the nature of returns to scale (fund-level vs strategy-level). Such learning process interacts with and amplifies the initial

[7] This statement is ignoring the off-diagonal elements of the covariance matrix of the key parameters (perceived by investors before observing fund alphas), but they tend to be small in magnitude.

beliefs that are relatively more biased about fund-level DRS than they are about skill levels or about strategy-level DRS, resulting in a very slow learning about fund-level decreasing returns.

We assess how well a purist learning model can reproduce the size dynamics in the data by first specifying investors' beliefs as of March 1993 (based on our estimates), which we update with actual data on observed returns. We then compute counterfactual fund sizes by invoking that the expected alpha on any fund receiving positive investment ought to be zero with respect to the counterfactual investors' beliefs; we compute the model-implied strategy size by adding up counterfactual fund sizes across all funds within a given strategy segment. In turn, we show that the aspects of learning in our model does surprisingly well in accounting for the aggregate size dynamics in the data—the R-squared from a panel regression of log actual aggregate sizes on log model-implied strategy sizes is about 0.96.

Our framework builds upon the influential models of Berk and Green (2004) and Pástor and Stambaugh (2012), amended to allow for learning about the nature of returns to scale and for the effect of entry and exit dynamics on aggregate fluctuations in capital allocation; Section 4.4.1 carefully relates our model to those of Berk and Green (2004) and Pástor and Stambaugh (2012). The main contribution of this paper is to show when such models, in which investors rationally learn from the history of observed returns, can quantitatively reproduce the observed dynamics of the aggregate allocation to mutual funds. Key to this result is that investors start with biased beliefs that fund-level decreasing returns to scale are much steeper than warranted by the data. This result can explain why the observed industry growth coincides with steady entry of new funds.

Another contribution of ours is to show that learning about fund-level DRS is especially slow. Investors persist in their beliefs about steeper-than-warranted fund-level DRS. This evidence is consistent with the inherent difficulty in estimating returns to scale. We estimate later in the paper that a $100 million increase in fund size (which is about 10% of the interquartile range) depresses performance by less than 0.01% per month, or 12 bps per year.

While this effect is clearly economically significant, it would be swamped by the performance variation due to differences in skill, which we estimate to be roughly distributed with mean 0.5% per month and standard deviation 0.5% per month. This, in turn, is swamped by portfolio volatility of around 2% per month. Against these fluctuations, a small decrease in performance due to fund-level DRS would be hard to detect, resulting in very slow undoing of investors' prior bias about the nature of returns to scale.

Nevertheless, we find that, in the new millennium, the aggregate allocation to mutual funds is broadly rational. The number of funds competing for investors grows large enough to ensure that investors' persistent bias about fund-level DRS has little influence on their perceptions of the total impact of strategy scale, which in turn are broadly consistent with the true values since 2001. This, coupled with our result discussed earlier that perceptions about the average fund's skill are broadly in line with the data, accentuates that the aggregate misallocation of capital to the mutual fund industry is small over the last two decades.

Our study relates to a large literature on learning in mutual funds.[8] Much of this literature focuses on learning about managerial skill. Baks, Metrick, and Wachter (2001), Pástor and Stambaugh (2002), and Avramov and Wermers (2006) accentuate the benefits of investing in active mutual funds from the Bayesian perspective of investors. Lynch and Musto (2003) explain that the observed convex flow-performance relationship[9] is consistent with investor learning, because bad past performance is less informative about future performance than good past performance, whereas Berk and Green (2004) show that the flow-performance relationship can be consistent with investor learning, even if past performance is not informative about future performance,[10] in the presence of fund-level decreasing returns to scale.[11]

[8] See Pastor and Veronesi (2009) for a review of the extensive literature that relies on parameter uncertainty and learning to explain empirical puzzles in mutual funds and beyond.

[9] See Chevalier and Ellison (1997) and Sirri and Tufano (1998), among others.

[10] See Malkiel (1995), Gruber (1996), Carhart (1997), and Fama and French (2010), among others, for evidence that performance is largely unpredictable. Some evidence of persistence in performance does exist at shorter horizons; see, for example, Bollen and Busse (2005) and Mamaysky, Spiegel, and Zhang (2008).

[11] Other studies that model learning in mutual funds with a focus on the determinants of flow-performance relationship include Huang, Wei, and Yan (2007) (investors' participation costs), Franzoni and Schmalz (2017) (market states), and Binsbergen, Kim, and Kim (2021) (fund-level decreasing returns to scale).

Pástor and Stambaugh (2012) explain that the popularity of active funds can persist even if they underperform, since learning about industry-level decreasing returns to scale is slow. Brown and Wu (2016) find that flows to a fund respond positively to the performance of the fund's family, which they show is consistent with cross-fund learning within groups. None of these studies consider learning about the nature of returns to scale, nor do they provide a formal account of whether a model based solely on Bayesian learning can quantitatively capture the historical fluctuations in capital allocation to the mutual fund industry.

On the other hand, there is evidence that investors do not react to new information as Bayesians. Berk and van Binsbergen (2016) and Barber, Huang, and Odean (2016) provide empirical evidence that investors do not respond immediately, as well as exponentially weighting past returns over long horizons when allocating capital across funds. Choi, Kahraman, and Mukherjee (2016) find that investors move capital in the right direction, but not sufficiently so, relative to a Bayesian benchmark; Roussanov, Ruan, and Wei (2022) find that investors over-weight recent performance and under-weight distant performance information more heavily than a Bayesian investor would. Roussanov, Ruan, and Wei (2021) document that marketing plays an important role in determining fund size, indicating that there exist substantial frictions in the market for mutual funds. We do not mean to suggest that such alternative explanations play no role in capital allocation to mutual funds. Rather, we simply highlight the joint-hypothesis problem: we can test whether investors learn as Bayesians only if we take a stand on what they learn about.

More broadly, this study adds to a growing literature addressing the evolution of the equilibrium size of the financial industry (e.g., Philippon 2015).[12] Our model, like most of those in the literature on learning in mutual funds, is partial equilibrium in the sense that we do not model the market for assets. Another strand of the mutual fund literature studies the asset pricing implications of asset management by abstracting from learning-related issues.

[12] In addition to Philippon (2015), recent examples of studies that empirically analyze various aspects of the size of the financial sector include Khorana, Servaes, and Tufano (2005), Greenwood and Scharfstein (2013), Cochrane (2013), and Philippon and Reshef (2013).

For example, Stambaugh (2020) analyze the amount of mispricing and the amount of active management under general equilibrium with many assets and costly trading. In his model, when managers become more skilled, they more accurately identify profit opportunities, but then active management in aggregate reduces mispricing, shrinking the profits those opportunities offer.[13]

2 The Model

The model has two types of agents: fund managers and investors. There are M_t active fund managers at time t who have the potential ability to identify and exploit opportunities to outperform passive benchmarks. There are S categories, which classify funds based on their investment styles. In other words, we assume that M_{st} active funds at time t are following investment strategy s, with $M_t = \sum_{s=1}^{S} M_{st}$. We abuse notation by writing $i \in M_{st}$ if fund i is following investment strategy s. We use the nine categories corresponding to Morningstar's 3×3 stylebox (large growth, mid-cap blend, etc.) for our empirical work. There is competitive provision of capital by investors to these active funds.

We model the rates of returns earned by investors in the funds belonging to category s at time $t-1$ as

$$\boldsymbol{R}_{st}^{n} = \boldsymbol{\alpha}_{st} + \boldsymbol{R}_{st}^{B} + \boldsymbol{u}_{st} \tag{1}$$

where $\boldsymbol{R}_{st}^{n}$ is the $M_{st-1} \times 1$ vector of fund returns in excess of the riskless rate, $\boldsymbol{\alpha}_{st}$ is the $M_{st-1} \times 1$ vector of fund alphas, $\boldsymbol{R}_{st}^{B}$ is the $M_{st-1} \times 1$ vector of excess returns on the passive benchmarks, and $\boldsymbol{u}_{st}$ is the $M_{st-1} \times 1$ vector of the residuals. The elements of the residual vector $\boldsymbol{u}_{st}$ have the following factor structure:

$$u_{ist} = x_{st} + \epsilon_{ist} \tag{2}$$

[13] Additional examples of recent studies in this strand of the literature include García and Vanden (2009), Stambaugh (2014), Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016), Gârleanu and Pedersen (2018), and Buffa, Vayanos, and Woolley (2021), among many others.

for $i \in M_{st-1}$, where all ϵ_{ist}'s have a mean of zero, a variance of $\sigma^2_{\epsilon,s}$, and zero correlation across funds and periods. For each category s the common factor x_{st} has mean zero and variance $\sigma^2_{x,s}$. We assume that these common factors are independently distributed across categories. The values of $\sigma_{x,s}$ and $\sigma_{\epsilon,s}$ for all categories s are constants known to investors.

The factor structure in equation (2) means that the benchmark-adjusted returns of funds in the same category are correlated as long as $\sigma_{x,s} > 0$, but those of funds from different categories are not correlated.[14] These assumptions are consistent with empirical evidence for active equity mutual funds. The average pairwise correlation in residuals between funds belonging to the same Morningstar Category is 0.31, but the average correlation between funds from different categories is only 0.07. As a result, investors update their beliefs about any given fund using not only its return history but also the histories of other funds in the same category.

Our key assumption is that funds face decreasing returns to scale both at the fund level and at the strategy level. That is, we assume that the expected benchmark-adjusted return received by investors in fund i belonging to category s at time t is given by:

$$\alpha_{ist+1} = a_{is} - b_s q_{ist} - c_s Q_{st}, \tag{3}$$

where q_{ist} is the size of the fund, with $Q_{st} = \sum_{i \in M_{st}} q_{ist}$ (i.e., the size of the fund's category measures the size of its competition).[15] The parameter a_{is} can be interpreted as the alpha on the first cent invested in active management through the fund, $b_s > 0$ is a parameter that captures the decreasing returns to fund scale the fund faces (which can vary by category),[16]

[14] Intuitively, funds in the same sector, which presumably follow similar investment strategies, are likely to identify the same opportunities to outperform, resulting in correlated performance.

[15] This assumption is consistent with empirical evidence regarding the nature of returns to scale. We empirically analyze the nature of returns to scale in active mutual fund management following Zhu (2018) in Appendix A. We find strong evidence of decreasing returns both at the fund level and at the strategy level. These findings are unaffected by the addition of industry size, as defined by Pástor, Stambaugh, and Taylor (2015) (i.e., *IndustrySize*), which itself is insignificant throughout. This result suggests that category size does a good job of capturing the adverse effect of aggregate scale on fund performance. In short, the relation (3) is a reasonable way of modeling returns to scale both at the fund level and at an aggregate level.

[16] See, for example, Chen et al. (2004), Yan (2008), Pollet and Wilson (2008), and Zhu (2018).

and $c_s > 0$ is a parameter that captures the decreasing returns to strategy scale the fund faces (which can vary by category).[17] The parameters a_{is} and b_s, c_s in equation (3) are unknown to investors. We denote the first and second moments of those parameters formed with respect to the *subjective* model, perceived by investors at time t, by

$$\widehat{E}\left(\left[\begin{array}{c} a_{is} \\ b_s \\ c_s \end{array}\right] \middle| I_t\right) = \left[\begin{array}{c} \widehat{a}_{ist} \\ \widehat{b}_{st} \\ \widehat{c}_{st} \end{array}\right], \tag{4}$$

$$\widehat{Var}\left(\left[\begin{array}{c} a_{is} \\ b_s \\ c_s \end{array}\right] \middle| I_t\right) = \left[\begin{array}{ccc} \widehat{\sigma}^2_{a,ist} & \widehat{\sigma}_{ab,ist} & \widehat{\sigma}_{ac,ist} \\ \widehat{\sigma}_{ab,ist} & \widehat{\sigma}^2_{b,st} & \widehat{\sigma}_{bc,st} \\ \widehat{\sigma}_{ac,ist} & \widehat{\sigma}_{bc,st} & \widehat{\sigma}^2_{c,st} \end{array}\right], \tag{5}$$

where I_t denotes the time t information set available to investors. These parameters are not the only unknowns to the investors: abilities of any other active funds in the same category, whether they are in business or not, are also unknown. So our investors learn about the skill parameters for all funds within a given category, including those that have yet to enter, as well as the parameters governing returns to scale for that category by observing the histories of all funds belonging to that category, as explained later in Section 2.1.

As in Berk and Green (2004), we assume that all funds must have net alphas of zero based on investors' perceptions of the parameters governing them in real time. This can be justified as long as investors competitively supply (remove) capital to (from) funds with positive (negative) excess expected returns from their subjective perspective, implying that investors perceive all funds earn an expected return commensurate with the risk of the fund. At each point in time, then, capital flows into and out of each fund so that the net alpha on

[17] See Wahal and Wang (2011), Pástor, Stambaugh, and Taylor (2015), and Hoberg, Kumar, and Prabhala (2018) for evidence that a fund's ability to outperform declines as the size of the fund's competition increases, which in turn motivates aggregate-level decreasing returns to scale as modeled here.

any surviving fund is zero, as perceived by investors:

$$\widehat{E}_t\left(\alpha_{ist+1}\right) = 0. \tag{6}$$

Condition (6) clearly implies that investors do not perceive any predictability in fund alphas and that the perceived alphas of all funds will be zero, regardless of their skill levels. The circumflex on the expectation operator indicates that the expectation is formed with respect to the investors' perceived model (as opposed to the true model, analyzed by an econometrician using the full sample).

Taking expectations of both sides of (3), and requiring net alphas of zero as in (6), gives

$$\widehat{a}_{ist} = \widehat{b}_{st} q_{ist} + \widehat{c}_{st} Q_{st} \tag{7}$$

for $i \in M_{st}$. Let $\boldsymbol{q}_{st}$ denote the $M_{st} \times 1$ vector of sizes of all funds belonging to category s at time t. As $\{\widehat{a}_{ist}\}$ and $\widehat{b}_{st}, \widehat{c}_{st}$ change, $\boldsymbol{q}_{st}$ changes to ensure that this equation is satisfied for all of these funds. The following proposition gives the equilibrium values of the key quantities, Q_{st} and $\boldsymbol{q}_{st}$, in terms of investors' posterior expectations, $\{\widehat{a}_{ist}\}$ and $\widehat{b}_{st}, \widehat{c}_{st}$.

Proposition 1 *In equilibrium, for each category s at any point in time,*

$$Q_{st} = \frac{\widehat{\overline{a}}_{st}}{\widehat{b}_{st}/M_{st} + \widehat{c}_{st}}, \tag{8}$$

where $\widehat{\overline{a}}_{st}$ is the average perceived skill of all incumbent funds within that category at time t, and in turn,

$$\boldsymbol{q}_{st} = \frac{\widehat{\boldsymbol{a}}_{st} - \gamma_{st}\widehat{\overline{a}}_{st}}{\widehat{b}_{st}}, \tag{9}$$

where $\gamma_{st} = \widehat{c}_{st} / \left(\widehat{b}_{st}/M_{st} + \widehat{c}_{st}\right)$.

Proof. *Aggregating equation (7) across all incumbent funds in a given category s at time t,* *and using the fact that the category size is given by* $Q_{st} = \sum_{i \in M_{st}} q_{ist}$*, gives*

$$\sum_{i \in M_{st}} \widehat{a}_{ist} = \widehat{b}_{st} Q_{st} + M_{st} \widehat{c}_{st} Q_{st} = M_{st} \left(\widehat{b}_{st}/M_{st} + \widehat{c}_{st} \right) Q_{st}. \tag{10}$$

Rewriting (10) now gives (8). Equation (7) can then be rewritten as

$$q_{ist} = \frac{\widehat{a}_{ist} - \widehat{c}_{st} Q_{st}}{\widehat{b}_{st}} = \frac{\widehat{a}_{ist} - \gamma_{st} \widehat{\overline{a}}_{st}}{\widehat{b}_{st}} \tag{11}$$

where the resulting expression in terms of perceived skills in (11) relies on (8). ■

Intuitively, the size of a category is given by the ratio of the average incumbent fund's perceived skill to the *total* decreasing returns to strategy scale the average fund is perceived to face. If the size of the category increases by \$1, so does the size of a typical fund belonging to that category by $\$1/M_{st}$. Such an increase in category size is associated with a total decrease in fund performance of c_s (directly from strategy-level decreasing returns to scale) plus b_s/M_{st} (indirectly through fund-level decreasing returns to scale).

Similarly, the size of a fund is given by the ratio of its perceived skill adjusted for strategy-level DRS to the perceived decreasing returns to its scale. The adjustment for strategy-level DRS amounts to a negative exposure, $-\gamma_{st}$, to the average skill of the fund's competitors. Note that condition (6) implies investors' belief that $b_s > 0$ is a necessary condition for the coexistence of funds with heterogeneous perceived skills. For $\widehat{b}_{st} \approx 0$, γ_{st} is close to one, which in turn would force the below-average funds out of business: Only the highest-skill funds can remains in business. This foreshadows that the subjective fund-level decreasing returns, as perceived by investors, play an important role in enabling the model to produce empirically realistic distributions of fund size, as shown later in Section 4.4.2.

2.1 Prior and Posterior Beliefs

In this subsection, we discuss the prior and posterior beliefs about the key parameters, $\{a_{is}\}$ and b_s, c_s, as well as the likelihood of a history of aggregate sizes (i.e., $\{Q_{st}\}_{t=1}^{T}$) implied by those parameters.

2.1.1 Prior Beliefs

We focus on prior beliefs in which the key parameters from different categories are not correlated. Technically, combined with (1)–(3), this assumption implies posterior beliefs in which the key parameters from different categories continue to be uncorrelated. The equilibrium allocations in each category can then be computed based only on the historical data for that category by Proposition 1. That is, we can estimate investors' prior beliefs category by category. Economically, such prior beliefs already allow for complex learning that would plausibly encompass the true learning problem investors face.

At the beginning of the sample, the investors' prior about the ability of any given fund within category s is that a_{is} is normally distributed with mean $\widehat{a}_{s0}$ and variance $\widehat{\sigma}^2_{a,s0}$. We allow investors to believe a priori that these skill parameters are correlated. The prior correlation in skills between funds within category s will be denoted $\widehat{\rho}_{s0}$. If $\widehat{\rho}_{s0} = 0$, then investors believe that a_{is} is idiosyncratic to the fund, just as in Berk and Green (2004), whereas if $\widehat{\rho}_{s0} = 1$, investors believe that all funds within a given category are created equal ($a_{is} = a_s$, where a_s is unknown), just as in Pástor and Stambaugh (2012). Our focus is on the case in which $\widehat{\rho}_{s0} \in (0, 1)$—investors believe that a fund's skill is a combination of a fund-specific component and a common component shared by all funds in the same category.[18] Note that the returns of funds in operation right now are informative about the skills of funds that have not yet entered as long as $\sigma_{x,s} > 0$. The learning problem faced by investors is then essentially equivalent to learning about the skills not only of the incumbent funds

[18] An example of a study analyzing the effect of cross-fund learning on fund flows is Brown and Wu (2016), although their source of cross-fund learning is common skill shared by funds in the same family.

$i \in M_{st}$, but also of a hypothetical fund (labeled "∞") that represents a potential entrant in the respective fund category, in addition to learning about the nature of returns to scale that we discuss in the next paragraph.

A common assumption in the literature is that the effects of scale are known to investors, except for Pástor and Stambaugh (2012) who consider learning about industry-level decreasing returns to scale. Pástor and Stambaugh's treatment of DRS corresponds to a setting in which $b_s = 0$ and c_s is an unknown positive value. While we consider such prior beliefs as well, our focus is on prior beliefs in which returns are decreasing in both fund scale and strategy scale at uncertain rates (i.e., both b_s and c_s are unknown positive values). We show in Section 4.4.2 that investors who believe a priori that $b_s > 0$ make very different allocation decisions than investors who believe that $b_s = 0$, even after observing exactly the same data.

To summarize, we specify a multivariate normal joint prior distribution for $\{a_{is}\}_{i \in M_{s0} \cup \{\infty\}}$ and b_s, c_s:

$$\begin{bmatrix} \{a_{is}\}_{i \in M_{s0} \cup \{\infty\}} \\ b_s \\ c_s \end{bmatrix} \sim N\left(\widehat{E}_{s0}, \widehat{V}_{s0}\right), \tag{12}$$

where $N\left(\widehat{E}_{s0}, \widehat{V}_{s0}\right)$ denotes a multivariate normal distribution with mean $\widehat{E}_{s0}$ and covariance matrix $\widehat{V}_{s0}$. Denote

$$\widehat{E}_{s0} = \begin{bmatrix} \widehat{\boldsymbol{a}}_{s0} \\ \widehat{a}_{\infty s0} \\ \widehat{b}_{s0} \\ \widehat{c}_{s0} \end{bmatrix} = \begin{bmatrix} \widehat{a}_{s0}\iota_{M_{s0}} + \boldsymbol{w}_{s0} \\ \widehat{a}_{s0} \\ \widehat{b}_{s0} \\ \widehat{c}_{s0} \end{bmatrix}, \tag{13}$$

$$\widehat{V}_{s0} = \begin{bmatrix} \widehat{\sigma}^2_{a,s0}\left(\widehat{\rho}_{s0}\iota_{M_{s0}}\iota'_{M_{s0}} + (1-\widehat{\rho}_{s0}) I_{M_{s0}}\right) & \widehat{\rho}_{s0}\widehat{\sigma}^2_{a,s0}\iota_{M_{s0}} & \widehat{\sigma}_{ab,s0}\iota_{M_{s0}} & \widehat{\sigma}_{ac,s0}\iota_{M_{s0}} \\ \widehat{\rho}_{s0}\widehat{\sigma}^2_{a,s0}\iota'_{M_{s0}} & \widehat{\sigma}^2_{a,s0} & \widehat{\sigma}_{ab,s0} & \widehat{\sigma}_{ac,s0} \\ \widehat{\sigma}_{ab,s0}\iota'_{M_{s0}} & \widehat{\sigma}_{ab,s0} & \widehat{\sigma}^2_{b,s0} & \widehat{\sigma}_{bc,s0} \\ \widehat{\sigma}_{ac,s0}\iota'_{M_{s0}} & \widehat{\sigma}_{ac,s0} & \widehat{\sigma}_{bc,s0} & \widehat{\sigma}^2_{c,s0} \end{bmatrix}, \tag{14}$$

where $\boldsymbol{w}_{s0}$ is the $M_{s0} \times 1$ vector of "belief shocks" (see below); $\iota_{M_{s0}}$ and $I_{M_{s0}}$ denote an M_{s0}-vector of ones and an $M_{j0} \times M_{j0}$ identity matrix, respectively. We specify $\widehat{\sigma}_{ab,s0} = \widehat{\sigma}_{ac,s0} = 0$ and $\widehat{\sigma}_{bc,s0} = 0$ for simplicity. We consider a wide range of prior beliefs, subject to the constraint that $\widehat{a}_{s0}, \widehat{b}_{s0}, \widehat{c}_{s0} > 0$ and $\widehat{\rho}_{s0} > 0$. Such beliefs reflect the notion that active funds do have the skill to generate alpha but they face decreasing returns, both at the fund level and at the strategy level, in deploying these skills and that funds in the same category share a common unobservable skill component.

2.1.2 Posterior Beliefs

To update their beliefs about $\{a_{is}\}_{i\in M_{st-1}\cup\{\infty\}}$ and b_s, c_s, investors conduct inference about these parameters based on the cross-sectional data on funds' returns and size. At the beginning of time t, the newly available data relevant for learning about the parameters governing fund alphas in category s consist of $\{\alpha_{ist}, q_{ist-1}\}_{i\in M_{st-1}}$ (from which the category size Q_{st-1} can be inferred). Investors' beliefs at time $t-1$ for $\{a_{is}\}_{i\in M_{st-1}\cup\{\infty\}}$ and b_s, c_s are given by a multivariate normal distribution, whose moments are $\widehat{E}_{st-1}$ and $\widehat{V}_{st-1}$. Those moments are updated by using standard results from Bayesian statistics on the multivariate normal distribution

$$\widetilde{V}_{st-1|t} = \widehat{V}_{st-1} - \widehat{V}_{st-1} Z'_{st-1} \widehat{Var}_{t-1} (\boldsymbol{\alpha}_{st})^{-1} Z_{st-1} \widehat{V}_{st-1}, \tag{15}$$

$$\widetilde{E}_{st-1|t} = \widehat{E}_{st-1} + \widehat{V}_{st-1} Z'_{st-1} \widehat{Var}_{t-1} (\boldsymbol{\alpha}_{st})^{-1} \boldsymbol{\alpha}_{st}, \tag{16}$$

where

$$\begin{aligned} Z_{st-1} &= \begin{bmatrix} I_{M_{st-1}} & \mathbf{0}_{M_{st-1}} & -\boldsymbol{q}_{st-1} & -Q_{st-1}\iota_{M_{st-1}} \end{bmatrix}, \\ \widehat{Var}_{t-1} (\boldsymbol{\alpha}_{st})^{-1} &= Z'_{st-1} \widehat{V}_{st-1} Z'_{st-1} + \sigma^2_{x,s} \iota_{M_{st-1}} \iota'_{M_{st-1}} + \sigma^2_{\epsilon,s} I_{M_{st-1}}, \end{aligned}$$

and $\mathbf{0}_{M_{st-1}}$ denotes an M_{st-1}-vector of zeros. Note that

$$\widetilde{E}_{st-1|t} = \begin{bmatrix} \widetilde{\boldsymbol{a}}_{st-1|t} \\ \widetilde{a}_{\infty st} \\ \widetilde{b}_{st} \\ \widetilde{c}_{st} \end{bmatrix}, \widetilde{V}_{st-1|t} = \begin{bmatrix} \widetilde{\Sigma}_{aa,st-1|t} & \widetilde{\Sigma}_{a\infty,st-1|t} & \widetilde{\Sigma}_{ab,st-1|t} & \widetilde{\Sigma}_{ac,st-1|t} \\ \widetilde{\Sigma}'_{a\infty,st-1|t} & \widetilde{\sigma}^2_{a,\infty st} & \widetilde{\sigma}_{ab,\infty st} & \widetilde{\sigma}_{ac,\infty st} \\ \widetilde{\Sigma}'_{ab,st-1|t} & \widetilde{\sigma}_{ab,\infty st} & \widetilde{\sigma}^2_{b,st} & \widetilde{\sigma}_{bc,st} \\ \widetilde{\Sigma}'_{ac,st-1|t} & \widetilde{\sigma}_{ac,\infty st} & \widetilde{\sigma}_{bc,st} & \widetilde{\sigma}^2_{c,st} \end{bmatrix}. \tag{17}$$

Having the updated moments $\widetilde{E}_{st-1|t}$ and $\widetilde{V}_{st-1|t}$ of the key parameters at time $t-1$ ($\{a_{is}\}_{i \in M_{st-1} \cup \{\infty\}}$ and b_s, c_s), we obtain the updated moments of the key parameters at time t ($\{a_{is}\}_{i \in M_{st} \cup \{\infty\}}$ and b_s, c_s) in two steps. First, we drop from $\widetilde{E}_{st-1|t}$ and $\widetilde{V}_{st-1|t}$ their elements that represent moments related to a_{is} for $i \in M_{st-1} \setminus M_{st}$ (i.e., exiting funds). Second, we add the moments related to a_{is} for $i \in M_{st} \setminus M_{st-1}$ (i.e., new funds), which we set equal to the corresponding moments related to $a_{\infty s}$. Taking into account the entry and exit of funds leads to the following moments

$$\widetilde{E}_{st|t} = \begin{bmatrix} \widetilde{\boldsymbol{a}}_{st|t} \\ \widetilde{a}_{\infty st} \\ \widetilde{b}_{st} \\ \widetilde{c}_{st} \end{bmatrix}, \widetilde{V}_{st|t} = \begin{bmatrix} \widetilde{\Sigma}_{aa,st|t} & \widetilde{\Sigma}_{a\infty,st|t} & \widetilde{\Sigma}_{ab,st|t} & \widetilde{\Sigma}_{ac,st|t} \\ \widetilde{\Sigma}'_{a\infty,st|t} & \widetilde{\sigma}^2_{a,\infty st} & \widetilde{\sigma}_{ab,\infty st} & \widetilde{\sigma}_{ac,\infty st} \\ \widetilde{\Sigma}'_{ab,st|t} & \widetilde{\sigma}_{ab,\infty st} & \widetilde{\sigma}^2_{b,st} & \widetilde{\sigma}_{bc,st} \\ \widetilde{\Sigma}'_{ac,st|t} & \widetilde{\sigma}_{ac,\infty st} & \widetilde{\sigma}_{bc,st} & \widetilde{\sigma}^2_{c,st} \end{bmatrix}, \tag{18}$$

which are the same as (17) but with $\widetilde{\boldsymbol{a}}_{st-1|t}$ and $\widetilde{\Sigma}_{aa,st-1|t}\,, \widetilde{\Sigma}_{a\infty,st-1|t}\,, \widetilde{\Sigma}_{ab,st-1|t}\,, \widetilde{\Sigma}_{ac,st-1|t}$ replaced by $\widetilde{\boldsymbol{a}}_{st|t}$ and $\widetilde{\Sigma}_{aa,st|t}\,, \widetilde{\Sigma}_{a\infty,st|t}\,, \widetilde{\Sigma}_{ab,st|t}\,, \widetilde{\Sigma}_{ac,st|t}$, respectively.

Finally, the posterior distribution of $\{a_{is}\}_{i\in M_{st}\cup\{\infty\}}$ and b_s, c_s is multivariate normal as in equation (12), except that $\widehat{E}_{s0}$ and $\widehat{V}_{s0}$ are replaced by $\widehat{E}_{st}$ and $\widehat{V}_{st}$, where

$$\widehat{E}_{st} = \widetilde{E}_{st|t} + \begin{bmatrix} \boldsymbol{w}_{st} \\ \mathbf{0}_3 \end{bmatrix}, \widehat{V}_{st} = \widetilde{V}_{st|t}\,, \tag{19}$$

and $\boldsymbol{w}_{st}$ is the $M_{st} \times 1$ vector of belief shocks. Equation (19) means that Bayesian learning determines the posteriors, up to shocks to expected skills of funds operating at time t. Combining equations (7) and (19) gives the shock to fund i's perceived skill at time t as

$$w_{ist} = \widetilde{b}_{st} q_{ist} + \widetilde{c}_{st} Q_{st} - \widetilde{a}_{ist} = \widehat{b}_{st} q_{ist} + \widehat{c}_{st} Q_{st} - \widetilde{a}_{ist}.$$

In other words, w_{ist} are wedges in the equilibrium condition (7) computed by using (18), i.e., investor beliefs based solely on Bayesian updating.

In any purely learning-based model of capital allocation, note that, for a given specification of priors about $\{a_{is}\}$ and b_s, c_s, Q_{st} is a deterministic function of data up to time t. As a result, the likelihood function for $\{Q_{st}; t = 1, \ldots, T\}$ would be degenerate. Following the DSGE model literature, we add belief shocks, which means that the likelihood function is nondegenerate because there are as many structural shocks as there are observables (i.e., panel data on fund size). Economically speaking, given the inherent complexity of inferring a fund's optimal size, it seems warranted that there are forces affecting capital allocation beyond the aspects of learning in our simple model: the belief shocks capture such forces.

2.1.3 Other Sources of Learning

There are many other sources of potential learning. For example, decreasing returns parameter can be heterogeneous (Barras, Scaillet, and Binsbergen et al.). Investors might also learn about betas.

2.1.4 Entry and Exit of Funds

We take as given funds' entry and exit (i.e., the supply side), with a focus on using them to identify what investors learn and how they allocate their capital (i.e., the demand side). Specifically, difference, if any, between the observed size and the size based on the investors' posteriors just after Bayesian updating due to entry and exit should manifest as belief shocks, which in turn are penalized in our estimation, as we discuss in the next section.

While endogenizing fund entry and exit would be a natural extension of our framework, note that, for the sake of understanding how investors allocate capital, it would simply add an additional dimension of learning for investors. Moreover, such models struggle to quantitatively reproduce salient features related to entry and exit in the data, including the number of funds over time, because there are few observations of entry or exit (i.e., the minority classes) for hundreds of observations of continuing incumbents (i.e., the majority class). This is an example of imbalanced classification problems, which are generally challenging.[19] Hence, such models deliver poor fits to the data on aggregate allocations to mutual funds for reasons unrelated to the learning process. We leave the task of modeling both the demand and supply sides of this market for future research.

[19] Fraud detection represents an important real-life example of imbalanced classification problems.

2.1.5 The Likelihood Function

We derive the likelihood function of our learning model for the multiple time series of aggregate allocations $\{Q_{st}\}$ across (mutually exclusive) S categories of the mutual fund industry. Recall from Proposition 1 that the equilibrium value of Q_{st} is:

$$Q_{st} = \frac{\widehat{\widetilde{a}}_{st}}{\widehat{b}_{st}/M_{st} + \widehat{c}_{st}} = \frac{\widetilde{\widetilde{a}}_{st} + \overline{w}_{st}}{\widehat{b}_{st}/M_{st} + \widehat{c}_{st}},$$

where $\widetilde{\widetilde{a}}_{st} = \frac{1}{M_{st}} \sum_{i \in M_{st}} \widetilde{a}_{ist}$ and $\overline{w}_{st} = \frac{1}{M_{st}} \sum_{i \in M_{st}} w_{ist}$ denote the average perceived skill of all incumbent funds belonging to category s based on investor beliefs at time t just after Bayesian updating and the average value of realized belief shocks across those funds, respectively.

To characterize the conditional distribution of Q_{st} given the histories for all funds within that category, we need to make a distributional assumption for w_{ist}. We shall proceed under the assumption that the conditional distributions of q_{ist} and in turn Q_{st} are normal:

$$w_{ist} \sim N\left(0, \sigma^2_{w,s}\right),$$

where $Cov\left(w_{ist}, w_{i's't'}\right)$ is equal to $\rho_{w,s}\sigma^2_{w,s}$ if $i' \neq i, s' = s, t' = t$ and zero otherwise, which implies $Var\left(\overline{w}_{st}\right) = \left(\left(1 - \frac{1}{M_{st}}\right)\rho_{w,s} + \frac{1}{M_{st}}\right)\sigma^2_{w,s}$.

We can now characterize the joint distribution of a sequence of observations $Q_{s1}, \ldots, Q_{sT}$. Let $\theta_s = \left[\widehat{a}_{s0}, \widehat{b}_{s0}, \widehat{c}_{s0}, \widehat{\sigma}_{a,s0}, \widehat{\rho}_{s0}, \widehat{\sigma}_{b,s0}, \widehat{\sigma}_{c,s0}\right]', \theta_{w,s} = \left[\sigma_{w,s}, \rho_{w,s}\right]'$ and define the time t information set $I_{st} = \{\alpha_{is\tau}, q_{is\tau-1}; i \in M_{s\tau-1}\}_{\tau=1}^{t}$. First, the conditional density of Q_{st} can be expressed as

$$
\begin{aligned}
p\left(Q_{st}\left|\theta_s, \theta_{w,s}, I_{st}\right.\right) &= \frac{1}{\sqrt{2\pi \frac{Var(\overline{w}_{st})}{\left(\widehat{b}_{st}/M_{st}+\widehat{c}_{st}\right)^2}}} \exp\left(-\frac{1}{2}\frac{\left(Q_{st}-\frac{\widetilde{\widetilde{a}}_{st}}{\widehat{b}_{st}/M_{st}+\widehat{c}_{st}}\right)^2}{\frac{Var(\overline{w}_{st})}{\left(\widehat{b}_{st}/M_{st}+\widehat{c}_{st}\right)^2}}\right) \\
&= \frac{\left|\widehat{b}_{st}/M_{st}+\widehat{c}_{st}\right|}{\sqrt{2\pi Var\left(\overline{w}_{st}\right)}} \exp\left(-\frac{1}{2}\frac{\overline{w}_{st}^2}{Var\left(\overline{w}_{st}\right)}\right)
\end{aligned}
$$

Note that $\{\widetilde{a}_{ist}\}_{i\in M_{st-1}}$ and $\widetilde{b}_{st}, \widetilde{c}_{st}$ are determined by the information set I_{st} at time t, so $\widetilde{\widetilde{a}}_{st}$ and $\widehat{b}_{st}, \widehat{c}_{st}$ (which are the same as $\widetilde{b}_{st}, \widetilde{c}_{st}$) are known conditional on I_{st}. Note also that M_{st} is known because we take as given funds' entry and exit. Thus, the log of the joint density of $Q_{s1:T}$—the log likelihood function—can be expressed as

$$
\begin{aligned}
\sum_{\tau=1}^{T} \log p\left(Q_{s\tau}\left|\theta_s, \theta_{w,s}, I_{s\tau}\right.\right) &= \sum_{\tau=1}^{T} \log\left(\left|\widehat{b}_{s\tau}/M_{s\tau}+\widehat{c}_{s\tau}\right|\right) - \frac{T}{2}\log\left(2\pi\right) \\
&\quad -\frac{1}{2}\sum_{\tau=1}^{T}\log\left(Var\left(\overline{w}_{s\tau}\right)\right) - \frac{1}{2}\sum_{\tau=1}^{T}\frac{\overline{w}_{s\tau}^2}{Var\left(\overline{w}_{s\tau}\right)}
\end{aligned}
\tag{20}
$$

Since posterior beliefs about the key parameters from different categories are uncorrelated, the log likelihood function for the multiple time series $\{Q_{st}; t = 1, \ldots, T\}_{s=1}^{S}$ is simply the sum of (20) across all the S categories of the mutual fund industry. The log likelihood function (20), combined with a prior distribution for θ_s (Section 3.1), forms the basis for our Bayesian inference.

3 Data and Priors

Our data come from CRSP and Morningstar. We require that funds appear in both CRSP and Morningstar, which allows us to validate data accuracy across the two databases. We merge CRSP and Morningstar based on funds' tickers, CUSIPs, and names. We then compare assets and returns across the two sources in an effort to check the accuracy of each match following Berk and van Binsbergen (2015) and Pástor, Stambaugh, and Taylor (2015). We refer the readers to the data appendices of those papers for the details.

Our sample contains 2,772 actively managed domestic equity mutual funds in the US between 1991 and 2014.[20,21] We start the sample in 1991, the first year in which CRSP provides monthly data on funds' size. We use Morningstar Category to categorize funds into nine groups corresponding to Morningstar's 3×3 stylebox (large value, mid-cap growth, etc.). We also drop any fund/month observations with expense ratios less than 0.1% per year, or with lagged fund size (in 2011 dollars) less than $15 million.

We now define the key variables used in our empirical analysis: fund performance and fund size. Summary statistics are in Table 1.

We use the standard risk-based approach to measuring fund performance. The recent literature finds that investors use the CAPM in making their capital allocation decisions (Berk and van Binsbergen 2016; Barber, Huang, and Odean 2016), so we adopt the CAPM. In this case the risk adjustment R^B_{ist} is given by:

$$R^B_{ist} = \beta_{is}\mathrm{MKT}_t,$$

where MKT is the market excess return and β_{is} is the market beta of fund i belonging to category s. We estimate β_{is} by regressing the fund's excess return to investors onto the market portfolio over the fund's lifetime. Our measure of fund performance is then α_{ist}, the realized return for the fund in month t less R^B_{ist}. The average of α_{ist} is +1.3 bp per month.[22]

Fund size (q_{ist-1}) equals the fund's AUM at the end of the previous month, inflated

[20] My Morningstar data was collected when I was at Princeton University. I moved to Emory University in 2016, and unfortunately Emory University does not have access to the Morningstar database. I am in the process of gaining access to the Morningstar database beyond 2014 through other channels.

[21] We use keywords in the Primary Prospectus Benchmark variable in Morningstar to exclude bond funds, international funds, target funds, real estate funds, sector funds, and other non-equity funds. We drop funds identified by CRSP or Morningstar as index funds, in addition to funds whose name contains "index."

[22] In earlier versions of the paper, we constructed fund performance by subtracting the benchmark return from the fund's return, effectively assuming that all funds have a benchmark beta of one (Pástor and Stambaugh 2012; Pástor, Stambaugh, and Taylor 2015) but this simple approach leads to the same conclusions.

to December 2011 dollars by using the ratio of the total stock market capitalization in December 2011 to its value at the end of the previous month. Adjusting AUM by inflation this way reflects the notion that the fund's relative (rather than absolute) size is relevant for capturing decreasing returns to scale it faces. There is considerable dispersion in fund size: the inner-quartile range is \$83 million to \$917 million.

The top panel of Figure 1 shows the number of funds in our sample over time. The number of funds increases from 338 in 1991 to 1,550 in 2014. The sustained entry of new funds holds across all three size-based categories: large-cap (the dashed blue line), mid-cap (the dotted green line), and small-cap (the dash-dotted red line). The middle panel shows the reciprocal of the Herfindahl index—the so-called "effective number of funds"—over time.[23] The effective number of funds increases from 74 in 1991 to 238 in 2014. This trend is consistent with the large cross-sectional variation in fund size.

As in the previous section, category size (Q_{st-1}) is the sum of fund sizes across all funds within a given category (i.e., the sum of AUM across all funds belonging to that category, divided by the total stock market capitalization in the same month, then multiplied by the total stock market capitalization at the end of 2011). The bottom panel of Figure 1 plots the industry size (i.e., the sum of fund sizes across all funds), as well as the aggregate sizes at the level of the three Morningstar size categories (large-cap, mid-cap, and small-cap). The figure scales these time series such that they represent the fraction of total stock market capitalization that the sample's mutual funds own at each point in time. It starts at 4.6% in January 1991, peaks at 13.6% in July 2008, and finishes at 12.9% in December 2014.

In sum, we find that the active mutual fund industry grew over the period 1991–2014. This growth in industry size is driven by increases in the number of funds, consistent with the evidence of Berk and van Binsbergen (2015) and Pástor, Stambaugh, and Taylor (2015).

[23] The Herfindahl index is a measure of the size of funds in relation to the category they belong to and an indicator of the degree of competition among them. It ranges from $1/N$ to one, where N is the number of funds in the market. Equivalently, its reciprocal can range from one (i.e., one fund dominates the market) up to N (i.e., all the funds in a given category have the same size).

Importantly, identification of the investors' learning problem in our model comes from the time-series variation in aggregate capital allocation due to changing competitive landscapes.

3.1 Prior Distributions

We adopt a Bayesian approach to estimating the model, which involves specifying the prior distribution for the model parameters θ_s (i.e., parameters governing investors' prior beliefs) and $\theta_{w,s}$ (i.e., parameters governing the distribution of belief shocks).

To obtain priors for the prior expectations about the key parameters ($\widehat{a}_{s0}$ and $\widehat{b}_{s0}, \widehat{c}_{s0}$), we estimate the following regression via Zhu's (2018) recursive demeaning procedure based on the full sample (1991–2014):[24]

$$\alpha_{ist} = \psi_{0,is} - \psi_1 q_{ist-1} - \psi_2 Q_{st-1} - \psi_3 Q_{st-1} * 1\,(LrgCap) + u_{ist}, \qquad (21)$$

where $1\,(LrgCap)$ is a dummy variable, equal to one if the fund is classified by Morningstar as a large-cap fund and zero otherwise. This regression is analogous to those in Table 2, except that we add the interaction of Q_{st-1} with $1\,(LrgCap)$.[25] We find that $Q_{st-1} \times 1\,(LrgCap)$ enters positively and significantly (i.e., the estimate of ψ_3 is negative), indicating that strategy-level decreasing returns to scale are less pronounced for large-cap funds presumably because such funds are likely to face smaller total price impact costs from trading large-cap stocks.

The estimates of ψ_1, ψ_2, ψ_3 are used to set the prior means for $\widehat{b}_{s0}, \widehat{c}_{s0}$ in our analysis.[26] We use the average of the estimated fund fixed effects $\psi_{0,is}$ across all funds to set the prior means for $\widehat{a}_{s0}$ at 0.5% per month. Priors for $\widehat{a}_{s0}$ and $\widehat{b}_{s0}, \widehat{c}_{s0}$ are chosen to be log-normally distributed. Such non-negative distributions reflect the notion that investors expected at

[24] The results based on the estimates from the main sample (March 1993–December 2014) are very similar.

[25] We also tried adding the interactions of q_{st-1} with indicators for the three Morningstar size categories (small, mid-cap, and large), but we did not find significant interactions.

[26] Specifically, we set the prior means for $\widehat{b}_{s0}$ to the estimate of ψ_1 for all categories s. We set the prior means for $\widehat{c}_{s0}$ to the sum of the estimates ψ_2, ψ_3 for the three large-cap categories and to the estimate of ψ_2 for all other categories.

the beginning of 1991 that active funds do have the skill to generate alpha but they face decreasing returns, both at the fund level and at the strategy level. For each of $\widehat{a}_{s0}, \widehat{b}_{s0}, \widehat{c}_{s0}$, the prior standard deviations are specified to be almost 3 times larger than their means so that the priors are fairly disperse.

The prior means for $\widehat{\sigma}_{a,s0}, \widehat{\sigma}_{b,s0}, \widehat{\sigma}_{c,s0}$ are set at one half of the prior means for $\widehat{a}_{s0}, \widehat{b}_{s0}, \widehat{c}_{s0}$, respectively. We choose prior distributions for $\widehat{\sigma}_{a,s0}$ and $\widehat{\sigma}_{b,s0}, \widehat{\sigma}_{c,s0}$ to be inverse gamma. Our priors for $\widehat{a}_{s0}, \widehat{b}_{s0}, \widehat{c}_{s0}$ and $\widehat{\sigma}_{a,s0}, \widehat{\sigma}_{b,s0}, \widehat{\sigma}_{c,s0}$ imply that the investors' priors admit small (2 and 4 percent) probabilities of (i) $a_{is} < 0$ and (ii) $b_s < 0$ or $c_s < 0$. To the extent that priors for $\widehat{\sigma}_{a,s0}, \widehat{\sigma}_{b,s0}, \widehat{\sigma}_{c,s0}$ are much more difficult to specify (because the "rational" amount of prior uncertainty cannot be observed), for each of these parameters, the prior distribution is parameterized such that it has infinite standard deviation.

Priors for the off-diagonal elements of the covariance matrix of the key parameters (perceived by investors a priori) are even more difficult to specify. Thus, we use very dispersed beta distributions centered around 0.5 for the priors of $\widehat{\rho}_{s0}$'s. On the other hand, recall that we specify $\widehat{\sigma}_{ab,s0} = \widehat{\sigma}_{ac,s0} = 0$ and $\widehat{\sigma}_{bc,s0} = 0$ for simplicity. This assumption is clearly an oversimplification. Thus, we estimate the model parameters governing investors' beliefs at the beginning of 1991 using only the period from March 1993 to December 2014 for the likelihood function (Section 2.1.5). This allows us to avoid the first few years, during which the oversimplification can have an unduly large impact on the equilibrium allocations, from affecting the estimation results.

The marginal prior distributions for the model parameters θ_s (i.e., parameters governing investors' prior beliefs) are summarized in the first three columns of Table 3. We assume all parameters to be a priori independent. As a final point, note that we specify the same prior distribution for θ_s across the nine categories (parameters governing investors' prior beliefs about strategy-level DRS being an exception). We simply let the data speak for any heterogeneity in the investors' priors across categories in our estimation procedure.

We still need to specify $\sigma_{w,s}$ and $\rho_{w,s}$ in $\theta_{w,s}$ (i.e., parameters governing the distribution of belief shocks). Conditional on the values of all other model parameters, we solve for the values of $\sigma_{w,s}$ and $\rho_{w,s}$ to maximize the likelihood function (Section 2.1.5), treating these parameters as "nuisance" parameters.

We end by tying down the model parameters that can be estimated directly from the data. For each category s we specify the volatility of the common factor ($\sigma_{x,s}$) roughly equal to the month-level standard deviation from fitting a random-effects model of the form (2) to the residuals u_{ist} (from the regression (21)) across all funds belonging to that category; similarly, we specify the volatility of idiosyncratic shocks ($\sigma_{\epsilon,s}$) roughly equal to the standard deviation of the idiosyncratic errors from the same random-effects model.

4 Estimation Results

The results shown below are based on two chains of 85,000 draws each from the Random Walk metropolis algorithm (see Appendix B). For each chain, the first 35,000 draws are discarded as burn-in and of the remaining 50,000 draws, one of every 20 draws is retained.[27]

4.1 Parameter Estimates

The last four columns of Table 3 summarize the posterior distribution of the model parameters, reporting posterior medians, standard deviations, and 5th and 95th percentiles computed with the draws of our posterior simulator. All parameters estimates, except for the skill correlation parameters ($\widehat{\rho}_{s0}$), are fairly tight. Importantly, they are strongly at odds with our priors for these parameters, which were specified using empirical estimates of the parameters governing fund alphas. The parameter estimates indicate that investors believed at the beginning of 1991 in average skill levels considerably higher than the data support

[27] See Appendix B.2 for evidence on the convergence of our posterior simulators.

and in decreasing returns to scale much steeper than warranted.[28]

Interestingly, while investors' priors are strongly at odds with empirical estimates of the parameters, they are considerably disperse as well. Investors' prior uncertainty about skills ($\widehat{\sigma}_{a,s0}$) are about half their prior skill expectations ($\widehat{a}_{s0}$) for large-cap funds; $\widehat{\sigma}_{a,s0}$'s are even larger than $\widehat{a}_{s0}$'s for mid-cap and small-cap funds. Thus, investors perceive a priori substantial uncertainty about skills. Investors' prior uncertainty about fund-level DRS ($\widehat{\sigma}_{a,s0}$) are about one-eighth of their prior fund-level DRS expectations ($\widehat{b}_{s0}$): investors *do* perceive high uncertainty about fund-level DRS (as they should), but their priors for fund-level DRS are relatively tighter than they are for skill levels or for strategy-level DRS. Investors' prior uncertainty about strategy-level DRS ($\widehat{\sigma}_{c,s0}$) are roughly 5 times larger than their prior strategy-level DRS expectations ($\widehat{c}_{s0}$): investors perceive a priori very substantial uncertainty about strategy-level DRS.

Finally, the fact that $\widehat{\rho}_{s0} < 1$ is not tightly estimated implies that investors who perceive a priori different degrees of skill heterogeneity make similar capital allocation decisions. We show in Section 4.4.2 that investors who believe a priori that $\widehat{\rho}_{s0} = 1$ (i.e., all funds in the same category are created equal) make very different allocation decisions because they learn about only one common skill parameter at the strategy level: what is important is that investors understand the presence (or absence) of skill heterogeneity, which involves learning about as many skill parameters as there are funds all the same regardless of how much skill heterogeneity investors perceive.

[28]Specifically, investors' prior expectations about skills ($\widehat{a}_{s0}$) are almost 330 times larger than the average estimated fund fixed effects for large-cap funds, while they are roughly 66 times larger than the average estimated fund fixed effects for mid-cap and small-cap funds. In addition, investors' prior expectations about fund-level DRS ($\widehat{b}_{s0}$) are 8 times the estimate of ψ_1 in the regression (21). Moreover, investors' prior expectations about strategy-level DRS ($\widehat{c}_{s0}$) are much steeper than warranted for large-cap funds, whereas they are relatively consistent with the empirical estimates for mid-cap and small-cap funds.

4.1.1 Estimated Investors' Beliefs as of March 1993

Recall that we estimate the parameters governing investors' prior beliefs at the beginning of 1991 by comparing the aggregate allocation of capital across categories from the model against the data for each month over the period from March 1993 to the end of 2014. We do not use the first few years for comparing the model against the data because the model-implied investors' posteriors in the early years of the sample are unduly sensitive to the specification of investors' priors, so they generally differ from the actual investors' posteriors: as time goes by, the actual investors' posteriors are well approximated by the model-implied investors' posteriors, which are increasingly driven by the data. Naturally, the more interesting objects are the estimated investors' beliefs as of March 1993.

The first two columns of Table 4 report the posterior estimates (i.e., medians) of investor expectations and uncertainty about the key parameters as of March 1993.[29] The reported numbers for $a_{is}, i \in M_{st}$ are respectively the average of investor expectations and uncertainty about a_{is} across all incumbent funds belonging to category s in March 1993.

Having actively learned about actively managed funds in their early years, investors as of March 1993 already held beliefs quite different from their beliefs at the beginning of 1991. While investors continue to believe in skills of large-cap funds higher than the data support, their expectations are much more in line with what the data show.[30] In addition, investors' expectations about skills of mid-cap and small-cap funds are consistent with the data, i.e., they are comparable in magnitude to the average estimated fund fixed effects. Overall, our estimates indicate that the skill expectations as of March 1993 are broadly unbiased.

In contrast, investors' expectations about decreasing returns to scale are still biased, though not as severely as before. They continue to believe in fund-level DRS much steeper

[29] The posterior standard deviations associated with these estimates are in parentheses.

[30] Specifically, investors believe as of March 1993 in skills of large-cap funds roughly 3 times larger than the average estimated fund fixed effects, whereas they believed in skills of large-cap funds almost 330 times larger than the average estimated fund fixed effects at the beginning of 1991.

than warranted.[31] On the other hand, investors now believe in strategy-level DRS that are smaller in magnitude compared to the empirical estimates of strategy-level DRS.[32] More crucially, our estimates indicate that investors as of March 1993 believe in decreasing returns driven mainly by fund size and not by competition with other funds.

Moreover, the last four columns of Table 4 report the posterior estimates (i.e., medians) of the off-diagonal elements of the correlation matrix of the key parameters, perceived by investors in March 1993. Our estimates indicate that investors as of March 1993 perceive (i) with high skills come steep decreasing returns to scale (i.e., $\{a_{is}\}$ are weakly positively correlated with b_s and c_s) and (ii) fund-level DRS and strategy-level DRS are competing hypotheses on the nature of returns to scale (i.e., b_s and c_s are weakly negatively correlated).

4.2 Accounting for the Aggregate Size Dynamics

We now use our model to decompose the aggregate size dynamics into fluctuations in investors' perceptions of various parameters governing fund alphas. Recall from Proposition 1 that the equilibrium value of Q_{st} is:

$$Q_{st} = \frac{\widehat{\overline{a}}_{st}}{\widehat{b}_{st}/M_{st} + \widehat{c}_{st}} = \frac{\widetilde{\overline{a}}_{st} + \overline{w}_{st}}{\widehat{b}_{st}/M_{st} + \widehat{c}_{st}},$$

where $\widetilde{\overline{a}}_{st} = \frac{1}{M_{st}} \sum_{i \in M_{st}} \widetilde{a}_{ist}$ and $\overline{w}_{st} = \frac{1}{M_{st}} \sum_{i \in M_{st}} w_{ist}$ denote the average perceived skill of all incumbent funds belonging to category s based on investor beliefs at time t just after Bayesian updating and the average value of realized belief shocks across those funds, respectively. In other words, the size of a category is given by the ratio of the average fund's perceived skill to the perceived *total* strategy-level decreasing returns to scale the fund faces.

[31] Specifically, investors believe as of March 1993 in fund-level DRS roughly 6 times larger than the estimate of ψ_1 in the regression (21).

[32] Specifically, investors believe as of March 1993 in strategy-level DRS less than half the empirical estimates of strategy-level DRS—the sum of the estimates ψ_2, ψ_3 for the three large-cap categories and the estimate of ψ_3 for all other categories (all from the regression (21)).

4.2.1 Evolution of the Average Perceived Skill

Figure 2 shows the evolution of the average perceived skill over time for each fund category. Panels A and B plot percentiles of the posterior distribution of the average perceived skill $\widetilde{\overline{a}}_{st}$ (just after Bayesian updating) and $\widehat{\overline{a}}_{st}$ (after accounting for belief shocks), respectively. There is virtually no difference between Panels A and B, indicating that belief shocks are very small in magnitude, which in turn suggests the Bayesian learning-and-allocating problem in our model does a great job of capturing key forces that affect the aggregate size dynamics.

Recall that the skill expectations as of March 1993 are broadly unbiased, except that investors believe large-cap funds have skills higher than the data support. As a result, the average perceived skill of large-cap funds generally trends downward over time, whereas the average perceived skill of all other funds remains roughly the same over time.[33]

Figure 2 also shows that learning about the skill of large-cap funds is slow. The posterior estimates (i.e., medians) of their average perceived skill decline only modestly from 1.5% per month in 1993 to 1.1% per month in 2014 over a period of more than 20 years. Such posteriors are still optimistic compared to the average estimated fund fixed effects of 0.5% per month. This result can be explained by the entry and exit patterns. First, it is well known that poor performance of a fund increases its exit probability:[34] past outperformance is more informative of the average skill across future incumbent funds than past underperformance. Second, fund entry has far exeeded exit over time (Figure 1): investors' beliefs about the average skill are persistent because they get refreshed regularly by the newly-entering funds.

Now, how much variation in aggregate size dynamics can be ascribed to fluctuations in perceived skill? To answer this question, we run a panel regression of $\log(Q_{st})$ on $\log\left(\widehat{\overline{a}}_{st}\right)$ for each draw of the model parameters.[35] Across posterior draws of the model parameters, the median and standard deviation of the R^2 from the regression are 0.63 and 0.09, respectively.

[33] The only exception is that investors as of March 1993 believe mid-cap blend funds have skills higher than the data support (just like large-cap funds), so the average perceived skill of such funds trend downward.

[34] See Brown and Goetzmann (1995) and Elton, Gruber, and Blake (1996).

[35] Using $\log\left(\widetilde{\overline{a}}_{st}\right)$ instead of $\log\left(\widehat{\overline{a}}_{st}\right)$ leads to the same conclusions.

This evidence is consistent with the important role of learning about managerial skill. But it is typically not enough: 37% of variation in aggregate size dynamics is attributable to fluctuations in the perceived total strategy-level decreasing returns—the denominator in (8).

In sum, we find that the skill expectations are broadly unbiased, except that investors were initially more optimistic about large-cap funds' skill. Such optimism is largely brought to an end by learning about managerial skill toward the end of the sample. We also find that investors' perceptions of skill play an important role in determining aggregate allocation, but just as important are their perceptions of the nature of returns to scale, which we turn to next.

4.2.2 Evolution of the Perceived Total Strategy-Level Decreasing Returns

Figure 3 shows the evolution of the perceived *total* strategy-level decreasing returns to scale, $\widehat{b}_{st}/M_{st} + \widehat{c}_{st}$, over time for each fund category. It plots posterior medians (solid blue line) and 5th and 95th percentiles (dashed red lines) of the perceived total strategy-level DRS. It also plots empirical estimates of the total strategy-level DRS—$0.799/M_{st} + 0.013$ for the three large-cap categories and $0.799/M_{st} + 0.053$ for all other categories (all numbers from estimating the regression (21))—using a solid black line.

Comparing the black and blue lines, we see that investors as of March 1993 perceived total strategy-level decreasing returns that are much steeper than the empirical estimates indicate. But by the end of 2000, the perceived total strategy-level decreasing returns are in line with the empirical estimates. There are exceptions to this pattern: for the mid-cap growth and mid-cap value categories, investors tend to perceive even milder total strategy-level DRS than the empirical estimates indicate, although the bias is typically small. Importantly, we see that, for all categories, the perceived total strategy-level decreasing returns have declined substantially over time, which helps explain the observed steady growth in aggregate size.

To understand this trend, Panels A and B of Figure 4 show the evolution of the perceived fund-level DRS, $\widehat{b}_{st}$, and the perceived *direct* strategy-level DRS, $\widehat{c}_{st}$, respectively, over time for each fund category. Recall that investors as of March 1993 believe in decreasing returns driven mainly by fund size and not by competition with other funds. Panel A of Figure 4 shows that not only do investors continue to believe in fund-level DRS much steeper than warranted, $\widehat{b}_{st}$ remains virtually constant over time, highlighting that learning about fund-level DRS is unusually slow. Panel B of Figure 4 shows that investors continue to believe in milder strategy-level DRS than the empirical estimates indicate,[36] although the bias is small compared to the bias in $\widehat{b}_{st}$. Moreover, $\widehat{c}_{st}$ fluctuates much more over time, highlighting that there is more learning about strategy-level DRS, albeit still slow.

Why is learning about the nature of returns to scale (in particular, the fund-level DRS) slow? Bayesian learning implies the % change in investors' perception of each parameter upon observing fund alphas is proportional to the product of investors' uncertainty about the parameter—its perceived variance over its perceived value—and alphas' weight on the parameter.[37] Now, the alpha of fund i at time $t+1$ has a weight of one on fund skill a_{is}, whereas it has weights equal to q_{ist} and Q_{st} on fund-level DRS b_s and strategy-level DRS c_s, respectively. Recall from Proposition 1 that $Q_{st} = \frac{\widehat{\overline{a}}_{st}}{\widehat{b}_{st}/M_{st}+\widehat{c}_{st}}$ and $q_{ist} = \frac{\widehat{\overline{a}}_{st}/M_{st}}{\widehat{b}_{st}/M_{st}+\widehat{c}_{st}}$ on average. Then, it is easy to see that, even if investors' uncertainty about each parameter were proportional to the magnitude of the perceived value of the parameter, their perception of b_s and c_s fluctuate by $\widehat{\overline{a}}_{st}\frac{\widehat{b}_{st}/M_{st}}{\widehat{b}_{st}/M_{st}+\widehat{c}_{st}}$ and $\widehat{\overline{a}}_{st}\frac{\widehat{c}_{st}}{\widehat{b}_{st}/M_{st}+\widehat{c}_{st}}$, respectively, compared to their perception of a_{is} that fluctuates by $\widehat{\overline{a}}_{st} \times 1$ on average. Learning about the nature of returns to scale is generally slower than learning about managerial skill because investors first decompose performance into skill and the total effects of scale, which they then decompose to analyze the nature of returns to scale (fund-level vs strategy-level). Such learning process interacts with and amplifies the initial beliefs that are relatively more biased about fund-level DRS than they are about skill levels or about strategy-level DRS (Tables 3 and 4), resulting in a very slow learning about fund-level decreasing returns.

[36]Specifically, the empirical estimates of strategy-level DRS are 0.013 for the three large-cap categories and 0.053 for all other categories.

[37]This statement is ignoring the off-diagonal elements of the covariance matrix of the key parameters (perceived by investors before observing fund alphas), but they tend to be small in magnitude.

Such beliefs translate the sustained entry of new funds into declines in the perceived total strategy-level DRS because the perceived total strategy-level decreasing returns have a weight equal to $1/M_{st}$ on fund-level DRS $\widehat{b}_{st}$. Since there are fewer funds in the early years (high $1/M_{st}$), the positive bias in $\widehat{b}_{st}$ imparts a substantial positive bias in the perceived strategy-level DRS. As fund entry has far exceeded exit over time (declining $1/M_{st}$), persistence of the positive bias in $\widehat{b}_{st}$ leads investors to perceive that the total strategy-level decreasing returns decline much more sharply than warranted. Interestingly, declining $1/M_{st}$ also means that the bias in $\widehat{b}_{st}$ imparts a smaller bias in the perceived strategy-level DRS over time. Indeed, by the end of 2000, there are sufficiently many funds (low enough $1/M_{st}$) that the perceived total strategy-level decreasing returns are in line with the empirical estimates, despite the enduring bias in their perceptions of the nature of returns to scale.

In sum, we find that learning about the nature of returns to scale is generally slower than learning about skill because fund alphas are less informative about the parameters governing returns to scale. We also find that learning about fund-level DRS is unusually slow precisely because investors begin with beliefs that the fund-level DRS are steeper than warranted. Such beliefs translate the sustained entry of new funds into declines in the perceived total strategy-level DRS, which helps explain the observed steady growth in aggregate size.

4.3 The Aggregate Size Dynamics with Truly Bayesian Investors

In this subsection, we present counterfactual paths of aggregate allocation were investors truly Bayesian. Recall that we add so-called belief shocks to capture forces that affect investors' capital allocation decisions beyond aspects of learning in our model. Technically, belief shocks are 1-step ahead forecast errors assuming investors update their posterior beliefs at time t based on the cross-section of observed returns at time t as Bayesians. As noted earlier, Figure 2 shows that belief shocks tend to be small. But since investors' posterior beliefs at time t themselves are a product of Bayesian learning and belief shocks, we need to be cautious in appealing to the magnitude of the belief shocks to conclude that the aggregate size dynamics can largely be explained by Bayesian learning.

To assess the extent to which a purist learning model can reproduce the size dynamics in the data, we first specify investors' beliefs as of March 1993 (computed with the draws of our posterior simulator), updating their beliefs with actual data on observed returns as truly Bayesian investors would by fixing $\boldsymbol{w}_{st}$ (an M_{st}-vector of belief shocks) to zero. We then compute counterfactual fund sizes $\{q^*_{ist}\}_{i\in M_{st}}$ by invoking (6), or that the expected alpha on any fund receiving positive investment ought to be zero with respect to the counterfactual investors' beliefs. Note that shutting down belief shocks, which would maintain consistency of investors' beliefs with actual fund sizes $\{q_{ist}\}_{i\in M_{st}}$, means that $\{q^*_{ist}\}_{i\in M_{st}}$ and $\{q_{ist}\}_{i\in M_{st}}$ generally differ, and that these differences will tend to amplify over time.

Figure 5 shows the evolution of the counterfactual size Q^*_{st} of each fund category over time were investors *not* subject to belief shocks, plotting posterior median (the solid blue line) and 5th and 95th percentiles (the dashed red lines), computed with the draws of our posterior simulator. For comparison, the solid black line in Figure 5 shows the actual size Q_{st} of each fund category over time.

Comparing the black line against the blue and red lines, we see that Q^*_{st} and Q_{st} indeed tend to be closely related. More formally, for each draw from the posterior of the model parameters, we calculate the R^2 from a panel regression of $\log(Q_{st})$ on $\log(Q^*_{st})$. Across the posterior draws, the average R^2 from this regression is 0.96, and the 5th and 95th percentiles are 0.94 and 0.98, respectively. This evidence is consistent with the aspects of learning in our model doing a great job of capturing aggregate size dynamics in the data.

Panel A of Figure 6 shows how the counterfactual industry size continues to be closely related to the actual industry size. While not "targeted" in our estimation, Panel B of Figure 6 shows the evolution of the counterfactual effective number of funds—the reciprocal of the Herfindahl index. Recall that the larger the cross-sectional variation in fund size, the smaller the effective number of funds compared to the number of incumbent funds.

We find an upward trend in the counterfactual effective number of funds that is much steeper than in the data, suggesting that investors in our model perceive smaller heterogeneity in skill than perceived by the investors in reality and, consequently, the cross-sectional variation in fund size in our model is smaller than that observed in the data. Importantly, this result does *not* imply a rejection of Bayesian rational learning. Instead, it simply suggests that we ought to explore additional aspects of learning about the parameters governing alphas of funds. For example, investors in our model assume that funds in the same category draw their skills from the same distribution, regardless of when they enter, but Pástor, Stambaugh, and Taylor (2015) provide suggestive evidence that active funds have become more skilled over time. Modeling investor awareness of such additional source of heterogeneity is likely to quantitatively reproduce the empirical fund size distribution on top of reproducing the aggregate size dynamics as in our model.

To summarize, the aspects of learning in our model does surprisingly well in accounting for the aggregate size dynamics in the data: Even if investors in our model were *not* subject to belief shocks, their aggregate allocation of capital in response to the history of observed returns closely approximates the observed equilibrium aggregate allocation.

4.4 Model Fit and Relation to the Literature

This section has two objectives. First, we relate our model to those of Berk and Green (2004) and Pástor and Stambaugh (2012), which motivates a number of alternative specifications as natural benchmarks for assessing our model. Second, we evaluate the fit of our model relative to these alternative specifications. As for the assessment of fit, we find that our baseline model substantially outperforms all alternative specifications.

4.4.1 Relation to the Literature

Central features of our model are that (i) active managers have differential abilities to generate alpha but they face decreasing returns to scale in deploying these abilities and (ii) investors learn not only about managerial ability but also about the nature of returns to scale from past data. Thus, our approach builds on the seminal works of Berk and Green (2004) and Pástor and Stambaugh (2012). But there are important differences. First, Berk and Green assume that decreasing returns apply at the fund level (i.e., $c_s = \widehat{c}_{s0} = 0$), while Pástor and Stambaugh assume that they apply to the active management industry as a whole (i.e., $b_s = \widehat{b}_{s0} = 0$). In contrast, we assume that decreasing returns apply both at the fund level and at an aggregate level.

A second difference in our treatment of prior beliefs is that our investors learn not only about managerial ability but also about both fund-level and industry-level decreasing returns to scale from panel data on mutual funds. In our parameterization of prior beliefs in (14), $\widehat{\sigma}_{b,s0}, \widehat{\sigma}_{c,s0} > 0$ and $\widehat{\rho}_{s0} \in (0,1)$. In contrast, Berk and Green's investors learn only about managerial ability (i.e., $\widehat{\sigma}_{b,s0} = \widehat{\sigma}_{c,s0} = 0$) fund by fund (i.e., $\widehat{\rho}_{s0} = 0$ and $\sigma_{x,s} = 0$), while Pástor and Stambaugh's investors learn also about industry-level decreasing returns to scale (i.e., $\widehat{\sigma}_{b,s0} = 0$), although they learn everything only from the mutual fund industry's time series (i.e., $\widehat{\rho}_{s0} = 1$). As discussed earlier, when $\widehat{\rho}_{s0} \in (0,1)$, investors face a cross-fund learning problem which cannot be handled fund by fund or only using aggregate time series.

Another difference from Berk and Green (2004) and Pástor and Stambaugh (2012) is that we use the entry and exit patterns in the data to help identify the investors' learning problem. When funds enter and exit, not only the average perceived skill of incumbent funds, but also the number of incumbent funds change, both of which affect equilibrium aggregate allocations $\{Q_{st}\}$—the "target" in our estimation. Berk and Green do model entry and exit of funds, but they focus on the stationary equilibrium, in which the distribution of skills across incumbent funds and the number of incumbent funds stay constant. In contrast, Pástor and Stambaugh do not model entry and exit of funds, and focus primarily on a perfectly competitive setting with infinite number of funds that have the same level of skill.

To summarize, our framework builds upon the models of Berk and Green (2004) and Pástor and Stambaugh (2012), amended to allow for learning about the nature of returns to scale and for the effect of entry and exit dynamics on aggregate fluctuations in capital allocation. In turn, we find that rational learning from past data is a quantitatively significant driver of aggregate capital allocation dynamics in the data.

4.4.2 Model Fit

Our previous discussion motivates a number of alternative specifications as natural benchmarks for assessing our model. First, to the extent that our focus on aggregate capital allocation follows Pástor and Stambaugh (2015), the learning problem faced by Pástor and Stambaugh's investors provides a natural benchmark, which corresponds to a specification in which $\widehat{b}_{s0} = \widehat{\sigma}_{b,s0} = 0$ and $\widehat{\rho}_{s0} = 1$ (hereafter, "PS model"). Amending this specification to allow for learning about fund-level DRS or learning about skill heterogeneity provides two additional benchmark specifications: (i) the PS model with $\widehat{b}_{s0} > 0, \widehat{\sigma}_{b,s0} > 0$ and (ii) the PS model with $\widehat{\rho}_{s0} \in (0, 1)$. In this respect, note that our model can be thought of as the PS model with $\widehat{b}_{s0} > 0, \widehat{\sigma}_{b,s0} > 0$ and $\widehat{\rho}_{s0} \in (0, 1)$.

We assess the fit of our model relative to these alternative specifications using the marginal likelihood (or marginal data density), which is just the posterior density with the model parameters integrated out. In a Bayesian framework, the marginal likelihood is the most comprehensive and accurate measure of fit, as it can be used to construct posterior probabilities on competing models. To this end, we use the modified harmonic mean method of Geweke (1999), which requires the user to specify a tuning parameter $\tau \in (0, 1)$.[38] The first four rows of Table 5 reports the log-marginal data density for our model, and the three alternative specifications with $\tau = 0.5$; the last four rows repeat this exercise with $\tau = 0.9$. As is evident, the values of the marginal likelihood are overwhelmingly in favor of our model, regardless of a particular value of τ.

Note that the two alternative specifications that amend the PS model to allow for learning about fund-level DRS (which requires estimating 6 parameters for each of 9 fund categories,

[38] Technical details about Geweke's modified harmonic mean estimator are presented in Appendix B.1.

or a total of 54 parameters) or learning about skill heterogeneity (which requires estimating a total of 45 parameters) are more complex than the PS model itself (which requires estimating a total of 36 parameters), but they have lower log marginal likelihoods than the PS model. This stems from the fact that Bayesian model comparison penalizes more richly parameterized models, and thus guards against overfitting. Such penalty is particularly pronounced in our application because the likelihood function peaks at a value that is at odds with the prior distribution of the model parameters: Recall that we constructed the priors by presuming that the investors' initial beliefs are in line with the true values of the parameters governing alphas, whereas our estimates suggest otherwise. Only our model that amends the PS model to allow for both learning about fund-level DRS and learning about skill heterogeneity (which requires estimating 63 parameters in total) delivers a substantially better fit to the data, enough to overcome the penalty for model complexity.

To see intuitively why these alternative specifications perform poorly in terms of fit, we plot the evolutions of the counterfactual industry size and the counterfactual effective number of funds *under the PS model* in Figure 7.[39] Panel A shows that, while the industry size generated from the PS model generally does vary over time, its fluctuations are significantly muted, compared to those of the actual industry size. This result underlines the fact that a Bayesian investor learning based solely on aggregate time series of active mutual funds is generally fast enough that it cannot quantitatively reproduce the aggregate size dynamics in the data. Consistent with this interpretation, we find that amending the PS model to allow for only learning about fund-level DRS, which still posits investor learning only using aggregate time series, generates very similar results. Less important but also noteworthy is the result (e.g., in Panel B) that, under these alternative specifications, there is a one-to-one correspondence between the effective number of funds and the number of incumbent funds

[39] Specifically, we estimate the PS model using Bayesian methods. We then specify investors' beliefs as of March 1993 (computed with the posterior draws of the PS model's parameters), updating their beliefs with actual data on observed returns as truly Bayesian investors who believe in the return generating process under the PS model would by shutting down belief shocks. As before, we derive the model's implications for the equilibrium size by invoking (6).

because investors perceive no heterogeneity across funds. This is clearly counterfactual, reinforcing the importance of allowing for skill heterogeneity. On the other hand, amending the PS model to allow for only learning about skill heterogeneity yields a poor fit for a different reason: Only a few funds with the highest perceived skill receive positive investment because investors perceive heterogeneity across funds, but perceive no fund-level decreasing returns to scale, which play a key role in equilibrating at the fund level.

To summarize, we find that both learning about skill heterogeneity and learning about fund-level DRS play quantitatively important roles in determining the aggregate capital allocation. Doing away with one or the other leads to a substantially worse fit to the data by rendering the speed of learning unrealistically fast or the market structure of the industry unrealistically concentrated.

5 Conclusion

In this paper, we investigate whether a model of capital allocation based solely on Bayesian learning can quantitatively capture the historical fluctuations in capital allocation to mutual funds. To this end, we estimate a model in which Bayesian investors learn not only about fund skill but also about the nature of returns to scale in real time and competitively allocate capital to funds, conditional on their current beliefs. We find that the model-implied aggregate allocation of capital in response to the history of observed returns closely approximates the observed equilibrium aggregate allocation. Key to this result is that investors start with biased beliefs that fund-level decreasing returns to scale are much steeper than warranted by the data, and investors continue to believe this despite rational learning because fund returns are not informative enough about the parameters governing returns to scale. Nevertheless, we also find that, in the new millennium, the number of funds competing for investors grows large enough to ensure that such beliefs about steeper-than-warranted fund-level DRS have little influence on the aggregate allocations, which in turn are rendered broadly rational.

Overall, our results support that, as a group, mutual fund investors are *not* naive.

References

[1] Avramov, A., and R. Wermers (2006): "Investing in mutual funds when returns are predictable," *Journal of Financial Economics,* 81(2), 339–377.

[2] Baks, K. P., A. Metrick, and J. Wachter (2002): "Should Investors Avoid All Actively Managed Mutual Funds? A Study in Bayesian Performance Evaluation," *Journal of Finance,* 56(1), 45–85.

[3] Barber, B. M., X. Huang, and T. Odean (2016): "Which Factors Matter to Investors? Evidence from Mutual Fund Flows," Review of Financial Studies, 29(10), 2600–2642.

[4] Ben-David, I., J. Li, A. Rossi, and Y. Song (2022): "What Do Mutual Fund Investors Really Care About?" Review of Financial Studies, 35(4), 1723–1774.

[5] Berk, J. B., and R. C. Green (2004): "Mutual Fund Flows and Performance in Rational Markets," Journal of Political Economy, 112(6), 1269–1295.

[6] Berk, J. B., and J. H. van Binsbergen (2015): "Measuring Skill in the Mutual Fund Industry," *Journal of Financial Economics,* 118(1), 1–20.

[7] Berk, J. B., and J. H. van Binsbergen (2016): "Assessing Asset Pricing Models Using Revealed Preference," *Journal of Financial Economics,* 119(1), 1–23.

[8] Berk, J. B., and J. H. van Binsbergen (2017): "Mutual Funds in Equilibrium," *Annual Review of Financial Economics,* 9, 147–167.

[9] Binsbergen, J. H. v., J. H. Kim, and S. Kim (2021): "Capital Allocation and the Market for Mutual Funds: Inspecting the Mechanism," Working Paper.

[10] Bollen, N. P. B., and J. A. Busse (2005): "Short-Term Persistence in Mutual Fund Performance," *Review of Financial Studies,* 18(2), 569–597.

[11] Brown, S. J., and W. N. Goetzmann (1995): "Performance Persistence," *Journal of Finance,* 50(2), 679–698.

[12] Brown, D. P., and Y. Wu (2016): "Mutual Fund Flows and Cross-Fund Learning within Families," *Journal of Finance,* 71(1), 383–424.

[13] Buffa, A. M., D. Vayanos, and P. Woolley (2021): "Asset Management Contracts and Equilibrium Prices," Working Paper.

[14] Carhart, M. M. (1997): "On Persistence in Mutual Fund Performance," *Journal of Finance,* 52(1), 57–82.

[15] Chen, J., H. Hong, M. Huang, and J. D. Kubik (2004): "Does Fund Size Erode Mutual Fund Performance? The Role of Liquidity and Organization," *American Economic Review,* 94(5), 1276–1302.

[16] Chevalier, J., and G. Ellison (1997): "Risk Taking by Mutual Funds as a Response to Incentives," *Journal of Political Economy,* 105(6), 1167–1200.

[17] Choi, D., B. Kahraman, and A. Mukherjee (2016): "Learning about Mutual Fund Managers," *Journal of Finance,* 71(6), 2809–2860.

[18] Cochrane, J. H. (2013): "Finance: Function Matters, Not Size," *Journal of Economic Perspectives,* 27(2), 29–50.

[19] Elton, E. J., M. J. Gruber, and C. R. Blake (1996): "Survivor Bias and Mutual Fund Performance," *Review of Financial Studies,* 9(4), 1097–1120.

[20] Fama, E. F., and K. R. French (2010): "Luck versus Skill in the Cross-Section of Mutual Fund Returns," *Journal of Finance,* 65(5), 1915–1947.

[21] Franzoni, F., and M. C. Schmalz (2017): "Fund Flows and Market States," *Review of Financial Studies,* 30(8), 2621–2673.

[22] García, D., and J. M. Vanden (2009): "Information Acquisition and Mutual Funds," *Journal of Economic Theory,* 144(5), 1965–1995.

[23] Gârleanu, N., and L. H. Pedersen (2018): "Efficiently Inefficient Markets for Assets and Asset Management," *Journal of Finance,* 73(4), 1663–1712.

[24] Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013): *Bayesian Data Analysis,* 3rd ed., Boca Raton, FL: Chapman & Hall/CRC.

[25] Geweke, J. (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews,* 18(1), 1–73.

[26] Greenwood, R., and D. Scharfstein (2013): "The Growth of Finance," *Journal of Economic Perspectives,* 27(2), 3–28.

[27] Gruber, M. J. (1996): "Another Puzzle: The Growth in Actively Managed Mutual Funds," *Journal of Finance,* 51(3), 783–810.

[28] Herbst, E. P., and F. Schorfheide (2015): *Bayesian Estimation of DSGE Models,* 1st ed., Princeton, NJ: Princeton University Press.

[29] Hoberg, G., N. Kumar, and N. Prabhala (2018): "Mutual Fund Competition, Managerial Skill, and Alpha Persistence," *Review of Financial Studies,* 31(5), 1896–1929.

[30] Huang, J., K. D. Wei, and H. Yan (2007): "Participation Costs and the Sensitivity of Fund Flows to Past Performance," *Journal of Finance,* 62(3), 1273–1311.

[31] Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2016): "A Rational Theory of Mutual Funds' Attention Allocation," *Econometrica,* 84(2), 571–626.

[32] Khorana, A., and H. Servaes (1999): "The Determinants of Mutual Fund Starts," *Review of Financial Studies,* 12(5), 1043–1074.

[33] Khorana, A., H. Servaes, and P. Tufano (2005): "Explaining the Size of the Mutual Fund Industry Around the World," *Journal of Financial Economics,* 78(1), 145–185.

[34] Lynch, A. W., and D. K. Musto (2003): "How Investors Interpret Past Fund Returns," *Journal of Finance,* 58(5), 2033–2058.

[35] Malkiel, B. G. (1995): "Returns from Investing in Equity Mutual Funds 1971 to 1991," *Journal of Finance,* 50(2), 549–572.

[36] Mamaysky, H, M. Spiegel, and H. Zhang (2008): "Estimating the Dynamics of Mutual Fund Alphas and Betas," *Review of Financial Studies,* 21(1), 233–264.

[37] Pástor, Ľ., and R. F. Stambaugh (2002): "Investing in Equity Mutual Funds," *Journal of Financial Economics,* 63(3), 351–380.

[38] Pástor, Ľ., and R. F. Stambaugh (2012): "On the Size of the Active Management Industry," *Journal of Political Economy,* 120(4), 740–781.

[39] Pástor, Ľ., R. F. Stambaugh, and L. A. Taylor (2015): "Scale and Skill in Active Management," *Journal of Financial Economics,* 116(1), 23–45.

[40] Pástor, Ľ., and P. Veronesi (2009): "Learning in Financial Markets," *Annual Review of Financial Economics,* 1, 361–381.

[41] Philippon, T. (2015): "Has the US Finance Industry Become Less Efficient? On the Theory and Measurement of Financial Intermediation," *American Economic Review,* 105(4), 1408–1438.

[42] Philippon, T., and A. Reshef (2013): "An International Look at the Growth of Modern Finance," *Journal of Economic Perspectives,* 27(2), 73–96.

[43] Pollet, J., and M. Wilson (2008): "How Does Size Affect Mutual Fund Behavior?," *Journal of Finance,* 63(6), 2941–2969.

[44] Robert, C. P., and G. Casella (2004): *Monte Carlo Statistical Methods,* 2nd ed., New York, NY: Springer.

[45] Roussanov, N., H. Ruan, and Y. Wei (2021): "Marketing Mutual Funds," *Review of Financial Studies,* 34(6), 3045–3094.

[46] Roussanov, N., H. Ruan, and Y. Wei (2022): "Mutual Fund Flows and Performance in (Imperfectly) Rational Markets?" Working Paper.

[47] Sirri, E. R., and P. Tufano (1998): "Costly Search and Mutual Fund Flows," *Journal of Finance,* 53(5), 1589–1622.

[48] Stambaugh, R. F. (2014): "Presidential Address: Investment Noise and Trends," *Journal of Finance,* 69(4), 1415–1453.

[49] Stambaugh, R. F. (2020): "Skill and Profit in Active Management," Working Paper.

[50] Wahal, S., and A. Wang (2011): "Competition Among Mutual Funds," *Journal of Financial Economics,* 99(1), 40–59.

[51] Yan, X. (2008): "Liquidity, Investment Style, and the Relation between Fund Size and Fund Performance," *Journal of Financial and Quantitative Analysis,* 43(3), 741–767.

[52] Zhu, M. (2018): "Informative Fund Size, Managerial Skill, and Investor Rationality," *Journal of Financial Economics,* 130(1), 114–134.

A Empirical Evidence on Returns to Scale

We empirically analyze the nature of returns to scale in active mutual fund management following Zhu (2018). Pástor, Stambaugh, and Taylor (2015) (PST) analyze the nature of returns to scale by developing a recursive demeaning procedure. While their estimates indicate decreasing returns to scale both at the fund level and at the industry level, they find that only the latter is statistically significant. Zhu (2018) improves upon their empirical strategy in PST (by including an intercept in the first-stage regression) and establishes decreasing returns to scale at the fund level with compelling statistical significance.

A fund's performance should be more closely related to the size of the fund's sector than to the size of the entire industry if decreasing returns to scale at an aggregate level are driven by competition with funds that follow similar investment strategies. To evaluate this idea, we check whether sector size exhibits a negative and significant relation with fund performance, as well as whether our finding is affected by including industry size as a control.[40]

To investigate returns to scale at the fund level, we run panel regressions of fund i's net alpha in month t, $\widehat{\alpha}_{ijt}$, on the fund's size at the end of the previous month, q_{ijt-1}. We test the null hypothesis that the slope on the lagged fund size is zero. We use the recursive demeaning (RD) approach taken in Zhu (2018). We refer the readers to Section 4 of that papers for the details. We report the results both before and after controlling for industry size in the first two columns of Table 2. Panel A reports the results from the full sample (1991–2014); Panel B focuses on our estimation sample (1993–2014).

The estimated effect of fund size on performance is highly statistically significant, with t-statistics of -3.7 (column 1 of Table 2). The estimates from Panel A indicates that a $100 million increase in fund size (which represents movement of about 10% of the interquartile range) depresses performance by 0.0099% per month, or 12 bp per year. In Panel B, the

[40] We measure the industry size in the same way as Pástor, Stambaugh, and Taylor (2015) by adding up the fund AUM across all funds in the sample and then dividing by the stock market capitalization (i.e., the sum of lagged size, q_{ijt-1}, across all sample funds, up to a constant).

same increase in fund size depresses performance by only 11 bp per year. These effects are economically significant. When industry size is included in the regression together with fund size (column 2), the coefficient on fund size is significantly negative, just like in column 1, whereas the coefficient on industry size is negative but insignificant.

To explore potential returns to scale at the sector level, we run panel regressions of $\widehat{\alpha}_{ijt}$ on the lagged sector size, Q_{jt-1}. We consider the same bias-free RD approach. The results before and after controlling for industry size are in column 3 through 4 of Table 2.

The evidence of decreasing returns to scale at the sector level is strong: the estimated coefficients on sector size are negative and highly significant, with t-statistics around -5 whether or not industry size is included in the regression. The effect is both economically and statistically significant. For example, a \$10 billion increase in sector size is associated with a sizable decrease in fund performance: 0.0105% per month, or almost 13 bp per year.[41] In our estimation sample, the effect is even stronger, about 14 bp per year.

In columns 5 through 6 of Table 2, we run the multiple regressions of $\widehat{\alpha}_{ijt}$ on both q_{ijt-1} and Q_{jt-1} before and after controlling for industry size. We consider two null hypotheses: that the slope coefficient on fund size is zero, and that the slope on sector size is zero. We find that the slope on fund size remains negative and significant, and it is only slightly smaller in magnitude compared to column 1. The slope on sector size also remains negative and significant. Even though its magnitude is smaller than in column 3, it is still substantial.

The results after controlling for industry size, which are reported in column 6 of Table 2, are quite similar to those from column 5. Even though the coefficients on fund size are slightly smaller in magnitude compared to column 5, they remain statistically significant in both panels. Interestingly, the addition of industry size makes both the slope coefficient on sector size and its t-statistic even more negative compared to column 5. As before, the coefficients on industry size are never statistically significant.

[41] According to Table 1, changing sector size by \$10 billion represents movement of about one thirtieth of the interquartile range.

To summarize, we find a strong negative relation between fund performance and fund size, and another between fund performance and sector size. These relations, which are both economically and statistically significant, are consistent with the presence of decreasing returns to scale both at the fund level and at an aggregate level. In addition, we find that industry size exhibits neither a consistent nor a significant relation with fund performance. These results suggest that our proxy for sector size does a good job of capturing the size of a fund's competition. Motivated by these findings, we focus on prior beliefs in which returns are decreasing in both fund scale and sector scale at uncertain rates.

B Estimation Algorithm

The estimation algorithm is a random walk Metropolis (RWM) MCMC procedure:

Algorithm 2 *For $i = 1$ to N:*

1. *Draw $\vartheta = \theta^{i-1} + \eta$, where η is mean zero with variance $c^2\widehat{\Sigma}$.*
2. *Set $\theta^i = \vartheta$ with probability*

$$\alpha = \min\left\{\frac{p\left(\vartheta\,|Y\right)}{p\left(\theta^{i-1}\,|Y\right)}, 1\right\}$$

and $\theta^i = \theta^{i-1}$ otherwise.

The algorithm constructs a Markov chain so that it converges to a unique stationary distribution that equals the posterior distribution of the model coefficients. Detailed textbook treatments can be found, for instance, in Robert and Casella (2004) or Gelman et al. (2013).

We first use a numerical optimization routine to find the posterior mode. This is done for 100 widely dispersed start points drawn at random around the prior mean to ensure convergence of this initial search to a global mode. We run two independent Markov chains.

Parameters for one chain are initialized at the global mode, while those for the other chain are initialized at a local mode with the next highest posterior density.

For each chain, we use a (partially) adaptive approach to set $\widehat{\Sigma}$: First, generate a set of posterior draws based on a diagonal matrix for $\widehat{\Sigma}$, i.e., the prior covariance matrix that is scaled to attain an acceptance rate close to 0.25. Second, compute the sample covariance matrix from the first sequence of posterior draws and use it as $\widehat{\Sigma}$ in a second run of the RWM algorithm. We then adjust this variance-covariance matrix to ensure an acceptance rate of about 0.25, as it is usually suggested. Appendix B.2 discusses convergence diagnostics which we apply to this posterior simulator.

B.1 Marginal Likelihood

To compute the marginal likelihood both for our baseline model and for competing models, we use the modified harmonic mean method of Geweke (1999).

Harmonic mean estimators are based on the identity

$$\frac{1}{p(Y)} = \int \frac{f(\theta)}{\mathcal{L}(Y|\theta)\, p(\theta)} p(\theta|Y)\, d\theta,$$

where $f(\theta)$ has the property that $\int f(\theta)\, d\theta = 1$. Given the choice of $f(\theta)$, a natural estimator of the marginal data density is

$$\widehat{p}(Y) = \left[\frac{1}{n_{sim}} \sum_{s=1}^{n_{sim}} \frac{f\left(\theta^{(s)}\right)}{\mathcal{L}\left(Y \middle| \theta^{(s)}\right) p\left(\theta^{(s)}\right)} \right]^{-1},$$

where $\theta^{(s)}$ is drawn from the posterior $p(\theta|Y)$. Geweke (1999) proposed to use the density of a truncated multivariate normal distribution,

$$\begin{aligned} f(\theta) &= \tau^{-1} (2\pi)^{-d/2} |V_\theta|^{-1/2} \exp\left[-0.5\left(\theta - \overline{\theta}\right)' V_\theta^{-1} \left(\theta - \overline{\theta}\right)\right] \\ &\quad \times I\left\{\left(\theta - \overline{\theta}\right)' V_\theta^{-1} \left(\theta - \overline{\theta}\right) \leq F_{\chi_d^2}^{-1}(\tau)\right\}. \end{aligned}$$

Here $\overline{\theta}$ and V_θ are the posterior mean and covariance matrix computed from the output of the posterior simulator, d is the dimension of the parameter vector, $F_{\chi^2_d}$ is the cumulative density function of χ^2 random variable with d degrees of freedom, and $\tau \in (0, 1)$. We refer the reader for a more detailed treatment to the book by Herbst and Schorfheide (2015).

B.2 Convergence Diagnostics

We assess the convergence of our posterior simulators using a battery of diagnostics. Recall that we launch two chains of our Metropolis simulator from different starting values (one at the global mode, the other at a local mode with the next highest posterior density). To check that these two chains agree in their characterization of the posterior distribution, we look at various sample moments within and across chains. We split each chain in half and check that all the resulting half-sequences delivered roughly identical results for means, medians, and posterior percentiles, as well as when looking at trace plots (available upon request).[42]

More formally, for the two chains used to generate the results in the paper, Table 6 reports potential scale reduction factors as defined in Gelman et al. (2013) (the first column). These numbers are very close to one and therefore well below the 1.1 benchmark widely used in practice as a threshold for convergence. In addition, the second column computes the effective sample size using Gelman et al.'s (2013) estimator for each coefficient. The average ESS across all parameters governing prior beliefs within a given sector ranges from 200 in small blend to 1998 in mid-cap growth. Gelman et al. (2013) suggest running the simulation until ESS is at least $5m$, where m is the number of chains (after splitting). We simulate two sequences, each of which we split into two parts, so that m is equal to 4; thus the reference suggests a "safe" threshold value of $ESS > 20$, which is completely met by all of our parameters. Therefore, we consider these simulations sufficiently converged.

Table 1: Summary Statistics

This table shows summary statistics for our sample of active equity mutual funds from 1991–2014. The unit of observation is the fund/month. All returns are in units of % per month. Net alpha (α_{ist}) equals the fund's net excess return (i.e., the fund's excess return to investors) minus the CAPM risk adjustment, which is the product of the fund's market beta (β_{is}) and the market excess return. We estimate β_{is} by regressing the fund's net excess return onto the market portfolio over the fund's lifetime. Fund size (q_{ist-1}) is the fund's AUM at the end of the previous month, inflated to December 2011 dollars by using the ratio of the total stock market capitalization in December 2011 to its value at the end of the previous month. Category size (Q_{st-1}) is the sum of AUM across all funds within a given category at the end of the previous month, divided by the total stock market capitalization in the same month, then multiplied by the total stock market capitalization at the end of 2011. We use the nine categories in Morningstar's 3×3 StyleBox. M_{st-1} is the number of funds belonging to category s at time $t-1$.

Panel A: Fund-Level Variables

Morningstar Category		Mean	Stdev.	Percentiles 1%	25%	50%	75%	99%
Large	α_{ist}	−0.0463	1.6452	−5.0078	−0.8531	−0.0509	0.7569	4.9472
	q_{ist-1}	1,643	3,683	17	91	325	1,224	20,549
Mid-Cap	α_{ist}	0.0649	2.2672	−6.3158	−1.1959	0.0684	1.3097	6.5549
	q_{ist-1}	1,009	2,223	16	80	266	936	11,885
Small	α_{ist}	0.0939	2.7084	−6.3158	−1.5707	0.0617	1.7734	6.5549
	q_{ist-1}	530	1,179	17	73	193	508	6,113

Panel B: Category-Level Variables

Morningstar Category		Mean	Stdev.	Percentiles 1%	25%	50%	75%	99%
Large	Q_{st-1}	355,420	111,540	103,560	301,710	388,900	418,450	554,110
	M_{st-1}	216	94	27	137	244	284	353
Mid-Cap	Q_{st-1}	99,580	44,610	22,280	70,950	84,050	139,890	194,850
	M_{st-1}	99	47	22	70	84	117	201
Small	Q_{st-1}	50,110	28,560	7,410	24,290	41,650	77,280	107,880
	M_{st-1}	95	51	11	59	82	135	192

Table 2: Relation Between Size and Fund Performance

The dependent variable in each regression model is α_{ist}, the fund's net alpha. Fund size (q_{ist-1}) is the fund's AUM at the end of the previous month, inflated to December 2011 dollars by using the ratio of the total stock market capitalization in December 2011 to its value at the end of the previous month. Category size (Q_{st-1}) is the sum of AUM across all funds within a given category at the end of the previous month, divided by the total stock market capitalization in the same month, then multiplied by the total stock market capitalization at the end of 2011. We use the nine categories in Morningstar's 3×3 StyleBox. *IndustrySize* is the total AUM of all active equity mutual funds divided by the total market cap of all stocks in CRSP. The RD2 estimator recursively forward-demeans all variables and instruments for forward-demeaned q_{ist-1} using q_{ist-1}. We multiply the slopes on q_{ist-1} and Q_{st-1} by 10^6 to make them easier to read. The reported slopes on q_{ist-1} and Q_{st-1} thus equal the change in α_{ist}, in units of bp per month, associated with a \$100 million increase in q_{ist-1} and Q_{st-1}. Heteroskedasticity-robust t-statistics clustered by fund and by month are in parentheses.

Panel A: Full Sample (January 1991 – December 2014)

q_{ist-1}	−0.992	−0.660			−0.850	−0.710
	(−3.65)	(−3.55)			(−4.15)	(−3.76)
Q_{st-1}			−0.0105	−0.0119	−0.00860	−0.0110
			(−5.04)	(−5.02)	(−4.07)	(−4.57)
IndustrySize		−0.0102		0.00887		0.0135
		(−0.94)		(0.70)		(1.05)
Observations	351041	351041	353438	353438	351041	351041
Estimator	RD2	RD2	RD2	RD2	RD2	RD2

Panel B: Main Sample (March 1993 – December 2014)

q_{ist-1}	−0.879	−0.674			−0.811	−0.714
	(−3.63)	(−3.50)			(−3.99)	(−3.68)
Q_{st-1}			−0.0112	−0.0125	−0.00973	−0.0116
			(−4.83)	(−5.10)	(−4.19)	(−4.68)
IndustrySize		−0.0108		0.0111		0.0142
		(−0.73)		(0.66)		(0.84)
Observations	341895	341895	344150	344150	341895	341895
Estimator	RD2	RD2	RD2	RD2	RD2	RD2

Table 3: Prior Densities and Posterior Estimates for the Model Parameters

The first three columns summarize marginal prior distributions for the model parameters θ_s (i.e., parameters governing investors' prior beliefs). Conditional on the values of these model parameters, we solve for the values of the model parameters $\theta_{w,s}$ (i.e., parameters governing the distribution of belief shocks) to maximize the likelihood function (Section 2.1.5). In the column marked "Density," LN stands for Log Normal, IG Inverse Gamma, and B Beta distribution. The last four columns summarize the posterior distribution of the model coefficients, reporting median, standard deviations and posterior percentiles from two chains of 85,000 draws each from the Random Walk metropolis algorithm, one initialized at the global mode, the other initialized at a local mode with the next highest posterior density. We discard the initial 35,000 draws and retain 1 in every 20 simulations from the remaining 50,000 draws. *Notes:* All $\widehat{a}_{s0}$ and $\widehat{\sigma}_{a,s0}$ are in units of fraction per month, and all $\widehat{b}_{s0}, \widehat{c}_{s0}$ and $\widehat{\sigma}_{b,s0}, \widehat{\sigma}_{c,s0}$ are in units of fraction per month, associated with a \$1 trillion increase in q_{ist-1} and Q_{st-1} (equivalently, in units of bp per month, associated with a \$100 million increase in q_{ist-1} and Q_{st-1}).

Morningstar Category		Prior			Posterior			
		Density	Mean	Std	Median	Std	5%	95%
Large Blend	$\widehat{a}_0$	LN	0.005	0.007	2.270	0.566	1.596	3.406
	$\widehat{b}_0$	LN	0.799	1.048	8.047	2.155	5.444	12.37
	$\widehat{c}_0$	LN	0.013	0.017	0.712	0.293	0.328	1.311
	$\widehat{\sigma}_{a,0}$	IG	0.003	∞	0.532	0.309	0.298	1.170
	$\widehat{\sigma}_{b,0}$	IG	0.400	∞	0.690	1.369	0.331	3.912
	$\widehat{\sigma}_{c,0}$	IG	0.006	∞	0.658	0.739	0.317	2.308
	$\widehat{\rho}_0$	B	0.500	0.224	0.529	0.228	0.138	0.880
Large Growth	$\widehat{a}_0$	LN	0.005	0.007	1.293	0.239	0.956	1.747
	$\widehat{b}_0$	LN	0.799	1.048	5.410	1.162	3.729	7.521
	$\widehat{c}_0$	LN	0.013	0.017	0.183	0.087	0.075	0.354
	$\widehat{\sigma}_{a,0}$	IG	0.003	∞	0.953	0.324	0.497	1.533
	$\widehat{\sigma}_{b,0}$	IG	0.400	∞	0.661	0.859	0.317	2.576
	$\widehat{\sigma}_{c,0}$	IG	0.006	∞	0.573	0.129	0.392	0.820
	$\widehat{\rho}_0$	B	0.500	0.224	0.359	0.198	0.103	0.745
Large Value	$\widehat{a}_0$	LN	0.005	0.007	1.334	0.364	0.822	2.012
	$\widehat{b}_0$	LN	0.799	1.048	6.949	1.965	4.001	10.39
	$\widehat{c}_0$	LN	0.013	0.017	0.230	0.125	0.087	0.478
	$\widehat{\sigma}_{a,0}$	IG	0.003	∞	0.729	0.406	0.391	1.442
	$\widehat{\sigma}_{b,0}$	IG	0.400	∞	0.688	0.700	0.339	2.070
	$\widehat{\sigma}_{c,0}$	IG	0.006	∞	0.685	0.373	0.344	1.504
	$\widehat{\rho}_0$	B	0.500	0.224	0.519	0.235	0.120	0.890

Morningstar Category		Prior			Posterior			
		Density	Mean	Std	Median	Std	5%	95%
Mid-Cap Blend	$\widehat{a}_0$	LN	0.005	0.007	0.422	0.140	0.262	0.701
	$\widehat{b}_0$	LN	0.799	1.048	6.566	2.154	4.270	10.84
	$\widehat{c}_0$	LN	0.053	0.070	0.060	0.033	0.025	0.128
	$\widehat{\sigma}_{a,0}$	IG	0.003	∞	0.531	0.215	0.338	0.945
	$\widehat{\sigma}_{b,0}$	IG	0.400	∞	0.833	1.077	0.378	2.897
	$\widehat{\sigma}_{c,0}$	IG	0.027	∞	0.280	0.071	0.202	0.422
	$\widehat{\rho}_0$	B	0.500	0.224	0.511	0.200	0.201	0.850
Mid-Cap Growth	$\widehat{a}_0$	LN	0.005	0.007	0.471	0.106	0.325	0.667
	$\widehat{b}_0$	LN	0.799	1.048	3.782	0.751	2.692	5.156
	$\widehat{c}_0$	LN	0.053	0.070	0.067	0.029	0.030	0.123
	$\widehat{\sigma}_{a,0}$	IG	0.003	∞	0.566	0.150	0.358	0.847
	$\widehat{\sigma}_{b,0}$	IG	0.400	∞	0.507	0.413	0.281	1.417
	$\widehat{\sigma}_{c,0}$	IG	0.027	∞	0.300	0.054	0.228	0.401
	$\widehat{\rho}_0$	B	0.500	0.224	0.622	0.190	0.278	0.901
Mid-Cap Value	$\widehat{a}_0$	LN	0.005	0.007	0.193	0.080	0.102	0.361
	$\widehat{b}_0$	LN	0.799	1.048	2.441	0.930	1.312	4.310
	$\widehat{c}_0$	LN	0.053	0.070	0.182	0.072	0.097	0.330
	$\widehat{\sigma}_{a,0}$	IG	0.003	∞	0.430	0.111	0.288	0.648
	$\widehat{\sigma}_{b,0}$	IG	0.400	∞	0.893	1.048	0.388	3.134
	$\widehat{\sigma}_{c,0}$	IG	0.027	∞	0.356	0.101	0.243	0.575
	$\widehat{\rho}_0$	B	0.500	0.224	0.460	0.204	0.171	0.842
Small Blend	$\widehat{a}_0$	LN	0.005	0.007	0.515	0.168	0.289	0.837
	$\widehat{b}_0$	LN	0.799	1.048	9.112	3.269	5.174	15.85
	$\widehat{c}_0$	LN	0.053	0.070	0.145	0.079	0.057	0.301
	$\widehat{\sigma}_{a,0}$	IG	0.003	∞	0.772	0.226	0.506	1.221
	$\widehat{\sigma}_{b,0}$	IG	0.400	∞	0.397	0.154	0.249	0.749
	$\widehat{\sigma}_{c,0}$	IG	0.027	∞	0.334	0.083	0.236	0.508
	$\widehat{\rho}_0$	B	0.500	0.224	0.676	0.167	0.350	0.904
Small Growth	$\widehat{a}_0$	LN	0.005	0.007	0.321	0.086	0.211	0.484
	$\widehat{b}_0$	LN	0.799	1.048	6.893	1.857	4.519	10.48
	$\widehat{c}_0$	LN	0.053	0.070	0.123	0.043	0.064	0.203
	$\widehat{\sigma}_{a,0}$	IG	0.003	∞	1.106	0.218	0.797	1.511
	$\widehat{\sigma}_{b,0}$	IG	0.400	∞	0.597	0.324	0.320	1.306
	$\widehat{\sigma}_{c,0}$	IG	0.027	∞	0.405	0.139	0.268	0.693
	$\widehat{\rho}_0$	B	0.500	0.224	0.524	0.163	0.278	0.810
Small Value	$\widehat{a}_0$	LN	0.005	0.007	0.057	0.024	0.025	0.104
	$\widehat{b}_0$	LN	0.799	1.048	5.506	1.141	3.817	7.556
	$\widehat{c}_0$	LN	0.053	0.070	0.026	0.013	0.011	0.052
	$\widehat{\sigma}_{a,0}$	IG	0.003	∞	0.546	0.080	0.426	0.690
	$\widehat{\sigma}_{b,0}$	IG	0.400	∞	0.582	0.374	0.331	1.300
	$\widehat{\sigma}_{c,0}$	IG	0.027	∞	0.557	0.101	0.409	0.741
	$\widehat{\rho}_0$	B	0.500	0.224	0.722	0.135	0.482	0.924

Table 4: Posterior Estimates of Investors' Initial Beliefs as of March 1993

The first two columns report the posterior estimates (i.e., medians) of investor expectations and uncertainty about the key parameters as of March 1993. The posterior standard deviations associated with these estimates are in parentheses. The reported numbers for $a_{is}, i \in M_{st}$ are respectively the average of investor expectations and uncertainty about a_{is} across all incumbent funds belonging to category s in March 1993. Moreover, the last four columns report the posterior estimates (i.e., medians) of the off-diagonal elements of the correlation matrix of the key parameters, perceived by investors in March 1993. *Notes:* In the first two columns, all reported numbers for $a_{is}, i \in M_{st}$ and $a_{\infty s}$ are in units of fraction per month, and all reported numbers for b_s and c_s are in units of fraction per month, associated with a \$1 trillion increase in q_{ist-1} and Q_{st-1} (equivalently, in units of bp per month, associated with a \$100 million increase in q_{ist-1} and Q_{st-1}).

Morningstar Category		Inv Mean	Inv Std	Inv Correlation Matrix			
				$a_{is}, i \in M_{st}$	$a_{\infty s}$	b_s	c_s
Large Blend	$a_{is}, i \in M_{st}$	0.016	0.001	0.215			
		(0.004)	(0.000)	(0.100)			
	$a_{\infty s}$	0.011	0.001	0.420	1		
		(0.003)	(0.001)	(0.206)			
	b_s	6.219	0.072	0.044	0.042	1	
		(1.707)	(0.007)	(0.025)	(0.024)		
	c_s	0.011	0.003	0.316	0.302	−0.074	1
		(0.004)	(0.002)	(0.163)	(0.161)	(0.032)	
Large Growth	$a_{is}, i \in M_{st}$	0.016	0.002	0.163			
		(0.003)	(0.001)	(0.079)			
	$a_{\infty s}$	0.006	0.002	0.303	1		
		(0.001)	(0.001)	(0.169)			
	b_s	4.698	0.119	0.061	0.053	1	
		(0.916)	(0.017)	(0.025)	(0.024)		
	c_s	0.003	0.004	0.056	0.048	−0.025	1
		(0.001)	(0.001)	(0.031)	(0.029)	(0.007)	
Large Value	$a_{is}, i \in M_{st}$	0.013	0.002	0.243			
		(0.004)	(0.001)	(0.099)			
	$a_{\infty s}$	0.007	0.002	0.466	1		
		(0.002)	(0.001)	(0.206)			
	b_s	5.630	0.069	0.023	0.022	1	
		(1.567)	(0.007)	(0.013)	(0.012)		
	c_s	0.002	0.004	0.088	0.082	−0.016	1
		(0.001)	(0.002)	(0.084)	(0.080)	(0.007)	

Morningstar Category		Inv Mean	Inv Std	Inv Correlation Matrix			
				$a_{is}, i \in M_{st}$	$a_{\infty s}$	b_s	c_s
Mid-Cap Blend	$a_{is}, i \in M_{st}$	0.009	0.001	0.235			
		(0.003)	(0.000)	(0.093)			
	$a_{\infty s}$	0.002	0.001	0.471	1		
		(0.001)	(0.000)	(0.193)			
	b_s	5.287	0.141	0.030	0.029	1	
		(1.722)	(0.015)	(0.011)	(0.011)		
	c_s	0.002	0.007	0.070	0.067	−0.025	1
		(0.002)	(0.002)	(0.039)	(0.038)	(0.006)	
Mid-Cap Growth	$a_{is}, i \in M_{st}$	0.004	0.001	0.300			
		(0.001)	(0.000)	(0.090)			
	$a_{\infty s}$	0.003	0.001	0.603	1		
		(0.001)	(0.000)	(0.187)			
	b_s	3.110	0.181	0.014	0.014	1	
		(0.611)	(0.066)	(0.008)	(0.008)		
	c_s	0.003	0.008	0.033	0.033	−0.008	1
		(0.002)	(0.001)	(0.013)	(0.013)	(0.004)	
Mid-Cap Value	$a_{is}, i \in M_{st}$	0.003	0.001	0.220			
		(0.001)	(0.000)	(0.096)			
	$a_{\infty s}$	0.001	0.001	0.446	1		
		(0.000)	(0.000)	(0.199)			
	b_s	2.173	0.155	0.012	0.012	1	
		(0.722)	(0.018)	(0.005)	(0.005)		
	c_s	0.009	0.009	0.018	0.017	−0.007	1
		(0.004)	(0.003)	(0.014)	(0.014)	(0.002)	
Small Blend	$a_{is}, i \in M_{st}$	0.005	0.002	0.305			
		(0.002)	(0.000)	(0.073)			
	$a_{\infty s}$	0.003	0.002	0.640	1		
		(0.001)	(0.001)	(0.164)			
	b_s	7.277	0.158	0.009	0.008	1	
		(2.612)	(0.060)	(0.005)	(0.005)		
	c_s	0.008	0.009	0.006	0.005	−0.001	1
		(0.004)	(0.002)	(0.003)	(0.003)	(0.000)	
Small Growth	$a_{is}, i \in M_{st}$	0.004	0.003	0.254			
		(0.001)	(0.000)	(0.069)			
	$a_{\infty s}$	0.002	0.003	0.501	1		
		(0.001)	(0.001)	(0.152)			
	b_s	5.480	0.229	0.010	0.009	1	
		(1.485)	(0.098)	(0.006)	(0.005)		
	c_s	0.006	0.011	0.010	0.009	−0.001	1
		(0.002)	(0.004)	(0.007)	(0.006)	(0.001)	
Small Value	$a_{is}, i \in M_{st}$	0.003	0.001	0.344			
		(0.001)	(0.000)	(0.064)			
	$a_{\infty s}$	0.001	0.001	0.709	1		
		(0.000)	(0.000)	(0.135)			
	b_s	4.461	0.221	0.006	0.006	1	
		(0.933)	(0.093)	(0.003)	(0.003)		
	c_s	0.001	0.015	0.009	0.009	−0.001	1
		(0.001)	(0.003)	(0.003)	(0.003)	(0.001)	

[42] This jointly tests mixing (if the chains have mixed well, the separate parts of the different chains will also mix) and stationarity (at stationarity, the two half-parts of each sequence will traverse the same distribution).

Table 5: Log-Marginal Data Densities for Baseline Model and Alternative Specifications

This table reports the log-marginal data density computed using the output of the MCMC simulators as described in Appendix B. Model favored by the data attains the highest marginal data density. The Bayes factor is the ratio of marginal data densities of our baseline model vs. an alternative specification. Assuming that all models are regarded as equally likely a priori, the Bayes factor is all we need to conduct model comparison. Parameter estimates for our baseline model is shown in Table 3. Full set of parameter estimates for the remaining models is available from the author upon request. *Notes:* The threshold τ is a tuning parameter for the modified harmonic mean method of Geweke (1999). We try different values of τ to assess the stability of Geweke's estimator.

Panel A: Geweke ($\tau = 0.5$)

Specification	Log marginal	Bayes factor
Baseline model	12807	1.00
Berk-Green model, amended to permit $\widehat{c}_{s0}, \widehat{\sigma}_{c,s0} > 0$	12268	exp (539)
Pastor-Stambaugh model (i.e., $\widehat{b}_{s0} = \widehat{\sigma}_{b,s0} = 0$ and $\widehat{\rho}_{s0} = 1$)	11756	exp (1051)
Pastor-Stambaugh model, amended to permit $\widehat{b}_{s0}, \widehat{\sigma}_{b,s0} > 0$	11754	exp (1053)
Pastor-Stambaugh model, amended to permit $\widehat{\rho}_{s0} < 1$	11678	exp (1129)

Panel B: Geweke ($\tau = 0.9$)

Specification	Log marginal	Bayes factor
Baseline model	12797	1.00
Berk-Green model, amended to permit $\widehat{c}_{s0}, \widehat{\sigma}_{c,s0} > 0$	12268	exp (539)
Pastor-Stambaugh model (i.e., $\widehat{b}_{s0} = \widehat{\sigma}_{b,s0} = 0$ and $\widehat{\rho}_{s0} = 1$)	11756	exp (1040)
Pastor-Stambaugh model, amended to permit $\widehat{b}_{s0}, \widehat{\sigma}_{b,s0} > 0$	11755	exp (1042)
Pastor-Stambaugh model, amended to permit $\widehat{\rho}_{s0} < 1$	11678	exp (1119)

Table 6: Convergence Diagnostics for the Model Parameters

This table reports the Gelman et al. (2013) potential scale reduction factor (PSRF) and the Gelman et al. (2013) effective sample size (ESS), computed for each element of the model parameter θ_s (i.e., parameters governing investors' prior beliefs) with the draws of our posterior simulator. Gelman et al. (2013) suggest running the simulation until PSRF is below 1.1 and ESS is at least $5m$, that is, until there are the equivalent of at least 10 independent draws per sequence (recall that m is twice the number of sequences, as we have split each sequence into two parts so that PSRF can assess stationarity as well as mixing). Note that $m = 4$, as we have simulated two sequences. The first column reports PSRF. The second column reports ESS. Both of these alternative measures of convergence suggest that there is an overall approximate convergence.

Morningstar Category		Potential scale reduction factor (PSRF)	Effective sample size (ESS)
Large Blend	$\widehat{a}_0$	1.00	782
	$\widehat{b}_0$	1.00	788
	$\widehat{c}_0$	1.00	829
	$\widehat{\sigma}_{a,0}$	1.00	634
	$\widehat{\sigma}_{b,0}$	1.05	55
	$\widehat{\sigma}_{c,0}$	1.01	185
	$\widehat{\rho}_0$	1.00	2831
Large Growth	$\widehat{a}_0$	1.00	688
	$\widehat{b}_0$	1.00	475
	$\widehat{c}_0$	1.00	2099
	$\widehat{\sigma}_{a,0}$	1.00	1415
	$\widehat{\sigma}_{b,0}$	1.03	113
	$\widehat{\sigma}_{c,0}$	1.00	698
	$\widehat{\rho}_0$	1.00	1256
Large Value	$\widehat{a}_0$	1.04	90
	$\widehat{b}_0$	1.05	84
	$\widehat{c}_0$	1.01	452
	$\widehat{\sigma}_{a,0}$	1.02	167
	$\widehat{\sigma}_{b,0}$	1.01	384
	$\widehat{\sigma}_{c,0}$	1.03	184
	$\widehat{\rho}_0$	1.01	287

Morningstar Category		Potential scale reduction factor (PSRF)	Effective sample size (ESS)
Mid-Cap Blend	$\widehat{a}_0$	1.00	1186
	$\widehat{b}_0$	1.00	657
	$\widehat{c}_0$	1.00	1375
	$\widehat{\sigma}_{a,0}$	1.00	513
	$\widehat{\sigma}_{b,0}$	1.00	88
	$\widehat{\sigma}_{c,0}$	1.00	1297
	$\widehat{\rho}_0$	1.00	1605
Mid-Cap Growth	$\widehat{a}_0$	1.00	2052
	$\widehat{b}_0$	1.00	2034
	$\widehat{c}_0$	1.00	2976
	$\widehat{\sigma}_{a,0}$	1.00	2282
	$\widehat{\sigma}_{b,0}$	1.00	573
	$\widehat{\sigma}_{c,0}$	1.00	2049
	$\widehat{\rho}_0$	1.00	2020
Mid-Cap Value	$\widehat{a}_0$	1.00	582
	$\widehat{b}_0$	1.00	235
	$\widehat{c}_0$	1.00	446
	$\widehat{\sigma}_{a,0}$	1.00	591
	$\widehat{\sigma}_{b,0}$	1.02	101
	$\widehat{\sigma}_{c,0}$	1.00	329
	$\widehat{\rho}_0$	1.00	308
Small Blend	$\widehat{a}_0$	1.01	73
	$\widehat{b}_0$	1.01	62
	$\widehat{c}_0$	1.00	326
	$\widehat{\sigma}_{a,0}$	1.01	69
	$\widehat{\sigma}_{b,0}$	1.01	140
	$\widehat{\sigma}_{c,0}$	1.00	290
	$\widehat{\rho}_0$	1.00	440
Small Growth	$\widehat{a}_0$	1.01	613
	$\widehat{b}_0$	1.02	193
	$\widehat{c}_0$	1.01	1049
	$\widehat{\sigma}_{a,0}$	1.01	472
	$\widehat{\sigma}_{b,0}$	1.00	1095
	$\widehat{\sigma}_{c,0}$	1.01	336
	$\widehat{\rho}_0$	1.00	484
Small Value	$\widehat{a}_0$	1.00	2487
	$\widehat{b}_0$	1.00	1606
	$\widehat{c}_0$	1.00	2626
	$\widehat{\sigma}_{a,0}$	1.00	1728
	$\widehat{\sigma}_{b,0}$	1.00	641
	$\widehat{\sigma}_{c,0}$	1.00	1893
	$\widehat{\rho}_0$	1.00	1917

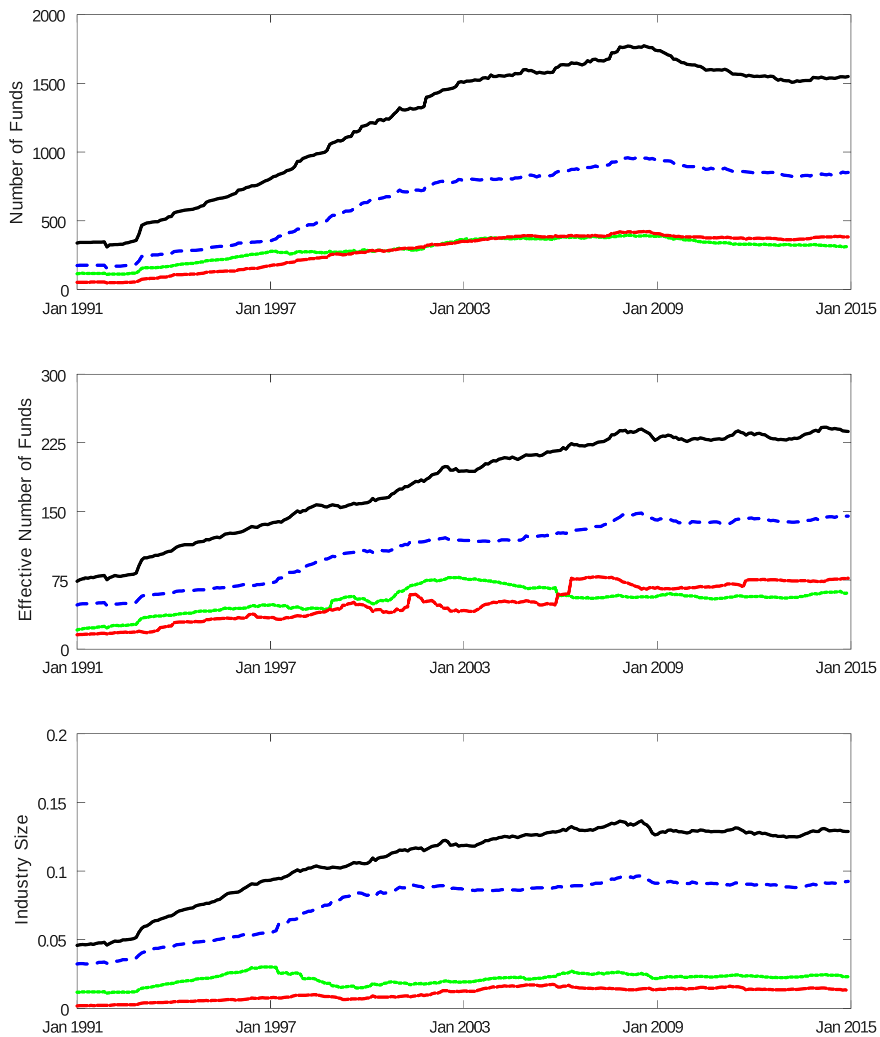

Figure 1: **Sample properties.** The top panel shows the number of funds in our sample over time in sum (solid black line), as well as across the three size-based categories: large-cap (dashed blue line), mid-cap (dotted green line), and small-cap (dash-dotted red line). The middle panel shows the reciprocal of the Herfindahl index—the so-called "effective number of funds"—over time. The bottom panel plots the industry size, as well as the aggregate sizes at the level of the three Morningstar size categories. *Notes:* In the bottom panel, we scale the time series such that they represent the fraction of total stock market capitalization that the sample's mutual funds own at each point in time.

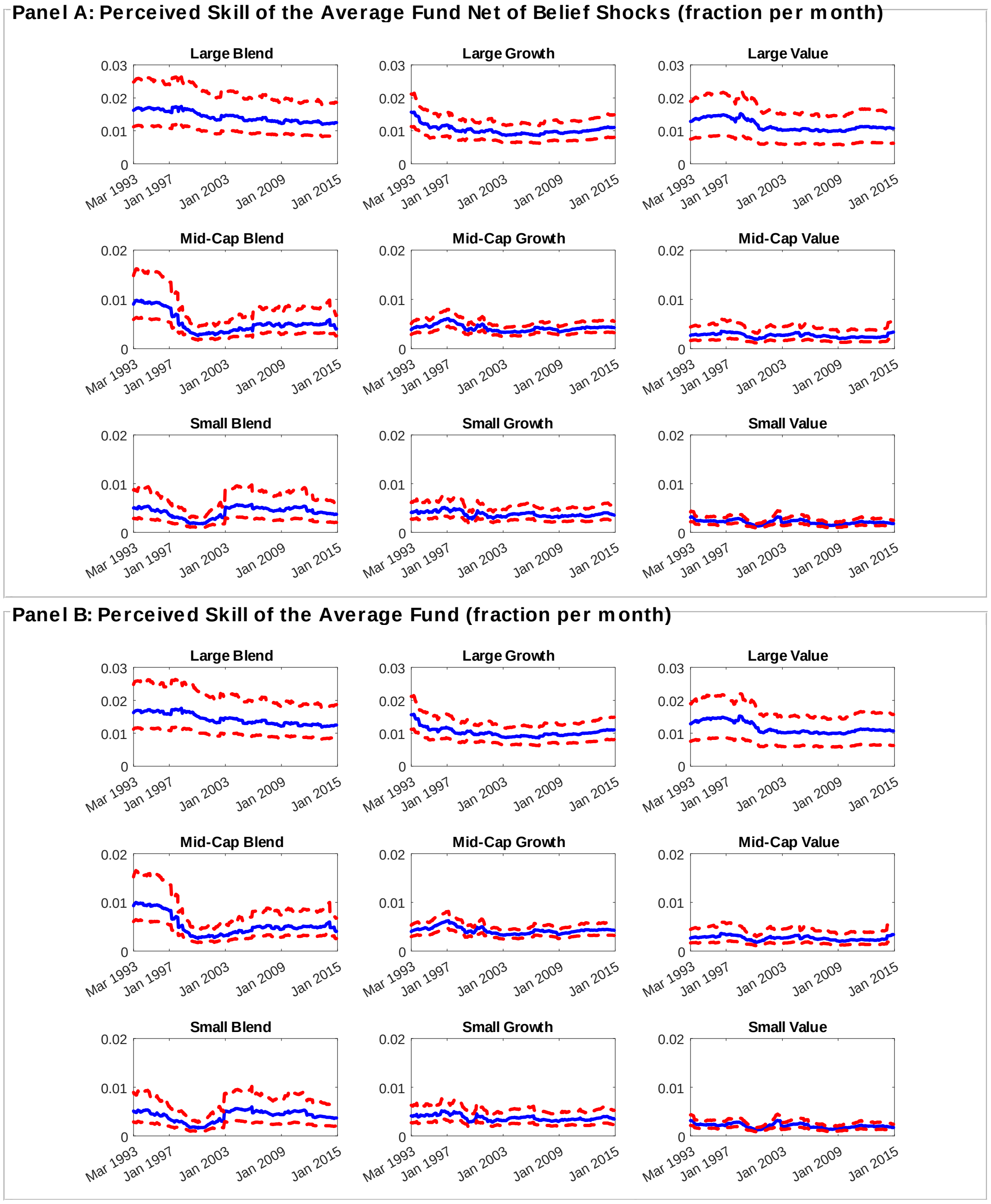

Figure 2: **Evolution of the Perceived Skill of an Average Fund.** This figure shows the evolution of the average perceived skill over time (March 1993–December 2014) for each fund category. Panels A and B plot percentiles of the posterior distribution of the average perceived skill $\widetilde{a}_{st}$ (just after Bayesian updating) and $\widehat{a}_{st}$ (after accounting for belief shocks), respectively. *Notes:* All $\widetilde{a}_{st}$ and $\widehat{a}_{st}$ are in units of fraction per month.

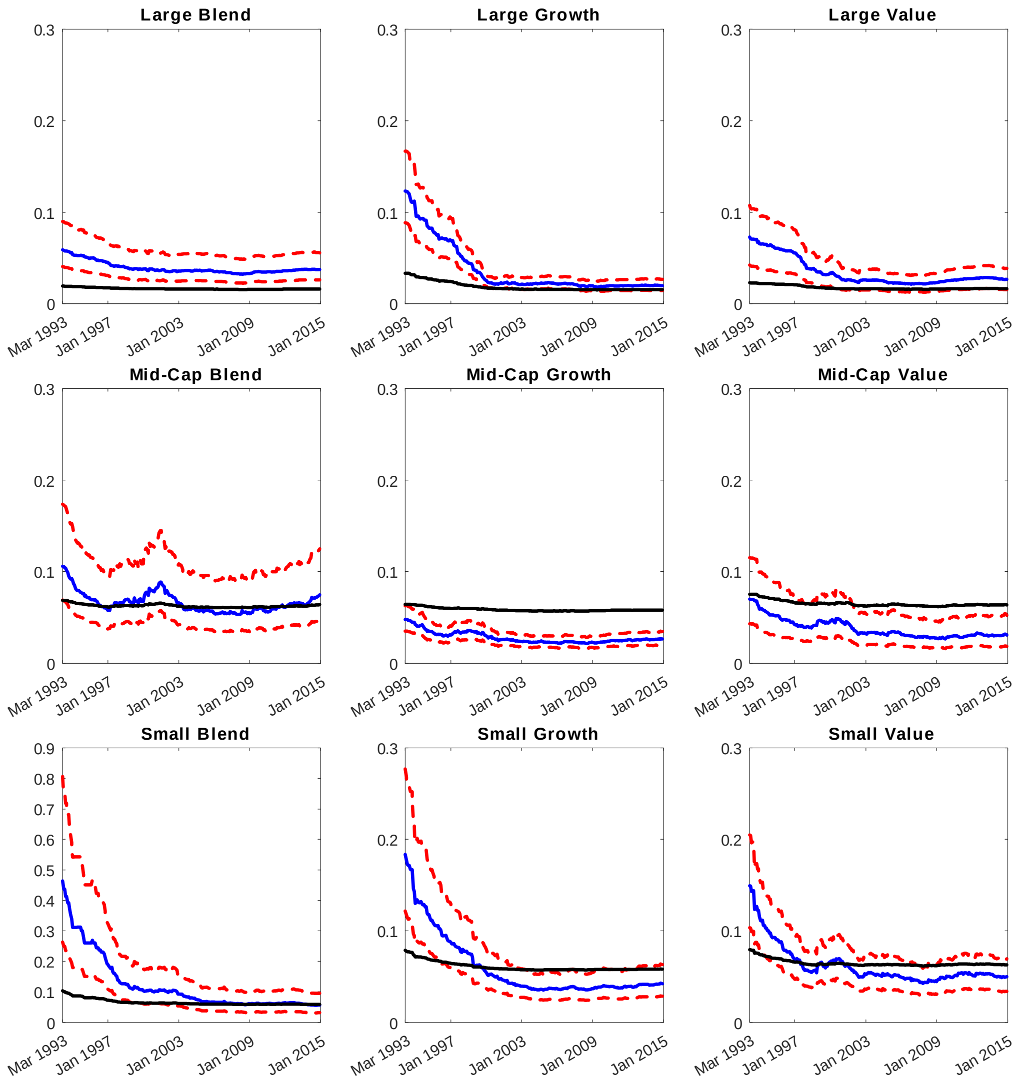

Figure 3: **Evolution of the Perceived Total Impact of Strategy-Scale.** This figure shows the evolution of the perceived total strategy-level decreasing returns to scale, $\widehat{b}_{st}/M_{st} + \widehat{c}_{st}$, over time (March 1993–December 2014) for each fund category. It plots posterior medians (solid blue line) and 5th and 95th percentiles (dashed red lines) of the perceived total strategy-level DRS. It also plots empirical estimates of the total strategy-level DRS—$0.799/M_{st} + 0.013$ for the three large-cap categories and $0.799/M_{st} + 0.053$ for all other categories—using a solid black line. *Notes:* All numbers are in units of fraction per month, associated with a \$1 trillion increase in Q_{st} (equivalently, in units of bp per month, associated with a \$100 million increase in Q_{st-1}).

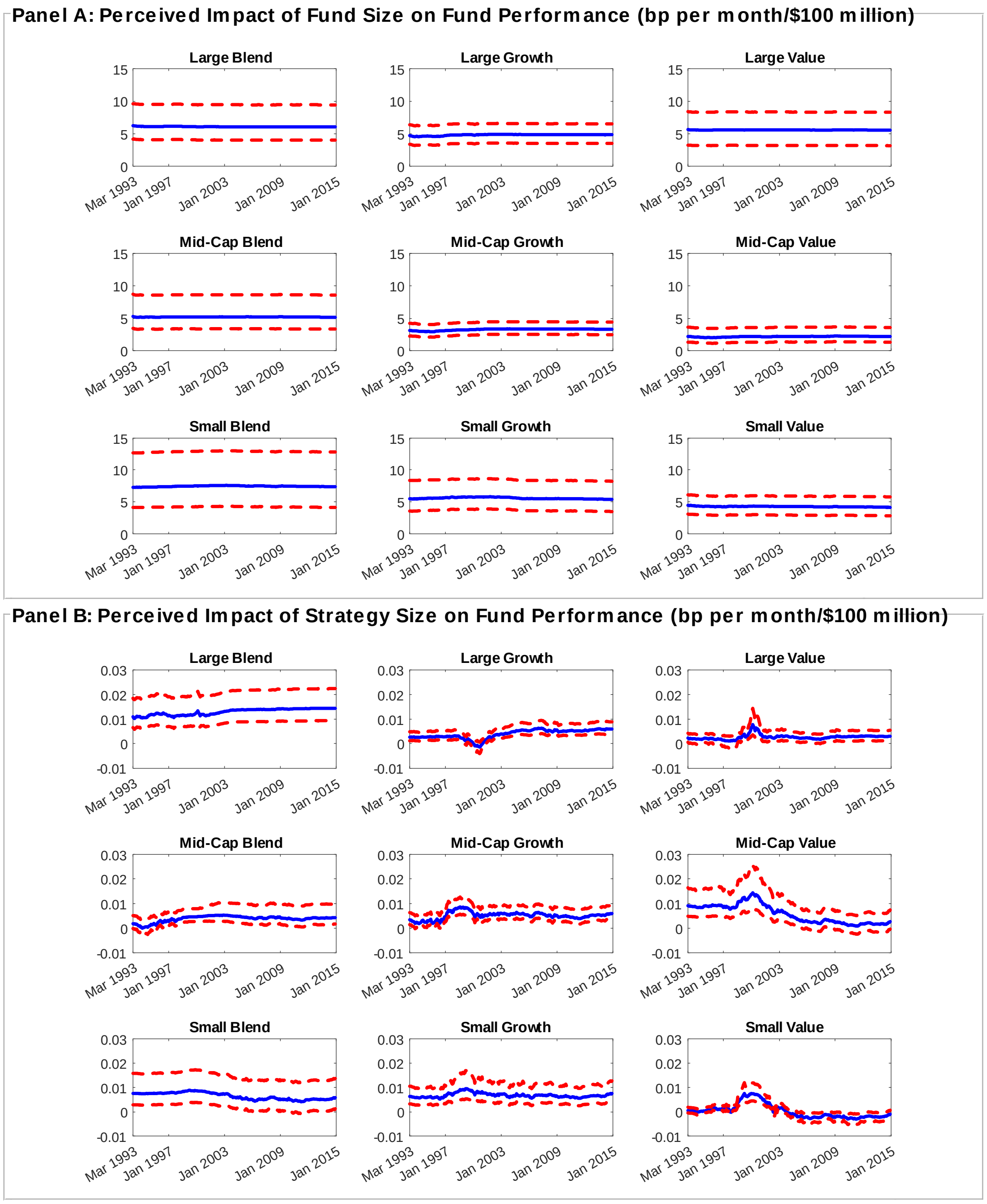

Figure 4: **Evolution of the Perceived Nature of Returns to Scale.** Panels A and B of this figure show the evolution of the perceived fund-level DRS, $\widehat{b}_{st}$, and the perceived direct strategy-level DRS, $\widehat{c}_{st}$, respectively, over time (March 1993–December 2014) for each fund category. *Notes:* All $\widehat{b}_{st}$ and $\widehat{c}_{st}$ are in units of fraction per month, associated with a \$1 trillion increase in q_{ist-1} and Q_{st-1} (equivalently, in units of bp per month, associated with a \$100 million increase in q_{ist-1} and Q_{st-1}).

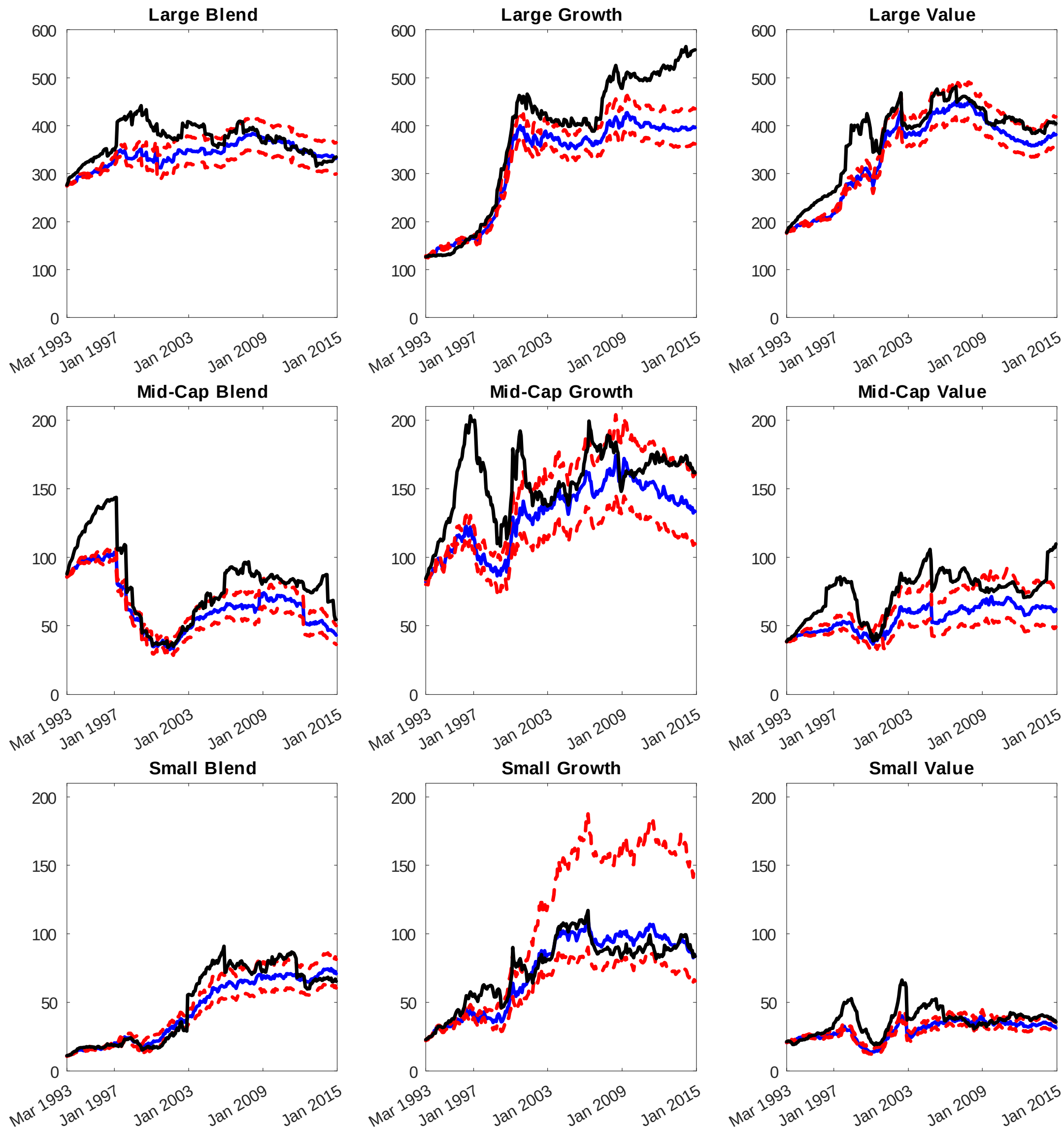

Figure 5: **The Aggregate Size Dynamics with Truly Bayesian Investors.** This figure shows the evolution of the counterfactual size Q^*_{st} of each fund category over time were investors not subject to belief shocks, plotting posterior median (the solid blue line) and 5th and 95th percentiles (the dashed red lines), computed with the draws of our posterior simulator. For comparison, the solid black line shows the actual size Q_{st} of each fund category over time. *Notes:* All Q^*_{st-1} and Q_{st-1} are in units of 2011 \$billions.

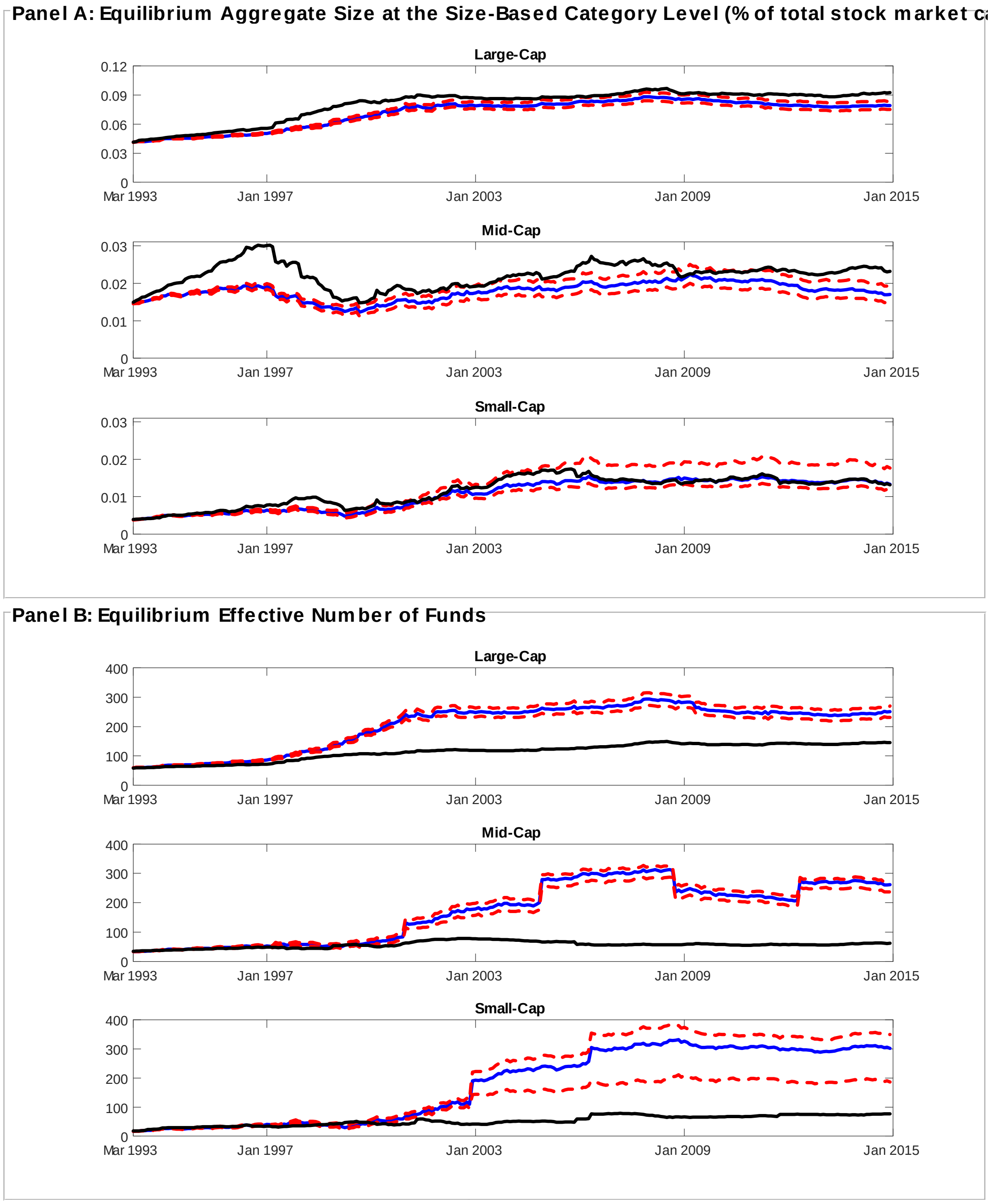

Figure 6: **The Aggregate Size Dynamics at the Size-Based Category Level.** Panel A shows the model-implied industry size at the size-based category level. Panel B shows the evolution of the model-implied effective number of funds—the reciprocal of the Herfindahl index. *Notes:* In the top panel, we scale the time series such that they represent the fraction of total stock market capitlization that the sample's mutual funds own at each point in time.

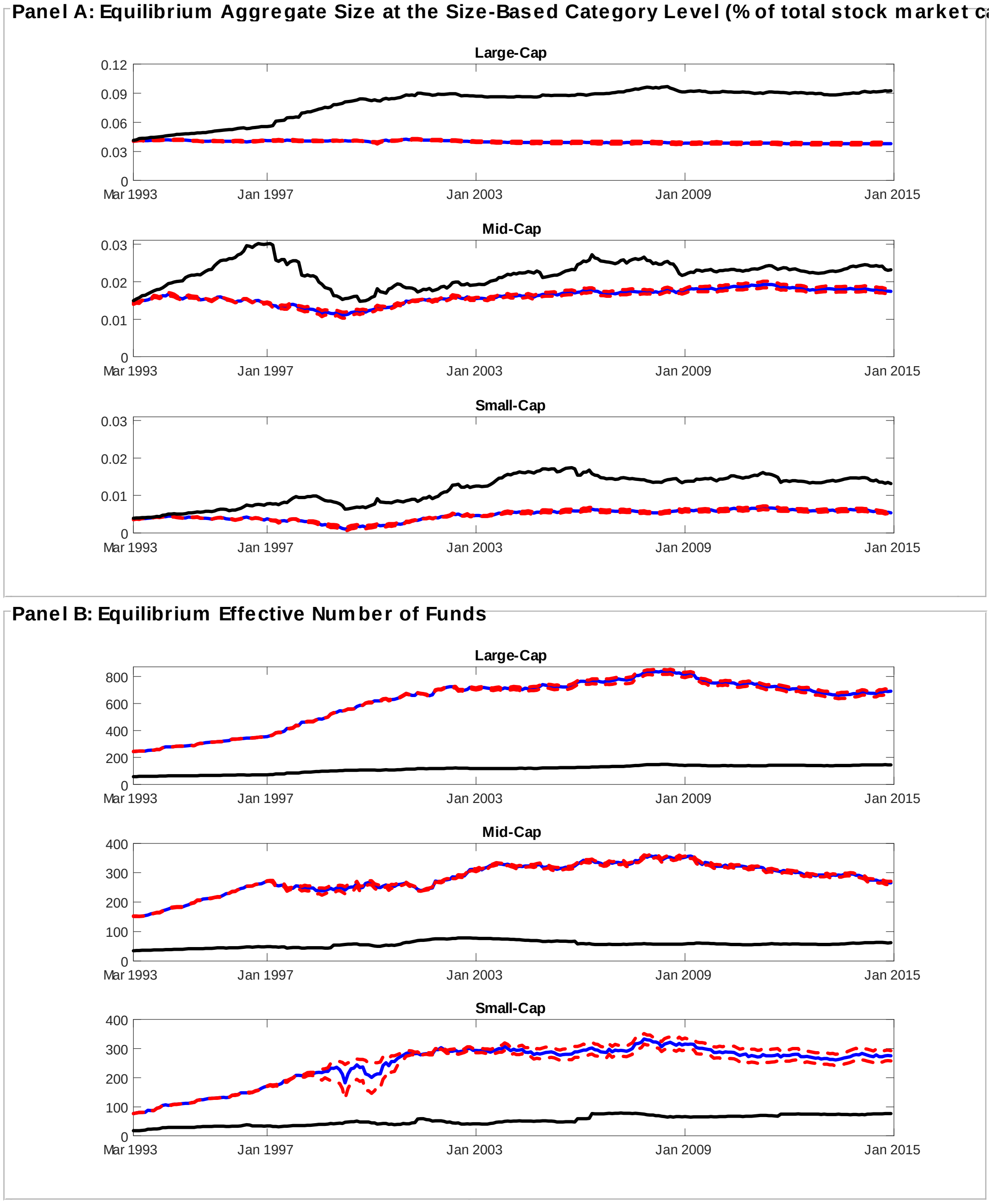

Figure 7: **The Aggregate Size Dynamics under the Simplest Alternative Specification.** This figure shows the the aggregate size dynamics for a special case of the model with $\widehat{b}_{s0} = \widehat{\sigma}_{b,s0} = 0$ and $\widehat{\rho}_{s0} = 1$. Panel A shows the model-implied industry size at the size-based category level. Panel B shows the evolution of the model-implied effective number of funds—the reciprocal of the Herfindahl index. *Notes:* In the top panel, we scale the time series such that they represent the fraction of total stock market capitlization that the sample's mutual funds own at each point in time.