When fund flows take the fun (alpha) away

Benjamin Golez and Sophie Shive*

August 10, 2015

Abstract

We provide new evidence for diseconomies of scale at the mutual fund level. Building on Berk and Green (2004) and allowing for gradual adjustment to equilibrium, we show that (quarterly) changes in fund performance are strongly negatively related to lagged predicted fund flows. We find that alphas would be more cross-sectionally dispersed and more persistent without the damping effect of flows. Thus, flows are an important factor behind the lack of predictability in mutual fund performance. This flow mechanism is strongest for smaller and more active funds with higher expense ratios, suggesting that it is related to the stock illiquidity and costly search for investment opportunities.

*Mendoza College of Business, University of Notre Dame, Notre Dame, IN 46556. Tel. 574-631-1458 / 574-631-8301. Email. bgolez@nd.edu / sshive1@nd.edu. We thank Martijn Cremers, Zhi Da, Paul Gao, Paul Schultz, Charles Trzcinka, and seminar participants at the University of Notre Dame.

1 Introduction

It is well established that money chases mutual fund performance,[1] but puzzling that past performance, as measured by fund alpha, is not a good predictor of future performance. The neoclassical view, advanced by Berk and Green (2004), argues that these observations can be reconciled with the simple assumption of decreasing returns to scale. Investors crowd into high-performing funds, diminishing their future performance, and poorly performing funds improve due to outflows. When rational investor flows attenuate performance in this way, managers can be skilled and yet we may observe little performance persistence.

The validity of this view hinges upon the assumption of decreasing returns to scale. To date, however, the evidence is mixed about whether individual fund returns vary with fund size. Empirical tests of diseconomies of scale face several statistical challenges because size is endogenous to a fund's performance, both directly through changes in portfolio value and indirectly through performance-induced flows (Pastor, Stambaugh, Taylor, 2004, McLemore, 2014).

We address this problem by modeling changes in fund performance and size rather than levels. Changes in fund size can be broken into components due to flows, and due to the growth of the fund through returns. In the Berk and Green (2004) model, prices are normalized to one and fund flows are the sole performance-equilibrating mechanism. Empirically as well, permanent changes in fund size are largely due to flows.[2] Thus, the test for decreasing returns to scale can alternatively be formulated by analyzing the relation between changes in performance and changes in size, most cleanly measured by flows. Because we only focus

[1] For example, Gruber (1996), Chevalier and Ellison (1997), Sirri and Tufano (1998), Spiegel and Zhang (2013).

[2] Because average fund beta is approximately one and average fund alpha is approximately zero, the change in market adjusted fund size must be ultimately driven mostly by flows.

on the part of the change in size that is due to flows, our empirical approach substantially diminishes the problem of endogeneity. Still, performance changes contain lagged performance, which may be related to lagged flows. To address the latter concern, we do not use lagged flows directly, but instrument lagged flows with further lags of flows and performance.

Another important benefit of our methodology is that an analysis of changes in performance is better suited to capture partial adjustments in performance than an analysis of levels. If per-period adjustment is partial, funds with high past alphas and high past inflows may still have relatively high current alphas. Analyzing changes may therefore be necessary to capture this partial adjustment due to flows.

We also examine an empirical model where we test both directions of the flow-performance relation simultaneously. This provides more efficient estimates and enables to quantify the overall importance of the flow-performance mechanism. Last, separating inflows and outflows allows to test for any asymmetries in performance response to changes in fund size.

We test our empirical model on domestic active funds from the CRSP mutual fund database from January 1999 to June 2014. Quarterly market model and four-factor alphas are estimated using daily fund returns (net-of-fees). We choose a quarterly horizon as this is the highest frequency at which alphas can be reliably estimated with daily data. To circumvent endogeneity issues, we instrument flows using four lags of past flows and past alphas.

Our key results are remarkably strong. Instrumented net fund flows are always negatively and statistically significantly related to changes in future fund alphas. This holds within the OLS model as well as within the logistic model where we relate flows to a binary variable that captures a positive change in fund alphas. The results are qualitatively the same for market model alphas and four-factor model alphas, and are robust to inclusion of control

variables and fixed effects.

Estimates of a logistic model suggest that a one percent increase in net flow is associated with a 0.44% to 0.60% decrease in the odds of a positive change in next-quarter alpha. OLS estimates imply that a one percent increase in net flows leads to approximately 0.02% decrease in next-quarter alpha.

When we examine inflows and outflows independently, we find, consistent with diseconomies of scale, that inflows are negatively related to subsequent changes in fund alphas, whereas outflows are positively related to changes in future fund alphas. Interestingly, outflows tend to have a somewhat stronger effect on fund performance than inflows.

Finally, we simultaneously estimate both directions of the flow-performance relationship (using a three-stage least squares procedure), which increases the efficiency of the estimated parameters. The results remain strong. The analysis of the reverse direction of the flow-performance relationship also confirms the well-documented observation that investors chase performance; flows are positively and significantly related to lagged flows and lagged fund alphas.

Although luck is always a concern in the analysis of mutual fund performance, recall that we mitigate this concern by using instrumented flows. Thus, while luck may affect our dependent variable, it does not co-determine flows. Also, all of our results are robust to the inclusion of proxies for the effects of luck, such as lagged changes in alphas and lagged returns.

Furthermore, we show that the flow mechanism is stronger for small-cap funds and is particularly pronounced among the more active small funds that have high expense ratios. This is consistent with the typical justification for decreasing returns to scale based on liquidity reasons and non-scalability of investment ideas (Berk and Green, 2004; Pollet and

Wilson, 2008; Pastor, Stambaugh, and Taylor, 2004). Finally, in a placebo test we show that the flow mechanism is nonexistent for large index funds.

To better understand the importance of the flow-performance mechanism, we next recalculate alphas by isolating the effect of flows. In particular, we calculate "no-flow" alphas by setting the coefficient on flows in our empirical model equal to zero. As expected, no-flow alphas are more cross-sectionally dispersed and farther from zero, as revealed by the increase in the sum of squared alphas. No-flow alphas are also more persistent than true alphas. We test for persistence using the standard procedure of first sorting funds into deciles based on their current alphas, and then testing for any pattern across deciles in the quarter-ahead alphas. In the case of market model alphas, the average alphas for extreme deciles are -0.65% and 0.57%, in comparison to -1.00% and 0.63% for the no-flow alphas. In the case of the four-factor model, the ranges for actual alphas and no-flow alphas are -0.90% to 0.37% and -0.99% to 0.47%. Note that these effects are only those of instrumented lagged flows on quarter-ahead alphas, and thus capture only a subset of the true relation between changes in size and changes in performance. Also, the dollar magnitudes are important, given the size of the mutual fund industry.

In sum, our results imply that flows have an important effect on alphas and suggest that lack of performance persistence in the mutual fund industry is at least partially due to the fund-level diseconomies of scale.

Our paper builds on the theoretical model of Berk and Green (2004) and contributes foremost to the study of returns to scale in the mutual fund industry. Initially, this literature focused on the cross-sectional evidence. For example, Chen, Hong, Huang, and Kubik (2004) and Yan (2008) document that larger funds underperform smaller ones, and attribute these adverse size effects to liquidity and trading costs. As pointed out by recent studies, however,

endogeneity between fund size and fund performance renders this evidence unreliable. To circumvent this endogeneity problem, Reuter and Zitzewitz (2013) consider a quasi-natural experiment where small differences in fund returns cause discrete changes in Morningstar ratings. These changes in turn induce flows (changes in fund size) that are unrelated to fund performance. Phillips et al. (2014) consider a set of instrumental variables that influence fund size but not fund expected returns. Finally, Pastor, Stambaugh, and Taylor (2014) provide a careful account of statistical issues and re-evaluate economies of scale using fund fixed effects and forward demeaning. All three studies find weak and insignificant evidence for diseconomies of scale at the fund level. Only McLemore (2014), who treats fund mergers as quasi exogenous events that are arguably unrelated to past performance, but have large impact on fund size, provides evidence consistent with diseconomies of scale. In sum, the evidence for scale effects at the individual fund level is at best mixed.

We differ from this literature in that we directly model changes in fund performance rather than its levels. To the extent that the adjustment each period from the equilibrating mechanism is not complete, alpha will adjust towards zero and the change will be related to the change in fund size, most cleanly measured by fund flows. Driven by this intuition, we document a strong and uniformly negative impact of flows on performance changes, and thus provide uniformly strong evidence for diseconomies of scale.[3] Our results are thus also related to Barras, Scaillet, and Wermers (2010), who document that, controlling for "false discoveries", the majority of mutual fund alphas are truly zero, aligned with the predictions of Berk and Green (2004).

Finally, note that the partial adjustment to equilibrium is consistent with the smart money effect, that is, funds that receive more money subsequently perform better than

[3]Our results are not driven by fund mergers.

those that lose money (Zheng 1999). Indeed, if the flow mechanism is slow, funds with high past alphas have high inflows and will continue to outperform the market until they receive enough flows to push their alpha to zero. The more skilled the fund managers are to cope with inflows, the longer they will manage to outperform the market. Zheng (1999) documents that the smart money is rather short lived. In a recent study, however, Berk and van Binsbergen (2015) document that skill as measured by the product between the gross alpha and the assets under management persists up to ten years. Thus, some fund managers appear to have the ability to overcome diseconomies of scale for an extended period of time.

2 Theoretical motivation

We base our empirical model on Berk and Green (2004). There is a source of differential skill across fund managers, $a_{i,t}$, for fund i in quarter t. Decreasing returns to scale are captured by the cost of managing a fund of size $q_{i,t}$, $\beta q_{i,t}$. The risk-adjusted, net-of-expenses return during the period is defined as:

$$\alpha_{i,t} = a_{i,t} + \beta q_{i,t-1} + \epsilon_{i,t}, \tag{1}$$

where $\epsilon_{i,t}$ is normally distributed with mean zero and β is negative, consistent with decreasing returns to scale.

Investors rationally respond to outperformance (positive α) with an inflow of funds, and to underperformance (negative α) with an outflow of funds. Thus, there is a two-way relation between flows and performance. Outperforming funds receive inflows, and these same inflows, by the means of decreasing returns to scale, cause funds future performance to deteriorate. Similarly, underperforming funds face outflows, which increases funds future performance. In

equilibrium, these relations imply that as investors learn about managerial ability, variations in α become increasingly due to noise, and its expected value and persistence are driven to zero.

To the extent that adjustment towards equilibrium is not complete each period, the last period's $\alpha_{i,t}$ is only a starting point for adjustment towards zero due to flows. To capture this adjustment mechanism, in our empirical model, we chose to model performance in terms of changes:

$$\Delta\alpha_{i,t} = \alpha_{i,t} - \alpha_{i,t-1} = (a_{i,t} - a_{i,t-1}) + \beta(q_{i,t-1} - q_{i,t-2}) + (\epsilon_{i,t} - \epsilon_{i,t-1}). \tag{2}$$

Assuming a fund manager's true skill a does not change during a period, in line with the existing empirical literature, the above relation simplifies to:

$$\Delta\alpha_{i,t} = \beta(q_{i,t-1} - q_{i,t-2}) + \nu_{i,t}. \tag{3}$$

Thus, the fund's risk-adjusted performance during a period may change either due to changes in fund size or due to random variation. Changes in fund size, $q_{i,t-1} - q_{i,t-2}$, can further be broken into components due to flows, $flow_{i,t-1}$, and due to the growth of the fund through returns, $ret_{i,t-1}$;

$$\Delta\alpha_{i,t} = \beta_{flow} flow_{i,t-1} + \beta_{ret} ret_{i,t-1} + \nu_{i,t}. \tag{4}$$

While growth through flows commands actual trading, either rescaling of existing positions or trading in new investment opportunities, organic growth does not necessitate trading (other than rebalancing). Also, given that average fund beta is approximately one and average fund alpha is approximately zero, the change in market adjusted fund size must

be ultimately driven mostly by flows. Similarly, in the Berk and Green (2004) model, the equilibrating mechanism is flows as funds are open-ended and prices are always one. In our empirical model, we thus focus on the relation between the change in alpha and lag flows and use lagged returns merely as a control variable.

Modeling changes in performance as a function of lagged flows also presents challenges. In particular, our dependent variable $\Delta\alpha_{i,t}$ contains lagged alpha. Insofar as lagged alpha is affected by luck, this may bias $\widehat{\beta}_{flow}$. To see this, consider a fund that outperforms in one period. If this outperformance is due to luck, then the next period alpha will likely decrease. At the same time, the contemporaneous flows increase as investors mistaken luck for skill and invest in the fund. This biases estimated $\widehat{\beta}_{flow}$ downwards. To circumvent this endogeneity problem, we instrument flows using lags of flows and alphas. We also control for lag changes in alpha and other standard control variables.

Thus, our base empirical model is as follows:

$$\Delta\alpha_{i,t} = A_0 + A_1 \widehat{flow_{i,t-1}} + A_2 \Delta\alpha_{i,t-1} + controls_{i,t-1} + \epsilon_{i,t}, \tag{5}$$

where $\widehat{flow_{i,t-1}}$ is predicted using lags of alphas and flows.

2.1 Logit models

We consider several variations of our base model. First, our aim is to test for the theoretically correct signs. We define our dependent variable as a binary variable, where $D^{\Delta\alpha_{i,t}}$ is a dummy variable that takes a value one if $\Delta\alpha_{i,t}$ is positive, and zero otherwise. We then use logit model to test whether the estimated parameter on flows is negative.

The theory also predicts that both inflows and outflows should impact fund performance

symmetrically, although with the opposite sign. To test this, we split predicted flows into predicted inflows and predicted outflows. The estimated parameter on inflows should be negative and the estimated parameter on outflows should be positive.

2.2 OLS models

Next, our aim is to estimate the magnitude of the flow mechanism. We thus estimate our base regression model in Eq. (5) with OLS. We use either the total predicted flows or flows decomposed into inflows and outflows. The expected signs on the estimated parameters are the same as in the above binary models.

2.3 3SLS models

Theory predicts a two-way relationship between flows and performance. Therefore, we next use 3SLS to estimate both directions of the flow-performance relationship simultaneously. Effectively we add a flow equation to our base empirical model.

$$\Delta\alpha_{i,t} = A_0 + A_{flow}\widehat{flow_{i,t-1}} + A_2\Delta\alpha_{i,t-1} + controls_{i,t-1} + \epsilon_{i,t}, \tag{6}$$

$$flow_{i,t} = B_0 + B_1 flow_{i,t-1} + B_2\alpha_{i,t-1} + controls_{i,t-1} + \epsilon_{i,t}. \tag{7}$$

The simultaneous estimation of both equations takes into account the correlation structure among residuals and thus leads to more efficient estimation.[4]

[4]The 3SLS method of Zellner and Theil (1962) uses instrumental variables to calculate consistent estimates when some equations contain endogenous variables, as well as generalized least squares to account for correlation in errors across equations.

3 Data

We use the CRSP Survivor-bias-free mutual fund database beginning in 1999 as that is when daily mutual fund returns were first reported. We restrict our sample to domestic funds and remove index funds and exchange-traded funds. We aggregate fund share classes using the CRSP class group and weight returns by total net assets at the end of the month. As in Chen, Hong, Huang, and Kubik (2004), we exclude the bottom quintile of funds each quarter.

Because we analyze active funds, we calculate performance using daily data to allow for risk loadings to change quarter-by-quarter. In particular, we estimate quarterly market model and Carhart (1997) 4-factor alphas using daily fund returns (net-of-fees) and daily factors from Ken French's website. We use quarterly data as a compromise. High frequency data is best to analyze equilibrating mechanisms, yet performance measures are less reliably estimated at higher frequencies.

We calculate quarterly flows as:

$$flow_{i,t} = \frac{TNA_{i,t} - TNA_{i,t-1}(1 + R_{i,t})}{TNA_{i,t-1}(1 + R_{i,t})},$$

where $R_{i,t}$ are fund returns and $TNA_{i,t}$ are total net assets. Total net assets are adjusted for mergers, so our results are not driven by those found in McLemore (2014). Flows are adjusted for fund mergers. Additionally, we estimate predicted flows $\widehat{flow}$, which we use as an instrument to circumvent endogeniety issues. We predict flows using their own lags as well as lagged alphas; lagged flows to capture flow persistence, lagged alphas to capture performance-chasing. We predict flows using either market model alphas or four-factor model alphas. In our main analysis, we use four lags of flows and alphas and require at least 20 quarters of data to estimate the parameters:

$$\widehat{flow_{i,t}} = \widehat{B_0} + \sum_{j=-4}^{-1} \widehat{B_1} flow_{i,t-j} + \sum_{j=-4}^{-1} \widehat{B_2} \alpha_{i,t-j}, \tag{8}$$

where $\widehat{B_1}$ and $\widehat{B_2}$ are vectors that collect the estimated parameters on lag flows and lag alphas.

To check for any asymmetry, we decompose flows into $\widehat{Inflow}$ that equals predicted flow when the predicted flow is positive, and zero otherwise; and $\widehat{Outflow}$ that equals negative the predicted flow when the predicted flow is negative, and zero otherwise.

Because CRSP coverage of mutual funds in 1999 is poor, the final sample period is from 2000Q3 to 2014Q2 and the sample contains 4,532 unique funds. To diminish the effect of outliers, we winsorize all variables related to performance (alphas) and flows at the 1% and 99% levels.

Table 1 reports summary statistics. Funds in our sample have on average total net assets of $1,350 million, with a median of $304 million. Quarterly fund returns average 2.1% per quarter with a median of 3.3% per quarter. Both risk-adjusted performance measures, however, are negative on average. The market model alpha is -0.2% per quarter and the 4-factor model alpha is -0.3% per quarter.

Funds in our sample experience 0.3% quarterly net flows with a standard deviation of 12.3%. Predicted flows are on average 1.2% per quarter and are somewhat less volatile than the actual flows, with a standard deviation of 9.5%. The average for predicted inflows is 3.6% per quarter and the average for predicted outflows is 2.4% per quarter. Inflows are also twice as volatile as outflows. Using market model alphas or four-factor model alphas in predicting flows does not affect the time series properties of predicted flows and the summary

statistics for predicted flows are closely similar. Funds in our sample have on average 93.0% of their value invested in common stock and 3.2% in cash, and their expense ratios average 1.2% per year.

4 Panel regressions: Equilibrating mechanism

Table 2 reports the main results for testing the equilibrating mechanism. We start by analyzing results for logistic regressions reported in Panel A. We first note that, in line with the argument for decreasing returns to scale, the estimated parameter on instrumented net flow, $\widehat{Flow}$, is negative and statistically significant. This holds when we measure performance in terms of the market model alpha or the 4-factor alpha. Results strengthen when we include control variables and time fixed effects. In the latter case, a one percent increase in $\widehat{Flow}$ is associated with a 0.53% (0.95%) decrease in the odds of a positive change in one quarter ahead market model alphas (four-factor alphas).

Separating the two, we find that inflows and outflows have the theoretically correct sign: $\widehat{Inflow}$ is negatively related to an increase in future fund performance and $\widehat{Outflow}$ is positively related to an increase in future fund performance. Interestingly, the coefficient for outflows is somewhat larger than the coefficient for inflows. Depending on how we measure alphas, a one percent increase in inflows leads to a 0.42% to 0.82% decrease in the odds of a positive change in alphas, whereas a one percent increase in outflows is associated with a 0.86% to 1.34% increase in the odds of an increase in alphas. Although the difference between these coefficients is not statistically significant, the asymmetry suggests that funds' strategy to cope with inflows is different from funds' reaction to outflows.

Examining the control variables, estimated parameters on lagged returns and lagged

changes in alphas are all negative and significant. Insofar as these two variables capture luck, our results confirm that a large part of variation in alphas is random. The estimated parameter on total assets under management is significant and negative, but we do not interpret the coefficient because the potential endogeneity between size and performance is not addressed in this variable. Changes in expense ratios and fund age are insignificantly related to changes in performance.

Next, we turn to the analysis of OLS estimates reported in Panel B. All the main variables of interest have the same sign as in the logistic regressions. Net flows are negatively and significantly related to future changes in market model alphas and four-factor alphas, and all results are robust to including control variables and fixed effects. Depending on how we measure alphas, a one percent increase in net flows leads to approximately 0.02% decrease in quarterly alphas. OLS estimates also confirm that outflows have a stronger effect on alphas than inflows. A one percent increase in inflows predicts 0.01% to 0.02% decrease in quarterly alphas, whereas a one percent increase in outflows leads to 0.03% to 0.04% increase in quarterly alphas. The differences in these coefficients are statistically significant, at the 4% level for market-model alphas and at the 1% level for 4-factor alphas.

In Panel C, we report results from three-stage least squares, (3SLS), where we estimate both sides of the flow-performance relationship simultaneously. Although the estimated parameters in the regression of changes in alphas on flows are similar same as the OLS estimates, the simultaneous estimation is more efficient, The simultaneous estimation also enables us to analyze the flip side of the flow-performance relation. Consistent with the literature, we find that investors chase performance; flows are positively and significantly related to lagged flows and lagged fund alphas, either market model alphas or the 4-factor model alphas.

All in all, we find strong support for the equilibrating mechanism. Flows work like a damper on a spring. Investors chase performance, but performance induced inflows decrease future performance. Similarly, underperformance leads to outflows and subsequent improvements in fund performance.

We perform several robustness checks on the data which remain untabulated. First, as expected, the results are much stronger if we use actual lagged flows rather than predicted lagged flows. Also, the results survive in the time-series when we aggregate funds, weighting by their assets under management. Third, using two-quarter change in alphas as the dependent variable instead of one-quarter change in alphas diminishes the coefficient somewhat but does not remove its statistical significance. Finally, we estimate fund alpha as fund return minus return on the benchmark as assigned by Morningstar. Since Morningstar does not provide data on benchmarks for all the funds in the CRSP database, the number of observation in our main model decreases by one third, but the results remain robust.

4.1 Cross-sectional variation

If funds are subject to decreasing returns to scale, net fund flows should correlate negatively with changes in future performance. Our results confirm just this, and thus provide evidence for decreasing returns to scale at the individual fund level. Importantly, we show that the performance not only deteriorates with an increase in fund size (inflows), but it also improves with a decrease in fund size (outflows). Further, the effect appears asymmetric and somewhat stronger for outflows.

Decreasing returns to scale, however, may not affect all funds equally. Therefore, we next verify that the flow mechanism is indeed strongest for funds that are most susceptible to diseconomies of scale.

Decreasing returns to scale are typically justified by liquidity reasons and non-scalability of investment ideas (Berk and Green 2004; Pollet and Wilson, 2008; Pastor, Stambaugh, and Taylor, 2014). When a fund grows, the price impact of its trades increases, reducing its returns from exploiting existing investment opportunities. It also means that funds that receive inflows but cannot simply rescale their positions due to illiquidity have to go down the list to their next-best investment ideas. Similarly, when funds receive outflows, fund managers partially rescale their existing portfolio and also take the opportunity to liquidate their least promising investment.

This suggests that we would expect diseconomies of scale to be primarily related to the liquidity of stocks in funds portfolios. Since small stocks are typically less liquid, small-cap funds should therefore be more prone to the flow mechanism than large-cap funds. To verify this, we first repeat our analysis on a subset of cap-based funds as classified by CRSP. We note that the main coefficient in Table 2 (Panel B, Column 1) is -0.011 and insignificant for large-cap funds, -0.021 and marginally significant for medium-cap funds, and -0.031 and significant for small- and micro-cap funds. Thus, as expected, small-cap funds indeed have the strongest flow mechanism.

The CRSP cap-based classification, however, leaves out all the sector and style funds, substantially reducing our sample. To circumvent this problem, we next build on the insight that funds that trade small stocks also tend to be smaller in size (CRSP large-cap funds are on average four times bigger than small- and micro-cap funds). This motivates fund size as a proxy for the liquidity of stocks in the fund portfolio.

The severity of diseconomies of scale may also be related to other fund characteristics. The more active the funds are, the more the liquidity considerations may affect their portfolio choice. Thus, for a given size, we would expect that funds that are more active are more

prone to the flow mechanism. For the same reason, we may expect that, for a given size, funds that invest more in research and charge higher fees would be more sensitive to the diseconomies of scale. This motivates two additional sorting variables: a measure of fund activeness and fund fees. We use our *Fit* variable, the R-square from the market model, as a measure of fund activeness (we run market model quarter by quarter on daily data). For fees, we use annual expense ratio.

For each quarter, we form terciles on each of the three sorting variables, always using the data from that particular quarter. Each fund is assigned to one of the resulting categories. We then estimate the flow mechanism within each of the categories. We consider single sorts, as well as double sorts on either fund size and fund fit or fund size and fund expense ratio. For exposition, we focus our attention on our main model where we regress changes in market model alphas on predicted flows (Table 2, Panel B, Column 1).[5]

Table 3 reports results for each category. Panel A reports results for single sorts. Panel B reports results for double sorts on fund size and fit and Panel C reports results for double sorts on fund size and expense ratio.

We note in Panel A that all the single sorts are aligned with the argument for decreasing returns to scale. Small funds have a much stronger flow mechanism than large funds. Whereas the estimated parameter of interest is -0.029 and sightly significant for small funds, it is -0.017 and only marginally significant for large funds. Similarly, more active have stronger flow mechanism. The estimated parameter in the low fit category is -0.040 and slightly significant and -0.005 and insignificant for the high fit category. We also observe similar monotonicity for the expense ratio sorts. Funds with low expense ratios have the estimated parameter of interest of -0.009 and insignificant, whereas this coefficient is -0.032

[5]Results based on 4-factor alphas are very similar.

and highly significant for the funds charging high fees.

Panel B further reveals that three-by-three sorts of fund size and fit are consistent with the single sorts. The estimated parameter is increasing (becoming less negative) in both size and fit. In Panel C, we also note that the estimated parameter of interest is increasing in size and decreasing in expense ratio.

Finally, we conduct a placebo test and re-do the analysis on a subset of index funds (which we exclude in the main analysis). As index funds are passive and simply track the performance of the benchmark, we expect to find no effect of the flow mechanism. Since alpha within the market model may be inaccurately measured for funds tracking sector and small-cap benchmarks, we restrict our analysis to large-cap index funds (Cremers, Petajiso, and Zitzewitz, 2012). As expected, we find no relation between changes in alphas and predicted flows. The main coefficient (Table 2, Panel B, Column 1) is even slightly positive 0.001, but insignificant.

All in all, these results confirm collectively that the flow mechanism indeed appears strongest among funds that are most susceptible to diseconomies of scale, further validating our methodology.

5 No flow alphas

Our analysis suggests that fund flows have an important effect on alphas. The natural question that arises is how alphas would look without the damping effect of flows. In this section, we address this question. To allow for between fund variation, we first estimate our empirical model on a fund-by-fund basis. Then we isolate the effect of fund flows by shutting down the flow-performance mechanism. Finally, we compare the distribution of the

hypothetical alphas without the effect of flows and the distribution of actual alphas.

5.1 Fund by fund regressions

For parsimony, given that we estimate coefficients for each fund, we adopt the following reduced model:

$$\Delta\alpha_{i,t} = A_{0,i} + A_{1,i}\widehat{flow_{i,t-1}} + A_{2,i}\ r_{i,t-1} + A_{3,i}\Delta\alpha_{i,t-1} \tag{9}$$

We require at least 20 quarterly observations for each fund. This leaves us with 2,151 funds for which we can estimate coefficients. As before, $\widehat{flow}$s are instrumented using four lags of flows and alphas.

Summary statistics for the coefficient of changes in alpha on flows $\widehat{A_1}$ are reported in Table 3. In the case of the market model alphas, the mean coefficient is -0.14 with a standard deviation of 0.37. In the case of the 4-factor model alphas, the mean coefficient is much smaller in absolute value at -0.07 and comes also with a lower standard deviation of 0.28. Although some funds have positive coefficients, the median for both market model alphas and four-factor alphas is negative, as in the earlier results.

5.2 No flow alphas

Next, we recalculate alphas by shutting down the flow-performance mechanism. In particular, we calculate the so called *No flow* alphas as:

$$\alpha_{i,t}^{Noflow} = \alpha_{i,t-1} + (\Delta\alpha_{i,t} - \widehat{A_{1,i}}\widehat{flow_{i,t-1}}), \tag{10}$$

where $\widehat{A_{1,i}}$ is the estimated parameter in Eq. (9).[6] Figure 1 plots the distribution of these hypothetical alphas without the effect of flows along with the actual alphas. The corresponding fund-level summary statistics are reported in Table 3.

Theoretically, if managers have differential ability, but the flow mechanism pushes actual alphas towards zero, we should observe that no-flow alphas better reflect managerial ability and are thus farther away from zero and more dispersed. As we note in Figure 1, this is indeed the case. The distribution of no-flow alphas, relative to the distribution of actual alphas, has more mass in the tails and less mass in the center.

To measure the effect more precisely, we calculate the sum of squared actual alphas and the sum of squared no-flow alphas. We then define the *Reversion ratio* that summarizes the effect of flows on alphas on a fund-by-fund basis within a single statistic. It is defined as the relative difference between the sum of squares of actual alphas and the sum of squares of alphas without the effect of flows:

$$Reversion\ ratio = \frac{SS(\alpha_{Noflow}) - SS(\alpha_{Actual})}{SS(\alpha_{Actual})},$$

where $SS(\alpha_{Noflow})$ is the sum of squares of no-flow alphas, and $SS(\alpha_{Actual})$ is the sum of squares of actual alphas.

The summary statistics for the reversion ratio are reported in Table 3. In the case of the market model, the mean for the *Reversion ratio* is 19.4%, whereas it is 14.1% in the case of

[6] We also recalculate no-flow alphas by separating inflows and outflows and obtain very similar results.

the four-factor model alphas. In untabulated results, we further find that the average of the cross-sectional average of the sum of squared alphas is higher for the no-flow alphas than for the actual alphas, 0.41% versus 0.36% in the case of the market model alphas, and 0.28% versus 0.26% in the case of the four-factor alphas. Thus, flows indeed play an important role for variation in alphas, consistent with the theory.

5.3 Flows and persistence in alphas

Through the flow mechanism, persistent skill attracts flows that destroy the persistence in alphas. This implies that removing the effect of flows should allow us to partially recover persistence in alphas.

To test this implication, we compare the persistence of actual alphas to that of our hypothetical no-flow alphas. We follow the standard procedure of first sorting funds in deciles based on current alphas, and then testing for any pattern across deciles in quarter-ahead alphas. Note, however, that our no-flow alphas are calculated period by period from actual alphas, which would render a pure time series analysis of no-flow alphas inconsistent. For this reason, we take a conservative approach and always sort on the actual alphas. In particular, each quarter, we sort mutual funds in deciles based on their actual alpha. We then calculate the mean actual alpha and the mean no-flow alpha for each decile in the following quarter. Repeating this procedure quarter by quarter, we obtain, for each decile, a time series of actual alphas and no-flow alphas. We plot the means for each time series in Figure 2.

Consistent with the literature, we note that there is little persistence in the actual alphas; where it exists, it is mostly on the negative side: underperforming funds keep underperforming. In the case of market model alphas, eight deciles have negative alphas, and extreme

decile alphas are -0.59% and 0.58%. In the case of four-factor alphas, the mean for alphas is even lower, with nine deciles exhibiting negative alphas and the alphas for extreme deciles ranging between -0.86% and 0.40%.

In comparison, the no-flow alphas are more persistent, with extreme deciles having greater range. Interestingly, the effect appears asymmetric and stronger among the worst performing funds than among the best performing funds. In the case of market model alphas, the range for the no-flow alphas is between -0.91% and 0.65%, whereas it is between -0.93% and 0.49% for the four-factor alphas.

Thus, consistent with the model, the flow mechanism is important for the persistence in alphas. Note that these effects are only those of instrumented lagged flows on quarter-ahead alphas, and thus capture only a subset of the true relation between changes in size and changes in performance. Note also that skill may not be constant as assumed in the model.

6 Conclusion

Motivated by the neoclassical view advanced by Berk and Green (2004) and addressing the possibility of slow adjustment, we create an empirical model that examines how changes in size relate to changes in performance. We find that lagged changes in size, instrumented by predicted flows, are strongly related to changes in mutual fund performance. The effect is present for both inflows and outflows. We estimate that fund alphas would be much more cross-sectionally dispersed and more persistent without the damping effect of flows on performance. While the evidence in prior literature is mixed, our results strongly support the notion of diseconomies of scale.

References

Barras, Laurent, Olivier Scaillet, and Russ Wermers, 2010, False discoveries in mutual fund performance: Meauring luck in estimated alphas, *The Journal of Finance* LXV, 179–216.

Berk, Jonathan, and Richard Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269–1295.

Berk, Jonathan, and Jules van Binsbergen, 2015, Measuring skill in the mutual fund industry, *Forthcoming, Journal of Financial Economics*.

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

Chen, Joseph, Harrison Hong, Ming Huang, and Jeffey Kubik, 2004, Does fund size erode mutual fund performance? the role of liquidity and organization, *The American Economic Review* 94, 1276–1302.

Chevalier, Judith, and Glen Ellison, 1997, Risk taking by mutual funds as a response to incentives, *The Journal of Political Economy* 105, 1167–1200.

Cremers, Martijn, Antti Petajisto, and Eric Zitzewitz, 2012, Should benchmark indices have alpha? revisiting performance evaluation, *Critical Finance Review* 2, 1–48.

Gruber, Martin, 1996, Another puzzle: The growth in actively managed mutual funds, *The Journal of Finance* LI, 783–810.

McLemore, Ping, 2014, Do mutual funds have decreasing returns to scale? evidence from fund mergers, *Working paper*.

Pastor, Lubos, Robert Stambaugh, and Lucian Taylor, 2014, Scale and skill in active management, *Working paper*.

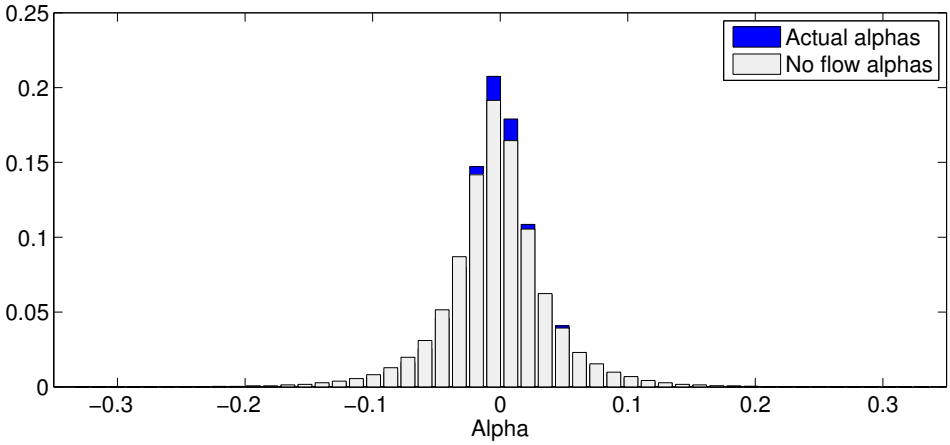
Reuter, Jonathan, and Eric Zitzewitz, 2013, How much does size erode mutual fund performance? a regression discontinuity approach, *Working paper*.

Sirri, Erik, and Peter Tufano, 1998, Costly search and mutual funds, *The Journal of Finance* LII, 1589–1622.

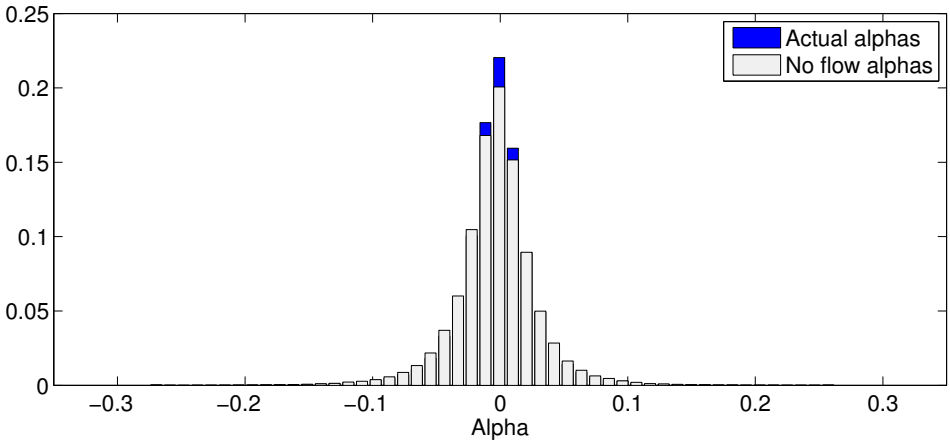
Spiegel, Matthew, and Hong Zhang, 2013, Mutual fund risk and market share-adjusted fund flows, *Journal of Financial Economics* 108, 506–528.

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

Zellner, A., and H. Theil, 1962, Three stage least squares: Simultaneous estimate of simultaneous equations, *Econometrica* 29, 54–78.

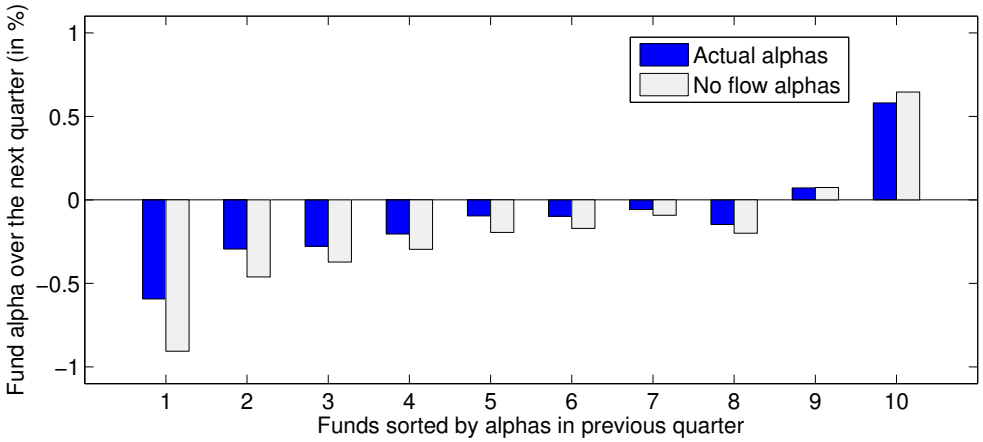


Panel A: Market model alphas

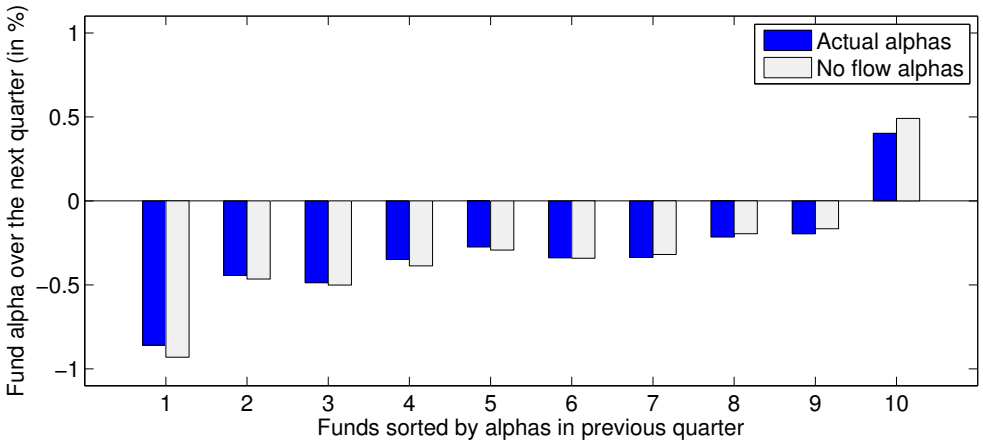


Panel B: Four-factor model alphas

Figure 1. Distribution of mutual fund alphas with and without the effect of flows. Quarterly alphas are computed using daily mutual fund return data and factors from Kenneth French's website. Panel A contains market model alphas and panel B uses 4-factor alphas.



Panel A: Market model alphas



Panel B: Four-factor model alphas

Figure 2. Persistence of mutual fund alphas with and without the effect of flows. Quarter by quarter, funds are sorted in deciles based on the actual alphas. The graph plots, for each decile, the mean for the one quarter ahead actual alphas and no flow alphas. Panel A contains market model alphas and panel B uses 4-factor alphas.

Table 1

Summary statistics

This table reports summary statistics for domestic equity funds from CRSP mutual funds database for the period 2000Q3 to 2014Q2. The data is quarterly and the sample consists of 4,532 unique funds. The smallest quintile of funds, index funds, and exchange-traded funds are removed from the sample. Alphas, changes in alphas, and flows are winsorized at the 1% level. The variable α^{MM} is the market model alpha and α^{4F} is the Carhart (1997) four-factor model alpha of the fund. Alphas are computed using daily fund returns and daily factors from Kenneth French's website. Flow is the percentage quarterly net flow computed by compounding monthly net flows of the fund. Flows are adjusted for fund mergers. The variable $\widehat{Flow}$ is netflow predicted using four lags of flows and four lags of alphas (market model or four-factor model alphas); $\widehat{Inflow}$ equals the predicted flow when the predicted flow is positive, and zero otherwise; $\widehat{Outflow}$ equals negative the predicted flow when the predicted flow is negative, and zero otherwise. Age is the number of years since the inception date of the fund. Expense ratios, and percentages of common and cash are from the CRSP Mutual Fund database. Fit is the R-squared from the regression of fund returns on the market factor.

	Mean	σ	p1	p50	p99	N
TNA	1,350.071	4,963.443	21.900	304.100	17,651.400	109,604
Quarterly return	0.021	0.102	-0.266	0.033	0.234	109,604
α^{MM}	-0.002	0.038	-0.125	-0.002	0.117	109,602
$\Delta\alpha^{MM}$	-0.001	0.053	-0.176	0.000	0.171	109,563
α^{4F}	-0.003	0.030	-0.107	-0.002	0.091	109,602
$\Delta\alpha^{4F}$	0.000	0.041	-0.133	0.000	0.131	109,563
Flow	0.003	0.123	-0.337	-0.014	0.559	109,604
$\widehat{Flow}^{MM}$	0.012	0.095	-0.191	-0.006	0.440	109,604
$\widehat{Inflow}^{MM}$	0.036	0.077	0.000	0.000	0.440	109,604
$\widehat{Outflow}^{MM}$	0.024	0.037	0.000	0.006	0.191	109,604
$\widehat{Flow}^{4F}$	0.012	0.095	-0.192	-0.006	0.439	109,604
$\widehat{Inflow}^{4F}$	0.036	0.076	0.000	0.000	0.439	109,604
$\widehat{Outflow}^{4F}$	0.024	0.037	0.000	0.006	0.192	109,604
Expense ratio	0.012	0.004	0.001	0.011	0.025	101,222
Age	13.047	10.696	2.307	10.282	59.067	109,604
Percent common	92.985	8.663	58.760	95.610	100.330	105,647
Percent cash	3.223	5.413	-3.110	1.940	25.020	105,647
Fit	0.864	0.151	0.150	0.910	0.994	109,604.00

Table 2

Panel regressions: Equilibrating mechanism

This table reports panel and time series regression results for domestic equity funds from CRSP mutual funds database from 2000Q3 to 2014Q2. Panel A presents panel logistic regressions of a binary variable for an increase in alpha (one for an increase in alpha, zero otherwise) on flows and control variables; Panel B reports OLS regressions of changes in alpha on flows and control variables; Panel C reports 3SLS regressions where changes in alphas on flows and control variables are estimated simultaneously with flows on lag flows and lag alphas. Market model alphas are denoted $\alpha_{i,t}^{MM}$, and Fama-French 4-factor model-adjusted alphas are denoted $\alpha_{i,t}^{4F}$. $\widehat{Flow}$ is flow predicted using four lags of flows and alphas; $\widehat{Inflow}$ equals predicted flow if predicted flow is positive, and zero otherwise; $\widehat{Outflow}$ equals negative the predicted flow if predicted flow is negative, and zero otherwise. Standard errors in Panel A and Panel B are clustered by fund quarter. P-values are reported in parentheses below the estimated coefficients. In brackets below the estimated parameter on predicted outflows is the p-value for the test that the estimated parameter on inflows equals negative the estimated parameter on outflows. In Panel C, we also report the correlation of the residuals of the two equations and the p-value for the Breusch-Pagan test of increased efficiency from simultaneous estimation.

Panel A: Logistic regressions

	$Y = D^{\Delta\alpha_{i,t}^{MM}}$			$Y = D^{\Delta\alpha_{i,t}^{4F}}$		
$\widehat{\text{Flow}_{i,t-1}}^{MM/4F}$	-0.535	-0.534		-0.407	-0.950	
	(0.012)	(0.003)		(0.002)	(0.000)	
$\widehat{\text{Inflow}_{i,t-1}}^{MM/4F}$			-0.422			-0.819
			(0.017)			(0.000)
$\widehat{\text{Outflow}_{i,t-1}}^{MM/4F}$			0.862			1.335
			(0.025)			(0.000)
			[0.258]			[0.122]
$\Delta\alpha_{i,t-1}^{MM/4F}$		-10.700	-10.720		-19.727	-19.745
		(0.000)	(0.000)		(0.000)	(0.000)
Total return$_{i,t-1}$		-23.573	-23.560		-8.012	-8.009
		(0.000)	(0.000)		(0.000)	(0.000)
log(TNA)$_{i,t-1}$		-0.020	-0.019		-0.013	-0.011
		(0.015)	(0.020)		(0.099)	(0.121)
$\Delta Expense_{i,t}$		-1.869	-1.802		-3.146	-3.043
		(0.811)	(0.818)		(0.660)	(0.671)
$Age_{i,t}$		0.000	0.000		-0.002	-0.002
		(0.798)	(0.841)		(0.159)	(0.192)
Time FE	NO	YES	YES	NO	YES	YES
Observations	109,563	101,053	101,053	109,563	101,053	101,053

Panel B: OLS regressions

	$Y = \Delta\alpha_{i,t}^{MM}$			$Y = \Delta\alpha_{i,t}^{4F}$		
$\widehat{\text{Flow}}_{i,t-1}^{MM/4F}$	-0.022	-0.015		-0.013	-0.023	
	(0.003)	(0.001)		(0.000)	(0.000)	
$\widehat{\text{Inflow}}_{i,t-1}^{MM/4F}$			-0.011			-0.017
			(0.005)			(0.000)
$\widehat{\text{Outflow}}_{i,t-1}^{MM/4F}$			0.028			0.039
			(0.011)			(0.000)
			[0.04]			[0.00]
$\Delta\alpha_{i,t-1}^{MM/4F}$		-0.251	-0.252		-0.431	-0.431
		(0.000)	(0.000)		(0.000)	(0.000)
Total return$_{i,t-1}$		-0.495	-0.495		-0.149	-0.149
		(0.000)	(0.000)		(0.000)	(0.000)
log(TNA)$_{i,t-1}$		-0.001	0.000		0.000	0.000
		(0.008)	(0.016)		(0.033)	(0.075)
$\Delta Expense_{i,t}$		-0.001	0.002		-0.069	-0.065
		(0.994)	(0.993)		(0.673)	(0.691)
$Age_{i,t}$		0.000	0.000		0.000	0.000
		(0.799)	(0.703)		(0.580)	(0.756)
Time FE	NO	YES	YES	NO	YES	YES
Observations	109,563	101,053	101,053	109,563	101,053	101,053
R-squared	0.002	0.469	0.469	0.001	0.323	0.323

Panel C: 3SLS regressions

$Y = \Delta\alpha$		$\alpha = \alpha_{i,t}^{MM}$			$\alpha = \alpha_{i,t}^{4F}$	
$\widehat{\text{Flow}_{i,t-1}}^{MM/4F}$	-0.033	-0.020		-0.018	-0.026	
	(0.000)	(0.000)		(0.000)	(0.000)	
$\widehat{\text{Inflow}_{i,t-1}}^{MM/4F}$			-0.013			-0.019
			(0.000)			(0.000)
$\widehat{\text{Outflow}_{i,t-1}}^{MM/4F}$			0.040			0.047
			(0.000)			(0.000)
			[0.000]			[0.000]
$\Delta\alpha_{i,t-1}^{MM/4F}$		-0.256	-0.257		-0.431	-0.432
		(0.000)	(0.000)		(0.000)	(0.000)
Total return$_{i,t-1}$		-0.488	-0.487		-0.149	-0.149
		(0.000)	(0.000)		(0.000)	(0.000)
log(TNA)$_{i,t-1}$		0.000	0.000		0.000	0.000
		(0.000)	(0.000)		(0.002)	(0.028)
$\Delta Expense_{i,t}$		-0.026	-0.022		-0.093	-0.087
		(0.819)	(0.849)		(0.357)	(0.384)
$Age_{i,t}$		0.000	0.000		0.000	0.000
		(0.413)	(0.194)		(0.327)	(0.710)
$Y = \text{Flow}_{i,t}$						
Flow$_{i,t-1}$	0.242	0.175	0.176	0.250	0.183	0.183
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Flow$_{i,t-2}$		0.108	0.109		0.112	0.112
		(0.000)	(0.000)		(0.000)	(0.000)
Flow$_{i,t-3}$		0.066	0.066		0.066	0.066
		(0.000)	(0.000)		(0.000)	(0.000)
Flow$_{i,t-4}$		0.059	0.059		0.058	0.058
		(0.000)	(0.000)		(0.000)	(0.000)
$\alpha_{i,t-1}^{MM/4F}$	0.568	0.420	0.420	0.552	0.487	0.488
	(0.000	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\alpha_{i,t-2}^{MM/4F}$		0.276	0.276		0.303	0.304
		(0.000)	(0.000)		(0.000)	(0.000)
$\alpha_{i,t-3}^{MM/4F}$		0.092	0.091		0.108	0.107
		(0.000)	(0.000)		(0.000)	(0.000)
$\alpha_{i,t-4}^{MM/4F}$		0.073	0.073		0.064	0.063
		(0.000)	(0.000)		(0.000)	(0.000)
Time FE	NO	YES	YES	NO	YES	YES
Observations	109,410	100,938	100,938	109,410	100,938	100,938
Weighted R-squared	0.076	0.369	0.369	0.066	0.268	0.269
Correlation of residuals	0.111	0.121	0.122	0.070	0.088	0.089
Breusch-Pagan test χ^2 p	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Table 3

Cross-sectional variation

This table reports category-specific regression slopes of changes in fund alphas on lagged predicted fund flows, analogous to Table 2, Panel B, Column 1. We consider three categories based on fund Size, Fit, and Expense ratio. Fund size (assets under management) and expense ratios are from the CRSP Mutual Fund database. Fit is the R-squared from the regression of fund returns on the market factor (we run market model quarter by quarter using daily data). Panel A reports single sorts; Panel B reports double sorts on Size and Fit; Panel C reports double sorts on Size and Expense ratio. P-values are reported in parentheses below the estimated coefficients.

Panel A: Single sorts

	Low	Medium	High
Size	-0.029	-0.017	-0.017
	(0.000)	(0.046)	(0.078)
Fit	-0.040	-0.013	-0.005
	(0.000)	(0.070)	(0.135)
Expense ratio	-0.009	-0.025	-0.032
	(0.132)	(0.003)	(0.001)

Panel B: Double sort on Size and Fit

	Fund size		
Fit	Low	Medium	High
Low	-0.053	-0.032	-0.031
	(0.000)	(0.017)	(0.047)
Medium	-0.018	-0.009	-0.007
	(0.006)	(0.310)	(0.437)
High	-0.006	-0.005	-0.006
	(0.061)	(0.246)	(0.185)

Panel C: Double sort on Size and Expense ratio

	Fund size		
Expense ratio	Low	Medium	High
Low	-0.018	-0.005	-0.004
	(0.007)	(0.520)	(0.568)
Medium	-0.032	-0.026	-0.016
	(0.000)	(0.019)	(0.148)
High	-0.038	-0.024	-0.030
	(0.000)	(0.063)	(0.070)

Table 4

Fund-level coefficients

This table reports different fund-level summary statistics for domestic equity funds from CRSP mutual funds database from 2000Q3 to 2014Q2; $\widehat{A}$ is the estimated coefficient in the regression of changes in alphas on predicted flows; α_{Actual} is the fund alpha estimated according to the market model (superscript MM) or the four-factor model (superscript $4F$); $\alpha_{\text{No flow}}$ is the alpha without the effect of fund flows; *Reversion ratio* is the relative difference between the sum of squared of actual alphas and the sum of squared of alphas without the effect of flows.

	Mean	σ	p1	p50	p99	N
			Market model alphas			
$\widehat{A_{1,i}}$	-0.136	0.367	-1.613	-0.070	0.646	2,151
$\alpha^{MM}_{\text{Actual}}$	-0.001	0.009	-0.021	-0.002	0.026	2,151
$\alpha^{MM}_{\text{No flow}}$	-0.003	0.017	-0.062	-0.002	0.039	2,151
Reversion ratio	0.194	0.846	-0.195	0.038	3.133	2,151
			4-factor model alphas			
$\widehat{A_{1,i}}$	-0.071	0.275	-1.069	-0.039	0.726	2,151
$\alpha^{4F}_{\text{Actual}}$	-0.003	0.008	-0.022	-0.003	0.023	2,151
$\alpha^{4F}_{\text{No flow}}$	-0.003	0.012	-0.039	-0.003	0.035	2,151
Reversion ratio	0.141	0.650	-0.198	0.023	2.032	2,151