Capital Allocation and the Market for Mutual Funds: Inspecting the Mechanism*

Jules H. van Binsbergen† Jeong Ho (John) Kim‡ Soohun Kim§

February 7, 2023

Abstract

We exploit heterogeneity in decreasing returns to scale parameters across mutual funds to analyze the importance of scalability for investors' capital allocation decisions. We find strong evidence that steeper decreasing returns to scale attenuate flow sensitivity to performance. We calibrate a rational model of active fund management and show that a large fraction of cross-sectional variation in assets-under-management is due to investors anticipating the effects of scale on return performance. We conclude that decreasing returns to scale play a key role in achieving equilibrium in the intermediated investment management market.

*We are grateful for comments from our discussants, Jun Kyung Auh, Laurent Barras, Jaewon Choi, Yi Li, Ľuboš Pástor, and Andrew Patton. We are also grateful for comments from Itay Goldstein, Nikolai Roussanov, Robert Stambaugh, and Youchang Wu, conference participants at the American Finance Association Annual Meeting, the Annual Hedge Fund Research Conference, the Australasian Finance & Banking Conference, the China International Risk Forum, the Conference on Financial Economics and Accounting, the Econometric Society World Congress, the Finance Down Under Conference, the Fixed Income and Financial Institutions Conference, and the Northern Finance Association Meetings, and seminar participants at Boston Fed, Emory University, and Georgia Tech.

†University of Pennsylvania and NBER. Email: julesv@wharton.upenn.edu

‡Emory University. Email: jkim230@emory.edu

§KAIST. Email: soohun.kimi@gmail.com

1 Introduction

An important determinant of the Net Present Value (NPV) of an investment project is its scalability. Even if the marginal profitability on a project is large at small scale, when the profitability deteriorates quickly with size, agents will choose not to commit much capital to such projects, and, in the presence of fixed costs, may choose to forgo them altogether. Despite the importance of scalability, surprisingly little empirical work has quantitatively evaluated how important its cross-sectional variation is for capital allocation. In this paper, we fill this void by focusing on the mutual fund market, where measuring the scalability of investment strategies has become commonplace. In particular, the literature has argued that decreasing returns to scale (DRS) play a key role in equilibrating the mutual fund market.[1] Consistent with this argument, we show that scalability is an important driver of investors' capital allocation decisions by exploiting heterogeneity across funds in DRS parameters: steeper decreasing returns to scale attenuate flow sensitivity to performance. Further, we calibrate a rational model of active fund management and show that 57% of the cross-sectional variation in fund size can plausibly be attributed to heterogeneity in decreasing returns to scale.

Our approach closely follows the insights from Berk and Green (2004) who pioneered a rational equilibrium model of active mutual fund management. Because the percentage fee that mutual funds charge changes infrequently, the bulk of the equilibration process operates through the size (or Assets Under Management (AUM)) of the fund. When good news about a mutual fund arrives, rational Bayesian updating will lead investors to view the fund as a positive NPV buying opportunity at its current size. In response, flows will go to that fund. As the fund grows, the manager of the fund finds it increasingly harder to put the new inflows to good use, leading to a deterioration of the performance of the fund. The flows into the fund will stop when the fund is no longer a positive NPV investment, and the fund's

[1] See Berk and Green (2004), Berk and van Binsbergen (2015), Pástor, Stambaugh, and Taylor (2020), and Barras, Gagliardini, and Scaillet (2022).

abnormal return to investors has reverted back to zero.

We investigate this equilibrating mechanism more closely as follows. If indeed the above-mentioned equilibration process is at work, we should expect to find that the degree of decreasing returns to scale (DRS) has implications for the flow sensitivity to performance (FSP). While there is much evidence that an active fund's ability to outperform its benchmark declines as its size increases,[2] there is surprisingly little empirical work devoted to whether investors account for the adverse effects of fund scale in making their capital allocation decisions.

We address this important question by formally deriving and empirically testing what a rational model for active management implies about the relation between returns to scale and flow sensitivity to performance. In the context of the theory model of Berk and Green (2004), we show that steeper decreasing returns to scale attenuate flow sensitivity to performance. In the model, investors rationally interpret high performance as evidence of the manager's superior skill, so good performance results in an inflow of funds. More relevant to our hypothesis, the magnitude of the capital response is primarily driven by the extent of decreasing returns to scale. As a fund's returns decrease in scale more steeply, the positive net alpha is competed away with a smaller amount of capital inflows, making flows less sensitive to performance.

To test this theoretical insight, one needs a source of heterogeneity in decreasing returns to scale. One also needs to observe investor reactions to this heterogeneity. Indeed, we demonstrate that there is a substantial amount of heterogeneity in DRS across individual funds,[3] with correspondingly heterogeneous flow sensitivity to performance across funds. Our approach can be interpreted as inferring how the subjective size-performance relation, perceived by investors in real time, is incorporated into the flow-performance relation going forward. We find that a steeper decreasing returns to scale parameter predicts a lower

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

[3] Barras, Gagliardini, and Scaillet (2021) also provide empirical evidence that not only skill but also scalability vary substantially across funds.

sensitivity of flows to performance, consistent with the main prediction of our model.

One of the challenges in estimating the effect of decreasing returns to scale on flow sensitivity to performance is the estimation error in fund-specific DRS. As a result, the point estimates of the DRS-FSP relation using DRS estimates from simple fund-by-fund regressions are likely to suffer from an errors-in-variables bias. To gauge the severity of this attenuation bias, we first adjust these simple estimates of the DRS-FSP relation for the errors-in-variable bias, assuming that the errors are of the classical type (i.e., independent to the actual DRS). As expected, the simple DRS-FSP relation estimates are indeed biased toward zero.

To address this issue, we estimate the DRS-FSP relation by instrumenting for the heterogeneity in decreasing returns to scale with a set of fund characteristics that are plausibly related to the scalability of the funds' investment strategies.[4] In particular, by regressing the fund-specific DRS estimates on these characteristics, we obtain fitted values that we use as a more robust way of obtaining cross-sectional variation. Importantly, we show that while the statistical significance of the DRS-FSP relation is unaffected by using the characteristic-based approach, the characteristic-based estimates of the DRS-FSP relation become substantially more negative, and their magnitudes are similar to those implied by the classical measurement error assumption, suggesting that the characteristic-based approach indeed has alleviated the errors-in-variables problem.

Next, we turn to the economic significance of our estimates. In particular, we assess how equilibrium fund size is affected by the cross-sectional variation in decreasing returns to scale parameters. This exercise does require model assumptions. We calibrate a rational model in the spirit of Berk and Green (2004). After simulating data in which investors know the DRS can vary by fund, we check how much of the simulated size can be explained by counterfactual

[4] We investigate a number of characteristics that seem relevant a priori (also from the previous literature) for heterogeneity in returns to scale: the number of managers, volatility, expense ratios, marketing expenses, international exposure, turnover, log fund size, as well as the loadings on the size, value, and momentum factors. For example, we find the degree of DRS is stronger for higher-volatility funds, sole-managed funds, small-cap funds, as well as funds charging higher fees.

fund sizes computed under the assumption that the investors believe the DRS is the same for all funds. We find that, on average, more than half (57%) of the variance of fund sizes across funds and periods can be related to cross-sectional variation in DRS parameters. More importantly, although we do not target the DRS-FSP relation in our calibration, our model produces DRS-FSP relation estimates that are quantitatively very similar to those obtained from the data. Thus, it appears that the magnitude of the empirical DRS-FSP relation estimates is consistent with what the model predicts. This result suggests that the model does a good job of capturing capital allocation patterns in the data.

Beyond implications for fund flows, steeper decreasing returns to scale have implications for fund size in equilibrium. In the model, equilibrium fund size is proportional to the ratio of perceived skill over diseconomies of scale, which predicts that, all else equal (holding the alpha earned on the first dollar fixed), the decreasing returns to scale parameter should be lower for larger funds. This prediction is confirmed in our empirical analysis. Moreover, if investors update their beliefs about skill as in the model, their perception of optimal size ought to converge to true optimal size as funds grow older. Consistent with this argument, we find that estimates for the optimal size largely explain capital allocation across older funds in the data. We measure (log) optimal size of a fund by the average ratio of the fund's net alpha (adjusted for returns to scale) to the fund's individual DRS parameter. We show that the sizes of older funds continue to be significantly related to the optimal sizes even when we control for an alternative measure of optimal size that assumes fund scale has the same effect on performance for all funds. Again, investors seem to account not only for average decreasing returns to scale, but also for the heterogeneity of decreasing returns to scale across funds.

Taken together, our results demonstrate that investors do account for the adverse effects of fund scale in making their capital allocation decisions, and that the rational expectations equilibrium does a reasonable job of approximating the observed equilibrium in the mutual fund market. In contrast, the previous literature has often deemed mutual fund investors as naive return chasers because fund flows respond to past performance even though performance is not persistent.[5] Furthermore, many papers in the mutual fund literature have documented that mutual fund returns show little evidence of outperformance.[6] While these findings led many researchers to question the rationality of mutual fund investors, Berk and Green (2004) argue that they are consistent with a model of how competition between rational investors determines the net alpha in equilibrium. We contribute to this debate by presenting findings that are hard to reconcile with anything other than the existence of rational fund flows.

Most closely related to our paper is Barras, Gagliardini, and Scaillet (2022), who also find that not only skill, but also scalability (i.e., the degree of fund-level decreasing returns to scale) vary substantially across funds. They find that the majority of funds *do* add value, consistent with theoretical models in which investors rationally allocate capital across funds taking into account fund-level diseconomies of scale. In contrast, we propose scalability as the primary determinant of the flow-performance relation based on such models,[7] a hypothesis that we test by exploiting the fact that scalability varies substantially across funds. Furthermore, contrary to their analysis, we quantify the importance of cross-sectional variation in decreasing returns to scale for capital allocation decisions.

2 Definitions and Hypotheses

To formally derive our hypothesis, we use the notation and setup presented in Berk and van Binsbergen (2016). Let R_{it}^n denote the return in excess of the risk free rate earned by investors in fund i at time t and let R_{it}^B denote the excess return of the manager's benchmark over the

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

[6] See Malkiel (1995), Gruber (1996), Carhart (1997), Fama and French (2010), and Del Guercio and Reuter (2013), among others.

[7] Studies that discuss other determinants of the flow-performance relation include Huang, Wei, and Yan (2007), Chen, Goldstein, and Jiang (2010), Ferreira et al. (2012), Huang, Wei, and Yan (2012), and Franzoni and Schmalz (2017).

same time interval. At time t, the investor observes the manager's net return outperformance,

$$\alpha_{it+1} \equiv R_{it}^n - R_{it}^B. \tag{1}$$

We assume throughout that α_{it}, or the manager's return outperformance, can be expressed as follows:

$$\alpha_{it} = a_i - h_i\left(q_{it-1}\right) + \epsilon_{it}, \tag{2}$$

where q_{it-1} denotes assets under management (AUM) of fund i at time $t-1$, a_i denotes a parameter that is the difference between fund i's gross alpha on the first dollar and the percentage fee that its manager charges, and ϵ_{it} is the noise in observed performance. Here $h_i\left(q\right)$ is a strictly increasing function of q that captures the decreasing returns to scale the manager faces, which can vary by fund.

At time t, investors use the time t information set I_t to update their beliefs on a_i resulting in the probability density $g_t\left(a_i\right)$ implying that the expectation of a_i at time t is:

$$\theta_{it} \equiv E\left[a_i \left| I_t\right.\right] = \int a_i g_t\left(a_i\right) da_i. \tag{3}$$

To avoid triviality, we assume that $g_t\left(\cdot\right)$ is not a degenerate probability density. Note that q_{it}, R_{it}^n and R_{it}^B are elements of I_t. Let $\overline{\alpha}_{it}\left(q\right)$ denote investors' subjective expectation of α_{it+1} when investing in fund i that has size q between time t and $t+1$, that is, fund i's *net alpha*:

$$\overline{\alpha}_{it}\left(q\right) = \theta_{it} - h_i\left(q\right). \tag{4}$$

In equilibrium, the size of the fund q_{it} adjusts to ensure that there are no positive net present value investment opportunities so $\overline{\alpha}_{it}\left(q_{it}\right) = 0$ and

$$\theta_{it} = h_i\left(q_{it}\right). \tag{5}$$

Now note that the manager's return outperformance, α_{it}, is a signal that is informative about a_i. The conditional probability density of α_{it} at time $t-1$, $f_{t-1}(\alpha_{it})$, satisfies the following condition in equilibrium:

$$E\left[\alpha_{it}\left|I_{t-1}\right.\right] = \int \alpha_{it} f_{t-1}(\alpha_{it})\, d\alpha_{it} = \overline{\alpha}_{it-1}(q_{it-1}) = 0. \tag{6}$$

In other words, the manager's return outperformance can be expressed as follows:

$$\alpha_{it} = s_{it} - h_i(q_{it-1}),$$

where $s_{it} = a_i + \epsilon_{it}$. Our hypothesis relies on the insight that good news (i.e., high s_{it}) implies good news about a_i and bad news (low s_{it}) implies bad news about a_i. The following lemma shows that this condition holds generally. That is, θ_{it} is a strictly increasing function of s_{it}.

Lemma 1 *If the likelihood ratio* $f_{t-1}\left(s_{it}\left|a_i\right.\right)/f_{t-1}\left(s_{it}\left|a_i^c\right.\right)$ *is monotone in* s_{it} *(increasing if* $a_i > a_i^c$ *and decreasing otherwise),*

$$\frac{\partial \theta_{it}}{\partial s_{it}} > 0. \tag{7}$$

Proof. *See Milgrom (1981).* ■

In addition, we assume that the costs that manager i faces in expanding the fund's scale is given by:

$$h_i(q) = b_i h(q), \tag{8}$$

where $b_i > 0$ is a parameter that captures the cross-sectional variation in the fund's returns to scale technology and $h(q)$ is a strictly increasing function of q, which essentially determines the form of decreasing returns to scale technology that is common across all funds. Using (8) to rewrite (5) now gives

$$q_{it} = h^{-1}\left(\frac{\theta_{it}}{b_i}\right). \tag{9}$$

The following lemma shows how the size of the fund q_{it} depends on the information in s_{it} or the parameter b_i.

Lemma 2

$$\frac{\partial q_{it}}{\partial s_{it}} = \frac{1}{b_i h'(q_{it})} \frac{\partial \theta_{it}}{\partial s_{it}} \tag{10}$$

and

$$\frac{\partial q_{it}}{\partial b_i} = -\frac{h(q_{it})}{b_i h'(q_{it})}. \tag{11}$$

Proof. *See Appendix A.* ■

Next, let the flow of capital into mutual fund i at time t be denoted by F_{it}, that is,

$$F_{it+1} \equiv \log(q_{it+1}/q_{it}).$$

Differentiating this expression with respect to s_{it+1},

$$\frac{\partial F_{it+1}}{\partial s_{it+1}} = \frac{1}{q_{it+1}} \frac{\partial q_{it+1}}{\partial s_{it+1}} = \frac{1}{q_{it+1}} \frac{1}{b_i h'(q_{it+1})} \frac{\partial \theta_{it+1}}{\partial s_{it+1}} > 0, \tag{12}$$

where the second equality follows from (10) and the inequality follows from Lemma 1, so good (bad) performance results in an inflow (outflow) of funds. This result is one of the important insights from Berk and Green (2004).

Given the importance of returns to scale technology in determining the size of a fund, a natural question to ask is, what is the implication of steeper decreasing returns to scale for the flow-performance relation? We answer this question by computing the derivative of the flow-performance sensitivity with respect to b_i: steeper decreasing returns to scale must lead to a smaller flow of funds response to performance if and only if

$$\frac{\partial}{\partial b_i}\left(\frac{\partial F_{it+1}}{\partial s_{it+1}}\right) < 0. \tag{13}$$

We show in Appendix A that condition (13) is equivalent to

$$\frac{\partial}{\partial q_{it+1}}\left(\frac{\partial \log\left(h\left(q_{it+1}\right)\right)}{\partial \log\left(q_{it+1}\right)}\right) < 0, \tag{14}$$

which means that the size elasticity of performance is decreasing in the size of a fund. This assumption is satisfied for many functional forms, including the commonly used specification in empirical studies that assume the function h is logarithmic ($h\left(q\right) = \log\left(q\right)$). In practice, such "concavity" of the fund's decreasing returns to scale technology can arise endogenously from funds changing their investment behavior as they grow: indeed, prior studies find that larger funds trade less and hold more-liquid stocks to mitigate the performance erosion due to diseconomies of scale.[8] This leads to the following proposition, i.e., steeper decreasing returns to scale lead to a weaker flow response to performance. We take this as our main hypothesis that we will take to the data.

Proposition 3 *Under condition (14), the derivative of the flow-performance sensitivity with respect to the decreasing returns to scale parameter is negative, that is,*

$$\frac{\partial}{\partial b_i}\left(\frac{\partial F_{it+1}}{\partial s_{it+1}}\right) < 0.$$

Proof. *See Appendix A.* ■

3 Data

Our data come from CRSP and Morningstar. We require that funds appear in both the CRSP and Morningstar databases, which allows us to validate data accuracy across the two. We merge CRSP and Morningstar based on funds' tickers, CUSIPs, and names. We then compare assets and returns across the two sources in an effort to check the accuracy of each

[8]See, for example, Pollet and Wilson (2008), Pástor, Stambaugh, and Taylor (2020), and Busse et al. (2021).

match following Berk and van Binsbergen (2015) and Pástor, Stambaugh, and Taylor (2015). We refer the readers to the data appendices of those papers for the details. Our mutual fund data set contains 3,066 actively managed domestic equity-only mutual funds in the United States between 1991 and 2014.[9]

We use Morningstar Category to categorize funds into nine groups corresponding to Morningstar's 3×3 stylebox (large value, mid-cap growth, etc.). We also use keywords in the Primary Prospectus Benchmark variable in Morningstar to exclude bond funds, international funds, target funds, real estate funds, sector funds, and other non-equity funds. We drop funds identified by CRSP or Morningstar as index funds, in addition to funds whose name contains "index." We also drop any fund observations before the fund's (inflation-adjusted) AUM reaches \$5 million.

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

3.1 Fund Performance

We take two approaches to measuring fund performance. First, we use the standard risk-based approach. The recent literature finds that investors use the CAPM in making their capital allocation decisions (Berk and van Binsbergen (2016)), and hence we adopt the CAPM. In this case the risk adjustment R_{it}^{CAPM} is given by:

$$R_{it}^{\text{CAPM}} = \beta_{it} \text{MKT}_t,$$

where MKT_t is the realized excess return on the market portfolio and β_{it} is the market beta of fund i. We estimate β_{it} by regressing the fund's excess return to investors onto the market portfolio over the sixty months prior to month t. Because we need historical data of sufficient length to produce reliable beta estimates, we require a fund to have at least two years of

[9]We start the sample in 1991, the first year in which CRSP provides monthly data on funds' size.

track record to estimate the fund's betas from the rolling window regressions.

Second, we follow Berk and van Binsbergen (2015) by taking the set of passively managed index funds offered by Vanguard as the alternative investment opportunity set.[10] We then define the Vanguard benchmark as the closest portfolio in that set to the mutual fund. Let R_t^j denote the excess return earned by investors in the j'th Vanguard index fund at time t. Then the Vanguard benchmark return for fund i is given by:

$$R_{it}^{\text{Vanguard}} = \sum_{j=1}^{n(t)} \beta_i^j R_t^j,$$

where $n(t)$ is the total number of index funds offered by Vanguard at time t and β_i^j is obtained from the appropriate linear projection of active mutual fund i onto the set of Vanguard index funds. As pointed out by Berk and van Binsbergen (2015), by using Vanguard funds as the benchmark, we ensure that this alternative investment opportunity set was marketed and tradable at the time. Again, we require a minimum of 24 months of data to estimate β_i^j's necessary for defining the Vanguard benchmark for fund i.

Our measures of fund performance are then $\widehat{\alpha}_{it}^{\text{CAPM}}$ and $\widehat{\alpha}_{it}^{\text{Vanguard}}$, the realized return for the fund in month t less $\widehat{R}_{it}^{\text{CAPM}}$ and $\widehat{R}_{it}^{\text{Vanguard}}$. The average of $\widehat{\alpha}_{it}^{\text{CAPM}}$ is $+1.0$ bp per month, whereas the average $\widehat{\alpha}_{it}^{\text{Vanguard}}$ is -1.7 bp per month.

3.2 Fund Size and Flows

We adjust all AUM numbers by inflation by expressing all numbers in January 1, 2000 dollars. Adjusting AUM by inflation reflects the notion that the fund's real (rather than nominal) size is relevant for capturing decreasing returns to scale in active management. That is, lagged real AUM corresponds to q_{it-1} in the model from Section 2. There is considerable dispersion in real AUM: the inner-quartile range is from \$44 million to \$621 million, while the 99th percentile is orders of magnitude larger at \$16 billion.

[10] See Table 1 of that paper for the list of Vanguard Index Funds used to calculate the Vanguard benchmark.

Fund flows are measured in two different ways. First, as in the model, we define fund flow F as the logarithmic change in real AUM, that is, the percentage change in fund size. Alternatively, we calculate flows for fund i in month t as:

$$F_{it} = \frac{AUM_{it} - AUM_{it-1}(1 + R_{it})}{AUM_{it-1}(1 + R_{it})},$$

where AUM_{it} is the nominal AUM of fund i at the end of month t, and R_{it} is the total return of fund i in month t.[11] Under this more traditional definition of F, flows represent the percentage change in new assets. The flow of fund data contain some implausible outliers, so we winsorize each of the two flow variables at its 1st and 99th percentiles. Mean percentage changes (per month) in fund size and in new assets are 0.8% and 0.5%, respectively.

4 Method

Our analysis relies on a theoretical link between decreasing returns to scale and flow sensitivity to returns. We discuss how we estimate each part in the following sections.

4.1 Fund-Specific Decreasing Returns to Scale (DRS)

Empirically, we assume that the net alpha that manager i generates by actively managing money is given by:

$$\alpha_{it} = a_i - b_i \log(q_{it-1}) + \epsilon_{it}, \tag{15}$$

where a_i is the fund fixed effect, b_i captures the size effect, which can vary by fund, and q_{it-1} is the dollar size of the fund.[12]

[11] Note that we use $AUM_{it-1}(1 + R_{it})$ in the denominator rather than AUM_{it-1}, which is typically used in much of the existing literature on fund flows. Unfortunately, this definition distorts the flow for very large negative returns, as shown by Berk and Green (2004): for example, liquidition of a fund, i.e., $AUM_{it} = 0$, implies a flow of $-(1 + R_{it})$. Our measure of the flow of funds is equal to, and correctly so, -1 in this case. Regardless, our findings are unaffected by using the more common definition of the flow.

[12] To the extent that fee changes are significant, it is possible that our results going forward might be sensitive to whether we use the net alpha or the gross alpha in equation (14). We report the former set of results but find that the latter results lead to the same conclusions. In fact, our results are stronger in the

The simple regression model in equation (15) corresponds to the model in Section 2. This model further assumes the form of the fund's decreasing returns to scale technology is logarithmic, which is often used to empirically analyze the nature of returns to scale due to severe skewness in dollar fund size.

We depart from much of the literature by allowing for heterogeneity in the size-performance relation across funds. Indeed, the effect of scale on a fund's performance is unlikely to be constant across funds. For example, a fund's returns should be decreasing in scale more steeply for those that have to invest in small and illiquid stocks.

We start our analysis by estimating fund-specific a_i and b_i parameters. It is well known that the OLS estimators of the coefficients b_i in (15) are subject to a small sample bias (Stambaugh (1999)). The small sample bias arises because changes in fund size tend to be positively correlated with unexpected fund returns. To address this bias, we follow Amihud and Hurvich (2004) and Barras, Gagliardini, and Scaillet (2022) and include a proxy for the size innovation $v_{i\tau}^c$ (see Appendix B).[13] Specifically, for each fund i at time t, we define the fund-specific DRS estimate $\widehat{b}_{it}$ to be the coefficient of $-\log(q_{i\tau-1})$ in the time-series regression of $\widehat{\alpha}_{i\tau}$ on $-\log(q_{i\tau-1})$ and $v_{i\tau}^c$ (including an intercept) using sixty months of data before time t. Estimating b_i fund-by-fund leads to imprecise estimates especially for funds with short track records, so we require at least three years of data to estimate fund-specific returns to scale of a mutual fund.

Note that the estimate of b_i, $\widehat{b}_{it}^m$, can be obtained using measures of the alpha estimated under model $m \in \{\text{CAPM}, \text{Vanguard}\}$. Intuitively, these estimates represent, for investors who use model m in making capital allocation decisions, their perception of the effect of size on performance for fund i at time t based on information prior to time t.

Panel A of Figure 1 shows how the cross-sectional distribution of $\widehat{b}_{it}$ using the CAPM

unreported results using the gross alpha in equation (14). This robustness is consistent with the evidence in the existing literature: fee changes are rare, so they are unlikely to play an important role in equilibrating the mutual fund market.

[13] Appendix B considers alternative proxies for $v_{i\tau}^c$ using approaches from the existing literature, as well as using a novel approach that relies on our model. Our results are robust to these alternative approaches.

alpha varies over time. For each month in 1991 through 2014, the figure plots the average as well as the percentiles of the estimated fund-specific b parameters across all funds operating in that month. The plot shows considerable heterogeneity in decreasing returns to scale across funds.[14] For example, the interquartile range is more than 4 times larger than the estimates' cross-sectional median in a typical month. We find that, for the average fund, one percent increase in fund size is typically associated with a sizeable decrease in fund performance of about 0.4 basis points (bp) per month. This evidence suggests that the subjective size-performance relation, perceived by investors in real time, provides identifying variation in the extent of decreasing returns to scale.

Panel B of Figure 1 shows the time evolution of $\widehat{b}_{it}$ when we take Vanguard index funds as the alternative investment opportunity set. Similar to our main measure in Panel A, the alternative measure exhibits a clear heterogeneity in diseconomies of scale across funds, though these estimates typically indicate milder decreasing returns to scale.

4.2 Fund-Specific Flow Sensitivity to Performance (FSP)

We estimate the fund-specific flow sensitivities to past performance by estimating the following regression fund by fund:

$$F_{it} = c_i + \gamma_i P_{it-1} + \upsilon_{it}, \tag{16}$$

where P_{it-1} is annual alpha for the year leading to month $t-1$, computed by compounding the monthly alphas as follows:

$$P_{it-1} = \prod_{s=t-12}^{t-1} \left(1 + R_{is}^{n} - R_{is}^{B}\right) - 1.$$

[14] Some of the heterogeneity in DRS could be attributable to estimation error. See Barras, Gagliardini, and Scaillet (2022) who develop a bias-adjusted approach to inferring the cross-sectional distribution of DRS, finding substantial heterogeneity in DRS even after adjusting for the bias arising from estimation error.

This regression is consistent with empirical evidence that investors do not respond immediately. For example, Berk and van Binsbergen (2016) and Barber, Huang, and Odean (2016) show that flows respond to recent returns, as well as distant returns. Parameter $\gamma_i > 0$ captures the positive time-series relation between performance and fund flows, which can vary by fund.

At time t, we calculate the fund's flow sensitivity to performance by estimating (16) using its data over the subsequent 5 years. For fund i, let $\widehat{FSP}_{it}^{m}$ be the estimated flow-performance regression coefficient of that model, where the performance can be estimated under model $m \in \{\text{CAPM}, \text{Vanguard}\}$. To avoid using imprecise estimates, we require these coefficient estimates to be obtained from at least three years of data. For the average fund, we observe that an increase of 1% in the annual CAPM alpha is associated with an increase of 0.1% in monthly flows next month.

Figures 2 and 3 display the evolution of the distributions of $\widehat{FSP}_{it}$ by plotting the average as well as the percentiles of the estimated flow sensitivities to performance at each point of time. In Figure 2, we estimate the FSP's using the change in fund size to capture flows; in Figure 3, we estimate the FSP's using the change in new assets to define F. Panel A in each figure shows the distribution using the CAPM alpha, and Panel B shows the distribution when net alpha is computed using Vanguard index funds as benchmark portfolios. Note that the results are very similar across the two figures, manifesting considerable heterogeneities in the flow-performance relation across funds. More importantly, these figures show that while the average $\widehat{FSP}_{it}$ for both versions of the flow variable do not exhibit any obvious trend, they are certainly time varying. As the red dashed lines in the figures make clear, the distributions remain roughly the same over our sample period, conditional on the median.

5 Results

5.1 DRS and Flow Sensitivity to Performance

To investigate whether fund-specific decreasing returns to scale parameters are related to capital allocation decisions, we run panel regressions of fund i's flow sensitivity to performance going forward in month t, $\widehat{FSP}_{it}$, on the fund's returns to scale estimated as of the previous month-end, $\widehat{b}_{it}$. We test the null hypothesis that the slope on $\widehat{b}_{it}$ is zero.[15] We report the results in Tables 2 and 3.[16] In Panel A, we report the results using the change in fund size to capture flows; in Panel B, we examine their robustness using the change in new assets to define F. The first two columns in each panel use the CAPM as the benchmark, while the last two columns use Vanguard index funds as the benchmark.

We show results based on raw estimates in Table 2. We focus on variation coming from the market equilibrating mechanism beyond differences in sensitivity across funds and over time by including month and fund fixed effects. The fund fixed effects absorb the cross-sectional variation in flow/performance sensitivity, for example, due to differences in investor clientele across funds, while the time fixed effects soak up variation in flow/performance sensitivity due to factors such as investor attention allocation over time. Indeed, there is evidence of clientele differences because some investors tend to update faster than others,[17] and Figures 2 and 3 show how the average as well as the median of flow-performance dynamics vary considerably over time.[18]

[15] Surely, not only the independent variable, but the dependent variable are measured imprecisely. The measurement error in $\widehat{b}_{it}$ will bias the OLS estimator toward zero. While the measurement error in $\widehat{FSP}_{it}$ will not induce bias in the OLS coefficients, it will render their variance larger. For now, we do not worry, as the errors-in-variables problem will work against us from finding a statistically significant relation that the model predicts.

[16] Tables 2 and 3 report the double clustered (by fund and time) standard errors.

[17] See Berk and Tonks (2007).

[18] Ferreira et al. (2012) discuss the role of economic, financial, and mutual fund industry development in determining the flow-performance relation across countries, while Franzoni and Schmalz (2017) document that mutual funds' flow-performance sensitivity depends on aggregate risk-factor realizations in a hump-shaped way. Note that our fixed-effect approach already controls for these factors, since time fixed effects subsume any potential time-series variation in FSP due to different stages of development in the US and/or across market states.

In the odd columns, we only include month and fund fixed effects. The results in Panel A are consistent with the main prediction of our model: the estimated coefficients on $\widehat{b}_{it}$ are negative and highly significant, with t-statistics of -6.3 in column 1 and -5.4 in column 3. These findings are unaffected by including a host of controls in the even columns, where we add proxies for participation costs, as considered by Huang, Wei, and Yan (2007),[19] as well as performance volatility and fund age.[20] The slopes on $\widehat{b}_{it}$ remain negative and highly significant, with t-statistics of -7.1 in column 2 and -5.2 in column 4, and their magnitudes are larger compared to odd columns where controls are excluded.

In Panel B, the same conclusions continue to hold when we consider $\widehat{FSP}_{it}$ estimated using the more traditional definition of F: the percentage change in new assets. Just like in Panel A, the coefficients on $\widehat{b}_{it}$ are significantly negative, and they increase in magnitude when we include a host of controls.

Table 3 repeats this exercise with percentile ranks in each month based on $\widehat{b}_{it}$ and $\widehat{FSP}_{it}$. In this case, we do not use month fixed effects, as percentile ranks already soak up any time variation in the flow-performance relation. In each column, the estimated coefficient on $\widehat{b}_{it}$ is significantly negative at the 1% confidence level. Again, the addition of other potential determinants of the flow-performance relationship makes the slope coefficients on $\widehat{b}_{it}$ even more negative (compare columns 1 and 3 against 2 and 4 in each panel, respectively).

To summarize, we find a strong negative relation between decreasing returns to scale and flow sensitivity to performance. This relation, which is statistically significant, is consistent with the presence of investors rationally accounting for the adverse effects of fund scale in making their capital allocation decisions. Unfortunately, these coefficient values are likely to be biased toward zero because of the measurement error in $\widehat{b}_{it}$. In Section 5.1.1, we first gauge the severity of attenuation bias under the classical measurement error assumption.

[19] Specifically, we use marketing expenses, star family affiliation, family size, as well as fund size, to proxy for the variation in investors' information costs across funds.

[20] Huang, Wei, and Yan (2012) find that the flow-performance sensitivity is weaker for funds with more volatile past performance and longer track records.

In Sections 5.1.2, we then exploit a set of fund characteristics that are plausibly related to the scalability of the funds' investment strategies as instruments for heterogeneity in DRS parameters across funds to address the attenuation bias associated with estimating the DRS-FSP relation. Finally, in Section 5.1.3, we propose a way of assessing the economic magnitude of these estimated coefficients by computing counterfactual fund sizes.

5.1.1 DRS-FSP Relation Under the Classical Measurement Error Assumption

We have estimated fund-specific b_i parameters based on a rolling estimation window. As noted earlier, estimating b_i fund by fund leads to imprecise estimates especially for funds with short track records. To gauge the severity of attenuation bias, we adjust the estimated coefficients on $\widehat{b}_{it}$ in Table 2 for the errors-in-variable (EIV) problem, assuming that the errors are of the classical type: they are purely random, have mean zero, and are uncorrelated with the regressors, including the actual b_i, and with the regression errors. Using the standard errors of $\widehat{b}_{it}$ to estimate the variance of measurement error in b_i, we can calculate the EIV-adjusted coefficients, reported in the last row of each panel.

As expected, the simple estimates of the DRS-FSP relation tend to be too small in magnitude. For example, when the DRS-FSP relation is estimated based on the CAPM to measure fund performance and on the change in fund size to capture flows controlling for other potential determinants of the flow-performance relationship (i.e., column 2 in Panel A), the coefficient becomes substantially more negative with the EIV adjustment: -16.11, compared to -1.30 without this adjustment. Bias is even more severe for estimates based on the Vanguard benchmark than those based on the CAPM. When the DRS-FSP relation is estimated using the Vanguard benchmark, the EIV-adjusted coefficients are over 26 times larger than their simple-estimate counterparts (see the last two columns of Table 2). Of course, these results are only true if the errors are indeed of the classical type, but they illustrate that our estimates of the DRS-FSP relation are likely to be severely biased against confirming our model prediction.[21] Thus, the fact that we find a strong relation between DRS and FSP despite this counterveiling effect of measurement error further strengthens the support for the model.

5.1.2 DRS-FSP Relation Using the Characteristic Component of DRS

In this section, we explore which fund characteristics are correlated with the observed heterogeneity in returns to scale. Based on this analysis, we obtain an economically interpretable component of $\widehat{b}_i$ based on fund characteristics, using which we re-estimate the DRS-FSP relation. The characteristic-based approach taken here exploits many fund characteristics that are relevant for identifying variation in DRS. The prior evidence of fund-level DRS depending on fund characteristics suggests that this method *is* likely to deliver a more accurate measure of b_i and thus is a reasonable way to mitigate the errors-in-variable problem. Indeed, when we conduct the analysis using the characteristic component of DRS, the estimates of the DRS-FSP relation become substantially more negative than in Table 2 and they are comparable in magnitude to those implied by the classical measurement error assumption.

Determinants of Fund-Level DRS We investigate a number of characteristics that seem relevant a priori (also from the previous literature) for heterogeneity in returns to scale: the number of managers, volatility, expense ratios, marketing expenses, an international exposure indicator, turnover, and log fund size. In analyzing the dependence of returns to scale on fund characteristics, we control for the loadings on the market, size, value, and momentum factors to capture fund style and risk.[22]

The first characteristic, $NMgr$, is the number of managers managing the fund. About 59% of our funds are multi-manager funds. The second characteristic, $Std\,(Alpha)$, is the standard deviation of a fund's alphas over the prior 1 year. The next two characteristics

[21] Barras, Gagliardini, and Scaillet (2021) make a closely related observation that the error-in-variable bias can have significant impact on the cross-sectional distribution of scale coefficient b_i.

[22] We estimate these risk exposures by regressing the fund's return on the four Fama-French-Carhart (FFC) factors over the prior sixty months.

we examine are the fund's expense ratios and marketing expenses. The fifth characteristic, $1(IntExp)$, is an indicator for funds with a high degree of international exposure, defined as follows. We test the null hypothesis that the coefficients on three Vanguard international index funds are 0.[23] For any given fund, the international exposure dummy is equal to one if we reject the null hypothesis at the 5% confidence level. Although we focus our attention on domestic funds, about 28% of them are highly exposed to international shocks. The sixth characteristic is the fund's average annual turnover (from CRSP).[24] Median turnover is 64% per year. We also examine whether the fund's log real AUM matters for its DRS technology.

We examine how these characteristics affect the impact of a fund's scale on its performance by running panel regressions of fund i's DRS parameter using only its observations prior to month t, $\widehat{b}_{it}$, on the fund's characteristics at the end of the previous month. Table 4 shows the estimation results.[25] Panel A reports the results using the CAPM as the benchmark; Panel B uses Vanguard index funds as the benchmark.

In both panels, we find significant relations between $\widehat{b}$ and three characteristics: the number of managers, volatility, and expense ratios (see the first three columns of Table 4). While the slope on marketing expenses (column 4) is consistently negative and the slope on fund size (column 7) is consistently positive, they are insignificant. On the other hand, we find a statistically insignificant relation between returns to scale and turnover (column 6) of mixed signs. Finally, we find that the relation between returns to scale and international exposure (column 5) is positive, but it is both statistically and economically insignificant.

When all seven fund characteristics are added at the same time, the estimated slopes on volatility and expense ratios are robust, indicating steeper decreasing returns to scale for higher-volatility funds and funds charging higher expense ratios. We continue to find a negative relation between $\widehat{b}$ and the number of managers, indicating steeper decreasing returns to

[23] Recall that we use a set of eleven Vanguard index funds to calculate the Vanguard benchmark. Three of these index funds are international: European Stock Index, Pacific Stock Index, and Small-Cap Value Index.

[24] We winsorize turnover at the 1st and 99th percentiles.

[25] Standard errors of these regressions are two-way clustered by fund and time.

scale for sole-manager funds, although the relation becomes statistically insignificant. Marketing expenses now enter with a consistently significantly negative slope, indicating that decreasing returns to scale are less pronounced for funds with higher marketing expenses. The relation between returns to scale and fund size remains statistically insignificantly positive. Finally, the slopes on turnover and on international exposure are still insignificant regardless of how one defines the benchmark, but their signs now flip to negative. Therefore, we focus on the results when the five fund characteristics whose signs are robust are added at the same time (see column 8 of Table 4).

While we leave the task of deriving these relations between fund characteristics and diseconomies of scale in an equilibrium model for future research, these results are consistent with the following interpretations. The division of labor within a fund might alleviate the negative impact of size on performance, so it is the fund's assets under management on a per-manager basis that matters for capturing decreasing returns to scale. If so, a multi-manager fund would be able to deploy capital more easily and, consequently, exhibit milder decreasing returns to scale. Pástor, Stambaugh, and Taylor (2015) offer a narrative for why higher-volatility funds might also exhibit steeper decreasing returns to scale, while steeper decreasing returns to scale for funds charging higher expense ratios are consistent with the model of Stambaugh (2020). Finally, gradual decreasing returns to scale for funds with high-marketing expenses are consistent with funds marketing to attract more flows only if they can manage the performance erosion associated with growing fund size.

Implications for DRS-FSP Relation Instead of using the coefficient estimates $\widehat{b_i}$ as before, we now use the estimates from column 8 of Table 4 to obtain an economically interpretable component of $\widehat{b_i}$ based on fund characteristics. This implementation choice assumes that all the funds with the same fund characteristics share the same b value. While ignoring variation might potentially lead to inaccuracy in quantifying fund-specific b, this method actually seems to increase the accuracy of the b_i estimate by dramatically reducing estimation errors. While about 36% of the funds in our sample end up with negative $\widehat{b}_i$, about 7% of their predicted values based on fund characteristics, denoted by $\widehat{b}_i^{Char}$, are negative. These results seem sensible since, theoretically, all mutual funds must face decreasing returns to scale in equilibrium.

To address the attenuation bias associated with estimating the DRS-FSP relation, we replace $\widehat{b}_i$ by $\widehat{b}_i^{Char}$ and rerun the regressions in Table 2, whose results are tabulated in Table 5. When we rerun our analysis in Table 2 with characteristic-based DRS, we obtain similar and even stronger results indicating that steeper decreasing returns to scale attenuate flow sensitivity. Table 5 shows that $\widehat{b}_i^{Char}$ has significantly negative slopes throughout, but the coefficients' estimated values become substantially more negative than in Table 2. For example, the estimated coefficients are typically more than 6 times larger when we use the change in fund size to capture flows (compare Panel A of Tables 2 and 5).

In summary, when we conduct the analysis using cleaner measures of decreasing returns to scale, our conclusions on the effects of decreasing returns to scale on capital allocation only become stronger. These estimates of the DRS-FSP relation are comparable in magnitude to those implied by the classical measurement error assumption.

5.1.3 Simulated DRS-FSP Relation

In this section, we use our model to ask how much capital is allocated the way it is because of these differences in decreasing returns to scale. Specifically, we compute counterfactual fund sizes by assuming the investors believe a priori that returns are decreasing in scale at the same (average) rate for all funds.

Two factors fully determine the magnitude of capital response to performance in a rational model — the degree of decreasing returns to scale, and the prior and posterior beliefs about managerial skill. This means that, for a given value of b in equation (15), the prior uncertainty about a, σ_0, can be inferred from the flow-performance relation, as long as investors update their posteriors with the history of returns as Bayesians.

We simulate benchmark-adjusted fund returns from equation (15). To this end, we assume that (i) investors' prior on a fund's ability is that a_i is normally distributed with mean θ_{i0} and standard deviation σ_0 and (ii) the error term, ϵ_{it}, is normally distributed with mean zero and variance σ^2. Then, it is straightforward to show that the mean of investors' posteriors will satisfy the following recursion:

$$\theta_{it} = \theta_{it-1} + \frac{\sigma_0^2}{\sigma^2 + t\sigma_0^2}\alpha_{it}.$$

Using (9), we compute fund size as follows:

$$q_{it} = \exp\left(\frac{\theta_{it}}{b_i}\right).$$

We begin by tying down the model parameters that can be set directly. Following Berk and Green (2004), we set $\sigma = 20\%$ per year, or 5.77% per month. Recall that investors believe a priori that a_i is normally distributed with mean θ_{i0} and standard deviation σ_0. Since investors are assumed to have rational expectations, this is also the distribution from which we draw each fund's skill. We shall also assume that funds shut down the first time $\theta_{it} < \overline{\theta}$, where we set $\overline{\theta} = 0$.[26] These parameter values are summarized in Panel A of Table 6. It is straightforward to see that the only remaining parameters that we need to set for simulating data are b_i, θ_{i0} and σ_0.

The empirical distribution of b is approximated by a scaled Beta distribution, from which we draw b randomly.[27] In that case, assuming that θ_0 is independent of b gives rise to distributions of fund size considerably more disperse than in our actual sample. Specifically, the simulated fund sizes tend to be too big for funds whose returns decrease in scale more

[26] Intuitively, managers incur fixed costs of operation each period. These costs can be, for example, overhead, back-office expenses, and the opportunity cost of the manager's time. Managers will optimally choose to exit when they cannot cover their fixed costs.

[27] Specifically, we fit the parameters for the scaled Beta distribution such that the mean, variance, skewness, and kurtosis of the simulated b_i are close to those of the fitted values of $\widehat{b}_{it}$ based on fund characteristics under the IV approach in the previous section (see the last column of Panel A of Table 4).

gradually, while the simulated fund sizes tend to be too small for those that exhibit steeper decreasing returns to scale. In turn, we model the prior mean as a linear function of b, $\theta_0(b)$. Our approach is to fit the parameters governing this function such that the simulated mean and standard deviation of log fund size essentially match the empirical benchmark values of 5.1 and 1.9, respectively.[28]

We set the prior uncertainty (σ_0) to match the average flow-performance relation in the data. To this end, we construct 10,000 samples of simulated panel data for 300 funds over 100 months. We simulate a given sample by first drawing each fund's DRS b_i randomly from the scaled Beta distribution consistent with the empirical distribution of b_i estimates, while drawing the fund's skill a_i from a normal distribution with mean $\theta_0(b_i)$ and standard deviation σ_0. Next, we simulate the values of ϵ_{it}, building up the panel data of α_{it} and q_{it}. For each fund in the sample, we run the following regression using data for just that fund to estimate its flow-performance sensitivity:

$$\log(q_{it}/q_{it-1}) = c_i + \gamma_i \alpha_{it} + \upsilon_{it}.$$

Given all other parameters, we set σ_0 so that the mean of the average $\widehat{\gamma}_i$ across simulated samples matches the average $\widehat{FSP}_{it}$ in our actual sample. Panel B of Table 6 shows the value of σ_0 that resulted from this process. It also contains the values of the parameters governing the scaled Beta distribution of DRS and those of the parameters governing the prior mean that we use in our simulation analysis. The last three columns of Panel B report all the moments that we target in our calibration, as well as their values in both the actual and simulated data. Note that the simulated moments in the model closely match the target moments.

[28]Note that there generally exist multiple ways prior mean as a function of b for which the simulated mean and standard deviation of log fund size can match the empirical benchmark values. To pick a single function, we impose the additional constraint that the simulated mean of log fund size is decreasing in b. This constraint is motivated by empirical evidence presented later in Section 5.2: steeper decreasing returns to scale shrink fund size.

Thus far, we have provided empirical evidence that steeper decreasing returns to scale imply less flow sensitivity to performance. For example, as shown in column 2 of Panel A of Tables 2 and 5, the estimates of the DRS-FSP relation with the EIV adjustment and using the characteristic component of DRS are -16.1 and -11.8, respectively. To assess the economic magnitude of such estimates, we estimate the DRS-FSP relation in each of the simulated samples. Panel A of Table 7 shows summary statistics of these estimates across simulated samples from the calibrated model. The DRS-FSP relation estimates in the calibrated model tend to be larger in magnitude compared to column 2 of Panel A of Tables 2 and 5. But importantly, the empirical estimates of the DRS-FSP relation lie comfortably within the 90% confidence interval for simulated DRS-FSP relation estimates and vice versa. Thus, it appears that the magnitude of the DRS-FSP relation estimates from the data is consistent with what the model predicts. This result suggests that the calibrated model does a good job of capturing capital allocation patterns in the data.

To quantitatively assess the role of heterogeneity in returns to scale in capital allocation, we must construct a counterfactual. We construct the counterfactual by assuming investors learn about skill based on distorted beliefs that the fund exhibits average decreasing returns to scale. Specifically, the counterfactual investors assume that $b_i = 0.0041$ for all funds.[29] Then, updating investors' beliefs with the history of its returns under the counterfactual assumption, we compute what the size of the fund would have been.

Again, we construct $10,000$ samples of simulated panel data for 300 funds over 100 months. To simulate a given sample, we first draw each fund's DRS b_i randomly from the scaled Beta distribution consistent with the distribution of fund-specific b estimates, while we draw the fund's skill a_i from a normal distribution with mean $\theta_0(b_i)$ and standard deviation

[29] Note that the counterfactual investors have distorted beliefs about the skill level consistent with their distorted beliefs that $b_i = 0.0041$ for all funds, i.e., they believe a priori that a_i is normally distributed with mean $\theta_{i0} = \theta_0(0.0041)$ for all funds. Alternatively, we can assume that the counterfactual investors have rational expectations about the skill level, i.e., they know the true $\theta_0(b_i)$ for each fund's DRS parameter b_i, but the counterfactual sizes under this assumption contrast even more sharply with the simulated outcomes because such investors would have beliefs about skill that are *inconsistent* with their beliefs about scalability.

σ_0. Next, we draw the random values of ε_{it}, building up the panel data of r_{it} and q_{it}. For every i and t, we compute the fund's size under the counterfactual, q_{it}^C, as detailed above. Finally, for each sample, we calculate the R^2 from a regression of $\log(q_{it})$ on $\log(q_{it}^C)$ to check the goodness of fit by the counterfactual. Here, 1 minus the R^2 can be interpreted as the fraction of capital allocation explained by individual heterogeneity in DRS.

We report the results from counterfactual simulations in Panel B of Table 7. The first two rows show summary statistics of the coefficient estimates from the regression of $\log(q_{it})$ on $\log\left(q_{it}^C\right)$ across simulated samples; the last row shows summary statistics of the R^2 from this regression across simulated samples.

The counterfactually computed fund sizes explain about 43% of the variation of simulated fund sizes. While counterfactual sizes are positively related to actual sizes, they are considerably larger than actual sizes and their distributions are substantially tighter than those of actual sizes. On the one hand, since the distribution of DRS is positively skewed and leptokurtic, what counterfactual investors believe about the size of funds in the right tail (i.e., funds exhibiting the steepest DRS) greatly influences their beliefs about the average fund size. The fact that fund size is inversely proportional to DRS implies that the counterfactual investors overestimate the size of the funds in the right tail and, in turn, the average fund size. On the other hand, since differences in DRS across funds is a major source of the cross-sectional variation in fund size, these counterfactual investors naturally underestimate the true dispersion in fund size. Thus, the counterfactuals ignoring heterogeneity in DRS are very different than the actual size. In this sense, we can interpret 1 minus the R^2 as a lower bound on the role of heterogeneity in returns to scale on capital allocation: more than half of the cross-sectional variance of fund sizes can be related to cross-sectional variation in decreasing returns to scale parameters, which is economically significant.

To summarize, Table 7 shows that a significant fraction of equilibrium capital allocation can be plausibly explained by investor response to differences in decreasing returns to scale. Not only are fund sizes in the data quantitatively consistent with what our simple model predicts they should be, the magnitude of empirical DRS-FSP relation estimates are consistent with what our simple model would predict.

5.2 DRS and Fund Size in Equilibrium

While the main implication of our model is that steeper decreasing returns to scale attenuate flow sensitivity to performance, another immediate implication is that steeper decreasing returns to scale shrink fund size. Recall that fund size in equilibrium is proportional to the ratio of perceived skill over diseconomies of scale (see equation (9)). This implies that large funds either earn a high gross alpha on the first dollar and/or implement strategies that are highly scalable. We now investigate the importance of the latter effect, while explicitly controlling for the former effect, as well as for fund style and fund age. Table 8 presents the results of this exercise.

To control for the effect of perceived skill, we first form quintile groups sorted on $\widehat{a}_{it}^{Char}$, which are fund fixed effects estimated as of month $t-1$. Specifically, $\widehat{a}_{it}^{Char}$ is equal to the average of fund i's alpha that is adjusted for fund-specific decreasing returns to scale, $\widehat{\alpha}_{i\tau} + \widehat{b}_{it}^{Char} \log(q_{i\tau-1})$, over the 60 months prior to time t, where $\widehat{b}_{it}^{Char}$ is the fund-specific DRS estimated as of month $t-1$ based on fund characteristics (see Section 5.1.2). Then, within each $\widehat{a}_{it}^{Char}$ quintile, we sort funds into five groups based on $\widehat{b}_{it}^{Char}$. We conduct double sorts of funds belonging to the same Morningstar category and to the same age category.[30] After forming the 5×5 $\widehat{a}_{it}^{Char}$ and $\widehat{b}_{it}^{Char}$ groups, we average fund sizes, as measured by log real AUM in month t, of each $\widehat{b}_{it}^{Char}$ quintile over the five $\widehat{a}_{it}^{Char}$ groups. This characteristic control procedure creates a set of quintile $\widehat{b}_{it}^{Char}$ groups with similar levels of perceived skill, and with near-identical distributions of fund style and fund age. Thus, these quintile $\widehat{b}_{it}^{Char}$ groups control for differences in skill, as well as fund style and fund age.

Panel A of Table 8 reports average fund sizes of the 25 $\widehat{a}_{it}^{Char} \times \widehat{b}_{it}^{Char}$ groups using the CAPM as the benchmark. The column labeled "Average" reports the average month-end

[30] Specifically, we assign funds to one of three samples based on fund age: $[0, 5]$, $(5, 10]$, and > 10 years.

fund sizes of the $\widehat{b}_{it}^{Char}$ quintiles, controlling for $\widehat{a}_{it}^{Char}$, fund style and fund age. The row labeled "High-low" reports the differences in average sizes between the first and fifth quintile $\widehat{b}_{it}^{Char}$ groups in each column.[31] The difference in average sizes in the bottom right entry of Panel A indicates that the sizes of funds that are perceived to face steepest decreasing returns to scale tend to be 86% smaller than those of funds that are perceived to be relatively immune to the adverse scale effects. This difference has a robust t-statistic around -15. Hence, steeper decreasing returns to scale shrink fund size, consistent with the above prediction of our model. Importantly, this effect is not only statistically but also economically significant. The patterns within each $\widehat{a}_{it}^{Char}$ quintile moving from low $\widehat{b}_{it}^{Char}$ to high $\widehat{b}_{it}^{Char}$ funds (reading down each column) are very similar. Panel B of Table 8 repeats the same exercise as Panel A, except we use Vanguard index funds as the benchmark. We find the same quantitative patterns.

In summary, steeper decreasing returns to scale shrink fund size in the data. This relation, which is not only statistically but also economically significant, is consistent with the presence of investors rationally accounting for the adverse effects of fund scale in making their capital allocation decisions. Less important but also noteworthy is that the sizes of funds with higher perceived skill tend to be larger than those of funds with lower perceived skill (reading from left to right in each panel), again consistent with our model.

5.2.1 DRS and Optimal Fund Size

Thus far, we have used heterogeneity in decreasing returns to scale across funds and over time to test whether investors respond to the adverse effects of fund scale in making their capital allocation decisions. If investors update their beliefs about skill as in the model, their perception of optimal size ought to converge to true optimal size over a fund's lifetime. This implies that the sizes of older funds should be more closely related to their optimal sizes

[31] To adjust for the strong persistence in fund size, we report standard errors of these differences in average fund sizes between quintile portfolio 5 (high $\widehat{b}_{it}^{Char}$) and quintile portfolio 1 (low $\widehat{b}_{it}^{Char}$) using 60 Newey-West lags.

based on the model than those of younger funds. In this section, we test this prediction and find empirical support for it.

We have estimated fund-specific b_i parameters based on a rolling estimation window. As noted earlier, estimating b_i fund by fund leads to imprecise estimates. In particular, about 36% of the funds in our sample end up with negative $\widehat{b}_i$.[32] While this is not an issue when focusing only on the *relative* steepness of the DRS technology as in the rest of the paper, it is a problem for computing the optimal fund size, which requires that $b_i > 0$ since, theoretically, all mutual funds must face decreasing returns to scale in equilibrium. A straightforward way to deal with this econometric shortcoming is to "shrink" the $\widehat{b}_i$ estimates toward their prior mean, i.e., the average fund-level DRS parameter in our sample, denoted by b^{RD2}, which we estimate using the recursive demeaning procedure of Zhu (2018).[33] Measuring performance using the CAPM, $\widehat{b}^{RD2}$ is statistically significant, indicating that an 1% increase in fund size is associated with a decrease in the fund's CAPM alpha of 0.47 bp per month.[34] All of the resulting fund-specific DRS values, denoted by $\widehat{b}_{it}^{Shr}$, are positive.[35] Then, the corresponding estimator of skill, denoted by $\widehat{a}_{it}^{Shr}$, is equal to the average of fund i's alpha that is adjusted for fund-specific DRS, $\widehat{\alpha}_{i\tau} + \widehat{b}_{it}^{Shr} \log(q_{i\tau-1})$, over the 60 months prior to time t. We employ the average value of the ratios $\widehat{a}_{it}^{Shr}/\widehat{b}_{it}^{Shr}$ over a fund's lifetime to get an estimate for the optimal fund size, $\log(\widehat{q}_i^*)$.[36]

[32] While substantially fewer of $\widehat{b}_i^{Char}$ (the predicted values of $\widehat{b}_i$ based on fund characteristics) turn out to be negative, still about 7% of them are negative.

[33] Pástor, Stambaugh, and Taylor (2015) analyze the nature of returns to scale by developing a recursive demeaning procedure. They find coefficients indicative of decreasing returns to scale both at the fund level and at the industry level, though only the latter is statistically significant. Zhu (2018) improves upon the empirical strategy in PST (by using more recent fund sizes as the instrument) and establishes strong evidence of fund-level diseconomies of scale.

[34] Using Vanguard index funds as benchmarks, the coefficient estimate is again statistically significant, indicating a decrease in fund performance of 0.0015% per month for an 1% increase in fund size.

[35] Formally, the shrinkage estimator is a weighted average of $\widehat{b}_{it}$ and $\widehat{b}^{RD2}$: $\widehat{b}_{it}^{Shr} = w_{it}\widehat{b}_{it} + (1 - w_{it})\,\widehat{b}^{RD2}$, where $w_{it} = \frac{1/\sigma_{b_{it}}^2}{1/\sigma_{b_{it}}^2 + 1/\kappa^2}$, $\sigma_{b_{it}}$ is the standard error of $\widehat{b}_{it}$, and κ is a constant controlling the amount of shrinkage toward $\widehat{b}^{RD2}$. We set the latter constant to ensure that all of the resulting $\widehat{b}_{it}^{Shr}$ values are positive (i.e., $\widehat{b}_{it}^{Shr} \geq 0.00001$).

[36] Note that the optimal fund size here is the size at which the benchmark-adjusted net return is expected to be zero. This is different than, but related to, the optimal amount the manager chooses to actively manage (Berk and van Binsbergen (2015)).

However, this measure of optimal fund size can be different than what investors believe to be optimal (ex-post) if they do not account for individual heterogeneity in decreasing returns to scale. In this case, the valid proxy for the optimal size, perceived by investors, would be the optimal size estimated assuming that the fund size has the same effect on performance across all funds. Once again, the estimator of skill corresponding to $\widehat{b}^{RD2}$, which we denote by $\widehat{a}_{it}^{RD2}$, is equal to the average of $\widehat{\alpha}_{i\tau} + \widehat{b}^{RD2} \log(q_{i\tau-1})$ over the 60 months prior to time t. The alternative measure of optimal fund size $\log\left(\widehat{q}_i^{*RD2}\right)$ is calculated as $\widehat{a}_{it}^{RD2}/\widehat{b}^{RD2}$.

To test the above prediction, we examine how the relation between log real AUM and our measure of optimal fund size depends on fund age. Specifically, we assign funds to one of three samples based on fund age: $[0, 5]$, $(5, 10]$, and > 10 years. In each age sample, we run panel regressions of fund i's log real AUM in month t on the fund's log optimal fund size estimate, $\log(\widehat{q}_i^*)$. We report the results in the first three columns of Table 9.[37] In Panel A, we report the results using the CAPM as the benchmark; in Panel B, we use Vanguard index funds as the benchmark.

Across all three age groups, the estimated coefficients on $\log(\widehat{q}_i^*)$ are positive and highly statistically significant. These coefficient values also increase over a typical fund's lifetime, indicating that this positive relation between the fund's size and its optimal size is stronger for older funds. As the fund ages, investors learn about its optimal size, implying that the equilibrium size is closer to this optimal size measure. In addition, the R^2 of the regressions confirm this insight. The R^2 in the > 10 age sample is the highest and the R^2 decreases monotonically as we move to the samples of ages $(5, 10]$ and $[0, 5]$ funds.

In columns 4 through 6, we run the multiple regression of $\log(q_{it})$ on both $\log(\widehat{q}_i^*)$ and $\log\left(\widehat{q}_i^{*RD2}\right)$ in all three age-sorted samples. We find that the slopes on $\log(\widehat{q}_i^*)$ in the ≤ 5 sample are rendered substantially smaller, and their magnitudes are close to zero (and their signs flip). The slopes on $\log(\widehat{q}_i^*)$ in samples of older funds are also smaller than before, but they *do* remain positive and significant.

[37] Table 9 reports the double clustered (by fund and time) standard errors.

More importantly, we continue to find that $\log\left(\widehat{q}_i^*\right)$ enters with slopes that increase over a typical fund's lifetime. In contrast, while $\log\left(\widehat{q}_i^{*RD2}\right)$ enters with significantly positive slopes across all age-sorted samples, the coefficients decrease in magnitude over a typical fund's lifetime. Taken together, these two results suggest that investors *do* account for the fund heterogeneity in decreasing returns to scale when allocating their capital to older funds, but they allocate their money to younger funds based on the simple version of optimal size. Consistent with this interpretation, the R^2 of the multiple regressions in the > 10 age sample (column 6) remains about the same as in column 3, and the R^2 improvement when we add $\log\left(\widehat{q}_i^{*RD2}\right)$ to the regression of log real AUM on $\log\left(\widehat{q}_i^*\right)$ gets steeper and steeper as we move to the samples of ages $(5, 10]$ and $[0, 5]$ funds.

Our results offer the following narrative. Investors want to account for heterogeneity in decreasing returns to scale, but need to learn about fund-specific values. Given that such fund-specific information is not yet available for young funds, investors use the sample-wide b instead. In particular, the investors only use the $\widehat{q}_i^*$ estimate (together with $\widehat{q}_i^{*RD2}$) in making their capital allocation decisions when a fund grows old enough such that the remaining Bayesian uncertainty on fund-specific b is relatively modest. Thus, it appears that investors in the data might be learning not only about skill but also about decreasing returns to scale.[38] We leave the task of examining the capital allocation implications of learning about fund heterogeneity in decreasing returns to scale technology for future research.

In short, the estimates of optimal size largely explains capital allocation to older funds. Both measures of optimal fund size matter, which is consistent with our narrative that investors account for not only the presence of decreasing returns to scale, but the heterogeneity of decreasing returns to scale across funds.

[38] For formal models that relate capital allocation to learning about returns to scale, see Pástor and Stambaugh (2012) and Kim (2022).

6 Conclusion

One important feature that determines the value of an investment opportunity is its degree of scalability. On the one hand, for the value of a positive net present value project to be finite, its scalability must be limited. On the other hand, in the presence of fixed costs, if the decreasing returns to scale are too large, investors may forgo the project altogether. In this paper, we empirically study the scalability of investment projects in the context of actively managed mutual funds. One common assumption in that literature is that all investment managers face the same degree of decreasing returns to scale while differing in the marginal profitability (gross alpha) on the first dollar invested. In this paper, we show that that assumption does not hold in the data. Heterogeneity in the degree of scalability is a key determinant of the capital allocation decision that investors make. Not only do we find that steeper decreasing returns to scale attenuate flow sensitivity to performance, we also find that differences in decreasing returns to scale across funds are quantitatively important for explaining capital allocation in the market for mutual funds. This heterogeneity is therefore an important driver of the cross-sectional distribution of fund size (AUM).

Appendix

A Proofs

A.1 Proof of Lemma 2

First, note that ε_{it} does not contain information about managerial ability that is not already contained in s_{it}. Because rescaling the fund's returns to scale technology (i.e., changing the parameter b_i) does not change the signal s_{it}, we can conclude that

$$\frac{\partial \theta_{it}}{\partial b_i} = 0. \tag{17}$$

Now differentiating (9) with respect to s_{it}, using the Inverse Function Theorem, and using the fact that these signals are independent of b_i (i.e., $\partial b_i / \partial s_{it} = 0$), gives

$$\frac{\partial q_{it}}{\partial s_{it}} = \frac{1}{h'(q_{it})} \frac{\partial (\theta_{it}/b_i)}{\partial s_{it}} = \frac{1}{b_i h'(q_{it})} \frac{\partial \theta_{it}}{\partial s_{it}},$$

Similarly, differentiate (9) with respect to b_i, use the Inverse Function Theorem, and use (17) to substitute for $\partial \theta_{it} / \partial b_i$ in this expression. This gives (11).

A.2 Proof of Proposition 3

$$\begin{aligned}
\frac{\partial}{\partial b_i}\left(\frac{\partial F_{it+1}}{\partial s_{it+1}}\right) &= \frac{\partial}{\partial b_i}\left(\frac{1}{q_{it+1}} \frac{1}{b_i h'(q_{it+1})}\right) \frac{\partial \theta_{it+1}}{\partial s_{it+1}} \\
&= -\frac{q_{it+1} h'(q_{it+1}) + \frac{\partial q_{it+1}}{\partial b_i}\left(b_i h'(q_{it+1}) + q_{it+1} b_i h''(q_{it+1})\right)}{q_{it+1}^2 \left(b_i h'(q_{it+1})\right)^2} \frac{\partial \theta_{it+1}}{\partial s_{it+1}} \\
&= -\frac{q_{it+1} h'(q_{it+1}) - h(q_{it+1})\left(1 + \frac{q_{it+1} h''(q_{it+1})}{h'(q_{it+1})}\right)}{q_{it+1}^2 \left(b_i h'(q_{it+1})\right)^2} \frac{\partial \theta_{it+1}}{\partial s_{it+1}},
\end{aligned} \tag{18}$$

where the first equality is implied by expression (12) and the fact that $\frac{\partial}{\partial b_i}\left(\frac{\partial \theta_{it+1}}{\partial s_{it+1}}\right) = 0$ (since

θ_{it+1} is solely a function of the history of realized signals and is not a function of b_i), and the last equality invokes expression (11). What (18) combined with Lemma 1 tells us is that steeper decreasing returns to scale must lead to a smaller flow of funds response to performance if and only if

$$q_{it+1}h'(q_{it+1}) - h(q_{it+1})\left(1 + \frac{q_{it+1}h''(q_{it+1})}{h'(q_{it+1})}\right) > 0. \tag{19}$$

Condition (19) is equivalent to

$$\frac{h'(q_{it+1})}{(h(q_{it+1}))^2} \times \left[q_{it+1}h'(q_{it+1}) - h(q_{it+1})\left(1 + \frac{q_{it+1}h''(q_{it+1})}{h'(q_{it+1})}\right)\right] > \frac{h'(q_{it+1})}{(h(q_{it+1}))^2} \times 0 \tag{20}$$

because $h(q)$ is a strictly increasing function of q, ensuring that $h'(q_{it+1}) > 0$. Notice that the left-hand side of (20) is equal to $-\frac{\partial}{\partial q_{it+1}}\left(\frac{\partial \log(h(q_{it+1}))}{\partial \log(q_{it+1})}\right)$, so (20) can be rewritten as

$$-\frac{\partial}{\partial q_{it+1}}\left(\frac{\partial \log(h(q_{it+1}))}{\partial \log(q_{it+1})}\right) > 0,$$

which is also equivalent to (14), completing the proof.

B Estimation Procedure for Fund-Specific DRS

This appendix describes the details of how we estimate fund-specific a_i and b_i parameters in the time-series regression $\alpha_{it} = a_i - b_i \log(q_{it-1}) + \epsilon_{it}$. It is well known that the OLS estimators of the coefficients b_i are subject to a small sample bias (Stambaugh (1999)). The small sample bias arises because of the flow-performance relation, which induces a positive correlation between the regression disturbance ϵ_{it} and the innovation in $\log(q_{it})$ (see below). Specifically, if $\log(q_{it})$ obeys a first-order autoregressive process,

$$\log(q_{it}) = \chi_i + \rho_i \log(q_{it-1}) + v_{it}, \tag{21}$$

Stambaugh (1999) shows that $\widehat{b}_i^{OLS}$ is upward biased, and proposes a first-order bias-corrected estimator of b_i. Amihud and Hurvich (2004, hereafter "AH") improve upon this estimator of Stambaugh by noting that adding a proxy v_{it}^c for the innovations in the autoregressive model can reduce the small-sample bias. The proxy v_{it}^c takes the form, $v_{it}^c = \log(q_{it}) - (\widehat{\chi}_i^c + \widehat{\rho}_i^c \log(q_{it-1}))$, where $\widehat{\chi}_i^c$ and $\widehat{\rho}_i^c$ are any estimators of χ_i and ρ_i constructed based on size data. Specifically, if we let $\widehat{b}_i^c$ be the coefficient of $\log(q_{it-1})$ in a time-series regression of α_{it} on $-\log(q_{it-1})$ and v_{it}^c, AH show that the bias of this estimator is given by:

$$E\left[\widehat{b}_i^c - b_i\right] = -\phi E\left[\widehat{\rho}_i^c - \rho_i\right], \tag{22}$$

where $\phi = \sigma_{\epsilon v}/\sigma_v^2$. We adopt this estimation procedure, except we use different estimators of ρ_i than AH, guided by a horse race among the various ρ_i estimators using simulations as further explained in the next subsection.

B.1 Reduced-Bias Estimators of the AR ρ_i

AH's bias expression (22) implies that the smaller the bias in the estimator $\widehat{\rho}_i^c$, the smaller the bias in the estimator $\widehat{b}_i^c$. AH suggest using a second-order bias-corrected estimator,

$$\widehat{\rho}_i^{c,AH} = \widehat{\rho}_i^{OLS} + \left(1 + 3\widehat{\rho}_i^{OLS}\right)/T_i + 3\left(1 + 3\widehat{\rho}_i^{OLS}\right)/T_i^2,$$

where $\widehat{\rho}_i^{OLS}$ is the coefficient of $\log(q_{it-1})$ in a time-series regression of $\log(q_{it})$ on $\log(q_{it-1})$ and T_i is the number of months. Barras, Gagliardini, and Scaillet (2022) improve upon this estimator by using $\widehat{\rho}_i^{c,BGS} = \min\left(\widehat{\rho}_i^{OLS} + \left(1 + 3\widehat{\rho}_i^{OLS}\right)/T_i + 3\left(1 + 3\widehat{\rho}_i^{OLS}\right)/T_i^2, 0.999\right)$. Andrews (1993) proposes a median-unbiased estimator (MUE) of the AR parameter ρ_i,

$$\widehat{\rho}_i^{c,MUE} = m_{TS}^{-1}\left(\widehat{\rho}_i^{OLS}\right),$$

where $m_{TS}(\rho_i)$ is the unique median of $\widehat{\rho}_i^{OLS}$ when ρ_i is the true parameter $\forall\rho_i \in (-1,1)$.[39] Phillips and Sul (2003) extend this estimator to the fixed-effects setting under homogeneous ρ across funds by using $\widehat{\rho}^{OLS\ FE}$ (the coefficient of $\log(q_{it-1})$ in a panel regression of $\log(q_{it})$ on $\log(q_{it-1})$ with fund fixed effects) to infer the true ρ. Similar to the MUE of Andrews, the panel median-unbiased estimator (PMUE) is $\widehat{\rho}^{c,PMUE} = m_{PNL}^{-1}\left(\widehat{\rho}^{OLS\ FE}\right)$, where $m_{PNL}(\rho)$ is the unique median of $\widehat{\rho}^{OLS\ FE}$ when ρ is the true parameter $\forall\rho \in (-1,1)$.[40]

Using our model, we first assess the performance of these various estimators of the AR parameter ρ_i for the sake of recovering the b_i coefficients, and we propose an improved estimator in the next subsection. As in Section 5.1.3, we simulate equilibrium alphas and sizes from the calibrated model to generate $10,000$ samples of panel data for 300 funds over 100 months. In each sample, we estimate $\widehat{b}_i^{c,AH}$, $\widehat{b}_i^{c,BGS}$, $\widehat{b}_i^{c,MUE}$, and $\widehat{b}_i^{c,PMUE}$, which denote the bias-corrected estimators of b_i corresponding to $\widehat{\rho}_i^{c,AH}$, $\widehat{\rho}_i^{c,BGS}$, $\widehat{\rho}_i^{c,MUE}$, and $\widehat{\rho}_i^{c,PMUE}$, respectively. In each sample, we then compute the positive biases in each of the $\widehat{b}_i^c$'s as the average difference between $\widehat{b}_i^c$ and b_i across funds. Another way to assess the performance of these bias-corrected estimators of b_i is to regress each of the $\widehat{b}_i^c$'s on b_i in each simulated sample: $\widehat{b}_i^c = \psi_0 + \psi_1 b_i + \varepsilon_i$. The results of these exercises are in columns 1–4 of Table 10. Panel A shows the mean biases of the b_i estimates across samples; Panel B shows the means of the ψ_0 and ψ_1 estimates across samples.

The first takeaway from this table is that the bias-corrected estimators of b_i corresponding to reduced-bias estimators of ρ_i from the existing literature *do* reduce the bias, compared to the OLS estimator whose bias is equal to 2.37. Further, the table shows that these estimators *do* approximate the b_i coefficients in the sense that they are positively related to the b_i coefficients. In particular, $\widehat{b}_i^{c,PMUE}$ performs the best. For this reason, as a first approach (and as a benchmark for the approach proposed in the next subsection), we use the

[39] Key to this estimator is the fact that the distribution of $\widehat{\rho}_i^{OLS}$ depends only on ρ_i (and the sample size, viz., $T_i + 1$) when (21) is correct. In particular, it does not depend on χ_i or σ_v^2 in (21).

[40] As with the MUE, key to this estimator is the fact that the distribution of $\widehat{\rho}^{OLS\ FE}$ depends only on ρ when the homogeneous dynamic panel model is correct.

PMUE to estimate fund-specific b_i parameters in our main analysis. Specifically, for each month t,

1. we use panel data on size $\{\log(q_{i\tau})\}$ across funds i and months τ from $t-61$ to $t-1$ to construct the PMUE $\widehat{\rho}_t^c$ of the size's persistence ρ_i and obtain the proxy for size innovations $\upsilon_{i\tau}^c = \log(q_{i\tau}) - (\widehat{\chi}_{it}^c + \widehat{\rho}_t^c \log(q_{i\tau-1}))$, where $\widehat{\chi}_{it}^c$ is chosen to ensure that $\{\upsilon_{i\tau}^c\}$ has zero mean for each fund i; and

2. we then obtain $\widehat{b}_{it}$ as the coefficient of $-\log(q_{i\tau-1})$ in the time-series regression of $\alpha_{i\tau}$ on $-\log(q_{i\tau-1})$ and $\upsilon_{i\tau}^c$, with intercept, using data from months $\tau = t-60, \ldots, t-1$ for each fund i.

In addition, we obtain $\widehat{a}_{it}$ as the average of fund i's alpha that is adjusted for fund-specific DRS, $\widehat{\alpha}_{i\tau} + \widehat{b}_{it}\log(q_{i\tau-1})$, over the 60 months prior to time t.

On the other hand, none of the bias-corrected estimators of b_i corresponding to reduced-bias estimators of ρ_i from the existing literature can completely eliminate the bias if data were generated in a Berk and Green equilibrium.

B.2 Novel Approach to Estimating the AR Parameter ρ_i

Given that none of the bias-corrected estimators of b_i corresponding to reduced-bias estimators of ρ_i from the existing literature can completely eliminate the bias if data were generated in a Berk and Green equilibrium, we propose a novel bias-corrected estimator of b_i motivated by AH's bias expression (22): set ρ_i to minimize the magnitude of the mean bias of the b_i estimates across simulated samples. This process results in the value of $\rho_i = 0.999863$, which we denote by $\widehat{\rho}_i^{c,*}$, with the corresponding bias-corrected b_i estimator, denoted by $\widehat{b}_i^{c,*}$. We consider the same simulation exercise approach as before: using the calibrated model, we generate $10,000$ samples of panel data for 300 funds over 100 months, in each of which we estimate $\widehat{b}_i^{c,*}$. In each sample, we then assess the performance of this bias-corrected estimator of b_i by (i) computing its bias (as the average difference between $\widehat{b}_i^{c,*}$ and b_i across funds) and (ii) regressing it on b_i ($\widehat{b}_i^{c,*} = \psi_0 + \psi_1 b_i + \varepsilon_i$). We report the results in the last column of Table 10. Again, Panel A reports the mean bias of $\widehat{b}_i^{c,*}$ across simulated samples; Panel B reports the means of the ψ_0 and ψ_1 estimates across simulated samples.

Fortunately, $\widehat{b}_i^{c,*}$ *does* eliminate the bias completely. Therefore, we also report our main results (Tables 2–5) based on the model-implied estimator of ρ_i (i.e., $\widehat{\rho}_i^{c,*} = 0.999863$) to estimate fund-specific b_i parameters in Tables 11–14. The latter set of tables lead to the same conclusions as the former tables.[41]

B.3 Further Discussion

We end this appendix by explaining how reduced-bias estimators of ρ_i from the existing literature cannot effectively eliminate the bias in the corresponding estimators of b_i if data were generated in a Berk and Green equilibrium, and yet our empirical results are robust to the choice of the bias-corrected estimator $\widehat{b}_i^c$. To this end, Panel C of Table 10 reports the means of the ρ_i estimates across simulated samples.

Consistent with AH's bias expression (22), the performance of a given estimator of b_i is monotonically worsening in the distance of the underlying estimator of ρ_i from $\widehat{\rho}_i^{c,*}$ (i.e., the "true" ρ_i in the simulated data).[42] But all of the reduced-bias estimators of ρ_i from the existing literature are found to be very close to $\widehat{\rho}_i^{c,*}$, which in turn implies that the poor performance of the b_i estimates corresponding to the ρ_i estimates from the existing literature must involve very high values of $\phi = \sigma_{\epsilon v}/\sigma_v^2$. Such high values of ϕ arise in our model for two reasons. The first is the fact that ϵ_{it} is perfectly correlated with v_{it}: a high fund return in period t is the sole reason for a higher fund size at the end of that period in a Berk and Green equilibrium. Therefore, $\sigma_{\epsilon v}$ is very high in our model. The second is the fact that $\sigma_v^2 = (1-\rho)^2 \frac{\sigma_\epsilon^2}{b_i^2}$ in our model:[43] a high persistence in fund size

[41] In fact, our results in Tables 8 and 9 are also very similar when we use the model-implied estimator of ρ_i (i.e., $\widehat{\rho}_i^{c,*} = 0.999863$) to estimate fund-specific b_i parameters. All of these results are available upon request.

[42] The PMUE of ρ_i in our actual data is 0.9988.

[43] Technically, ρ is a function of the fund's age in our model, but the variation in ρ that is due to differences

coincides with tighter beliefs about skill in a Berk and Green equilibrium and thus a weaker flow-performance relationship—smaller changes in fund size in response to unexpected fund returns. Therefore, σ_v^2 is very small in our model because fund size is highly persistent. Taken together, $\phi_i = \frac{b_i}{1-\rho}$ in our model, and is very high because the true ρ is close to one.

On the other hand, while a flow-performance relation exists, a substantial fraction of flows remain unexplained by performance in the data (Berk and van Binsbergen (2016)).[44] Consistent with this observation, the average of the ϕ estimates (based on the CAPM) in our actual data is 0.115. This value implies that the bias of 0.00496 in $\widehat{\rho}_i^{c,PMUE}$ would produce a bias of 0.000570 in $\widehat{b}_i^{c,PMUE}$, which is about 14% of 0.0041 (i.e., the mean of b_i). Therefore, our empirical results are robust to the choice of the bias-corrected estimator because small biases in $\widehat{\rho}_i^c$ *do* lead to small biases in the corresponding $\widehat{b}_i^c$ in the actual data. Importantly, this observation suggests that the estimation error in fund-specific DRS are mostly of the classical type (i.e., mean zero and independent to the actual DRS parameter) in the actual data, so correcting the estimated coefficients on $\widehat{b}_{it}$ in Table 2 under the classical measurement error assumption in Section 5.1.1 ought to do a good job of recovering the true DRS-FSP relation. Indeed, these estimates of the DRS-FSP relation are comparable to those implied by the characteristic-based approach, as well as the calibrated model.

in age is essentially negligible.

[44] Similarly, Barber, Huang, and Odean (2016) find that less than 18% of flows are explained by performance and a variety of controls.

References

[1] AMIHUD, Y., AND C. M. HURVICH (2004): "Predictive Regressions: A Reduced-Bias Estimation Method," *Journal of Financial and Quantitative Analysis*, 39(4), 1897–1923.

[2] ANDREWS, D. W. K. (1993): "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models," *Econometrica*, 61(1), 139–165.

[3] BARBER, B. M., X. HUANG, AND T. ODEAN (2016): "Which Factors Matter to Investors? Evidence from Mutual Fund Flows," *Review of Financial Studies*, 29(10), 2600–2642.

[4] BARRAS, L., P. GAGLIARDINI, AND O. SCAILLET (2021): "Skill, Scale, and Value Creation in the Mutual Fund Industry," *Journal of Finance*, Forthcoming.

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

[6] BERK, J. B., AND I. TONKS (2007): "Return Persistence and Fund Flows in the Worst Performing Mutual Funds," Working Paper, 13042, NBER.

[7] BERK, J. B., AND J. H. VAN BINSBERGEN (2015): "Measuring Skill in the Mutual Fund Industry," *Journal of Financial Economics*, 118(1), 1–20.

[8] BERK, J. B., AND J. H. VAN BINSBERGEN (2016): "Assessing Asset Pricing Models using Revealed Preference," *Journal of Financial Economics*, 119(1), 1–23.

[9] BUSSE, J. A., T. CHORDIA, L. JIANG, AND Y. TANG (2021): "Transaction Costs, Portfolio Characteristics, and Mutual Fund Performance," *Management Science*, 67(2), 1227–1248.

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

[11] CHEN, J., H. HONG, M. HONG, AND J. D. KUBIK (2004): "Does Fund Size Erode Mutual Fund Performance? The Role of Liquidity and Organization," *American Economic Review*, 94(5), 1276–1302.

[12] CHEN, Q., I. GOLDSTEIN, AND W. JIANG (2010): "Payoff Complementarities and Financial Fragility: Evidence from Mutual Fund Outflows," *Journal of Financial Economics*, 97(2), 239–262.

[13] CHEVALIER, J., AND G. ELLISON (1997): "Risk Taking by Mutual Funds as a Response to Incentives," *Journal of Political Economy*, 105(6), 1167–1200.

[14] DEL GUERCIO, D., AND J. REUTER (2013): "Mutual Fund Performance and the Incentive to Generate Alpha," *Journal of Finance*, 69(4), 1673–1704.

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

[16] FERREIRA, M. A., A. KESWANI, A. F. MIGUEL, AND S. B. RAMOS (2012): "The Flow-Performance Relationship Around the World," *Journal of Banking & Finance*, 36(6), 1759–1780.

[17] FRANZONI, F., AND M. C. SCHMALZ (2017): "Fund Flows and Market States," *Review of Financial Studies*, 30(8), 2621–2673.

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

[19] HUANG, J., K. D. WEI, AND H. YAN (2007): "Participation Costs and the Sensitivity of Fund Flows to Past Performance," *Journal of Finance*, 62(3), 1273–1311.

[20] HUANG, J., K. D. WEI, AND H. YAN (2012): "Investor Learning and Mutual Fund Flows," Working Paper, University of Texas at Austin.

[21] KIM, J. H. (2022): "Investor Learning and the Aggregate Allocation of Capital to Active Management," Working Paper, Emory University.

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

[23] MILGROM, P. R. (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12(2), 380–391.

[24] PÁSTOR, Ľ., AND R. F. STAMBAUGH (2012): "On the Size of the Active Management Industry," *Journal of Political Economy*, 120(4), 740–781.

[25] PÁSTOR, Ľ., R. F. STAMBAUGH, AND L. A. TAYLOR (2015): "Scale and Skill in Active Management," *Journal of Financial Economics*, 116(1), 23–45.

[26] PÁSTOR, Ľ., R. F. STAMBAUGH, AND L. A. TAYLOR (2020): "Fund Tradeoffs," *Journal of Financial Economics*, 138(3), 614–634.

[27] PHILLIPS, P. C. B., AND D. SUL (2003): "Dynamic panel estimation and homogeneity testing under cross section dependence," *The Econometrics Journal*, 6(1), 217–259.

[28] POLLET, J. M., AND M. WILSON (2008): "How Does Size Affect Mutual Fund Behavior?" *Journal of Finance*, 63(6), 2941–2969.

[29] SIRRI, E. R., AND P. TUFANO (1998): "Costly Search and Mutual Fund Flows," *Journal of Finance*, 53(5), 1589–1622.

[30] STAMBAUGH, R. F. (1999): "Predictive Regressions," *Journal of Financial Economics*, 54(3), 375–421.

[31] STAMBAUGH, R. F. (2020): "Skill and Profit in Active Management," Working Paper, 26027, NBER.

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

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

Table 1: Summary Statistics

This table shows summary statistics for our sample of active equity mutual funds from 1991–2014. The unit of observation is the fund/month. All returns are in units of fraction per month. Net return is the return received by investors. Net alpha equals net return minus the return on benchmark portfolio, calculated using the CAPM or using a set of Vanguard index funds. Fund size is the fund's total AUM aggregated across share classes, adjusted by inflation. The numbers are reported in Y2000 $ millions per month. The first version of Flows is the logarithmic change in real AUM (i.e., the percentage change in fund size); the second version of Flows is the monthly change in the fund's net assets not attributable to its return gains or losses. $\widehat{\beta}_{it}^{\cdot}$ are fund i's estimated risk exposures from the regression of the fund's return on the four FFC factors over the sixty months prior to month t. # of managers is the number of managers managing the fund in a given month. Volatility is the standard deviation of a fund's alphas, calculated over the prior 1 year. All expenses are in units of fraction per year. Marketing expenses is a fund's total fee ratio, defined as the annual expense ratio plus one-seventh of the up-front load fees. Fund age is the number of years since the fund's first offer date (from CRSP or, if missing, from Morningstar). Turnover is in units of fraction per year. $\widehat{b}_{it}$ is the fund's decreasing returns to scale estimated as of the previous month-end; $\widehat{b}_{it}^{Char}$ is the characteristic component of $\widehat{b}_{it}$ estimated from the specification in column 8 of Table 4. $\widehat{FSP}_{it}$ is the fund's flow sensitivity to performance going forward.

Panel A: Fund-Level Variables

	# of obs.	Mean	Stdev.	Percentiles 25%	50%	75%
Net return	394, 351	0.0075	0.0498	−0.0196	0.0122	0.0384
Net alpha (CAPM Risk Adj.)	331, 450	0.0001	0.0225	−0.0105	−0.0002	0.0103
Net alpha (Vanguard BM)	389, 721	−0.0002	0.0162	−0.0082	−0.0002	0.0078
Fund size (in 2000 $millions)	391, 615	1031	4147	44	164	621
Flows (v.1)	391, 611	0.0078	0.0735	−0.0288	0.0096	0.0447
Flows (v.2)	391, 611	0.0049	0.0504	−0.0134	−0.0020	0.0136
$\widehat{\beta}_{it}^{mkt}$	331, 450	0.9670	0.1429	0.8910	0.9707	1.0480
$\widehat{\beta}_{it}^{smb}$	331, 450	0.2157	0.3376	−0.0613	0.1314	0.4764
$\widehat{\beta}_{it}^{hml}$	331, 450	−0.0046	0.3100	−0.2214	−0.0020	0.2020
$\widehat{\beta}_{it}^{mom}$	331, 450	0.0146	0.1374	−0.0598	0.0062	0.0774
# of managers	382, 531	2.41	2.16	1	2	3
Volatility (CAPM Risk Adj.)	303, 147	0.0188	0.0116	0.0105	0.0158	0.0239
Volatility (Vanguard BM)	360, 609	0.0138	0.0080	0.0084	0.0119	0.0172
Expense ratio	391, 631	0.0125	0.0043	0.0097	0.0120	0.0148
Marketing expenses	258, 905	0.0182	0.0115	0.0143	0.0195	0.0218
Fund age	393, 909	12.92	13.04	4.48	8.95	15.97
Turnover	375, 293	0.837	0.705	0.350	0.642	1.100

Panel B: Estimated DRS and FSP

	# of obs.	Mean	Stdev.	Percentiles 25%	50%	75%
$\widehat{b}$ (CAPM Risk Adj.)	252, 433	0.0038	0.0153	−0.0027	0.0029	0.0100
$\widehat{b}$ (Vanguard BM)	300, 963	0.0023	0.0105	−0.0021	0.0017	0.0064
$\widehat{b}^{Char}$ (CAPM Risk Adj.)	152, 007	0.0041	0.0030	0.0020	0.0037	0.0056
$\widehat{b}^{Char}$ (Vanguard BM)	176, 756	0.0025	0.0016	0.0014	0.0023	0.0033
$\widehat{FSP}$ (CAPM Risk Adj., v.1)	266, 376	0.0560	0.2758	−0.0846	0.0432	0.1907
$\widehat{FSP}$ (Vanguard BM, v.1)	293, 895	0.0993	0.3938	−0.0988	0.0940	0.2973
$\widehat{FSP}$ (CAPM Risk Adj., v.2)	266, 376	0.1045	0.1910	0.0140	0.0756	0.1694
$\widehat{FSP}$ (Vanguard BM, v.2)	293, 895	0.1487	0.2898	0.0171	0.1094	0.2499

Table 2: Relation Between DRS and Flow Sensitivity to Performance

The dependent variable in each regression model is $\widehat{FSP}_{it}$, the fund's flow sensitivity to performance going forward, where flow is defined as the % change in fund size in Panel A and as the % change in new assets in Panel B. $\widehat{b}_{it}$ is the fund's decreasing returns to scale estimated as of the previous month-end. In the odd columns, we only include month and fund fixed effects; in the even columns, we add proxies for participation costs, as well as performance volatility and fund age. Standard errors, two-way clustered by fund and by month, are in parentheses. We report the EIV-adjusted coefficients in the last row of each panel.

	Panel A:	**Flow as % Change in Fund Size**		
		$\widehat{FSP}_{it}$		
$\widehat{b}_{it}$	-0.937	-1.299	-1.551	-1.714
	(0.148)	(0.182)	(0.288)	(0.328)
Fund FE & Month FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	$182,675$	$114,623$	$221,743$	$136,365$
Performance relative to	CAPM	CAPM	Vanguard BM	Vanguard BM
EIV Adj. Coefficient	-9.143	-16.112	-40.984	-77.384

	Panel B:	**Flow as % Change in New Assets**		
		$\widehat{FSP}_{it}$		
$\widehat{b}_{it}$	-0.262	-0.386	-0.370	-0.549
	(0.087)	(0.117)	(0.199)	(0.182)
Fund FE & Month FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	$182,675$	$114,623$	$221,743$	$136,365$
Performance relative to	CAPM	CAPM	Vanguard BM	Vanguard BM
EIV Adj. Coefficient	-2.560	-4.788	-9.772	-24.781

Table 3: Relation Between DRS and FSP Based on Their Percentile Ranks

This table is the same as Table 2 but, instead of using the coefficient estimates $\widehat{b}_{it}$ and $\widehat{FSP}_{it}$, uses their percentile ranks in each month.

	Panel A:	**Flow as % Change in Fund Size**		
	Pctl. rank based on $\widehat{FSP}_{it}$			
Pctl. rank based on $\widehat{b}_{it}$	-0.1083	-0.1251	-0.0918	-0.0951
	(0.0109)	(0.0129)	(0.0095)	(0.0117)
Fund FE	Yes	Yes	Yes	Yes
Month FE	No	No	No	No
Controls	No	Yes	No	Yes
Observations	182,675	114,623	221,743	136,365
Performance relative to	CAPM	CAPM	Vanguard BM	Vanguard BM

	Panel B:	**Flow as % Change in New Assets**		
	Pctl. rank based on $\widehat{FSP}_{it}$			
Pctl. rank based on $\widehat{b}_{it}$	-0.0659	-0.0753	-0.0560	-0.0593
	(0.0095)	(0.0117)	(0.0086)	(0.0108)
Fund FE	Yes	Yes	Yes	Yes
Month FE	No	No	No	No
Controls	No	Yes	No	Yes
Observations	182,675	114,623	221,743	136,365
Performance relative to	CAPM	CAPM	Vanguard BM	Vanguard BM

Table 4: Determinants of Decreasing Returns to Scale

The dependent variable in all regressions is $\widehat{b}_{it}$, the fund's adverse scale effect computed using only its data prior to the month. $NMgr$ is the number of managers managing the fund. $Std\,(Alpha)$ is the standard deviation of monthly outperformance for each fund over the prior 1 year. $ExpRatio$ and $MktgExp$ are the fund's expense ratios and marketing expenses, respectively. $1\,(IntExp)$ is an indicator for funds with a high degree of international exposure, defined as follows. We test the null hypothesis that the coefficients on three Vanguard international index funds are 0. For any given fund, the international exposure dummy is equal to one if we reject the null hypothesis at the 5% confidence level. $Turnover$ is the fund's average annual turnover (from CRSP). Finally, $\log(q_{it-1})$ is the log of the fund's lagged real AUM. All regressions control for fund style and risk, as proxied by the loadings on the four Fama-French-Carhart (FFC) factors ($\widehat{\beta}_{:it}$). Panel A reports the results using the CAPM as the benchmark; Panel B uses Vanguard index funds as the benchmark. Standard errors, two-way clustered by fund and time, are in parentheses.

	Panel A: Performance Relative to the CAPM							
	Dependent Variable: $\widehat{b}_{it}$							
$NMgr$	−0.000254							−0.000063
	(0.000070)							(0.000088)
$Std\,(Alpha)$		0.1906						0.1828
		(0.0226)						(0.0262)
$ExpRatio$			0.1622					0.2128
			(0.0457)					(0.0689)
$MktgExp$				−0.0338				−0.1323
				(0.0358)				(0.0440)
$1\,(IntExp)$					0.000057			
					(0.000351)			
$Turnover$						0.000458		
						(0.000287)		
$\log(q_{it-1})$							0.000041	0.000160
							(0.000109)	(0.000135)
Style & Risk Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	248,076	252,361	252,427	154,998	252,433	247,256	252,413	152,007

Panel B: Performance Relative to the Vanguard BM

	Dependent Variable: $\widehat{b}_{it}$							
$NMgr$	-0.000078							-0.000047
	(0.000037)							(0.000059)
$Std\,(Alpha)$		0.0512						0.0703
		(0.0185)						(0.0217)
$ExpRatio$			0.0528					0.0780
			(0.0308)					(0.0461)
$MktgExp$				-0.0344				-0.0818
				(0.0278)				(0.0340)
$1\,(IntExp)$					0.000006			
					(0.000236)			
$Turnover$						-0.000037		
						(0.000186)		
$\log\left(q_{it-1}\right)$							0.000045	0.000042
							(0.000069)	(0.000089)
Style & Risk Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	294,922	300,883	300,957	180,913	300,963	294,666	300,941	176,756

Table 5: DRS-FSP Relation Using the Characteristic Component of DRS

This table is the same as Table 2 but, instead of using the coefficient estimate $\widehat{b}_{it}$ as before, uses its characteristic component estimated from the specification in column 8 of Table 4, denoted by $\widehat{b}_{it}^{Char}$.

	Panel A:	**Flow as % Change in Fund Size**		
		$\widehat{FSP}_{it}$		
$\widehat{b}_{it}^{Char}$	-6.472	-11.751	-5.679	-8.497
	(1.453)	(2.015)	(3.433)	(3.912)
Fund FE & Month FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	112,975	112,071	134,117	132,802
Performance relative to	CAPM	CAPM	Vanguard BM	Vanguard BM

	Panel B:	**Flow as % Change in New Assets**		
		$\widehat{FSP}_{it}$		
$\widehat{b}_{it}^{Char}$	-3.828	-6.482	-3.612	-7.568
	(0.948)	(1.275)	(2.047)	(2.266)
Fund FE & Month FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	112,975	112,071	134,117	132,802
Performance relative to	CAPM	CAPM	Vanguard BM	Vanguard BM

Table 6: Parameter Values Used for Simulations

Panel A summarizes the model parameters that we set directly and their parameter values. The empirical distribution of $\widehat{b}_{it}^{Char}$ is approximated by a scaled Beta distribution, from which we draw b_i randomly. In that case, assuming that θ_{i0} is independent of b_i gives rise to distributions of fund size considerably more disperse than in our actual sample. Therefore, we model the prior mean as a linear function of b_i, $\theta_{i0} = \theta_0(b_i) = \theta_{0a} + \theta_{0b}b_i$. Our approach is to fit the parameters θ_{0a} and θ_{0b} by essentially matching the simulated mean and standard deviation of log fund size to their empirical counterparts. We set the prior uncertainty (σ_0) to match the average flow-performance relation in the data. Panel B shows the value of σ_0 that resulted from this process. It also contains the values of the parameters governing the scaled Beta distribution of DRS and those of the parameters governing the prior mean that we use in our simulation analysis. The last three columns of Panel B report all the moments that we target in our calibration, as well as their values in both the actual and simulated data.

	Panel A:		**Parameters Set Directly**
Variable	Symbol	Value	
Return standard deviation	σ	5.77%	
Exit mean	$\overline{\theta}$	0%	

	Panel B:		**Calibrated Parameters**		
Variable	Symbol	Value	Target	Emp. Value	Sim. Value
$x \sim Beta(A, B), b_i = Cx$					
First shape parameter	A	1.5	Mean of b_i	0.0041	0.0041
Second shape parameter	B	11.9	Std dev of b_i	0.0030	0.0030
Scale parameter	C	0.036	Skewness of b_i	1.04	1.20
			Kurtosis of b_i	5.56	4.68
$\theta_0(b_i) = \theta_{0a} + \theta_{0b}b_i$					
Prior mean for CRS funds	θ_{0a}	0.15%	Mean of $\log(q_{it})$	5.1	5.1
Prior mean slope on DRS	θ_{0b}	4.2	Std dev of $\log(q_{it})$	1.9	1.9
Prior uncertainty	σ_0	0.06%	Mean of $\widehat{\gamma}_i$	0.56	0.55

Table 7: Simulated DRS-FSP Relation

We construct $10,000$ samples of simulated panel data for 300 funds over 100 months. We simulate a given sample by first drawing each fund's DRS b_i randomly from the scaled Beta distribution consistent with the distribution of fund-specific b estimates $(\widehat{b}_{it}^{Char})$, while drawing the fund's skill a_i from a normal distribution with mean $\theta_0(b_i)$ and standard deviation σ_0. Next, we simulate ϵ_{it} as independent draws across funds and periods, building up the panel data of α_{it} and q_{it}. For each fund in the sample, we run the following regression using data for just that fund to estimate its FSP: $\log(q_{it}/q_{it-1}) = c_i + \gamma_i \alpha_{it} + \upsilon_{it}$. We then estimate the DRS-FSP relation in each of the simulated samples. Panel A shows summary statistics of these estimates across simulated samples from the calibrated model. To quantitatively assess the role of heterogeneity in returns to scale in capital allocation, we construct a counterfactual by assuming investors learn about skill based on distorted beliefs that $b_i = 0.0041$ for all funds. Then, updating investors' beliefs with the history of its returns under the counterfactual assumption, we compute the fund's size under the counterfactual, q_{it}^C, for every i and t. For each sample, we calculate the R^2 from a regression of $\log(q_{it})$ on $\log(q_{it}^C)$ to check the goodness of fit by the counterfactual. We report the results from counterfactual simulations in Panel B. The first two rows show summary statistics of the coefficient estimates from the regression of $\log(q_{it})$ on $\log\left(q_{it}^C\right)$ across simulated samples; the last row shows summary statistics of the R^2 from this regression across simulated samples.

Panel A: Simulated DRS-FSP Relation

$\widehat{\gamma}_i = k + \lambda b_i + u_i$

	Mean	Percentiles 1%	10%	25%	50%	75%	90%	99%
$\widehat{\lambda}$	-16.9	-49.0	-22.6	-18.0	-14.8	-12.7	-11.2	-9.33
Data	-11.8							

Panel B: Capital Allocation Explained by Counterfactual

$\log(q_{it}) = \pi_0 + \pi_1 \log\left(q_{it}^C\right) + e_{it}$

	Mean	Percentiles 1%	10%	25%	50%	75%	90%	99%
$\widehat{\pi}_0$	-421	-743	-664	-548	-405	-288	-217	-136
$\widehat{\pi}_1$	93.5	30.9	48.8	64.1	90.0	121	147	164
R^2	0.430	0.138	0.227	0.301	0.413	0.556	0.668	0.749

Table 8: Relation Between DRS and Fund Size

We first form quintile groups sorted on fund fixed effects estimated as of month $t-1$, $\widehat{a}_{it}^{Char}$. Within each $\widehat{a}_{it}^{Char}$ quintile, we sort funds into five groups based on fund-specific DRS estimated as of month $t-1$, $\widehat{b}_{it}^{Char}$. We conduct double sorts of funds belonging to the same Morningstar category $\times$ month $\times$ age category. After forming the 5×5 $\widehat{a}_{it}^{Char}$ and $\widehat{b}_{it}^{Char}$ groups, we average fund sizes, as measured by log real AUM, in month t, of each $\widehat{b}_{it}^{Char}$ quintile over the five $\widehat{a}_{it}^{Char}$ groups. Panel A reports average fund sizes of the 25 $\widehat{a}_{it}^{Char} \times \widehat{b}_{it}^{Char}$ groups using the CAPM as the benchmark; Panel B repeats the same exercise, except we use Vanguard index funds as the benchmark. The column labeled "Average" reports the average month-end fund sizes of the $\widehat{b}_{it}^{Char}$ quintiles, controlling for $\widehat{a}_{it}^{Char}$, fund style and fund age. The row labeled "High-low" reports the differences in average sizes between the first and fifth quintile $\widehat{b}_{it}^{Char}$ groups in each column. We report standard errors of these differences using 60 Newey-West lags.

Panel A: Performance Relative to the CAPM

	$\widehat{a}_{it}^{Char}$ Quintiles					
Group	1 Low	2	3	4	5	Average
1 Low $\widehat{b}_{it}^{Char}$	5.452	6.463	6.915	7.280	7.484	6.689
2	5.667	6.158	6.399	6.763	7.256	6.419
3	5.318	5.349	5.682	6.108	6.768	5.808
4	4.802	4.877	5.137	5.678	6.615	5.390
5 High $\widehat{b}_{it}^{Char}$	3.870	3.982	4.403	4.965	6.559	4.705
High-low	−1.582	−2.480	−2.513	−2.316	−0.925	−1.983
	(0.182)	(0.095)	(0.191)	(0.165)	(0.156)	(0.132)

Panel B: Performance Relative to the Vanguard BM

	$\widehat{a}_{it}^{Char}$ Quintiles					
Group	1 Low	2	3	4	5	Average
1 Low $\widehat{b}_{it}^{Char}$	5.203	6.075	6.531	6.946	7.279	6.361
2	5.490	5.892	6.266	6.655	7.064	6.237
3	5.147	5.189	5.570	5.950	6.575	5.656
4	4.647	4.691	5.033	5.577	6.502	5.247
5 High $\widehat{b}_{it}^{Char}$	3.818	3.893	4.338	4.955	6.392	4.628
High-low	−1.385	−2.182	−2.193	−1.991	−0.887	−1.733
	(0.085)	(0.091)	(0.164)	(0.124)	(0.132)	(0.092)

Table 9: Relation Between Optimal Size and Fund Size

The dependent variable in each regression model is the fund's log real AUM in \$ millions (base year is 2000). A fund's optimal size, $\log(\widehat{q}_i^*)$, is the average ratio of its net alpha (adjusted for decreasing returns to scale) to its individual DRS parameter; the alternative measure of a fund's optimal size, $\log\left(\widehat{q}_i^{*RD2}\right)$, is calculated assuming that the effect of scale on performance is the same for all funds. Details of these optimal fund size measures are in Section 5.2.1. In Panel A, we report the results using the CAPM as the benchmark; in Panel B, we use Vanguard index funds as the benchmark. Standard errors are double clustered (by fund and time) and are reported in parentheses.

Panel A: Performance Relative to the CAPM

	Dependent Variable: Log Real AUM					
$\log(\widehat{q}_i^*)$	0.367	0.674	0.905	−0.052	0.271	0.597
	(0.016)	(0.015)	(0.012)	(0.021)	(0.023)	(0.037)
$\log\left(\widehat{q}_i^{*RD2}\right)$				0.585	0.517	0.332
				(0.025)	(0.023)	(0.038)
R^2	0.24	0.63	0.78	0.39	0.71	0.79
Observations	84,310	100,048	176,375	84,310	100,048	176,375
Fund ages	$[0,5]$ yr.	$(5,10]$ yr.	>10 yr.	$[0,5]$ yr.	$(5,10]$ yr.	>10 yr.

Panel B: Performance Relative to the Vanguard BM

	Dependent Variable: Log Real AUM					
$\log(\widehat{q}_i^*)$	0.203	0.416	0.659	0.031	0.181	0.361
	(0.011)	(0.028)	(0.024)	(0.011)	(0.025)	(0.038)
$\log\left(\widehat{q}_i^{*RD2}\right)$				0.300	0.354	0.340
				(0.014)	(0.024)	(0.038)
R^2	0.17	0.44	0.65	0.31	0.54	0.68
Observations	98,088	102,325	176,505	98,088	102,325	176,505
Fund ages	$[0,5]$ yr.	$(5,10]$ yr.	>10 yr.	$[0,5]$ yr.	$(5,10]$ yr.	>10 yr.

Table 10: Horse Race Among the Various Estimators of DRS Using Simulations

This table reports the results of a horse race among the various estimators of the AR parameter ρ_i governing the size process $\log(q_{it}) = \chi_i + \rho_i \log(q_{it-1}) + v_{it}$ for the sake of recovering the b_i coefficients using our model. We simulate equilibrium alphas and sizes from the calibrated model to generate 10,000 samples of panel data for 300 funds over 100 months. In each sample, we estimate $\widehat{b}_i^c$ as the coefficient of $\log(q_{it-1})$ in a time-series regression of α_{it} on $-\log(q_{it-1})$ and v_{it}^c, where v_{it}^c is a proxy for the size innovation. The proxy v_{it}^c takes the form, $v_{it}^c = \log(q_{it}) - (\widehat{\chi}_i^c + \widehat{\rho}_i^c \log(q_{it-1}))$, where $\widehat{\chi}_i^c$ and $\widehat{\rho}_i^c$ are any estimators of χ_i and ρ_i constructed based on size data. To obtain the proxy v_{it}^c, we consider four popular reduced-bias estimators of ρ_i from the existing literature, details of which are in Appendix B.1. Importantly, we propose yet another bias-corrected estimator of b_i motivated by AH's bias expression (22): set ρ_i to minimize the magnitude of the mean bias of the b_i estimates across simulated samples. This process results in the value of $\rho_i = 0.999863$, which we denote by $\widehat{\rho}_i^{c,*}$, with the corresponding bias-corrected b_i estimator, denoted by $\widehat{b}_i^{c,*}$. In each sample, we then compute the non-negative biases in each of the $\widehat{b}_i^c$'s as the average difference between $\widehat{b}_i^c$ and b_i across funds. Panel A shows the mean biases of the b_i estimates across simulated samples. Another way to assess the performance of these bias-corrected estimators of b_i is to regress each of the $\widehat{b}_i^c$'s on b_i in each sample: $\widehat{b}_i^c = \psi_0 + \psi_1 b_i + \varepsilon_i$. Panel B shows the means of the ψ_0 and ψ_1 estimates across simulated samples. Panel C reports the means of the ρ_i estimates (averaged across funds in each sample) across simulated samples.

	Panel A: Bias in estimators of b_i corresponding to				
	reduced-bias estimators of ρ_i from the existing literature				our approach to estimating ρ_i
Bias in	$\widehat{b}_i^{c,AH}$	$\widehat{b}_i^{c,BGS}$	$\widehat{b}_i^{c,MUE}$	$\widehat{b}_i^{c,PMUE}$	$\widehat{b}_i^{c,*}$
Mean	0.570	1.043	1.095	0.520	0.000

	Panel B: Relation between b_i and its estimators corresponding to				
	reduced-bias estimators of ρ_i from the existing literature				our approach to estimating ρ_i
	$\widehat{b}_i^{c,AH}$	$\widehat{b}_i^{c,BGS}$	$\widehat{b}_i^{c,MUE}$	$\widehat{b}_i^{c,PMUE}$	$\widehat{b}_i^{c,*}$
Mean $\widehat{\psi}_0$	0.003	0.003	0.003	0.000	0.000
Mean $\widehat{\psi}_1$	139.2	254.7	267.2	127.8	1.000

	Panel C: Mean estimated ρ_i across funds corresponding to				
	reduced-bias estimators of ρ_i from the existing literature				our approach to estimating ρ_i
	$\widehat{\rho}_i^{c,AH}$	$\widehat{\rho}_i^{c,BGS}$	$\widehat{\rho}_i^{c,MUE}$	$\widehat{\rho}_i^{c,PMUE}$	$\widehat{\rho}_i^{c,*}$
Mean	0.9873	0.9769	0.9758	0.9949	0.999863

Table 11: Relation Between DRS and Flow Sensitivity to Performance

This table is the same as Table 2 but, in lieu of $\widehat{b}_{it}$ (the bias-corrected b_i estimator based on the PMUE of ρ_i), uses $\widehat{b}_{it}^{c,*}$ (the bias-corrected b_i estimator based on the model-implied estimator of ρ_i (i.e., $\widehat{\rho}_i^{c,*} = 0.999863$). Details of these b_i estimators (as well as other b_i estimators from the existing literature) are in Appendix B.

	Panel A:	**Flow as % Change in Fund Size**		
		$\widehat{FSP}_{it}$		
$\widehat{b}_{it}$	-0.943 (0.148)	-1.308 (0.182)	-1.551 (0.288)	-1.709 (0.327)
Fund FE & Month FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	182,675	114,623	221,743	136,365
Performance relative to	CAPM	CAPM	Vanguard BM	Vanguard BM
EIV Adj. Coefficient	-9.812	-17.989	-50.572	-134.929

	Panel B:	**Flow as % Change in New Assets**		
		$\widehat{FSP}_{it}$		
$\widehat{b}_{it}$	-0.263 (0.087)	-0.386 (0.117)	-0.384 (0.197)	-0.559 (0.181)
Fund FE & Month FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	182,675	114,623	221,743	136,365
Performance relative to	CAPM	CAPM	Vanguard BM	Vanguard BM
EIV Adj. Coefficient	-2.738	-5.314	-12.512	-44.160

Table 12: Relation Between DRS and FSP Based on Their Percentile Ranks

This table is the same as Table 3 but, in lieu of $\widehat{b}_{it}$ (the bias-corrected b_i estimator based on the PMUE of ρ_i), uses $\widehat{b}_{it}^{c,*}$ (the bias-corrected b_i estimator based on the model-implied estimator of ρ_i (i.e., $\widehat{\rho}_i^{c,*} = 0.999863$). Details of these b_i estimators (as well as other b_i estimators from the existing literature) are in Appendix B.

	Panel A:	**Flow as % Change in Fund Size**		
	Pctl. rank based on $\widehat{FSP}_{it}$			
Pctl. rank based on $\widehat{b}_{it}$	-0.1084 (0.0109)	-0.1250 (0.0128)	-0.0914 (0.0095)	-0.0943 (0.0117)
Fund FE	Yes	Yes	Yes	Yes
Month FE	No	No	No	No
Controls	No	Yes	No	Yes
Observations	182,675	114,623	221,743	136,365
Performance relative to	CAPM	CAPM	Vanguard BM	Vanguard BM

	Panel B:	**Flow as % Change in New Assets**		
	Pctl. rank based on $\widehat{FSP}_{it}$			
Pctl. rank based on $\widehat{b}_{it}$	-0.0657 (0.0094)	-0.0750 (0.0117)	-0.0563 (0.0085)	-0.0597 (0.0108)
Fund FE	Yes	Yes	Yes	Yes
Month FE	No	No	No	No
Controls	No	Yes	No	Yes
Observations	182,675	114,623	221,743	136,365
Performance relative to	CAPM	CAPM	Vanguard BM	Vanguard BM

Table 13: Determinants of Decreasing Returns to Scale

This table is the same as Table 4 but, in lieu of $\widehat{b}_{it}$ (the bias-corrected b_i estimator based on the PMUE of ρ_i), uses $\widehat{b}_{it}^{c,*}$ (the bias-corrected b_i estimator based on the model-implied estimator of ρ_i (i.e., $\widehat{\rho}_i^{c,*} = 0.999863$). Details of these b_i estimators (as well as other b_i estimators from the existing literature) are in Appendix B.

	Panel A: Performance Relative to the CAPM							
	Dependent Variable: $\widehat{b}_{it}$							
$NMgr$	-0.000251							-0.000059
	(0.000070)							(0.000088)
$Std\,(Alpha)$		0.1882						0.1782
		(0.0224)						(0.0258)
$ExpRatio$			0.1661					0.2167
			(0.0458)					(0.0691)
$MktgExp$				-0.0250				-0.1230
				(0.0359)				(0.0437)
$1\,(IntExp)$					0.000055			
					(0.000350)			
$Turnover$						0.000452		
						(0.000288)		
$\log\,(q_{it-1})$							0.000040	0.000174
							(0.000109)	(0.000135)
Style & Risk Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	248,076	252,361	252,427	154,998	252,433	247,256	252,413	152,007

Panel B: Performance Relative to the Vanguard BM

	Dependent Variable: $\widehat{b}_{it}$							
$NMgr$	-0.000077							-0.000048
	(0.000036)							(0.000059)
$Std\,(Alpha)$		0.0474						0.0623
		(0.0182)						(0.0210)
$ExpRatio$			0.0539					0.0797
			(0.0307)					(0.0461)
$MktgExp$				-0.0301				-0.0766
				(0.0278)				(0.0338)
$1\,(IntExp)$					0.000001			
					(0.000235)			
$Turnover$						-0.000040		
						(0.000185)		
$\log\,(q_{it-1})$							0.000043	0.000046
							(0.000069)	(0.000089)
Style & Risk Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	294,922	300,883	300,957	180,913	300,963	294,666	300,941	176,756

Table 14: DRS-FSP Relation Using the Characteristic Component of DRS

This table is the same as Table 5 but, in lieu of $\widehat{b}_{it}$ (the bias-corrected b_i estimator based on the PMUE of ρ_i), uses $\widehat{b}_{it}^{c,*}$ (the bias-corrected b_i estimator based on the model-implied estimator of ρ_i (i.e., $\widehat{\rho}_i^{c,*} = 0.999863$). Details of these b_i estimators (as well as other b_i estimators from the existing literature) are in Appendix B.

	Panel A:	**Flow as % Change in Fund Size**		
		$\widehat{FSP}_{it}$		
$\widehat{b}_{it}^{Char}$	-6.908	-12.470	-5.829	-8.586
	(1.494)	(2.076)	(3.550)	(3.974)
Fund FE & Month FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	112,975	112,071	134,117	132,802
Performance relative to	CAPM	CAPM	Vanguard BM	Vanguard BM

	Panel B:	**Flow as % Change in New Assets**		
		$\widehat{FSP}_{it}$		
$\widehat{b}_{it}^{Char}$	-3.964	-6.669	-3.714	-7.596
	(0.975)	(1.303)	(2.108)	(2.297)
Fund FE & Month FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	112,975	112,071	134,117	132,802
Performance relative to	CAPM	CAPM	Vanguard BM	Vanguard BM

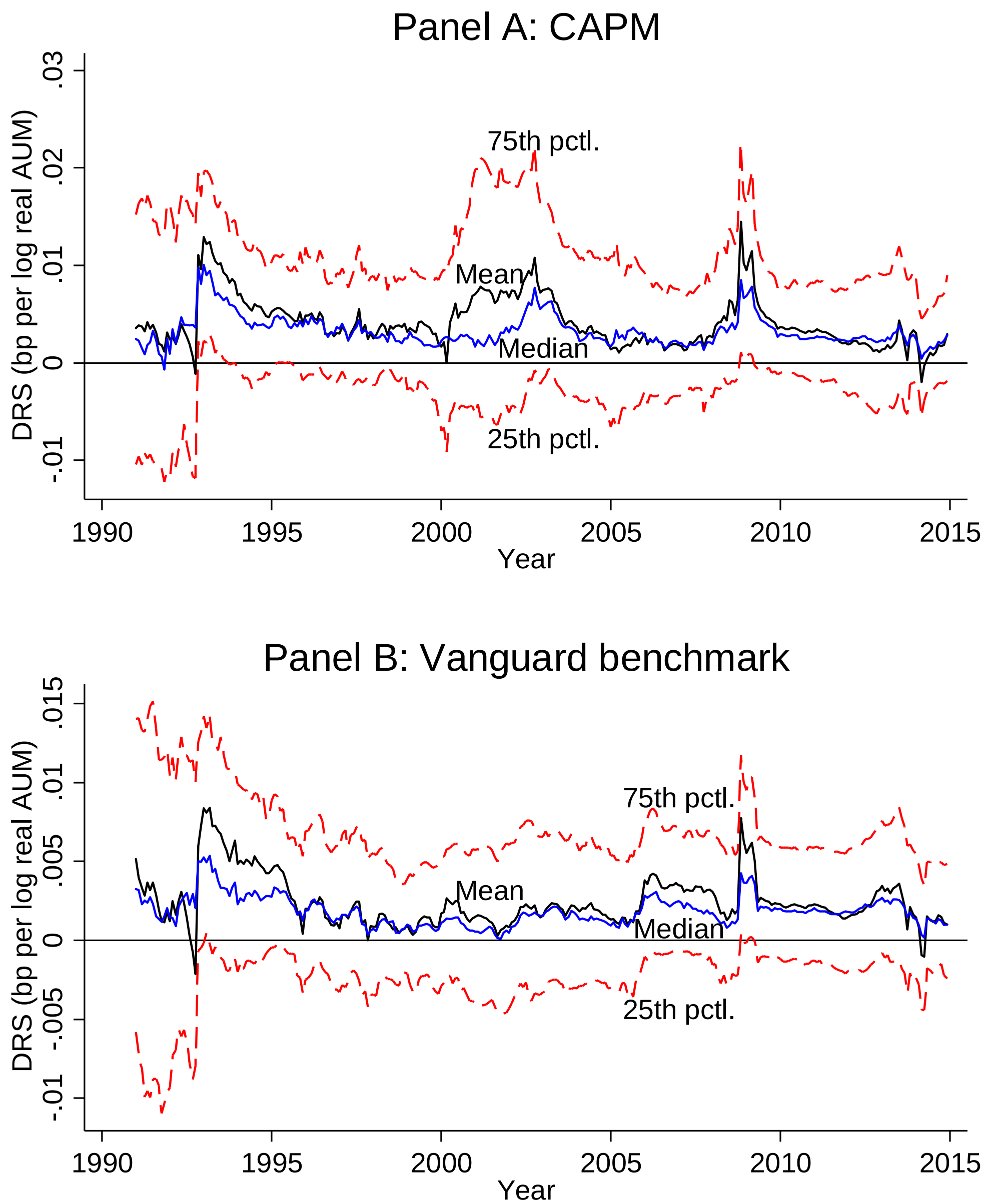


Figure 1: **Distribution of individual decreasing returns to scale (DRS) parameters over time:** The figure plots each month's mean and percentiles of estimated size effect on performance $(\widehat{b}_{it})$ across all funds operating during that month. Panel A estimates DRS when we calculate outperformance relative to the CAPM. Panel B estimates DRS when we calculate outperformance relative to the Vanguard benchmark.

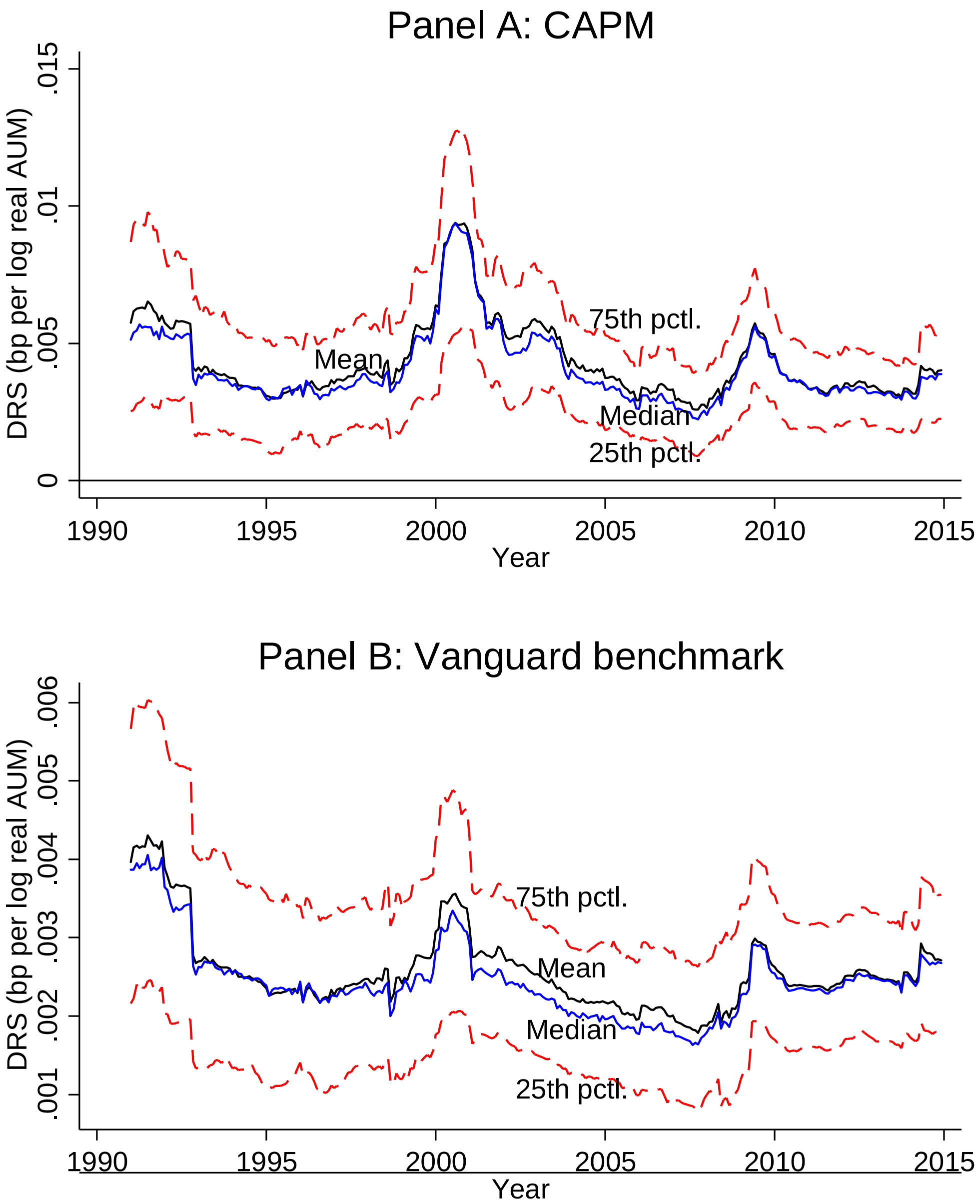


Figure 2: **Distribution of individual DRS parameters' characteristic component over time:** The figure plots each month's mean and percentiles of estimated size effect on performance $(\widehat{b}_{it})$ *explained by fund characteristics* $(\widehat{b}_{it}^{Char})$ across all funds operating during that month. Panel A estimates DRS when we calculate outperformance relative to the CAPM. Panel B estimates DRS when we calculate outperformance relative to the Vanguard benchmark.

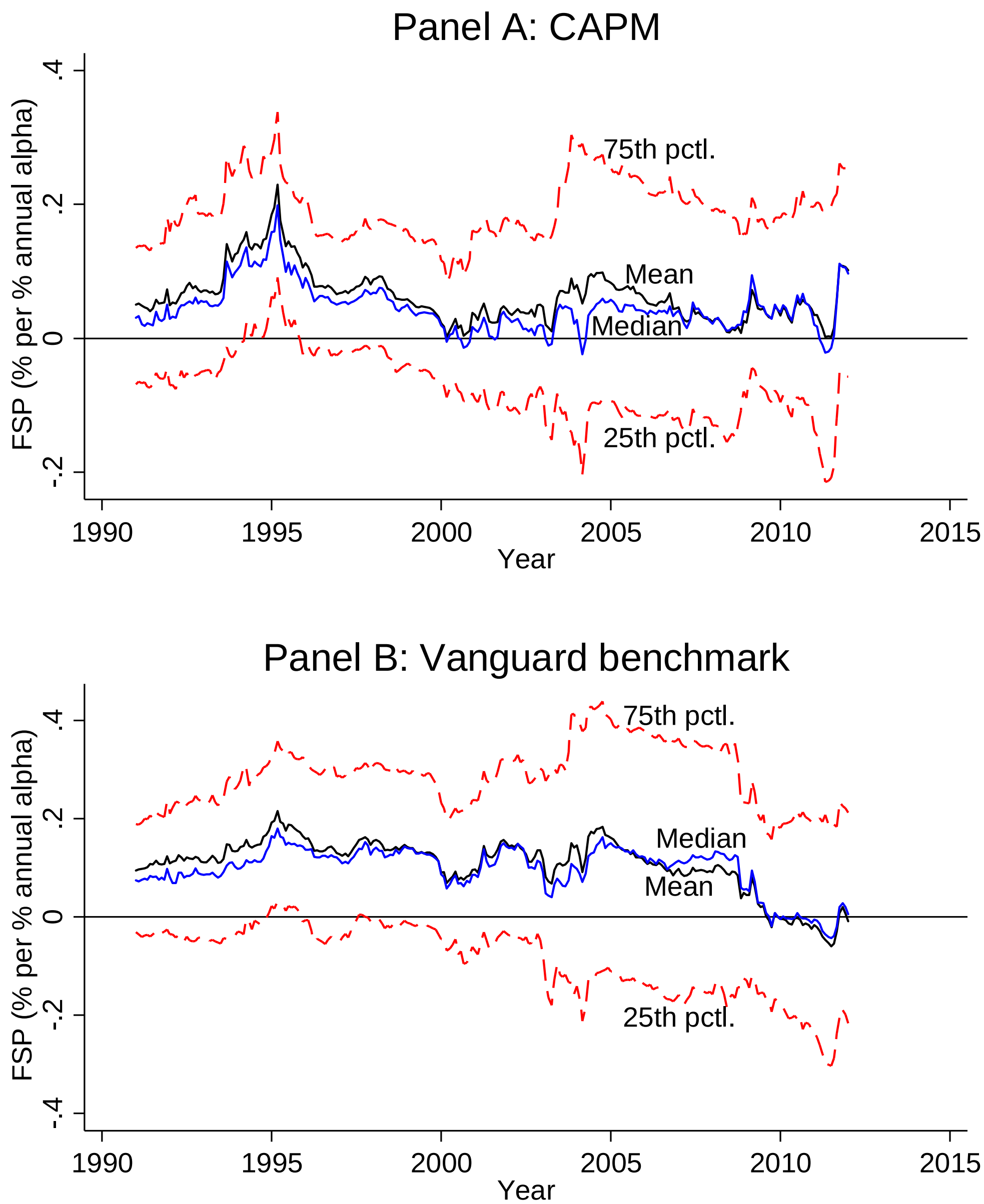
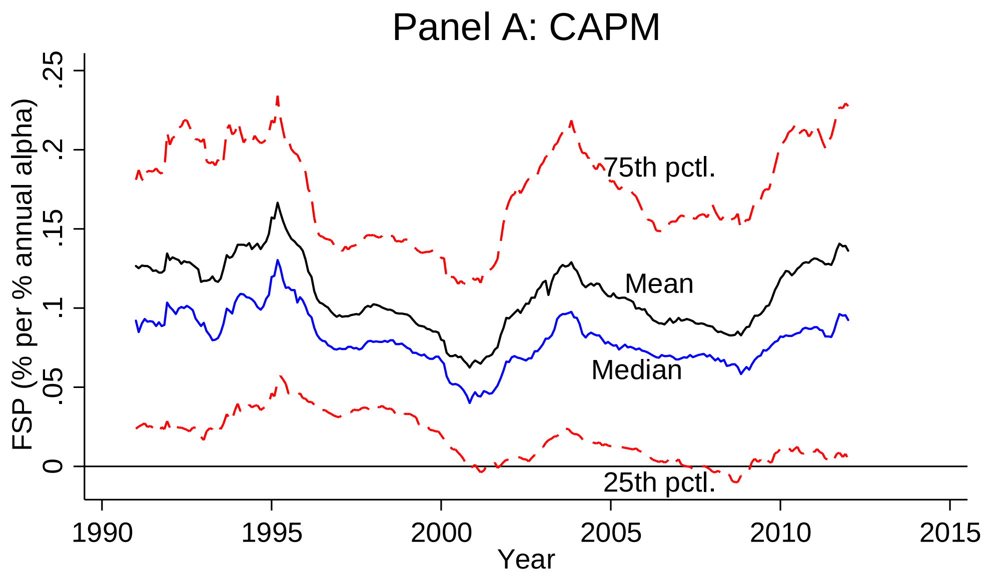


Figure 3: **Distribution of individual flow sensitivity to performance (FSP, v.1) over time:** The figure plots each month's mean and percentiles of estimated % change in real AUM due to performance across all funds operating during that month. Panel A estimates FSP when we calculate outperformance relative to the CAPM. Panel B estimates FSP when we calculate outperformance relative to the Vanguard benchmark.



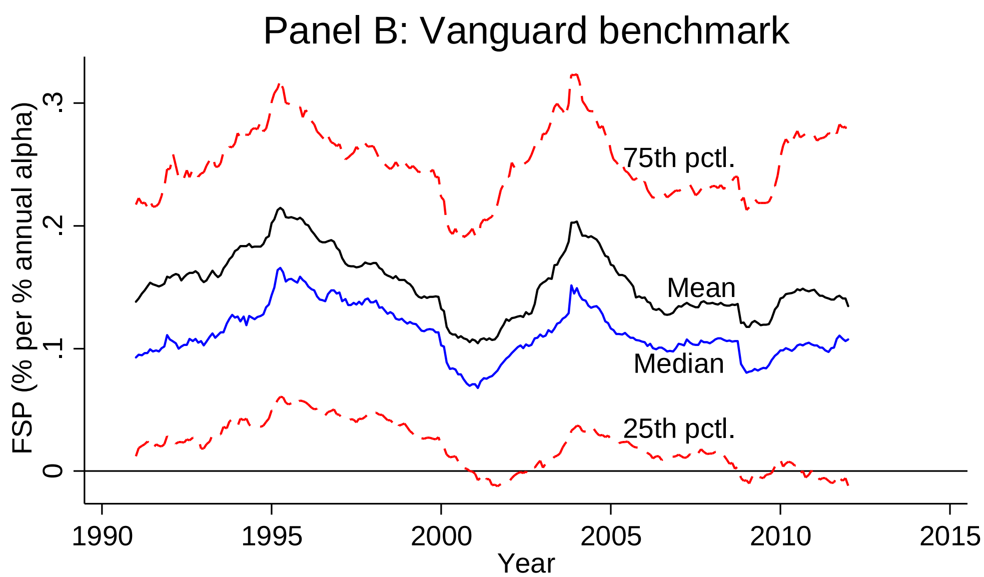


Figure 4: **Distribution of individual flow sensitivity to performance (FSP, v.2) over time:** The figure plots each month's mean and percentiles of estimated % change in new assets due to performance across all funds operating during that month. Panel A estimates FSP when we calculate outperformance relative to the CAPM. Panel B estimates FSP when we calculate outperformance relative to the Vanguard benchmark.