

Why Has Active Asset Management Grown?

Jeong-Ho (John) Kim*

October 21, 2016

Abstract

This paper provides an explanation for the growth of active asset management, despite the industry's poor track record. I present a model in which investors rationally learn about the parameters governing fund returns in real time and optimally allocate to active management, conditional on their current beliefs. Investors' prior beliefs are characterized by optimism about fund-level parameters and pessimism about aggregate-level parameters. Industry can grow too big because learning about fund-level parameters is slower. The model accounts remarkably well for the evolution of investors' beliefs, allocation to active management, and the importance of extensive margin (more funds) in growing the industry.

*Emory University. Email: jkim230@emory.edu. I am deeply indebted to Chris Sims, Wei Xiong, and Motohiro Yogo for their guidance and advice. I am also grateful to Markus Brunnermeier, Maryam Farboodi, Valentin Haddad, Harrison Hong, Jakub Kastl, Atif Mian, Stephen Morris, Olivier Darmouni, Ji Huang, Felix Matthys, Terry Moon, Michael Sockin, and all seminar participants at Princeton University for useful suggestions and comments.

1 Introduction

One of the central facts facing financial economics is that during the last 30 years, the financial services sector has grown enormously. Research into the growth of the financial sector found that a large part of this growth came from the dramatic increase in the value of financial assets under professional management, with the total fees charged to manage these assets growing at approximately the same pace (see Greenwood and Scharfstein 2013; Malkiel 2013). One social benefit of active management is more accurate (“efficient”) securities prices, which enable firms to raise new capital at prices that better reflect their fundamental value. On the other hand, numerous studies have shown that these funds have provided investors with average returns significantly below those on passive benchmarks.¹ Although it may be socially beneficial for active managers to acquire information, it is puzzling that they are able to attract funds despite their underperformance. Moreover, what is less clear is whether we need nearly as much active money management as exists.

The purpose of this paper is to answer the question, why has active asset management grown? By definition, industry size fluctuates over time as a result of changes in the number of active mutual funds and changes in the average fund size. As illustrated in Figure 1, we observe that the active management industry has grown steadily over time. This industry growth coincides with sustained entry of skilled competitors. The top panel of Fig. 1 shows the number of funds over time. The number of funds increases from 104 in 1980 to 1,279 in 2014. The bottom panel of Fig. 1 plots the industry’s size as a fraction of total stock market capitalization over time. It starts at 1.7% in January 1980, peaks at 15.4% in July 2008, and finishes at 13.2% in December 2014. Indeed, the time-series correlation between industry size and the number of active mutual funds is 0.98, whereas the time-series correlation between industry size and the average fund size, however, is -0.78 . Therefore, the increase in the measure of industry size was itself driven essentially by an increase in the number of active mutual funds.

These facts raise interesting new questions. First, why has active asset management grown in spite of its historical underperformance? Second, what is the mechanism behind the almost perfect positive correlation between industry size and the number of active mutual funds?

¹See Jensen (1968), Malkiel (1995), Gruber (1996), Fama and French (2010), and others.

In order to address these questions, I develop a model of active management based on learning, which can quantitatively capture the historical fluctuations in the industry size. Interestingly, the model can rationalize the growth of actively managed mutual funds not only in spite of their historical underperformance, but also in the presence of net fund entry over time.

Recently, Pástor and Stambaugh (2012) show that the presence of industry-level decreasing returns to scale presents investors with an inference problem that can involve slow learning about the degree of decreasing returns and thus slow convergence to the correct allocation. Pástor and Stambaugh analyze this learning problem with a constant true underlying relation between the size of the active management industry and active management's alpha. They argue that the popularity of active management is not puzzling despite its poor track record if investors must arrive at the appropriate investment level by inferring empirically the effect of scale on performance. In that case, adjustment to the correct level of investment can be slow.

We argue that the simple model of Pástor and Stambaugh (2012) abstracts from various important considerations in explaining the observed steady growth in industry size. To highlight these issues, we simulate historical paths of industry size from our estimated version of the Pástor and Stambaugh model. We find that their simple model cannot qualitatively (or quantitatively) capture the historical fluctuations in the industry size.²

One reason is difference in how we measure the industry's size. They measure the active management industry's size as a fraction of the total amount managed actively and passively, whereas we measure it as a fraction of total stock market capitalization that active mutual funds own at that time. Their measure of industry size trends downward for the full sample period, and this trend is monotonic. Empirically, we find that the active management industry has grown over time, which is the stylized fact we focus on. This industry growth coincides with an even stronger growth of indexing,

²Pástor and Stambaugh (2012) do not hope that their simple model can quantitatively capture the historical fluctuations in the industry size and write that "our model delivers a year-by-year dependence between [the active return] r_A and [the equilibrium active allocation] (S/W), generally implying that an unexpectedly high r_A in a given year causes a higher (S/W) going into the next year. In principle, one could also look for this dependence in the year-by-year historical data in table 1, but we do not believe that such an exercise would be very informative. Indexing was novel when it emerged on the investment landscape during the 1970s. Understanding subsequent year-by-year fluctuations in its share relative to active management must surely have much to do with the dissemination and adoption of financial innovation, which we cannot hope to capture in our simple model" (p. 771), so the fact that their model deliver a rather poor fit of the data is understandable.

which explains why the time trend in industry size depends on how we measure it. While active management's negative track record relative to passive benchmarks and investors' growing awareness of indexing render the reduction of actively managed funds relative to index funds not quite puzzling, we consider its growth relative to the stock market the real "active-management puzzle".

On the other hand, there also are economically significant reasons behind these shortcomings. Key to overcoming these shortcomings is to realize that the active management industry has faced a sustained entry of fresh competitors, and the implications of this fact on the growth of active management industry depends critically on the nature of learning problem that investors face with respect to not only the heterogeneity in skills across funds, but also the nature of decreasing returns to scale.

Incorporating heterogeneity in skills, carefully modeling returns to scale, and learning about them help us understand the growing popularity of active management. Investors are uncertain about parameters governing active funds' alphas, and they learn about them from realized returns. After observing an active fund's negative performance, investors infer that it is due to either the particular fund's skill being lower than expected or the value of active investing more generally being lower than expected.

If investors believe *a priori* that fund skills are different across funds, i.e., fund skills are not perfectly correlated, they will interpret the fund's underperformance as evidence that this fund manager lacks skill rather than that the industry as a whole lacks skill. At the same time, the sustained entry of new funds effectively mutes away the negative expectations about incumbent funds' idiosyncratic parameters. Effectively, investors' conditional expectations about the average fund skill do not fluctuate much and if their subjective prior expectations were optimistic, this aspect of posterior beliefs stay optimistic, at the expense of disappointment of funds in the past.

Decreasing returns to scale at the fund level also help us understand the growing popularity of active management. In the presence of fund-level decreasing returns to scale, the equilibrium industry size increases as the number of funds increases when investors' beliefs about the parameters governing expected gross returns at an aggregated level are held constant. In the limit with infinite number of managers, the industry's fully competitive equilibrium size is determined by the posterior means of the average of the fund fixed effects (net of proportional costs and managerial com-

pensation) across all funds operating in that period in each sector and the posterior means of sector-level decreasing returns to scale for each sector. If investors' prior beliefs about fund fixed effects were more optimistic and they recognized the effects of fund sizes on returns, the industry's fully competitive equilibrium size would be too big and the entry of new funds over time induces a steady growth in industry size. Taken together, sustained entry of new funds, in the presence of investors learning about fund heterogeneity and fund-level returns to scale, will lead the industry to grow bigger.

To explore the quantitative implications of the above mechanisms, we develop a model of active management featuring investors who competitively supply funds to managers based on their perceptions of skill and decreasing returns to scale. We model skill and decreasing returns to scale, with investors learning about unknown parameters both at an aggregated level and at the fund level. We derive the model's implications for the equilibrium sizes of each segment of the active management industry, measured in relative terms. Quantitatively, our simulation exercise shows that the model succeeds in rationalizing the salient trends in the growth of active management industry in absolute terms.

We also find that the implications of net fund entry on the industry's equilibrium size depends critically on the subjective learning model of active returns that investors use to make their investment decisions. To clearly delineate this point, we will specify an exogenous entry probability and an exit process with the process parameter as functions of the lagged number of funds and time, to mimick the entry and exit observed in the data. Taking as given the entry and exit processes, we compare multiple specifications of investors' learning problem to infer the key mechanisms for the observed relationship between intensity of competition and sizes in the active management industry.

Interestingly, if investors did not have to learn and rather, they knew the true values of the parameters governing active returns, the equilibrium industry size would have been significantly smaller than the size in reality. In particular, our counterfactual exercise shows that the industry size would have been about half its actual size at the end of 2014. We interpret this result as a telling evidence that the increased complexity of investors' inference problem is the key reason why the industry has historically grown and as a suggestive evidence that this industry is too big. This, in turn, has clear policy implications.

Both our model and that of Pástor and Stambaugh (2012) build on the influential work of Berk and Green (2004). While the model of Pástor and Stambaugh assumes homogeneity across funds and in turn, is necessarily inconsistent or indeterminate with respect to cross-sectional facts on mutual funds, our model allows for heterogeneity in skill, similar to Berk and Green, so that it is at least consistent with cross-sectional facts which Berk and Green reproduce. For example, fund flows respond to past performance in my model, but not in the Pástor and Stambaugh model. In addition, as mentioned already, the increase in the number of funds has been an important driver of growth in this industry. In the Pástor and Stambaugh model, this extensive margin of growth is completely muted in driving the aggregate dynamics.

Quintessentially, our contributions are the following. First, we propose a new learning mechanism that help us understand the historic growth of active management industry relative to the stock market. Second, we show the importance of not simply industry-level decreasing returns to scale, but fund-level decreasing returns to scale in understanding the fluctuations in the active management industry size.

Taken together, our results are consistent with the following narrative. When a fund's performance turns out disappointing, it may reflect that the particular manager is incapable, or it may reflect upon the value of active management in general. Investors blame the past poor performance on the existing fund managers rather than on the notion of active management, and this, combined with the entry of new funds, prevent investors from undoing their optimistic assessment of the industry size. On the other hand, fund-level decreasing returns to scale allows the observed industry size to grow to this optimistic competitive limit size, generating the dramatic growth of assets under management. These effects would not be there, were it not for investor learning.

We are not alone in trying to explain the puzzling growth of active management in spite of its poor track record. In our story, similar to those in Berk and Green (2004) or Pástor and Stambaugh (2012), investors do not expect negative past performance to persist, but in other explanations they do. Gruber (1996) suggests that some "disadvantaged" investors are influenced by advertising and brokers, institutional restrictions, or tax considerations. Glode (2011) investors expect negative future performance as a fair trade-off for active management policies that insure investors against bad states of the economy. We do not imply that such alternative explanations

play no role in explaining the growth. We simply suggest that investor learning in the active management industry is a critical element and such learning models featuring decreasing returns to scale in recent literature are missing two interesting aspects of learning and returns to scale, which are our contribution.

A number of studies address learning about managerial skill, but none of them analyze the size of the activemanagement industry. Baks, Metrick, and Wachter (2001) analyze mutual-fund performance from an investor’s perspective and find that even extremely skeptical prior beliefs about skill would lead to economically significant allocations to active managers. Other studies that model learning about managerial skill with a focus different from ours include Lynch and Musto (2003), Berk and Green (2004), Huang, Wei, and Yan (2012), and Brown and Wu (forthcoming).

There are also a few studies using mutual fund flows to infer investor preferences, similar to how we use them to infer the parameters governing investors’ learning problem. Berk and van Binsbergen (forthcoming) and Barber, Huang, and Odean (2015) use fund flows to infer investor risk preferences and find that investors use the CAPM. While my paper and these are about investors’ subjective model, there is a large literature that estimates the objective regression models for mutual fund returns and the true distribution of managerial skills. A recent example is Pástor, Stambaugh, and Taylor (2015), which empirically analyzes the nature of returns to scale in active mutual fund management. Other examples include Chen et al. (2004), Fama and French (2010), and Ferreira et al. (2015).

Our study relates to a number of other directions in recent research. More broadly, the study adds to a growing literature addressing contentious issues related to the size of the financial industry (e.g., Philippon 2015; Bolton, Santos, and Scheinkman forthcoming). Our approach is partial equilibrium, similar to that in Berk and Green (2004), Pástor and Stambaugh (2012), and He and Xiong (2013), in the sense that asset prices are not determined endogenously in the model. On the other hand, Gârleanu and Pedersen (2015) introduce asset managers into the Grossman-Stiglitz model, so that the efficiency of asset prices is linked to the efficiency of the asset management market.³ Finally, Khorana, Servaes, and Tufano (2005) empirically analyze the determinants of the size of the mutual fund industry across countries.

³Other recent examples of studies that analyze the implications of delegated portfolio management on equilibrium asset prices include García and Vanden (2009), Cuoco and Kaniel (2011), and Guerrieri and Kondor (2012).

The paper is organized as follows. Section 2 presents our model. After describing the general model, we show how the models of Pástor and Stambaugh (2012) are obtained as special cases of our model. Section 3 describes our mutual fund dataset and sketches the estimation technique, based on Bayesian Markov chain Monte Carlo (MCMC) method. Sections 4 discuss the estimation results and provides interpretations of our results. Section 5 conducts a number of robustness checks. First, we compare the fit of our baseline model relative to alternative specifications, including the model of Pástor and Stambaugh. Second, we simulate the industry dynamics from these model specifications to study their quantitative success at explaining the data. Section 6 presents conclusions.

2 The Model

If decreasing returns to scale are driven by competition, a fund’s performance should be more closely related to the sizes and performance of funds in the same sector, which presumably follow similar investment strategies, than to the size and performance of a typical fund. To model this idea, we use the nine sectors corresponding to Morningstar’s 3×3 stylebox (small growth, mid-cap value, etc.) to label funds’ investment strategies. We assume that, for any given sector $j \in N$ and period t , $M_{j,t}$ active mutual fund managers construct (presumably similar) investment portfolios from the primitive assets in the economy. Investors reward managers by paying a given time-varying percent of assets under management every period, where $\mathbf{f}_{j,t}$ is the $M_{j,t} \times 1$ vector of rates at which managers in sector $j \in N$ at time t charge proportional fees.

All participants in the model are symmetrically informed. Funds differ in their managers’ ability to generate expected returns in excess of those provided by a passive benchmark—an alternative investment opportunity available to all investors. The model is partial equilibrium. Managers’ actions do not affect the benchmark returns, and we do not model other investors at the expense of whose decisions the managers’ potential outperformance comes. A manager’s ability to beat this benchmark is unknown to both the manager and investors, who learn about this ability by observing the history of returns of the fund and other funds in the same sector. Let

$$R_{i,j,t+1} = \alpha_{i,j,t} + u_{i,j,t+1} \tag{1}$$

denote the return, in excess of the passive benchmark, on actively managed fund i in sector j , without costs and fees. This is not the return actually earned and paid out by the fund, which is net of costs and fees (see below). The fund alpha $\alpha_{i,j,t}$ is the source of differential performance across funds at time t . The error terms, $u_{i,j,t+1}$, have the following factor structure:

$$u_{i,j,t+1} = x_{j,t+1} + \epsilon_{i,j,t+1} \quad (2)$$

for $i \in M_{j,t}, j \in J$, where all $\epsilon_{i,j,t+1}$'s have a mean of zero, a variance of $\sigma_{j,\epsilon}^2$, and zero correlation with each other. The common factor $x_{j,t+1}$ for sector j has mean zero and variance $\sigma_{j,x}^2$. The values of $\sigma_{j,x}$ and $\sigma_{j,\epsilon}$ are constants known to both investors and managers.

The factor structure in equation (2) means that the benchmark-adjusted returns of skilled managers in the same sector are correlated as long as $\sigma_{j,x} > 0$ (which, indeed, is empirically true). Multiple skilled managers who follow similar investment strategies are likely to identify the same opportunities, so multiple managers in the same sector are likely to hold some of the same positions, resulting in correlated benchmark-adjusted returns among funds belonging to the same sector. As a result, investors' posteriors on the abilities of active managers in the same sector will exhibit more and more correlation over time.

Participants learn about the parameters governing equilibrium alphas by observing the realized excess returns the managers produce. This learning is the source of the relationship between performance and the flow of funds as in several recent models of learning about managerial skill.

Assume that fund manager i in sector j is simply paid a fixed management fee, $f_{i,j,t+1}$, expressed as a fraction of the assets under management, $q_{i,j,t}$. (We shall discuss further this assumption shortly.) Managers accordingly accept and invest all the funds investors are willing to allocate to them. The amount investors will invest in the fund depends on their subjective assessment of the managers' ability and on the costs they perceive managers face in expanding the fund's scale.

The excess total payout to investors over what would be earned on the passive benchmark is

$$TP_{i,j,t+1} = q_{i,j,t} (R_{i,j,t+1} - f_{i,j,t+1}).$$

Let $r_{i,j,t}$ denote the excess return over the benchmark that investors in fund i in sector

j receive in period t . Then

$$r_{i,j,t+1} = \frac{TP_{i,j,t+1}}{q_{i,j,t}} = R_{i,j,t+1} - f_{i,j,t+1}. \quad (3)$$

The return $r_{i,j,t+1}$ corresponds to the return empirically observed.

We assume that investors supply capital with infinite elasticity to funds that have positive excess expected returns from their subjective perspective. This can be justified as long as investors are all risk neutral. Our assumption of risk neutrality is conservative in that, given active management's popularity despite its historical underperformance, allowing for risk aversion and estimating the risk-aversion parameter would presumably lead to risk neutrality anyhow, if not risk seeking. Similarly, they remove all funds from any fund that has a negative excess expected return from their subjective perspective. At each point in time, then, funds flow to and from each fund so that the expected excess return to investing in any surviving fund is zero with respect to investors' subjective probability distribution of next period's returns:

$$\widehat{E}_t(r_{i,j,t+1}) = 0. \quad (4)$$

The circumflex on the expectation operator indicates that the expectation is taken with respect to the probability distribution generated by investors' perceived model.

We can complete the description of a learning model of active management by specifying the subjective model of next period's returns (and fees), which investors base their investment decisions on and whose parameters they learn about.

2.1 The Benchmark Model: Pástor and Stambaugh (2012)

We begin by describing the model of Pástor and Stambaugh (2012) that we take as the benchmark model against which to evaluate our general model. We shall then combine the model with two new elements, which are necessary to capture the quantitative trends in industry-size time series and to reproduce the time-series correlation between the number of funds and industry size.

Pástor and Stambaugh model decreasing returns to scale as follows:

$$\alpha_{i,j,t} = a_j - b_j \left(\frac{Q}{W} \right)_{j,t}, \quad (5)$$

where $\alpha_{i,j,t}$ is the expected return gross of fees and costs at time t in excess of passive benchmarks generated by fund i in sector j and $(Q/W)_{j,t}$ is the sector size as a fraction of total stock market capitalization. Sector-level decreasing returns to scale are captured by $b_j > 0$. The parameters a_j and b_j in equation (5) are unknown. We denote their (subjective) first and second conditional moments by

$$\widehat{E} \left(\begin{bmatrix} a_j \\ b_j \end{bmatrix} \middle| D_{t-1} \right) = \begin{bmatrix} \phi_{j,t-1} \\ \widetilde{b}_{j,t-1} \end{bmatrix}, \quad (6)$$

$$\widehat{Var} \left(\begin{bmatrix} a_j \\ b_j \end{bmatrix} \middle| D_{t-1} \right) = \begin{bmatrix} \sigma_{a,j,t-1}^2 & \sigma_{ab,j,t-1} \\ \sigma_{ab,j,t-1} & \sigma_{b,j,t-1}^2 \end{bmatrix}, \quad (7)$$

where D_{t-1} denotes the set of information available to investors at time $t - 1$.

The parameter a_j represents the expected return on the initial small fraction of wealth invested in sector j of the active management industry, without proportional costs and managerial compensation. It seems likely that $a_j > 0$, although we do not preclude $a_j < 0$. If no money were invested in sector j , some opportunities to outperform the passive benchmarks by following a typical investment strategy in sector j would likely be present, so it is likely to have a positive expected benchmark-adjusted return.

The parameter b_j determines the degree to which the expected benchmark-adjusted return for any fund declines as the relative size of the fund's sector increases. We allow $b_j > 0$ to capture decreasing returns to scale at the sector level. As more money chases opportunities to outperform, prices move (unless markets are perfectly liquid), making such opportunities more elusive. If decreasing returns to scale at an aggregated level are driven by competition with other funds, a fund's performance should be related to the size of the fund's sector rather than the size of the entire industry.

Their primary focus is on the fully competitive case ($M_{j,t} \rightarrow \infty$). They compare the model-implied equilibrium size of the active management industry in the fully competitive case with the actual size.

Let $r_{j,t+1}$ denote the benchmark-adjusted net return on the aggregate portfolio of all funds:

$$r_{j,t+1} = a_j - b_j \left(\frac{Q}{W} \right)_{j,t} - f_{j,t+1} + x_{j,t+1} + \frac{1}{M_{j,t}} \sum \epsilon_{i,j,t+1}, \quad (8)$$

using (1), (2), and (3). Thus, as $M_{j,t} \rightarrow \infty$,

$$r_{j,t+1} = a_j - b_j \left(\frac{Q}{W} \right)_{j,t} - f_{j,t+1} + x_{j,t+1} \quad (9)$$

since the variance of the last term in (8) goes to zero. It follows from (9) that

$$\widehat{E}(r_{j,t+1}|D_t) = \phi_{j,t} - \widetilde{b}_{j,t} \left(\frac{Q}{W} \right)_{j,t} \quad (10)$$

Equation (4) can then be rewritten as

$$\left(\frac{Q}{W} \right)_{j,t} = \frac{\phi_{j,t}}{\widetilde{b}_{j,t}} \quad (11)$$

We specify a bivariate normal joint prior distribution for a_j and b_j :

$$\begin{bmatrix} a_j \\ b_j \end{bmatrix} \sim N(E_{j,0}, V_{j,0}), \quad (12)$$

where $N(E_{j,0}, V_{j,0})$ denotes a bivariate normal distribution with mean $E_{j,0}$ and covariance matrix $V_{j,0}$. Denote

$$E_{j,0} = \begin{bmatrix} \phi_{j,0} \\ \widetilde{b}_{j,0} \end{bmatrix}, \quad V_{j,0} = \begin{bmatrix} \sigma_{a,j,0}^2 & \sigma_{ab,j,0} \\ \sigma_{ab,j,0} & \sigma_{b,j,0}^2 \end{bmatrix}. \quad (13)$$

To capture beliefs that the industry faces decreasing returns to scale, we specify $\widetilde{b}_{j,0} > 0$. We also specify $\sigma_{ab,j,0} = 0$ for simplicity.

These moments are updated by using standard results for the conditional distributions of a multivariate normal

$$V_{j,t} = V_{j,t-1} - V_{j,t-1} Z'_{j,t-1} (Z_{j,t-1} V_{j,t-1} Z'_{j,t-1} + \sigma_x^2)^{-1} Z_{j,t-1} V_{j,t-1}, \quad (14)$$

$$E_{j,t} = E_{j,t-1} + V_{j,t-1} Z'_{j,t-1} r_{j,t} + z_{j,t}, \quad (15)$$

where $Z_{j,t-1} = \begin{bmatrix} 1 & -(Q/W)_{j,t-1} \end{bmatrix}$. Here, $z_{j,t}$ is a belief shock. Given the complexity of this industry, it is most probable that there are other factors influencing investors' perception of the parameters governing this industry beyond the simple learning mechanism described above. To capture this and to proceed with likelihood

estimation of the parameters, we encapsulate these other variations in $z_{j,t}$. Technically, it serves to avoid stochastic singularity, much like measurement error or shocks do in macroeconomic DSGE models. The subjective posterior distribution of a_j and b_j is bivariate normal as in equation (12), except that $E_{j,0}$ and $V_{j,0}$ are replaced by $E_{j,t}$ and $V_{j,t}$ from equations (14) and (15).

In the model of Pástor and Stambaugh (2012), the industry dynamics is fully described by equations (9), (11), and (13)-(15). Note that the number of funds do not enter in any of these equations. Hence it follows straightforwardly that there is no relationship between the growth in the number of competitors and the growth of the industry size, which is inconsistent with the empirical fact documented in the introduction. We will now embed this simple model into our model, which can not only generate industry size growth, but also relate this growth to the growth in the number of funds.

2.2 Heterogeneity in Skill and Fund-Level Returns to Scale

Our model combine the quintessential spirit of the Pástor and Stambaugh model and two other elements. There is heterogeneity in skill across funds, which investors learn about. Finally, there is decreasing returns to scale at the fund level, which is also subject to investor learning.

Our key assumption is that $\alpha_{i,j,t}$ is decreasing in $(Q/W)_{j,t}$ and $(q/W)_{i,j,t}$, where Q_j is the sum of fund sizes across all funds within a given sector and $q_{i,j}$ is the size of fund i in sector j . Here W is equal to the total stock market capitalization in the same period.⁴ Dividing Q_j and $q_{i,j}$ by W reflects the notion that the relative (rather than absolute) sizes are relevant for capturing decreasing returns to scale in active management. In order to obtain closed-form equilibrium results, we assume

⁴An earlier version of the model included decreasing returns at the industry level. Empirically, we do not find evidence of decreasing returns at the industry level. The estimation results indicate that industry-level decreasing returns to scale play an insignificant role in determining the industry's size. All things considered, the addition of decreasing returns to scale at the industry level do not affect the conclusions discussed in this paper.

Both sector-level and industry-level decreasing returns to scale are plausible hypotheses of how a fund's performance could depend on other funds, due to competition among active funds. Moreover, these alternative hypotheses are not mutually exclusive. These results suggest essentially that our proxy for sector size accurately measures the size of a fund's competition.

the functional relation

$$\alpha_{i,j,t} = a_{i,j} - b_j \left(\frac{Q}{W} \right)_{j,t} - c_j \left(\frac{q}{W} \right)_{i,j,t}, \quad (16)$$

with $S_{j,t} = \sum_{i \in M_{j,t}} s_{i,j,t}$.⁵ Analogously, S is the aggregate size of the active management industry, with $S_t = \sum_{j \in J} S_{j,t}$. The parameters $a_{i,j}$ and b_j, c_j in equation (16) are unknown. We denote their first and second conditional moments by

$$\begin{aligned} \mathbb{E} \left(\begin{bmatrix} \{a_{i,j}\}_{i \in M_{j,t-1}} \\ b_j \\ c_j \end{bmatrix} \middle| D_{t-1} \right) &= \begin{bmatrix} \{\phi_{i,j,t-1}\}_{i \in M_{j,t-1}} \\ \tilde{b}_{j,t-1} \\ \tilde{c}_{j,t-1} \end{bmatrix}, \quad (17) \\ \text{Var} \left(\begin{bmatrix} \{a_{i,j}\}_{i \in M_{j,t-1}} \\ b_j \\ c_j \end{bmatrix} \middle| D_{t-1} \right) &= \begin{bmatrix} \{\sigma_{i_1 i_2, j, t-1}\}_{i_1, i_2 \in M_{j,t-1}} & \{\sigma_{bi, j, t-1}\}_{i \in M_{j,t-1}} & \{\sigma_{ci, j, t-1}\}_{i \in M_{j,t-1}} \\ \{\sigma_{bi, j, t-1}\}_{i \in M_{j,t-1}} & \sigma_{b, j, t-1}^2 & \sigma_{bc, j, t-1} \\ \{\sigma_{ci, j, t-1}\}_{i \in M_{j,t-1}} & \sigma_{bc, j, t-1} & \sigma_{c, j, t-1}^2 \end{bmatrix}, \quad (18) \end{aligned}$$

where D_{t-1} denotes the set of information available to investors at time $t - 1$.

The parameter $a_{i,j}$ represents, thus, the expected return on the initial small fraction of wealth invested in actively managed fund i in sector j , without proportional costs and managerial compensation. Again, it seems likely that $a_{i,j} > 0$, although we do not preclude $a_{i,j} < 0$.⁶

The parameter c_j determines the degree to which the expected benchmark-adjusted return for any fund in sector j declines as the relative size of the fund increases. We allow $c_j > 0$ to capture decreasing returns to scale at the fund level. Like sector-level decreasing returns to scale, fund-level decreasing returns to scale has been motivated by liquidity constraints. At the fund level, a larger fund's trades have a larger impact on asset prices, eroding the fund's performance. Like the parameter b_j governing sector-level returns to scale, we assume that the costs associated with fund-level de-

⁵We specify the relation (??) exogenously, but decreasing returns to aggregate scale can also arise endogenously in a richer model. In canonical rational-expectations equilibrium models of financial markets (see, e.g., Grossman and Stiglitz 1980; Garleanu and Pedersen 2015), agents choose whether to pay the cost of becoming informed, and the benefit from informed trading is decreasing in the proportion of informed traders, just like equation (??).

⁶If no money were invested in a skilled active fund, some opportunities (to outperform the passive benchmarks) would likely be present for that manager whoever he is (given that he essentially is an informed trader), so it is likely to have a positive expected benchmark-adjusted return.

creasing returns to scale faced by the funds are subject to investor learning. We allow learning about b_j so as to embed the Pastor and Stambaugh model. Our novel mechanism is how learning about fund-specific factors interact with learning about returns to scale, so we build a sufficiently flexible model which also accomodates the Pastor and Stambaugh model as a special case and let the data speak for itself the prominence of this aspect of investor learning.

Assume that the fund managers do not optimally choose the fee they charge at each point in time and, instead, are simply paid an exogenous time-varying management fee, $f_{i,j,t}$, expressed as a fraction of the assets under management, $q_{i,j,t}$. Each element of $\mathbf{f}_{j,t}, j \in N$ evolves (independently) according to the following AR(1) stochastic processes:

$$\log f_{i,j,t} = \rho_f \log f_{i,j,t-1} + \nu_{i,j,t}, \quad i \in M_{j,t}, j \in N, \quad (19)$$

where $\nu_{i,j,t} \sim N(0, \sigma_\nu^2)$. As discussed in Berk and Green (2004), the fixed fee contract comes closer to the institutional setting for retail mutual funds, which is the source of data for our empirical analysis. Empirical evidence suggests that mutual funds show relatively little variation in fees, through time and across funds.⁷ On the basis of monthly data for the January 1993–December 2014 period, the natural logarithm of fund expense ratio in CRSP has a beta of 0.98 with respect to its one-month-lag variables. Our assumption of the exogenous time-varying expense ratio throughout the sample is clearly appropriate in that the R -squared (R^2) for the regression is 0.99.

Then

$$r_{i,j,t+1} = a_{i,j} - k_j \left(\left(\frac{Q}{W} \right)_{j,t}, \left(\frac{q}{W} \right)_{i,j,t} \right) - f_{i,j,t+1} + u_{i,j,t+1}, \quad (20)$$

where

$$k_j \left(\left(\frac{Q}{W} \right)_{j,t}, \left(\frac{q}{W} \right)_{i,j,t} \right) = b_j \left(\frac{Q}{W} \right)_{j,t} + c_j \left(\frac{q}{W} \right)_{i,j,t} \quad (21)$$

denotes the extent to which the fund's gross alpha is eroded by decreasing returns to scale. For any given sector j , investors learn about $\{a_{i,j}\}_{i \in M_{j,t}}$ and b_j, c_j by observing realized returns of and active allocations to all funds belonging to that sector.

We do not impose the assumption that investors have rational expectations. At

⁷Christoffersen (2001) describes both historical practice and regulatory constraints which limit the ability of retail mutual funds to use performance-based fees.

the birth of a fund in sector j at time t , the participants' prior about the skill of a typical fund in that sector is that $a_{i,j}$ is normally distributed with mean $\phi_{0,j,t}$ and variance $\sigma_{0,j,t}^2$. At the birth of sector j , they also believed *a priori* that the skills of two different funds in the same sector are correlated such that $\widehat{Corr}(a_{h,j}, a_{i,j}) = \rho_{j,0}$. Since investors are not restricted to have rational expectations, this is not necessarily the distribution of fund-specific skills across new funds. Investors (and the managers) update their posteriors on the basis of the history of observed returns as Bayesians. In particular, if investors believed a priori that fund skills are uncorrelated, the performance of old funds has no implications on the skills of newly entering funds so that $\phi_{0,j,t}$ and $\sigma_{0,j,t}^2$ do not fluctuate over time. On the other hand, if investors believed a priori that fund skills are correlated, they update their expectations about newly entering funds based on the performance of old funds so that we need to put the time subscript on investors' prior about the fund skill at time t . Let the posterior mean of management ability of incumbent fund i in sector j at time t be denoted

$$\phi_{i,j,t} \equiv \widehat{E}(a_{i,j}|D_t).$$

The timing convention is as follows.

1. Each incumbent fund $i \in M_{j,t-1}$ in sector $j \in N$ enters period t with $q_{i,j,t-1}$ funds under management and subjective estimates of parameters governing fund returns, $\left\{ \phi_{0,j,t-1}, \left\{ \phi_{i,j,t-1} \right\}_{i \in M_{j,t-1}}, \tilde{b}_{j,t-1}, \tilde{c}_{j,t-1} \right\}_{j \in N}$.
2. Managers and investors observe $\{r_{i,j,t}, f_{i,j,t}\}_{i \in M_{j,t-1}, j \in N}$ (from which they can infer $\{R_{i,j,t}\}_{i \in M_{j,t-1}, j \in N}$) and update their estimates of the parameters governing fund returns by calculating $\left\{ \phi_{0,j,t}, \left\{ \phi_{i,j,t} \right\}_{i \in M_{j,t-1}}, \tilde{b}_{j,t}, \tilde{c}_{j,t} \right\}_{j \in N}$.
3. Each incumbent fund $i \in M_{j,t-1}$ in sector $j \in N$ makes exit decision.
4. The number of entering funds in each sector $j \in N$ is determined, and the set of time- t incumbents, $M_{j,t}$, is determined.
5. Belief shocks $\{z_{i,j,t}\}_{i \in M_{j,t}, j \in N}$ are drawn for each incumbent funds at time t .
6. Then capital flows into or out of existing funds to determine $\left\{ (q/W)_{i,j,t} \right\}_{i \in M_{j,t-1}, j \in N}$.

Next, we shall calculate this flow of funds explicitly, and derive from the model's equilibrium relationships a likelihood function which the estimation is based on.

Recall that there is competitive provision of capital by investors to mutual funds, imposing the participation condition, (4).

Taking expectations of both sides of (20), and requiring expected excess returns of zero as in (4), gives

$$\phi_{i,j,t} - E_t [f_{i,j,t+1}] = E_t \left[k_j \left(\left(\frac{Q}{W} \right)_{j,t}, \left(\frac{q}{W} \right)_{i,j,t} \right) \right] = \tilde{b}_{j,t} \left(\frac{Q}{W} \right)_{j,t} + \tilde{c}_{j,t} \left(\frac{q}{W} \right)_{i,j,t}, \quad (22)$$

As $\left\{ \left\{ \phi_{i,j,t} \right\}_{i \in M_{j,t}}, \tilde{b}_{j,t}, \tilde{c}_{j,t} \right\}_{j \in N}$ change, $\left\{ (q/W)_{i,j,t} \right\}_{i \in M_{j,t}}$ change to ensure that this equation is satisfied for all funds at all points in time.

In equilibrium, then $(Q/W)_{j,t}$ is given by the (unique) real positive solution to the equation

$$\frac{1}{M_{j,t}} \sum_{i \in M_{j,t}} (\phi_{i,j,t} - E_t [f_{i,j,t+1}]) = \left(\tilde{b}_{j,t} + \tilde{c}_{j,t}/M_{j,t} \right) \left(\frac{Q}{W} \right)_{j,t} \quad (23)$$

Equation (23) can then be rewritten as

$$\begin{aligned} \left(\frac{Q}{W} \right)_{j,t} &= \frac{\frac{1}{M_{j,t}} \sum_{i \in M_{j,t}} (\phi_{i,j,t} - E_t [f_{i,j,t+1}])}{\tilde{b}_{j,t} + \tilde{c}_{j,t}/M_{j,t}}, \\ &= \frac{\phi_{j,t} - E_t [f_{j,t+1}]}{\kappa_{j,t}}, \end{aligned}$$

where

$$\begin{aligned} \phi_{j,t} &= \frac{1}{M_{j,t}} \sum_{i \in M_{j,t}} \phi_{i,j,t}, \\ E_t [f_{j,t+1}] &= \frac{1}{M_{j,t}} \sum_{i \in M_{j,t}} E_t [f_{i,j,t+1}], \\ \kappa_{j,t} &= \tilde{b}_{j,t} + \tilde{c}_{j,t}/M_{j,t}. \end{aligned}$$

Similarly, we can compute individual fund sizes by rewriting (22),

$$\left(\frac{q}{W}\right)_{i,j,t} = \frac{1}{\tilde{c}_{j,t}} \left(\phi_{i,j,t} - E_t [f_{i,j,t+1}] - \frac{\tilde{b}_{j,t}}{\kappa_{j,t}} (\phi_{j,t} - E_t [f_{j,t+1}]) \right). \quad (24)$$

Recall that investors' priors (before drawing belief shock) is a multivariate normal distribution. These prior specifications, together with the fact that each active return is a linear transformation of the parameter vector, imply that investors' posteriors at each point of time is also a multivariate normal distribution.

The moments for $\{a_{i,j}\}_{i \in M_{j,t-1}}$ and b_j, c_j (i.e., fund-specific skills for funds that operated at time $t - 1$ and the parameters governing returns to scale) are updated using standard results for the conditional distributions of a multivariate normal

$$\text{Var} \left(\begin{bmatrix} \{a_{i,j}\}_{i \in M_{j,t-1}} \\ b_j \\ c_j \end{bmatrix} \middle| D_t \right) = V_{j,t-1} - V_{j,t-1} Z'_{j,t-1} V_{R,t-1}^{-1} Z_{j,t-1} V_{j,t-1} \quad (25)$$

$$\begin{aligned} \text{E} \left(\begin{bmatrix} \{a_{i,j}\}_{i \in M_{j,t} \cap M_{j,t-1}} \\ \{a_{i,j}\}_{i \in M_{j,t} \setminus M_{j,t-1}} \\ b_j \\ c_j \end{bmatrix} \middle| D_t \right) &= E_{j,t-1} + V_{j,t-1} Z'_{j,t-1} V_{R,t-1}^{-1} (\mathbf{R}_{j,t} - E_{t-1}(\mathbf{f}_{j,t})) \\ &+ \begin{bmatrix} \{z_{i,j,t}\}_{i \in M_{j,t} \cap M_{j,t-1}} \\ \{z_{i,j,t}\}_{i \in M_{j,t} \setminus M_{j,t-1}} \\ 0 \\ 0 \end{bmatrix} \end{aligned} \quad (26)$$

where $Z_{j,t-1} = \begin{bmatrix} \iota_{M_{j,t-1}} & \left[1 \quad -(Q/W)_{j,t-1} \right] \\ & I_{M_{j,t-1}} \end{bmatrix}$ and

$$V_{R,t-1} = Z_{j,t-1} V_{j,t-1} Z'_{j,t-1} + \sigma_x^2 \iota_{M_{j,t-1}} \iota'_{M_{j,t-1}} + \sigma_\epsilon^2 I_{M_{j,t-1}}. \quad (27)$$

2.3 Entry and Exit of Funds

Our primary focus is on the nature of investors' inference problem in this industry with respect to skill heterogeneity and the nature of decreasing returns to scale. Indeed,

depending on the particular specification of this investor learning model, implications of fund entry and exit on the equilibrium industry size dynamics changes substantially. In order to focus on the roles of investor learning in shaping the active management industry's popularity independent of the entry and exit dynamics, we will keep the entry and exit processes simple and exogenous.

Our focus is on analyzing the nature of investors' subjective learning model that decomposes the growth of the industry size into fluctuations in its two components, the number of funds and the average fund size. To do this, we will assume that entry and exit are exogenous stochastic processes and determine the industry and average fund sizes in equilibrium.

At the beginning of time t , each incumbent fund $i \in M_{j,t-1}$, which was in operation at time $t-1$, will make an exit with probability p_j . At the same time, a cohort of $N_{j,t}$ new funds enter the industry. For simplicity, we will assume that $N_{j,t}$ is a Poisson random variable with intensity

$$\log(\lambda_{j,t}) = \beta_{j,0} + \beta_{j,1}t + \beta_{j,2}t^2 + \beta_{j,M} \log(M_{j,t-1}).$$

We will estimate the coefficients $\beta_{j,0}, \beta_{j,1}, \beta_{j,2}, \beta_{j,M}$ by running Poisson regressions in our data.

This specification for entry and exit assumes that entry and exit processes are not affected by fund returns or fund sizes. In particular, it simplifies our problem substantially, for it allows us to separately estimate the entry-exit processes and the parameters governing investors' learning problem. This allows us to study investors' learning problem in isolation, and the specification of entry and exit becomes important only when we do quantitative simulations. There, we will see that this simple specification captures reasonably the entry and exit time series observed in the data.

2.4 Likelihood Function

The model allows us to estimate the parameters regarding the evolution of investors' subjective beliefs separately from other model parameters. This involves two sets of parameters. First, the parameters governing the first and second moments of investors' prior beliefs, i.e., $\left\{ \phi_{0,j,0}, \tilde{b}_{j,0}, \tilde{c}_{j,0}, \sigma_{a,j,0}, \sigma_{b,j,0}, \sigma_{c,j,0}, \rho_{0,j,0} \right\}_{j \in N}$. Second, standard deviations of the belief shocks, $\left\{ \sigma_{z,j,0}, \rho_{z,j,0} \right\}_{j \in N}$. Collectively, I will denote this vector of parameters of the model as θ . Our goal is to evaluate the likelihood function

of a sequence of realizations of the sector sizes $\left\{ (Q/W)_{j,t} \right\}_{j \in N, t=1, \dots, T}$ to estimate the parameter vector θ :

$$L \left(\left\{ (Q/W)_{j,t} \right\}_{j \in N, t=1, \dots, T}; \theta \right) = p \left(\left\{ (Q/W)_{j,t} \right\}_{j \in N, t=1, \dots, T}; \theta \right)$$

Note that

$$\left(\frac{Q}{W} \right)_{j,t} = \frac{\tilde{E}_t [a_j] - E_t [f_{j,t+1}]}{\kappa_{j,t}} + z_{j,t}$$

where $z_{j,t} = \sum_{i \in M_{j,t}} z_{i,j,t} / M_{j,t}$ is the aggregated belief shock for sector j at time t . On the right hand side, all terms, except the aggregated belief shock, are pinned down after the realizations of net returns and fees at the beginning of time t . Quintessentially,

$$p \left(\left\{ \left(\frac{Q}{W} \right)_{j,t} \right\}_{j \in N} \mid D_{t-1}, \theta \right) = \prod_{j \in N} p(z_{j,t}; \theta)$$

and in turn,

$$p \left(\left\{ (Q/W)_{j,t} \right\}_{j \in N, t=1, \dots, T}; \theta \right) = \prod_{t=1}^T p \left(\left\{ \left(\frac{Q}{W} \right)_{j,t} \right\}_{j \in N} \mid D_{t-1}, \theta \right) = \prod_{t=1}^T \prod_{j \in N} p(z_{j,t}; \theta).$$

3 Inference

3.1 The Data

The data come from CRSP and Morningstar. The sample contains 2,922 actively managed domestic equity-only mutual funds from the United States between 1980 and 2014. Pástor, Stambaugh, and Taylor (2015) create a cross-validated dataset of actively managed US equity mutual funds, building on the work of Berk and van Binsbergen (2014). We follow closely the Data Appendix to Pástor, Stambaugh, and Taylor (2015) in creating our dataset, reconciling between CRSP and Morningstar the key data items: returns and fund size.

We now define the variables used in our analysis. Summary statistics are in Table 1.

Our measure of fund performance is $NetR$, the fund's monthly benchmark-adjusted net return, which corresponds to $r_{i,j,t}$ in Section 2. $NetR$ equals the fund's gross

return minus its monthly expense ratio minus the return on the benchmark index portfolio for actively managed US equity mutual funds. We take expense ratios from CRSP because Morningstar is ambiguous about their timing. The average of $NetR$ is -9 bps per month, and the average benchmark-adjusted gross return is $+1$ bps per month.

The benchmark against which we judge a fund’s performance is the CRSP value-weighted market portfolio. For the market portfolio to represent a fair benchmark for active funds, we need to take into account the small but nontrivial cost of holding the market. We do so by subtracting 15 basis points from each annual market return.⁸ We follow Pástor and Stambaugh (2012) in choosing our benchmark.

We construct $NetR$ by subtracting the index benchmark return from the fund’s gross return, effectively assuming that the fund’s benchmark beta is equal to one.⁹

$FundSize$ corresponds to $(q/W)_{i,j,t-1}$ in Section 2. $FundSize$ equals the fund’s AUM at the end of the previous month, divided by the total market value of all stocks at the end of the previous month. We fill in missing values of $FundSize$ by taking the fund’s most recent reported size and updating it by using interim realized total fund returns. There is considerable dispersion in $FundSize$: The coefficient of variation is 344%.

We measure $SectorSize$ by adding up fund sizes across all funds within a given sector, divided by the total market value of all stocks (i.e., the sum of $FundSize$ across all sample funds within a given sector). We use the nine sectors corresponding to Morningstar’s 3×3 stylebox (small growth, mid-cap value, etc.). The number of funds in these sectors ranges from 88 in small value to 479 in large growth.

$IndustrySize$ is the sum of AUM across all funds in our sample, divided by the total market value of all stocks (i.e., the sum of $FundSize$ across all sample funds). It is the fraction of total stock market capitalization that the sample’s mutual funds own at that time.

The variables defined above— $NetR$, $ExpenseRatio$, $FundSize$, $SectorSize$, and $IndustrySize$ —are the main variables used in our maximum likelihood estimation of the model presented in Section 2.

The average pairwise correlation in $GrossR$ between funds belonging to the same

⁸asdf

⁹This assumption is consistent with empirical evidence for active equity mutual funds. On the basis of monthly data for the January 1980–December 2014 period, US actively managed mutual funds have an average beta of 1.00 with respect to the value-weighted market index.

Morningstar Category is 0.17. To account for these cross-sectional correlations in our subsequent regressions, we cluster standard errors by Morningstar Category month. The average correlation between funds from different categories is only 0.02. Therefore, we do not cluster by month to avoid adding noise to standard errors.

3.2 Bayesian Inference

No serious quantitative model of the size of the active management industry, similar to dynamic equilibrium models in macroeconomics, has, as far as I know, been developed or examined. Therefore, following a growing literature in macroeconomics, we adopt a Bayesian approach to inference, integrating the sample information with weakly informative priors, which summarize additional information about the parameters (see, for instance, Levin et al. 2005, Del Negro et al. 2007, or Justiniano and Primiceri 2008). One advantage of this approach is that it ameliorates common numerical problems related to both the flatness of the likelihood function in some regions of the parameter space and the existence of multiple local maxima. On the other hand, it also allows us to pick some parameter values to ensure that the prior beliefs of investors are reasonable.

The estimation algorithm is a random walk Metropolis MCMC procedure based on An and Schorfheide (2007):

1. Use a maximization algorithm (specifically, a simulated annealing algorithm) to maximize $\ln \mathcal{L}(\theta|Y) + \ln p(\theta)$. This is done for multiple initial values drawn at random from our prior to ensure convergence of this initial search to a unique mode. Denote the posterior mode by $\tilde{\theta}$.
2. Obtain an inverse Hessian at the posterior mode $\tilde{\theta}$, which then becomes the dispersion measure for our proposal distribution. Denote the inverse Hessian by $\tilde{\Sigma}$.
3. Draw $\theta^{(0)}$ from $\mathcal{N}(\tilde{\theta}, c_0^2 \tilde{\Sigma})$.
4. For $s = 1, \dots, n_{sim}$, draw from the proposal distribution $\mathcal{N}(\theta^{(s-1)}, c^2 \tilde{\Sigma})$, where we scale $\tilde{\Sigma}$ to attain an acceptance rate close to 0.25, as it is usually suggested. The jump from $\theta^{(s-1)}$ is accepted ($\theta^{(s)} = \vartheta$) with probability $\min \left\{ 1, r \left(\theta^{(s-1)}, \vartheta | Y \right) \right\}$

and rejected ($\theta^{(s)} = \theta^{(s-1)}$) otherwise. Here

$$r\left(\theta^{(s-1)}, \vartheta|Y\right) = \frac{\mathcal{L}(\vartheta|Y)p(\vartheta)}{\mathcal{L}(\vartheta|Y)p\left(\theta^{(s-1)}\right)}.$$

As it is standard in the literature, for the computation of the marginal likelihood of these models, we use the modified harmonic mean method of Gelfand and Dey (1994) and Geweke (1999). Appendix 7.2 discusses checks for the convergence of the algorithm.

3.3 Priors

Priors for the nondiversifiable and idiosyncratic risks of the managed portfolios' returns ($\sigma_{j,x}$ and $\sigma_{j,\epsilon}$), the parameters that govern the subjective distribution of skill level ($\phi_{j,0}$, $\sigma_{a,j,0}$) perceived *a priori* by investors, as well as investors' prior perception of the parameter governing returns to scale ($\tilde{b}_{j,0}$, $\sigma_{b,j,0}$) and the fund-level decreasing returns to scale ($\tilde{c}_{j,0}$, $\sigma_{c,j,0}$), are quantified based on panel regressions of funds' benchmark-adjusted returns on lagged fund size and sector size.

We begin by tying down the model parameters that can be inferred directly from the data. In particular, we set the sector-level return volatility ($\sigma_{j,x}$) to an average of 0.016, or 1.6 percent per month, which is approximately equal to the average of monthly cluster-level standard deviation component from our performance regressions in Appendix A, where we cluster standard errors by Morningstar Category \times month. Moreover, we set the fund-specific return volatility ($\sigma_{j,\epsilon}$) to an average of 0.019, or 1.9 percent per month, which is approximately equal to the average of monthly residual standard deviation from the same panel regressions in Appendix A. In the special case of the model, in which we assume that investors know the true fund-level decreasing returns to scale, we will set the parameter (c_j) to the slope coefficient on *FundSize*.

Table 2 report our priors for the parameters of the model discussed above. First, priors for the parameters that govern investors' initial prior over managerial ability ($\phi_{a,j,0}$ and $\sigma_{a,j,0}$) are then centered at the average and standard deviation of the estimated fund fixed effects across all funds operating in our sample, respectively. While the prior for $\sigma_{a,j,0}$ is relatively disperse, we use the standard error of the mean fixed effect for the standard deviation of the prior for $\phi_{a,j,0}$. These priors reflect our view that the regression analysis are informative about investors' subjective prior estimates

(which we impose ought to be reasonable), but that they substantially understate investors' subjective prior uncertainty about each fund's skill $a_{i,j}$ (or at least produce a rough proxy for rational prior uncertainty over managerial ability). Second, prior for the parameters that govern investors' subjective assessment of sector-level returns to scale $(\tilde{b}_j, \sigma_{b,j,0})$ are centered at the estimated slopes on *SectorSize*. Analogously, prior for the parameters that govern investors' subjective assessment of fund-level decreasing returns $(\tilde{c}_j, \sigma_{c,j,0})$ are centered at the estimated slopes on *FundSize*.

Next, we need to specify priors for the subjective prior correlation, $\rho_{a,j,0}$, between fund fixed effects across two different funds. We will let the prior be fairly disperse and the same across investment styles so as to let the algorithm reveal what investors believed a priori about how similar different funds are within each sector.

Finally, following Del Negro et al. (2007), the priors for the standard deviations of the belief shocks are fairly disperse and chosen in order to generate realistic volatilities for the endogenous variables.

4 Estimation Results

4.1 Parameter Estimates

Table 3 summarizes the posterior distribution of the model coefficients, reporting posterior medians, standard deviations, and fifth and ninety-fifth percentiles computed with the draws of our posterior simulator. All coefficient estimates are fairly tight and seem sensible.

We make three observations with respect to the parameter estimates.

First, I find that investors' subjective prior beliefs feature pessimism about the extent to which fund-level decreasing returns to scale impact performance. That is, investors initially believed that increasing the fund size is significantly more costly in terms of performance than it really is in the data. Second, I find that investors' subjective priors about funds' raw skills are optimistically higher than average fund fixed effects in the data. Taken together, investors initially believed that funds start out with higher skills than they really do, but they also thought the skills are eroded much faster than they really do. Thus, it is interesting to note that investors were not unambiguously optimistic, contrary to what one might have expected *a priori* based on anecdotal understanding of the industry.

Finally, I find that the standard deviations of belief shocks are small relative to the magnitude of fund fixed effects or investors' estimates about funds' raw skills. I interpret this as telling us that my model allowing for learning about heterogeneity in skills across funds and learning about the nature of decreasing returns to scale well summarizes the sources of fluctuations in investors' subjective beliefs.

4.2 The Growth in Actively Managed Mutual Funds

Using the posterior distribution of the model parameters, we now study the evolution of specific aspects of investors' posterior beliefs that shed light on the sources of the growth in actively managed mutual funds. In equilibrium, the size of this industry is determined by investors' posterior expectations about two things, the average raw skill of incumbent funds and the effective degree of decreasing returns to scale. Figures 7.2 and 7.2, respectively, plot the evolution of investors' posterior expectations about the average raw skill of incumbent funds for each sector and investors' posterior expectations about the effective degree of decreasing returns to scale.

Qualitatively, Figure 7.2 shows that investors' conditional expectation about the average raw skill of incumbent funds does not fluctuate much. Given that investors started with subjectively optimistic beliefs about average fund skill, this implies that investors' optimism about fund skills dissipates only very slowly and persists. On the other hand, Figure 7.2 highlights that investors' conditional expectation about the effective degree of decreasing returns to scale is subject to significant amount of learning and in turn, declines much more over time. Thus, over time, investors come to recognize that they were too pessimistic about the extent to which scale impacts performance. Taken together, the optimistic aspect of investors' subjective prior stays, but the pessimistic aspect evaporates over time. This suggests that the growth in active asset management is too optimistic and too much.

Now that we have a mechanical understanding of this growth, what is the economic intuition underlying it?

After observing a fund's negative performance, investors can infer one of two things, that the fund manager's raw skill is lower than expected or that the value of active investing more generally is lower than expected. Since investors believed a priori that fund-specific raw skills are not strongly correlated across funds, they do not infer that low skills of old funds imply new funds will also have low skills. As a

result, sustained entry of new funds keeps refreshing investors' optimism about the average raw skills of incumbent funds. Basically, optimism about the industry as a whole persists at the expense of disappointment about individual funds in the past.

On the other hand, since investors started with pessimism about the extent of decreasing returns to scale, they learn that the scale impacts performance less than they expected over time and in turn, they do more active investing, growing active management as a whole. In addition, fund-level decreasing returns to scale imply that the average unit cost associated with investing in active management declines as the number of funds increases, making the growth of this industry more dramatic.

Taken together, we find that learning about heterogeneity in skills across funds and about fund-level decreasing returns to scale are key to accounting for the growth in actively managed mutual funds in spite of its historical underperformance and in the presence of fund proliferation.

5 Model Fit and Robustness Issues

5.1 Model Fit

In this section, we evaluate the fit of our model relative to a few restricted versions of the model, including the Pástor-Stambaugh model specification. Another restricted version of the model, in addition to the Pástor-Stambaugh model, impose that investors are not learning about fund-level decreasing returns to scale and instead, know the actual degree of fund-level returns to scale. The fit of the flexible model relative to this restriction allows us to show the key role of learning about fund-level decreasing returns to scale. The fit of these models relative to the Pástor-Stambaugh model restriction, i.e., no heterogeneity in fund skills, allows us to show the key role of learning about heterogeneity in skills across funds.

From a Bayesian perspective, the marginal likelihood is the most comprehensive and accurate measure of fit, as it can be used to construct posterior odds on competing models. The first column corresponding to each model specification in Table 4 reports the log-marginal data density for our baseline model and two restricted versions of it. As it is evident from the table, the value of the log-marginal likelihood is conclusively in favor of our model that allows for investors learning about both heterogeneity in skills across funds and fund-level decreasing returns to scale.

Beyond the likelihood ratio comparison, we simulate the size dynamics, again, under the three different model specifications. For a given model specification, we will use our data from January 1980 to December 1992 to initialize the economy, and then simulate 3,000 samples of data for the periods January 1993-December 2014.

We will set the model parameters to the posterior mode of the parameter vector. We then update investors' prior beliefs at the beginning of January 1980 to initialize investors' posterior beliefs at the end of December 1992 regarding the parameters governing returns to scale and fund-specific parameters for incumbent funds. We also draw fund fixed effects for incumbent funds in December 1992.

Step 0 Estimate the parameters b_j, c_j governing true returns to scale from the fixed effect regression of benchmark-adjusted fund returns on both sector size and fund size. Estimate, further, the probability p_j of exit by an incumbent fund in sector j from the average life length of funds in sector j , and estimate the Poisson regression

$$\log(E_{t-1}(M_{j,t})) = \beta_{j,0} + \beta_{j,1}t + \beta_{j,2}t^2 + \beta_{j,M}M_{j,t-1}$$

Then in each period during January 1993-December 2014, the following events happen sequentially:

1. Each incumbent fund $i \in M_{j,t-1}$ in sector $j \in N$ enters period t with $q_{i,j,t-1}$ funds under management and subjective estimates of parameters governing fund returns, $\left\{ \left\{ \phi_{i,j,t-1} \right\}_{i \in M_{j,t-1}}, \tilde{b}_{j,t-1}, \tilde{c}_{j,t-1} \right\}_{j \in N}$.
2. Draw returns according to

$$R_{i,j,t} = a_{i,j} - b_j \left(\frac{Q}{W} \right)_{j,t} - c_j \left(\frac{q}{W} \right)_{j,t} + x_{j,t} + \epsilon_{i,j,t}$$

where $a_{i,j}$ is a randomly drawn fund-fixed effect, and the standard deviations for $x_{j,t}, \epsilon_{i,j,t}$ are set from the standard errors of the same regression. Draw, in addition, fees according to

$$\log(f_{i,j,t}) = \rho_f \log(f_{i,j,t-1}) + \nu_{i,j,t}$$

where the standard deviation of $\nu_{i,j,t}$ is set from running this regression in the data.

3. Managers and investors observe $\{R_{i,j,t}, f_{i,j,t}\}_{i \in M_{j,t-1}, j \in N}$ and update their estimates of the parameters governing fund returns by calculating $\left\{ \left\{ \phi_{i,j,t} \right\}_{i \in M_{j,t-1}}, \tilde{b}_{j,t}, \tilde{c}_{j,t} \right\}_{j \in N}$ (using formulas (25)-(27)).
4. Each incumbent fund $i \in M_{j,t-1}$ in sector $j \in N$ makes an exit with probability p_j .
5. Draw the number of entering funds in each sector $j \in N$ according to the Poisson distribution with intensity

$$\log(\lambda_{j,t}) = \beta_{j,0} + \beta_{j,1}t + \beta_{j,2}t^2 + \beta_{j,M}M_{j,t-1}.$$

6. Then capital flows into or out of existing funds to determine $\left\{ (q/W)_{i,j,t} \right\}_{i \in M_{j,t}, j \in N}$ using equation (24).

We report the simulation results in Figures 7.2 and 7.2, and in Table 4. In Table 4, beyond the first two columns, we report the simulated time series correlation between the industry size and the number of incumbent funds under competing model specifications. Comparing to the sample time series correlation of 0.97 in the data, only the Pastor-Stambaugh model fail at generating any positive correlation. Intuitively, if investors perceive no heterogeneity in skills across funds, the inference problem simplifies to learning about two parameters, so that there are little investor learning and, in turn, little fluctuations in the equilibrium size of this industry. In addition, in this simple model, the number of incumbent funds play no role in determining the equilibrium size, so it is easily anticipated that it cannot generate the positive comovement between the industry size and the number of incumbent funds.

On the other hand, if investors are aware of heterogeneity in skills across funds and of fund-level decreasing returns to scale, the inference problem grows significantly more complex. In addition, the number of incumbent funds play an important role in refreshing investors' belief regarding the average skill of incumbent funds and in reducing the effective level of decreasing returns to scale. This allows the two models allowing for investors' learning about heterogeneity in skills to successfully generate the positive comovement between the industry size and the number of incumbent

funds. In particular, learning about fund-level returns to scale is not necessary for understanding this fact.

However, Figure 7.2 clarifies that learning about fund-level returns to scale is important for quantitatively understanding the dramatic growth of active asset management. This is because, as we explained in Section 4.2, simply allowing for investors' optimism regarding typical fund's skill does not, by itself, suffice to rationalize the growth of this industry in the presence of poor average performance. Thus, it is necessary to allow for investors learning about some other aspect of this industry, of which they start and do away with pessimism through rational learning. Learning about fund-level returns to scale does precisely this, and helps us quantitatively understand the growth of active asset management.

Finally, in Figure 7.2, the Pastor-Stambaugh model generates no time trend in active management industry size. This is not surprising given our discussion up to now. After observing negative performance, investors can only infer that they ought to do less active investing, which would shrink active management as a whole. So the best that the model can do is to produce no significant time trend in industry size.

Figure 7.2 reports the simulated industry size dynamics from our baseline model, i.e., its median dynamics and the associated 5th and 95th posterior bands. In spite of its simplicity, the model does a surprisingly good job of capturing the actual industry size dynamics. In conclusion, our proposed model is the simplest investor learning model specification necessary to quantitatively understand the observed dynamics of demand for active asset management in the data in the presence of negative historical performance and of net fund entry over time.

5.2 Counterfactual Experiment: No Learning

In this subsection, I consider a counterfactual scenario, in which I assume that investors know all the parameters governing active returns. This means that investors know the actual raw skills of all incumbent funds (which have mean equal to the average fund fixed effect and variance equal to the variance of fund fixed effects). This means, further, that investors invest as if the parameters governing decreasing returns to scale are the estimates of decreasing returns. This scenario, of course, does not square with the prominent stylized fact of performance-flow relationship, i.e., flows into and out of mutual funds are strongly related to lagged measures of excess returns

(see, e.g., Chevalier and Ellison (1997) or Sirri and Tufano (1998)). Nonetheless, it provides a natural way of measuring how big the industry ought to be based on the empirical determinants of mutual fund performance.

We do a back-of-the-envelope calculation, where we calculate the industry size as follows:

$$\left(\frac{Q}{W}\right)_t = \sum_{j \in N} \frac{a_{j,t} - E_t(f_{j,t+1})}{b_j + c_j/M_{j,t}},$$

where $a_{j,t}$ is the estimated average fund fixed effect, b_j is the estimated slope on *SectorSize*, and c_j is the estimated slope on *FundSize*. Using the fees data and the number of incumbent funds in December 2014, it is straightforward to calculate the industry size with no learning.

I find that if learning had not been necessary, industry size would have been about 6.8% (compared to its observed size of 13.2%) in 2014. In this sense, the model suggests that there is probably too much active asset management. This result has clear policy implications. For example, this means that investors were overpaying for active management by over \$14 billions, or almost 100% in 2014. There is clearly some socially useful role for active management as a type of informed trading, but I would say that we do not need nearly as much active money management as it exists to ensure an optimal amount of price discovery.

6 Conclusion

The size of active management industry is a big part of recent literature on the "size of finance". In this paper, we developed a model of active management, building in investors' learning about heterogeneity in skills across funds and about the nature of decreasing returns to scale, in order to understand the sources of historical growth of this industry. By estimating our flexible learning model, I found that two aspects of learning are key to explaining the growth in spite of historical underperformance of active funds and in the presence of net fund entry. First, learning about heterogeneity in skills render investors' inference problem to be significantly more complex and in turn, learning to be slow. Second, fund-level decreasing returns to scale and investors learning about it are sources of growth in an environment that is otherwise steady. Interestingly, such a seemingly simple model of rational learning can be quantitatively consistent with the growth in actively managed funds in the data. Furthermore,

I found that the link from steady entry of new funds to the dramatic growth of assets under management is highly sensitive to the specification of investors' learning problem.

Finally, the fact that the industry size would have significantly smaller without learning (according to our counterfactual experiment) shows that investors' learning about the correct allocation to active management is complex and slow, due to the proliferation of new funds. This has clear policy implications that there should be active efforts to increase investors' awareness of the parameters governing active fund returns, which would help the industry size converge to the correct level of allocation much faster.

7 Appendix

7.1 Data-Cleaning Steps

We require that funds appear in both CRSP and Morningstar, which allows us to check data accuracy by comparing the two databases, as detailed below. Morningstar assigns each fund a category (e.g., large growth, mid-cap blend, small value), which we use to categorize funds. We start the sample in 1980, the first decade after Vanguard introduced its S&P 500 index fund (i.e., May 1975). We merge CRSP and Morningstar using funds' tickers, CUSIPs, and names. We check the accuracy of each match by comparing assets and returns across the two databases.

We use keywords in the Primary Prospectus Benchmark variable to exclude bond funds, money market funds, international funds, funds of funds, industry funds, real estate funds, target retirement funds, and other non-equity funds. We also exclude funds identified by CRSP or Morningstar as index funds, as well as funds whose name contains the word "index". We exclude fund-month observations with expense ratios below 0.1% per year, since it is extremely unlikely that any actively managed funds would charge such low fees. Finally, we exclude fund-month observations with lagged fund size below \$15 million in 2011 dollars.

First, we follow Berk and van Binsbergen in reconciling return data between CRSP and Morningstar. Returns differ across the two databases by at least 10 bps per month in 1.6% of observations. By applying Berk and van Binsbergen's algorithm we reduce the discrepancy rate to 0.5%. We set the remaining return discrepancies to missing.

Similarly, total assets under management (AUM) differ between CRSP and Morningstar in 7.0% of observations. The average of these discrepancies is \$31.3 million. AUM differs by at least \$100,000 and 5% across databases in 1.1% of observations. We set these AUM values to missing; otherwise we use CRSP's value.

We use FundID in Morningstar to aggregate share classes.¹⁰

Finally, we exclude the returns of funds younger than three years to address the incubation bias¹¹ documented by Evans (2010). Evans (2010) reports that "removing the first three years of return data for all funds eliminates the bias" (p. 1584).

7.2 Convergence Diagnostics

We assess the convergence of our posterior simulators using a battery of diagnostics. We launch multiple chains of the Random Walk metropolis algorithm from different starting values (drawn randomly around the posterior mode). To check that these multiple chains agree in their characterization of the posterior distribution, we look at various sample moments within and across chains. Several chains initialized in this manner delivered roughly identical results when looking at means, medians, and posterior percentiles, as well as trace and kernel plots.

More formally, for the multiple chains used to generate the results in the paper, Table 5 reports potential scale reduction factors proposed by Brooks and Gelman (1998) both for variances and 90 percent intervals. These numbers are close to one and therefore well below the 1.2 benchmark widely used in practice as an upper bound for convergence.

References

- [1] An, Sungbae, and Frank Schorfheide. 2007. "Bayesian Analysis of DSGE Models." *Econometric Reviews* 26:113-72.

¹⁰Many mutual funds offer multiple share classes, which represent claims on the same underlying assets but have different fee structures. Different share classes of the same fund have the same Morningstar FundID. We aggregate all share classes of the same fund. Specifically, we compute a fund's AUM by summing AUM across the fund's share classes, and we compute the fund's returns and expense ratios by asset-weighting across share classes.

¹¹Incubation is a strategy for initiating new funds, where multiple funds are started privately, and, at the end of an evaluation period, some are opened to the public. Evans finds that the incubated funds outperform the non-incubated funds during the incubation period.

- [2] Baks, Klaas P., Andrew Metrick, and Jessica Wachter. 2001. "Should Investors Avoid All Actively Managed Mutual Funds? A Study in Bayesian Performance Evaluation." *Journal of Finance* 56: 45-85.
- [3] Barber, Brad M., Xing Huang, and Terrance Odean. 2015. "Which Risk Factors Matter to Investors? Evidence from Mutual Fund Flows." Manuscript. Berkeley: University of California, Berkeley, Haas School of Business.
- [4] Barberis, Nicholas, and Andrei Shleifer. 2003. "Style Investing." *Journal of Financial Economics* 68:161-99.
- [5] Berk, Jonathan B., and Richard C. Green. 2004. "Mutual Fund Flows and Performance in Rational Markets." *Journal of Political Economy* 112:1269-95.
- [6] Berk, Jonathan B., and Jules H. van Binsbergen. 2015. "Measuring Skill in the Mutual Fund Industry." *Journal of Financial Economics* 118:1-20.
- [7] Berk, Jonathan B., and Jules H. van Binsbergen. Forthcoming. "Assessing Asset Pricing Models using Revealed Preference." *Journal of Financial Economics*.
- [8] Bolton, Patrick, Tano Santos, and Jose A. Scheinkman. Forthcoming. "Cream Skimming in Financial Markets." *Journal of Finance*.
- [9] Brooks, Stephen P., and Andrew Gelman. 1998. "General Methods for Monitoring Convergence of Iterative Simulations." *Journal of Computational and Graphical Statistics* 7:434-55.
- [10] Brown, David P., and Yuchang Wu. Forthcoming. "Mutual Fund Flows and Cross-Fund Learning Within Families." *Journal of Finance*.
- [11] Chen, Joseph, Harrison Hong, Ming Huang, and Jeffrey D. Kubik. 2004. "Does Fund Size Erode Mutual Fund Performance? The Role of Liquidity and Organization." *American Economic Review* 94:1276-302.
- [12] Cheng, Si, Massimo Massa, and Hong Zhang. 2015. "Short-Sale Constraints and the Pricing of Managerial Skills." Manuscript. London: Centre for Economic Policy Research.
- [13] Cochrane, John H. 2013. "Finance: Function Matters, Not Size." *Journal of Economic Perspectives* 27:29-50.

- [14] Christoffersen, Susan E. K. 2001. "Why Do Money Fund Managers Voluntarily Waive Their Fees?" *Journal of Finance* 56:1117-40.
- [15] Cuoco, Domenico, and Ron Kaniel. 2011. "Equilibrium Prices in the Presence of Delegated Portfolio Management." *Journal of Financial Economics* 101:264-96.
- [16] Del Negro, Marco, Frank Schorfheide, Frank Smets, and Rafael Wouters. 2007. "On the Fit of New Keynesian Models." *Journal of Business & Economic Statistics* 25:123-43.
- [17] Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake. 2001. "A First Look at the Accuracy of the CRSP Mutual Fund Database and a Comparison of the CRSP and Morningstar Mutual Fund Databases." *Journal of Finance* 56:2415-30.
- [18] Evans, Richard B. 2010. "Mutual Fund Incubation." *Journal of Finance* 65:1581-611.
- [19] Fama, Eugene F., and Kenneth R. French. 2010. "Luck versus Skill in the Cross-Section of Mutual Fund Returns." *Journal of Finance* 65:1915-47.
- [20] Ferreira, Miguel A., Aneel Keswani, Antonio F. Miguel, and Sofia B. Ramos. 2015. "Testing the Berk and Green model around the world." Manuscript. Portugal: Nova School of Business and Economics.
- [21] Franzoni, Francesco, and Martin C. Schmalz. 2014. "Performance Measurement with Uncertain Risk Loadings." Manuscript. Ann Arbor: University of Michigan, Ross School of Business.
- [22] French, Kenneth R. 2008. "Presidential Address: The Cost of Active Investing." *Journal of Finance* 63:1537-73.
- [23] Garcia, Diego, and Joel M. Vanden. 2009. "Information Acquisition and Mutual Funds." *Journal of Economic Theory* 144:1965-95.
- [24] Gârleanu, Nicolae, and Lasse Heje Pedersen. 2015. "Efficiently Inefficient Markets for Assets and Asset Management." Manuscript. New York: New York University, Stern School of Business.

- [25] Gelfand, Alan E., and Dipak K. Dey. 1994. "Bayesian Model Choice: Asymptotics and Exact Calculations." *Journal of the Royal Statistical Society, Series B (Methodological)* 56:501-14.
- [26] Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny. 2015. "Money Doctors." *Journal of Finance* 70:91-114.
- [27] Geweke, John. 1999. "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication." *Econometric Reviews* 18:1-73.
- [28] Glode, Vincent. 2011. "Why Mutual Funds "Underperform" ." *Journal of Financial Economics* 99:546-59.
- [29] Greenwood, Robin, and David Scharfstein. 2013. "The Growth of Finance." *Journal of Economic Perspectives* 27:3-28.
- [30] Grossman, Sanford J., and Joseph E. Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets." *American Economic Review* 70:393-408.
- [31] Gruber, Martin J. 1996. "Another Puzzle: The Growth in Actively Managed Mutual Funds." *Journal of Finance* 51:783-810.
- [32] Guerrieri, Veronica, and Péter Kondor. 2012. "Fund Managers, Career Concerns, and Asset Price Volatility." *American Economic Review* 102:1986-2017.
- [33] Haddad, Valentin, Paul Ho, and Erik Loualiche. 2014. "Detail Disagreement and Innovation Booms." Manuscript. Princeton: Princeton University, Bendheim Center for Finance.
- [34] He, Zhiguo, and Wei Xiong. 2013. "Delegated Asset Management, Investment Mandates, and Capital Immobility." *Journal of Financial Economics* 107:239-58.
- [35] Hortagsu, Ali, and Chad Syverson. 2004. "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds." *Quarterly Journal of Economics* 119:403-56.
- [36] Huang, Jennifer, Kelsey D. Wei, and Hong Yan. 2012. "Investor Learning and Mutual Fund Flows." Manuscript. Austin: University of Texas at Austin, McCombs School of Business.

- [37] Jensen, Michael C. 1968. "The Performance of Mutual Funds in the Period 1945-64." *Journal of Finance* 23:389-416.
- [38] Justiniano, Alejandro, and Giorgio E. Primiceri. 2008. "The Time-Varying Volatility of Macroeconomic Fluctuations." *American Economic Review* 98:604-41.
- [39] Kacperczyk, Marcin T., Stijn Van Nieuwerburgh, and Laura Veldkamp. 2014. "Time-Varying Fund Manager Skill." *Journal of Finance* 69:1455-84.
- [40] Khorana, Ajay, Henri Servaes, and Peter Tufano. 2005. "Explaining the Size of the Mutual Fund Industry around the World." *Journal of Financial Economics* 78:145-85.
- [41] Levin, Andrew T., Alexei Onatski, John C. Williams, and Noah Williams. 2006. "Monetary Policy Under Uncertainty in Micro-Founded Macroeconometric Models." In *NBER Macroeconomics Annual 2005*, edited by Mark Gertler and Kenneth Rogoff. Cambridge, MA: MIT Press.
- [42] Lynch, Anthony W., and David K. Musto. 2003. "How Investors Interpret Past Fund Returns." *Journal of Finance* 58:2033-58.
- [43] Malkiel, Burton G. 1995. "Returns from Investing in Equity Mutual Funds 1971 to 1991." *Journal of Finance* 50:549-72.
- [44] Malkiel, Burton G. 2013. "Asset Management Fees and the Growth." *Journal of Economic Perspectives* 27:97-108.
- [45] Nanda, Vikram, M.P. Narayanan, and Vincent A. Warther. 2000. "Liquidity, Investment Ability, and Mutual Fund Structure." *Journal of Financial Economics* 57:417-43.
- [46] Pástor, Ľuboš, and Robert F. Stambaugh. 2002. "Mutual Fund Performance and Seemingly Unrelated Assets." *Journal of Financial Economics* 63:315-49.
- [47] Pástor, Ľuboš, and Robert F. Stambaugh. 2012. "On the Size of the Active Management Industry." *Journal of Political Economy* 120:740-81.
- [48] Pástor, Ľuboš, Robert F. Stambaugh, and Lucian A. Taylor. 2015. "Scale and Skill in Active Management." *Journal of Financial Economics* 116:23-45.

- [49] Philippon, Thomas. 2015. "Has the US Finance Industry Become Less Efficient? On the Theory and Measurement of Financial Intermediation." *American Economic Review* 105:1408-38.
- [50] Philippon, Thomas, and Ariell Reshef. 2013. "An International Look at the Growth of Modern Finance." *Journal of Economic Perspectives* 27:73-96.
- [51] Scheinkman, José A., and Wei Xiong. 2003. "Overconfidence and Speculative Bubbles." *Journal of Political Economy* 111:1183-219.
- [52] Stambaugh, Robert F. 2014. "Presidential Address: Investment Noise." *Journal of Finance* 69:1415-53.

Table 1
Summary Statistics

This table shows summary statistics for our sample of active equity mutual funds from 1980–2014. The unit of observation is the fund/month. All returns and expense ratios are in units of fraction per month. $NetR$ is the return received by investors. Benchmark-adjusted net return equals net return minus the return on benchmark portfolio. Gross return is the benchmark-adjusted net return plus 1/12th of the annual expense ratio. $FundSize$ is the fund’s total AUM aggregated across share classes, divided by the total stock market capitalization in the same month, imputing $FundSize$ when missing. $IndustrySize$ is the sum of all funds’ AUM divided by the total stock market capitalization in the same month. $FundAge$ is the number of years since the fund’s first offer date. $SectorSize$ is the sum of AUM (in nominal dollars) across all funds within a given sector, divided by the total stock market capitalization in the same month. We use the nine sectors in Morningstar’s 3×3 StyleBox.

Variable	Mean	Std	Percentiles					
			1st	25th	50th	75th	99th	
Net return	0.006876	0.051678	-0.147376	-0.019901	0.011533	0.037475	0.127383	
Benchmark-adj. $NetR$	-0.000924	0.024404	-0.067661	-0.011832	-0.001265	0.009454	0.068755	
Benchmark-adj. $GrossR$	0.000072	0.024362	-0.06644	-0.010863	-0.000326	0.010434	0.069824	
Expense ratio	0.000996	0.000763	0.000233	0.000779	0.000954	0.001171	0.001983	
$FundSize$	0.000125	0.000431	0.000001	0.000007	0.000025	0.000081	0.001848	
$IndustrySize$	0.118938	0.033086	0.017509	0.10616	0.133244	0.141828	0.151461	
$SectorSize$	0.021329	0.016539	0.000557	0.005314	0.020215	0.031650	0.052676	

Table 2

Prior Densities for the Model Parameters

This table shows the priors for the model parameter, which we use for the Bayesian MCMC algorithm. The first six model parameters govern investors' subjective prior beliefs and the last two model parameters govern return volatilities. They are quantified based on panel regressions of benchmark-adjusted gross returns ($GrossR$) on sector size ($SectorSize$) and fund size ($FundSize$), controlling for investment style via the nine sectors in Morningstar's 3×3 StyleBox. We do not specify standard deviations for the parameters governing return volatilities because we take them as given in the Bayesian MCMC algorithm.

Morningstar Category	Coefficient Density	$\phi_{j,0}$ L	$\tilde{b}_{j,0}$ L	$\tilde{c}_{j,0}$ L	$\sigma_{a,j,0}$ I	$\sigma_{b,j,0}$ I	$\sigma_{c,j,0}$ I	$\sigma_{j,x}$	$\sigma_{j,\epsilon}$
Large Blend	Mean	.00123	0.067	1.86	.00123	0.067	1.86	0.00671	0.01679
	Std	.00081	0.030	0.34	.00123	0.067	1.86		
Large Growth	Mean	.00167	0.047	1.06	.00167	0.047	1.06	0.00774	0.01956
	Std	.00115	0.032	0.18	.00167	0.047	1.06		
Large Value	Mean	.00562	0.236	1.00	.00562	0.236	1.00	0.01281	0.01418
	Std	.00420	0.144	0.49	.00562	0.236	1.00		
Mid-Cap Blend	Mean	.00409	1.162	17.4	.00409	1.162	17.4	0.01121	0.02300
	Std	.00196	0.761	6.60	.00409	1.162	17.4		
Mid-Cap Growth	Mean	.00817	0.685	12.6	.00817	0.685	12.6	0.01801	0.02205
	Std	.00275	0.297	3.05	.00817	0.685	12.6		
Mid-Cap Value	Mean	.00555	0.653	2.81	.00555	0.653	2.81	0.01118	0.01682
	Std	.00275	0.390	1.23	.00555	0.653	2.81		
Small Blend	Mean	.00967	2.029	24.3	.00967	2.029	24.3	0.01915	0.01898
	Std	.00385	1.008	5.46	.00967	2.029	24.3		
Small Growth	Mean	.00924	1.821	10.2	.00924	1.821	10.2	0.02112	0.02349
	Std	.00512	1.093	8.34	.00924	1.821	10.2		
Small Value	Mean	.00683	2.797	22.5	.00683	2.797	22.5	0.01874	0.01930
	Std	.00326	1.852	6.58	.00683	2.797	22.5		

Table 3

Posterior Estimates for the Model with Fund-Level Heterogeneity in Skill and Returns to Scale

This table summarizes the posterior distribution of the model parameters. Median, standard deviations and posterior percentiles from four chains of 70,000 draws each from the Random Walk metropolis algorithm initialized from dispersed starting values around the mode. We discard the initial 20,000 draws.

Morningstar Category	Coefficient	$\phi_{0,j,0}$	$\tilde{b}_{j,0}$	$\tilde{c}_{j,0}$	$\sigma_{0,j,0}$	$\sigma_{b,j,0}$	$\sigma_{c,j,0}$	$\rho_{a,j,0}$	$\sigma_{j,zx}$	$\sigma_{j,ze}$
Large Blend	Median	.0056	.1062	2.090	.0008	.0194	16.37	.1010	.000049	.000083
	Std	.0009	.0237	0.415	.0004	.0031	4.340	.0976	.000008	.000060
	5th	.0043	.0710	1.487	.0004	.0148	10.29	.0049	.000037	.000033
Large Growth	95th	.0072	.1480	2.851	.0016	.0248	24.10	.2788	.000057	.000162
	Median	.0050	.0340	1.052	.0010	.0114	6.113	.2557	.000047	.000150
	Std	.0008	.0116	0.181	.0004	.0023	1.218	.0052	.000008	.000047
Large Value	5th	.0040	.0203	.7720	.0006	.0083	3.996	.2434	.000037	.000085
	95th	.0064	.0567	1.383	.0020	.0154	7.566	.2600	.000061	.000217
	Median	.0114	.1886	4.038	.0111	.0329	.7389	.8690	.000053	.000891
Mid-Cap Blend	Std	.0035	.0168	1.594	.0010	.0117	.3634	.1195	.000007	.000360
	5th	.0099	.1646	.6572	.0105	.0208	.2096	.7139	.000042	.000578
	95th	.0223	.4892	4.798	.0204	.0500	1.509	.9419	.0001	.0015
Mid-Cap Growth	Median	.0037	.2970	49.50	.0012	.3454	12.30	.2972	.000042	.000304
	Std	.0003	.1001	4.816	.0002	.0462	3.115	.0581	.000011	.000037
	5th	.0032	.1669	41.90	.0009	.2775	8.011	.1819	.000026	.000268
Mid-Cap Growth	95th	.0042	.3992	57.82	.0016	.4256	18.20	.3794	.000062	.000361
	Median	.0060	.2854	12.80	.0018	.2276	21.02	.2726	.000121	.000066
	Std	.0046	.1257	3.307	.0009	.0419	.9158	.0391	.000084	.000006
Mid-Cap Growth	5th	.0045	.2126	10.73	.0016	.2195	20.75	.2536	.000105	.000053
	95th	.0123	.3906	17.26	.0028	.3548	22.28	.3489	.000252	.000068

Table 3
Posterior Estimates for the Model with Fund-Level Heterogeneity in Skill and Returns to Scale
 (continued)

Morningstar Category	Coefficient	$\phi_{0,j,0}$	$\tilde{b}_{j,0}$	$\tilde{c}_{j,0}$	$\sigma_{0,j,0}$	$\sigma_{b,j,0}$	$\sigma_{c,j,0}$	$\rho_{a,j,0}$	$\sigma_{j,zx}$	$\sigma_{j,z\epsilon}$
Mid-Cap Value	Median	.0049	.3927	24.71	.0019	.1779	2.492	.3808	.000100	.000116
	Std	.0010	.1104	6.782	.0005	.0705	.8611	.2042	.000029	.000090
	5th	.0036	.2203	17.33	.0013	.1134	1.256	.1392	.000062	.000036
Small Growth	95th	.0069	.5849	32.03	.0025	.2466	3.337	.6372	.000161	.000228
	Median	.0055	1.235	24.06	.0033	.7380	45.35	.6012	.000166	.000033
	Std	.0006	.6114	6.131	.0010	.2826	6.951	.1262	.000099	.000009
Small Value	5th	.0046	.9786	17.61	.0025	.5955	34.55	.4633	.000138	.000023
	95th	.0066	1.784	36.51	.0048	1.002	57.38	.8408	.000211	.000050
	Median	.0136	1.601	111.4	.0042	7.676	7.057	.0800	.000365	.000041
Small Blend	Std	.0040	.7485	5.817	.0031	7.145	3.059	.0706	.000166	.000008
	5th	.0092	.7227	101.7	.0019	.6868	3.396	.0071	.000178	.000031
	95th	.0202	2.755	118.0	.0070	18.34	12.22	.1788	.000568	.000047
	Median	.0038	.5865	66.01	.0019	.5654	14.20	.4545	.000027	.000313
	Std	.0004	.1160	7.093	.0003	.0641	3.628	.1629	.000006	.000034
	5th	.0033	.4182	56.26	.0014	.0016	9.175	.2325	.000018	.000281
	95th	.0045	.7934	79.58	.0024	.6771	21.40	.7895	.000039	.000363

Table 4

Likelihood Ratio Comparison between Our Baseline Model and the Pastor-Stambaugh Model

The first column reports the log-likelihood of each model at the median parameter estimates. The rest of the columns report the simulated time-series correlation between the industry size and the number of funds. The actual value of correlation in the data is 0.97. The two tests of fit show that our flexible model uncontroersially performs better in terms of explaining the data.

	<i>(log) Likelihood at median</i>	Median	Std	$Corr(\{(Q/W)_t\}, \{M_t\})$
Our Baseline Model	18535	.96	.044	[.91 , .98]
Model with No Learning About Fund-Level Returns to Scale	18462	.93	.064	[.80 , .98]
Pástor and Stambaugh (2012) Model	17355	.11	.596	[-.82 , .92]

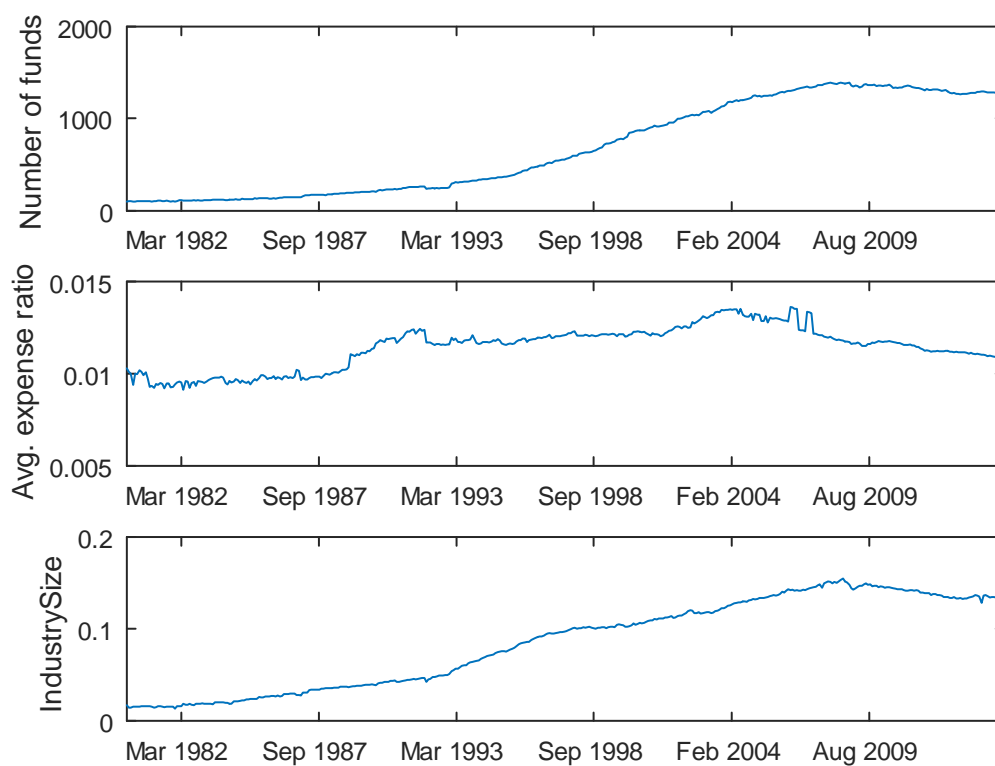
Table 5
Convergence Diagnostics for the Baseline Model

Potential scale reduction factors for multiple chains below 1.2 (and close to 1) are regarded as indicative of convergence. For each Morningstar Category, the first row reports PSRF using within and between variances, and the second row reports PSRF based on empirical 90 percent interval lengths (length of total-sequence interval to mean length of within-sequence intervals), as proposed by Brooks and Gelman (1998).

Morningstar Category	Potential scale reduction factors (PSRF)	$\phi_{0,j,0}$	$\tilde{b}_{j,0}$	$\tilde{c}_{j,0}$	$\sigma_{0,j,0}$	$\sigma_{b,j,0}$	$\sigma_{c,j,0}$	$\rho_{a,j,0}$	$\sigma_{j,zz}$	$\sigma_{j,z\epsilon}$
Large Blend	Variances	1.00	1.00	1.00	1.03	1.00	1.07	1.02	1.00	1.14
	Interval lengths (90 percent)	0.99	1.00	1.00	0.99	1.00	1.06	1.03	1.00	1.01
Large Growth	Variances	1.18	1.16	1.00	1.04	1.02	1.05	1.13	1.06	1.01
	Interval lengths (90 percent)	1.05	1.08	1.02	1.04	1.00	0.99	1.16	1.03	1.01
Large Value	Variances	1.02	1.04	1.05	1.12	1.04	1.00	1.16	1.17	1.01
	Interval lengths (90 percent)	1.01	1.12	1.03	1.05	1.01	1.05	1.06	1.10	1.19
Mid-Cap Blend	Variances	1.00	1.04	1.05	1.05	1.00	1.16	1.18	1.11	1.10
	Interval lengths (90 percent)	0.99	1.01	1.03	1.17	0.99	1.03	1.07	1.00	1.20
Mid-Cap Growth	Variances	1.03	1.15	1.18	1.01	1.11	1.05	1.16	1.06	1.11
	Interval lengths (90 percent)	1.02	1.04	1.11	1.19	1.19	1.02	1.11	1.08	1.05
Mid-Cap Value	Variances	1.00	1.02	1.03	1.04	1.00	1.00	1.02	1.01	1.18
	Interval lengths (90 percent)	1.00	1.01	1.05	1.00	1.00	1.00	1.00	1.00	0.98
Small Blend	Variances	1.07	1.05	1.03	1.01	1.00	1.03	1.05	1.20	1.14
	Interval lengths (90 percent)	0.94	0.87	0.90	0.90	0.88	1.03	1.06	0.92	1.12
Small Growth	Variances	1.01	1.01	1.19	1.01	1.01	1.12	1.07	1.01	1.01
	Interval lengths (90 percent)	0.89	1.03	1.20	1.05	0.88	1.13	1.09	1.14	1.20
Small Value	Variances	1.00	1.00	1.00	1.01	1.00	1.01	1.01	1.00	1.00
	Interval lengths (90 percent)	1.01	1.00	1.00	0.99	0.98	1.00	1.00	1.01	0.92

Figure 1. Sample Properties

The first figure plots the number of funds over time in our sample for the period 1980-2014. The second figure plots the average expense ratio, which has historically fluctuated significantly less than the sizes. The third figure plots the industry size (measured relative to the stock market capitalization), which has grown enormously over time.



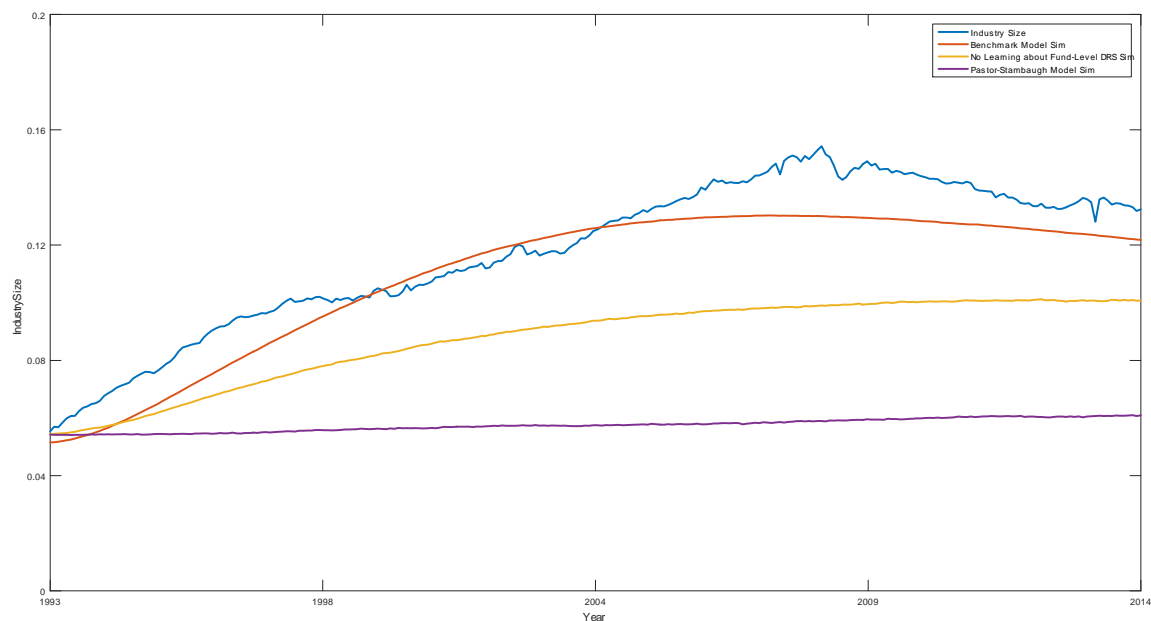


Figure 2. Simulated Industry Size under Different Model Specifications

This figure plots the time-series of actual industry size and the simulated industry sizes from three different versions of the model, including both our baseline model and the Pástor-Stambaugh model.

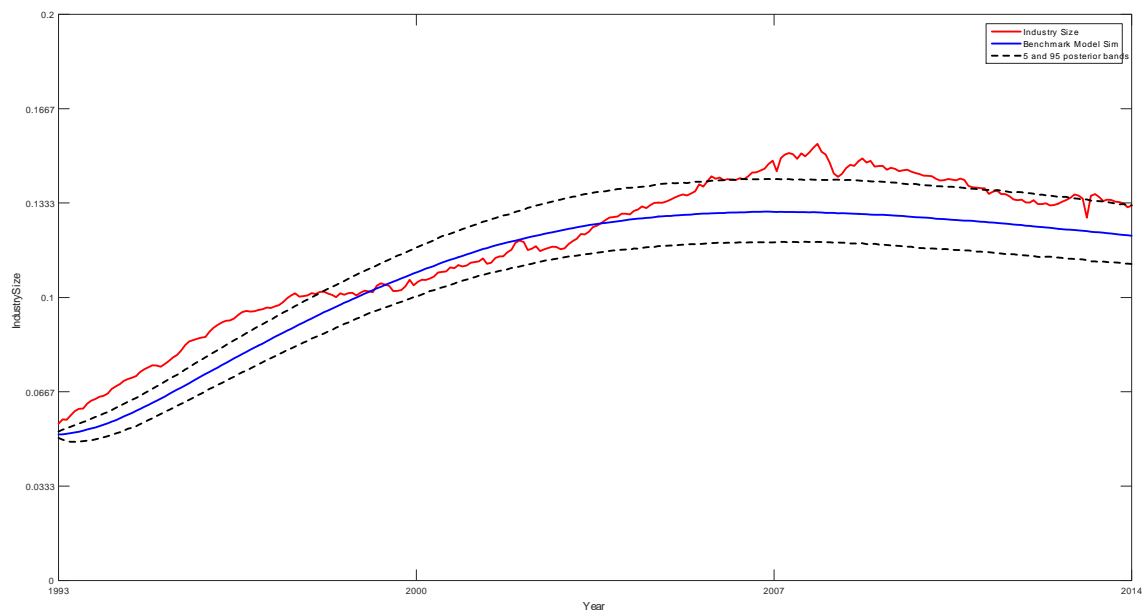


Figure 3.

This figure plots the actual size and the simulated industry size from our baseline model featuring learning about heterogeneity in fund skills and about fund-level returns to scale. The median and 5 – 95 percentiles of the time series are computed with the draws generated by the model under the posterior mode of its parameters.

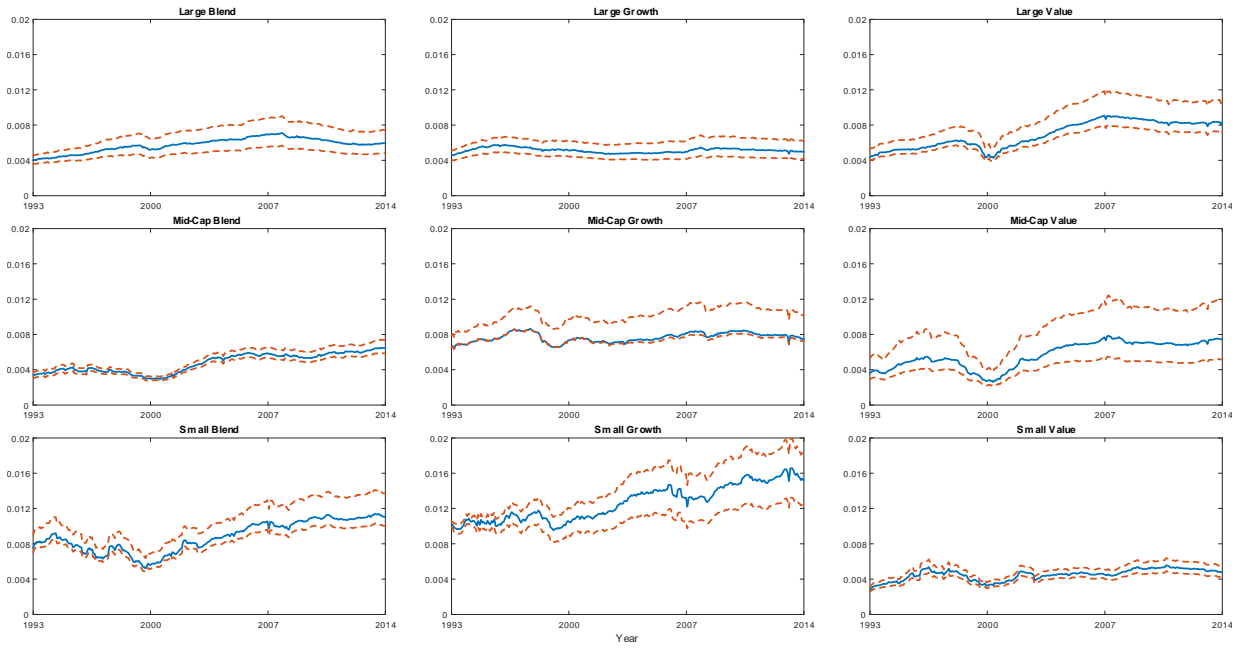


Figure 4.

This figure plots the evolution of investors' posterior expectation over the average raw skills of incumbent active fund managers. Compared to their posterior expectation about the effective degree of decreasing returns to scale, this expectation remains stable and modestly increasing.

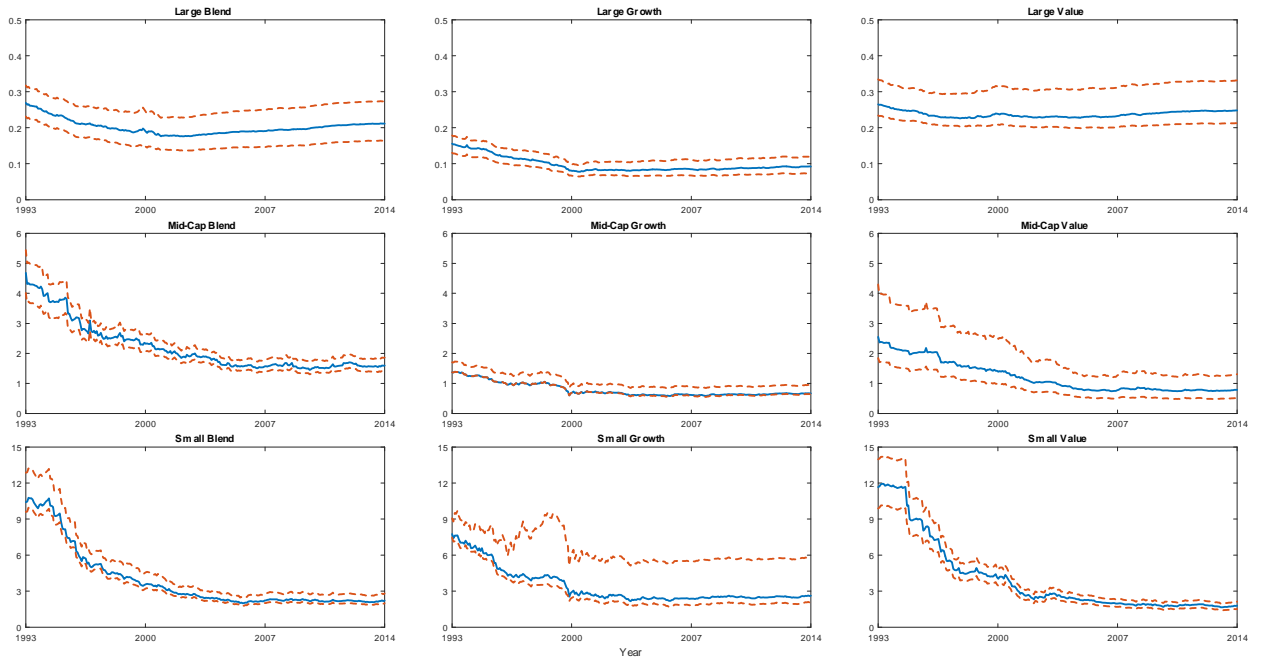


Figure 5.

This figure plots the evolution of investors' posterior expectation over the effective degree of decreasing returns to scale. Since investors started with pessimism about decreasing returns, investors grow (rationally more) optimistic about the extent to which scale erodes performance.