NBER WORKING PAPER SERIES

HOW MUCH DOES SIZE ERODE MUTUAL FUND PERFORMANCE? A REGRESSION DISCONTINUITY APPROACH

Jonathan Reuter
Eric Zitzewitz

Working Paper 16329
http://www.nber.org/papers/w16329

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2010

We thank Edie Hotchkiss, Jeff Pontiff, Phil Strahan, and participants at the 2009 Financial Research Association annual meeting for helpful comments on this project. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

How Much Does Size Erode Mutual Fund Performance? A Regression Discontinuity Approach
Jonathan Reuter and Eric Zitzewitz
NBER Working Paper No. 16329
September 2010
JEL No. G14,G23,G24

ABSTRACT

The two main stylized facts in the mutual fund literature are that funds exhibit little ability to persistently outperform their peers, but new money flows into funds with the highest past returns. The traditional interpretations of these facts are that fund managers are unskilled and fund investors are unsophisticated. Berk and Green (2004) use a model that combines skilled managers with diseconomies of scale in asset management to challenge these interpretations. They argue that more-skilled managers will manage more assets but—-precisely because they manage more assets—-will generate the same expected future returns as less-skilled managers. In their model, standard cross-sectional regressions of fund returns on fund size will significantly underestimate diseconomies of scale. To identify the causal impact of mutual fund flows on performance, we exploit the fact that small differences in mutual fund returns can cause discrete changes in Morningstar ratings and, thereby, cause discrete differences in mutual fund flows. The diseconomies of scale that we estimate using this regression discontinuity approach are larger than those estimated in standard regressions, but generally smaller than assumed in Berk and Green—-or than are required to explain the low observed levels of performance persistence.

Jonathan Reuter
Carroll School of Management
Boston College
224B Fulton
140 Commonwealth Avenue
Chestnut Hill, MA 02467
and NBER
reuterj@bc.edu

Eric Zitzewitz
Dartmouth College
6016 Rockefeller Center
Hanover, NH 03755
Eric.Zitzewitz@Dartmouth.edu

The mutual fund literature is responsible for two well-known facts. The first fact, based on more than thirty years of research, is that actively managed mutual fund returns exhibit little ability to persistently outperform their peers (e.g., Jensen (1968) and Carhart (1997)). The second, newer fact is that new money flows disproportionately into those actively managed funds with the highest past returns (e.g., Chevalier and Ellison (1997) and Sirri and Tufano (1998)). The traditional (academic) interpretations of these facts are that fund managers are unskilled and fund investors are unsophisticated.

Berk and Green (2004) challenge these interpretations. They argue that both stylized facts are consistent with a model that combines skilled managers with diseconomies of scale in asset management. In their model, rational investors chase performance to the point that expected future returns are equalized across funds. In equilibrium, more-skilled managers manage more assets but—precisely because of the diseconomies of scale associated with managing more assets—earn the same expected future return as their less-skilled peers. Berk (2005) goes further, arguing that the traditional interpretations of these stylized facts are "myths" and that the Berk-Green model shows that "most active managers are skilled." Berk and Green's interpretations have quite different implications for our view of financial markets (i.e., easier to beat than we thought) and investors (i.e., harder to fool than we thought), which, in turn, have important implications for public policy, and for the evaluation of fund managers. However, the empirical relevance of the Berk-Green model depends crucially on the degree of scale diseconomies in asset management.

Our goal in this paper is to measure the causal impact of fund size on fund performance. To motivate our empirical strategy, it is helpful to view the existing evidence through the lens of Berk and Green's (2004) model. In a study that is both representative and widely cited, Chen, Hong, Huang, and Kubik (2004, hereafter CHHK) regress mutual fund returns on lagged fund size and other observable characteristics. They find that a fund that is a log order of magnitude larger earns risk-adjusted returns that are 2 to 3 basis points per month lower.[1] If we were to interpret this difference as the causal effect of fund size on returns, we would conclude that diseconomies could not be masking a meaningful amount of performance persistence. First, we know that a fund that outperforms its peers by one log percentage point this year will be 2-5 log percentage points larger next year (one log percentage point from returns mechanically increasing assets, and the other 1-4 log percentage points from the inflow performance relationship).[2] Second, CHHK's estimate implies that a fund that is one log percentage point larger will earn returns that are about 0.003 log percentage points lower over the next 12 months. Combining these two estimates implies that a fund that outperforms its peers by one percentage point this year will suffer a 0.6-1.5 basis point penalty next year. In other words, if we interpret CHHK's estimate as an estimate of the causal effect of fund size on performance, the effect described in Berk and Green will cause us to underestimate an annual AR(1) coefficient by about 0.006 to 0.015. Given that we estimate the AR(1) coefficient range to be approximately 0.1, the estimated diseconomies of scale are approximately 10 times too small to meaningfully affect our views about manager skill.

However, it is important to note that the calculation above is not an appropriate test of the Berk-Green model. If fund size is endogenously related to expected future returns, in equilibrium, fund size will be uncorrelated with future returns, thereby frustrating standard approaches to estimate diseconomies of scale. Even if fund sizes are out of equilibrium, to the extent that

[1] More recently, Chen, Hong, and Kubik (2008) and Massa, Reuter, and Zitzewitz (2009) estimate similar partial correlations between fund size and fund returns, although neither of these papers is focused on the relation between fund size and returns.

[2] We take our range from the graphs of the inflow-performance relationship for the "young" (<2 years) and "old" (>10 years) funds in Chevalier and Ellison (1997), but these slopes have been replicated in many other studies.

larger funds have more skilled managers, the estimates in CHHK (and other studies) will underestimate the actual diseconomies of scale.[3] To identify the scale-return relationship, we require a natural experiment—something that causes an increase in assets for reasons that are related to future returns only through diseconomies of scale.

To shed new light on diseconomies of scale in asset management, we take a regression discontinuity approach. Our insight is that small changes in fund returns can have discontinuous impacts on fund flows through their impact on the fund's Morningstar rating. For example, as a fund's within-category Morningstar performance ranking increases from the 89^{th} percentile to the 90^{th} percentile, its Morningstar rating increases from four stars to five stars. Under the assumption that manager skill varies continuously across each of the Morningstar rating thresholds, we can use high frequency data on Morningstar performance rankings to estimate the impact of Morningstar rating thresholds on fund inflows. Then, because this source of fund inflows is uncorrelated with manager skill (and other factors affecting future returns), we can use it to identify the causal impact of fund size on fund performance. In other words, we use small deviations from the rational flows assumed by Berk and Green's model to test for diseconomies of scale.

We have four main empirical findings, based on monthly data from Morningstar that covers December 1996 to August 2009. First, in our first stage regressions, we show that mutual funds just above the threshold for a Morningstar rating receive incremental net flows over the next six months that are approximately 1.1 percentage points higher than mutual funds just below the threshold. Second, looking out over the next 6-24 months, we find little evidence of diseconomies of scale. Our reduced form estimates of the effect of these incremental inflows on

[3] Note that controlling for additional fund characteristics, as most studies comparing large and small funds do, does not change the fundamental prediction that the partial correlation between fund size and expected returns should be zero, even in the presence of scale diseconomies. When observable fund characteristics impact expected returns, investors should allocate flows such that expected returns are equal conditional on those characteristics.

returns are actually slightly *positive* during the first six months and slightly negative during the subsequent period, but few of the estimates are statistically different from zero.

Third, when we focus on subsamples of mutual funds (e.g., all large-cap equity funds or all municipal bond funds), we continue to find limited evidence of diseconomies of scale. In fact, in only seven of the 27 specifications that we estimate, can we reject the hypothesis that the IV estimates derived from our first stage and reduced form regressions are statistically different from those obtained via standard OLS regressions. Moreover, in only four of these seven cases are the IV estimates consistent with diseconomies of scale. Evidence of diseconomies of scale is strongest for sector funds, although we also find diseconomies of scale in the sample of all equity funds. In contrast, municipal bond funds exhibit positive economies of scale.

Finally, when we adjust estimates of performance persistence for diseconomies of scale, the median AR(1) estimate goes from 0.096 to 0.151, with the largest increases for large-cap equity and sector funds. However, the upper bound of the confidence interval for the corrected persistence coefficient ranges from 0.12 to 0.25 in the full sample of funds, and this bound is even tighter in some subsamples. Therefore, while we cannot rule out the possibility that Berk-Green effects are large for some categories of funds (like sector funds), we can rule out biases that are large enough to warrant significantly revising our interpretation of the performance persistence literature.

The remainder of our paper is organized as follows. In Section I, we describe the process that Morningstar uses to determine ratings, as well as our data. In Section II, we outline our empirical strategy and discuss our identifying assumption. In Section III, we show that share classes (and funds) with return patterns that place them just above a Morningstar ratings threshold receive higher flows than share classes (and funds) with return patterns that place them just below the same Morningstar ratings threshold. In Section IV, we use the findings from Section III to test for diseconomies of scale, both overall and within specific asset classes. In Section V, we adjust estimates of return persistence for diseconomies of scale. In Section VI, we conclude.

I. Morningstar Ratings and Fund Characteristics

Our identification strategy relies both on the discrete nature of Morningstar ratings, and on the fact that, because they are based on past returns, we can identify funds near ratings thresholds. In this section, we describe how Morningstar ratings are determined. We then describe our sample, and report summary statistics for funds with different Morningstar ratings.

A. Morningstar Ratings

Morningstar rates mutual fund share classes on a scale that ranges from one star (the lowest possible rating) to five stars (the highest possible rating). The rating assigned to each mutual fund share class depends on its relative performance within its Morningstar-determined investment category over the prior 3 years, 5 years, and 10 years, "after adjusting for risk and accounting for all sales charges."[4] Morningstar does not rate mutual fund share classes that are less than three years old.

For mutual fund share classes between the age of three and five years, the Morningstar rating depends entirely on its relative performance over the prior 36 months. "Within each Morningstar Category, the top 10% of funds receive five stars, the next 22.5% four stars, the

[4] Morningstar changed various detailed of its ratings process in June 2002. See Blume (1998) for a description of the rating system used from 1996-2002 and http://quicktake.morningstar.com/DataDefs/FundRatingsAndRisk.html for Morningstar's description of their current ratings process. The most significant change was that the number of Morningstar Categories increased from four on May 2002 (Domestic Equity, International Equity, Taxable Bonds, and Munipal Bonds) to 48 on June 2002, eventually growing to 81 in August 2009. The new Morningstar Categories better reflect actual investment styles (e.g., distinguishing domestic equity funds that focus on large-cap growth from those that focus on small-cap value). Morningstar also changed the exact method used for risk-adjusting returns, made the relative importance of 5 and 10-year performance depend on whether a fund had experienced style drift, and made several more minor changes.

middle 35% three stars, the next 22.5% two stars, and the bottom 10% receive one star."[5] Therefore, small differences in past returns, such as going from the 10th percentile to the 11th percentile, or from the 89th percentile to the 90th percentile, result in discrete changes in Morningstar ratings. These discrete changes are evident in Figure 1, in which we plot Morningstar ratings for all share classes that are less than 5 years old against Morningstar's risk-adjusted, within category return percentile. Figure 1 also provides graphical evidence that (residual) flows increase sharply around ratings thresholds.[6] (We present more formal evidence in Section III.)

*** FIGURE 1 HERE ***

For share classes between the age of 5 and 10, Morningstar determines separate ratings based on the prior 36 months and the prior 60 months, and "averages" the underlying ratings to calculate an overall integer rating. In Figure 2, we show how relative performance over the prior 36 and 60 months maps into a share classes' overall rating. The pattern reveals that Morningstar calculates a fund's overall rating as a 60-40 average of the 5-year and 3-year integer ratings, causing it to "round up" when the better performance is over the longer horizon.[7] For example, a share class with a 36-month return that puts it at the 89th percentile (four stars) and a 60-month return that puts it at the 90th percentile (five stars), receives an overall rating of five stars. In contrast, a share class with a 36-month return that puts it at the 90th percentile (five stars) and a 60-month return that puts it at the 89th percentile (four stars), receives an overall rating of four stars. To the extent that we're willing to assume that the managers of these two funds are similarly skilled (conditional on current assets under management), and that the five-star fund receives

[5] See http://quicktake.morningstar.com/DataDefs/FundRatingsAndRisk.html.
[6] The residual flows in Figure 1 come from versions of the baseline inflow regression in Section III that omit the Morningstar within-category percentile ranking and discontinuity dummy variable.
[7] After June 2002, Morningstar began giving older history less weight when funds had experienced style drift. To make Figure 2 more transparent, we exclude these funds from the picture. Depending on how much style drift was experienced and when it was experienced, a fund's 3-year history can receive more than 50 percent of the weight, causing the rounding to occur in the other direction.

higher residual flows, we can study the impact of these incremental flows on future returns.

*** FIGURE 2 HERE ***

While the staircase boundaries between overall ratings may strike readers as an unusual methodological choice by Morningstar, it is helpful from the perspective of our research, since this approach increases the number of funds that are very close to a rating boundary. For share classes that are more than 10 years old, Morningstar's overall rating depends on the average of the 3-year, 5-year, and 10-year ratings. For these share classes, thresholds between ratings are conceptually similar to those in Figure 2. However, because these thresholds relate to three underlying ratings, they must be plotted in three dimensions.

B. Sample Construction

To study the impact of mutual fund flows on mutual fund returns, we obtain data from Morningstar Principia CDs. Our sample consists of all open-end mutual funds that have at least one share class rated by Morningstar. Because Morningstar does not rate share classes that are less than three years old, mutual funds enter our sample when their oldest share class reaches three years of age. The fact that we only study funds in the time period in which they appear on a Morningstar CD limits the influence of incubation bias (Evans (2010)) on our results. While incubation bias might help explain why funds appearing on a Morningstar CD for the first time have average Morningstar ratings about a quarter point above older funds, our analysis of future inflows and performance uses only non-back-filled data. Consequently, our estimates of scale diseconomies should be unaffected by incubation bias.

Our data begin in December 1996 and end in August 2009.[8] Because mutual funds can earn different Morningstar ratings and experience different inflows in their different share

[8] Currently, we lack data for 12 of the 36 months between January 1997 and December 1999. The missing months are January 1997, February 1997, April 1997, May 1997, July 1997, August 1997, October 1997, November 1997, January 1998, July 1998, January 1999, and November 1999.

classes, we use share classes as the unit of observation in our initial analysis of inflows. As any scale diseconomies would occur at the fund (portfolio) level, however, in most of our analysis we aggregate variables to the fund level, weighting each share class in proportion to its assets under management in the prior month. In practice, the exact approach we take to weighting share classes has little influence on the results because the average fund gets 84 percent of its assets from its largest share class.

Finally, because Morningstar within-category percentile rankings do not distinguish between actively and passively managed mutual funds, we include the share classes of index funds in our sample when calculating within-category percentile rankings. However, we exclude index funds from all inflow and return regressions.

C. Summary Statistics

In Table 1, we report fund-level summary statistics for the full sample of 491,863 fund-month observations. We also use asset-weighted average Morningstar ratings to assign fund-level ratings, and report summary statistics for each fund-level rating category. Looking across these categories, we see that funds with higher ratings tend to be larger and come from larger families. Funds with higher ratings also tend to charge lower average fees (both in month t and month t+12), offer fewer share classes, and are less likely to charge a sales load. Of course, differences in fees and sales loads follow, at least in part, from the fact that Morningstar ratings are based on returns measured net of fees and loads.

*** TABLE 1 HERE ***

The most interesting differences between funds with higher and lower ratings involve future flows and future returns. Consistent with investors either responding to Morningstar ratings or to the return histories underlying them, we find that funds with higher ratings receive higher net flows over the next 24 months. Relative to other funds in their Morningstar category, the typical five-star funds grows by 23 log percentage points over this period, while the typical one-star funds shrinks by 18 log percentage points. The results presented later imply that of this 41 log percentage point difference, about 9 log percentage point represents a causal effect of Morningstar on flows, with the remainder being due to investors responding directly to observable fund characteristics included in Morningstar's ratings (e.g., past returns, risk, and loads), other observable characteristics correlated with the ratings (e.g., low expenses), or unobservable characteristics correlated with the ratings (e.g., marketing efforts).[9]

Consistent with prior work on the predictive power of Morningstar ratings (e.g., Blake and Morey (2009)), we find that one-star funds underperform other funds over the next 24 months, but find little difference in the future performance of other funds. The fact that 5-star and 2-star funds perform approximately as well in the future despite 5-star funds experiencing greater inflows does not necessarily imply the absence of scale diseconomies, however. In the Berk-Green model, the 5-star funds attract more inflows because they have more skilled managers, and this skill allows the funds to match the 2-star funds' returns despite managing more assets. For a test for scale diseconomies to be valid in the Berk-Green model, it needs to exploit a source of variation in inflows that is not produced by or correlated with manager skill. Fortunately, the discontinuities in the Morningstar ranking function generate this type of variation.

II. Overview of RD and our Identification Strategy

In order to measure the causal impact of fund size on fund performance, we must identify flows that are uncorrelated with manager skill. We use a regression discontinuity approach that exploits the fact that mutual funds with past returns immediately above a Morningstar rating

[9] Prior work examining the relationship between fund inflows and Morningstar ratings uses observable variables to control for these factors. For example, when they include fund fixed effects, Del Guerico and Tkac (2008) continue to find a positive association between stars and flows.

threshold receive a discretely higher rating than mutual funds with past returns immediately below the threshold. To the extent that investors place positive weight on Morningstar ratings, funds with risk-adjusted returns immediately above a ratings threshold are likely to receive significantly more inflows than funds with risk-adjusted returns immediately below the threshold.[10]

Our analysis proceeds in two stages. In the first stage regressions, we estimate the impact of rating thresholds on future flows. Then, we use reduced form regressions to estimate the impact of rating thresholds on future returns. The identifying assumption is that while inflow will vary sharply at each threshold, the other fund characteristics that might be related to future returns will vary continuously.[11] Under this assumption, our first stage and reduced form estimates allows us to measure the extent of diseconomies of scale.

More formally, our analysis focuses on actively managed mutual funds just above and below each rating threshold. For example, with respect to the threshold between four stars and five stars, our first stage regression predicts log net flows as function of the within-category percentile ranking used to determine Morningstar ratings, a dummy variable that indicates whether the within-category percentile ranking the share class *i* of fund *j* in month *t* is above the five star rating threshold, and controls, including multiple controls for past performance and past flows.

$$\text{Flow}_{i,j,t+1} = \delta_{1st}\ \text{threshold}_{i,j,t} + \lambda_{1st}\ \text{ranking}_{i,j,t} + \beta_{1st}\ \text{X}_{i,j,t} + \varepsilon_{i,j,t} \qquad (1)$$

[10] Technically, a firm like Morningstar does not need to exist in the Berk-Green model, as all investors use risk-adjusted past returns to directly infer manager skill, and thus there are no threshold effects in inflow-return relationship. Our thought experiment can be thought of as a Berk-Green model where many investors observe only a discrete number of Morningstar stars, and make inferences about managers' ability based on the average characteristics of funds with that rating. In this model, there would be flow discontinuities at rating boundaries, funds just over a boundary would have similar managerial skill to those just beneath it, and thus the extra assets would cause them to underperform. Average performance would still be equalized across the different Morningstar ratings though. Savvy investors who did observe returns would invest in funds that just below ratings boundaries, so for the flow discontinuities we observe in the data to exist, their numbers would have to be limited..

[11] Imbens and Lemieux (2008) and Lee and Lemieux (2009) provide excellent overviews of the regression discontinuity approach.

where δ_{1st} measures the discontinuous flow effect associated with the ratings threshold.[12]

In many RD settings, the "forcing variable", which determines whether an observation is above or below the threshold, is exogenous.[13] In our setting, the within-category percentile ranking is not exogenous. However, our identifying assumption is that, because all managers are trying to maximize relative performance, manager skill will vary continuously across the threshold for a higher rating. In other words, while we allow for the possibility that managers with slightly higher returns are slightly more skilled, our identification strategy assumes that skill does not jump in a discontinuous way at the threshold between ratings. The fact that thresholds for different Morningstar ratings depend on within-category performance rankings over as many as three investment horizons, increases our confidence that the distribution of manager skill is smoother than the distribution of Morningstar ratings.

To estimate fund-level flows, we focus on the discontinuity measure for the fund's largest share class. Then, because the estimated coefficient on the fund-level discontinuity measure is positive and statistically significant, we estimate a reduced form regression

$$\text{Return}_{i,j,t+1} = \delta_{\text{rf}} \ \text{threshold}_{\text{i, j, t}} + \lambda_{\text{rf}} \ \text{ranking}_{\text{i, j, t}} + \beta_{\text{rf}} \ \text{X}_{\text{i, j, t}} + \eta_{i,j,t} \qquad (2)$$

where δ_{rf} measures the causal effect of ratings thresholds on returns. Under the assumption that the causal effect of ratings thresholds on flows is unrelated to differences in manager skill, δ_{rf} will capture any diseconomies of scale. Finally, we can estimate the causal impact of flows on returns as the ratio of δ_{rf} to δ_{1st}. The more negative this IV estimate, the larger the implied diseconomies of scale.

[12] Following the advice in Imbens and Lemieux (2008), we experimented with more flexible approaches to controlling for the ranking variable, but found that the results varied little from the local linear approach.

[13] For example, to study the impact of the Sarbanes-Oxley Act on firms costs and earnings, Iliev (2010) exploits the fact that U.S. firms with a public float below $75 million in 2002, 2003, or 2004 were allowed to delay compliance with Section 404 until well after the November 2004 date on which slightly larger firms were required to comply.

III. Impact of Morningstar Ratings on Flows

In this section, we present evidence that Morningstar ratings have a causal impact on investor flows. Because our identification strategy exploits the discreteness of Morningstar ratings, and because different share classes of the same mutual fund can receive different Morningstar ratings, we begin by studying the impact of Morningstar ratings on net flows at the share class level. Consistent with equation (1), our general approach is to regress log net flows of share class *i* in month *t+1* on its Morningstar percentile ranking in month *t*, which is our local linear control, and a dummy variable that indicates whether share class *i* is above the threshold for a particular rating in month *t*. Under the assumption that manager skill varies continuously across the rating threshold, the dummy variable will capture incremental flows into the higher-rated fund that are uncorrelated with manager skill.

To quantify these discontinuous flow effects, we estimate separate regressions for each rating thresholds (i.e. one star versus two stars, …, four stars versus five stars), and a pooled regression that combines all four thresholds. In each case, the sample is restricted to those share classes that are within five percentiles of a rating threshold. For example, when we focus on the threshold between four and five stars, we restrict the sample to share classes with Morningstar rankings between the 85^{th} and 95^{th} percentiles. We further restrict the sample to actively managed funds by excluding any fund that Morningstar identifies as an index fund.

In addition to the variables that we report in Table 2, "Baseline" regressions control for the lagged log size of the share class, portfolio, and family, portfolio turnover, expense ratio, and the presence of loads (front, deferred, and trailing). Because our sample includes the full range of Morningstar categories (i.e., includes large-cap equity, sector funds, corporate bond funds, etc.), we also include a separate fixed effect for each Morningstar category each month, thereby comparing each fund to other funds with the same investment style in the same month. In the regressions with "Additional Controls", we supplement the Morningstar percentile ranking variable with controls for Morningstar's measure of risk-adjusted returns, lagged log returns from *t-12* to *t-1*, *t-24* to *t-13*, and *t-36* to *t-25*, and lagged log inflows from *t-12* to *t-1* and *t-3* to *t-1*. Because mutual funds with multiple share classes can appear multiple times in the same month, we cluster standard errors on fund.[14]

*** TABLE 2 HERE ***

The estimated coefficients on the discontinuity dummy variable are positive and statistically significant for each of the four ratings thresholds, and for the regression that includes all four ratings thresholds. In the baseline regressions, the estimates range from 0.337 log percentage points at the threshold between 1 star and 2 stars (significant at the 5-percent level) to 0.946 log percentage points at the threshold between 4 stars and 5 stars (significant at the 1-percent level). When we include additional controls for past returns and past flows, the estimated coefficients decline, but only slightly. For example, within the stacked regression, the estimated coefficient falls from 0.518 to 0.432 log percentage points, but remains statistically significant at the 1-percent level. In other words, share classes that are just above the Morningstar ratings threshold *this month* receive an additional 0.432 log percentage points in net flow *next month*, compared to share classes that are just below the threshold.

If a share class were to maintain its Morningstar percentile ranking just above the threshold for an entire year, this would translate into an additional annual net flow of 5.154 log per-

[14] Given the number of regressions that are estimated in the study and the category-time fixed effects included in the models, it was not practical to cluster by both fund and month in every regression (e.g., following Petersen (2009)). In a unreported estimations of a few of our models, we found that adding a time dimension of clustering did not affect standard errors meaningfully, which is likely because the regressions include time fixed effects and because Morningstar ratings by design do not cluster within time periods. We also experimented with clustering by family instead of by fund and found that it did not meaningfully affect results.

centage points. However, this persistence would also call into question our assumption that manager skill varies continuously across rating thresholds. To shed light on this assumption, we change the dependent variable from log net flows in month *t+1* to log net flows in month *t* in the last two columns of Table 2. When we focus on current-month flows instead of next-month flows, only five of the ten estimated coefficients on the discontinuity dummy variable are positive, and only one is statistically significant from zero (at the 10-percent level). These results strongly suggest that there is no discontinuity in inflow-producing fund characteristics at the start of month *t*, and thus, crucially for our identification strategy, that the discontinuity in flows in month *t+1* can be plausibly attributed to the Morningstar rating. Overall, we view the results in Table 2 as providing the "first stage" that we need to study the causal impact of flows on performance.

*** TABLE 3 HERE ***

Of course, to test for diseconomies of scale, we need to study the impact of fund-level flows on fund-level performance. In Table 3, we study the impact of Morningstar ratings on log net flows at the fund level. Because many funds have more than one share class, we need a measure of incremental flows that is aggregated across all of fund *j*'s share classes. Most funds have a main share class that contains most of the assets (often the "A" class for load funds or the "Investor" class for no-load funds). Because other share classes have the same return gross of fees and expenses, differences in returns (and Morningstar percentile rankings) reflect differences in fees and expenses. The largest share classes' Morningstar rating is generally the one that is marketed to potential investors, as other share classes either have lower ratings due to higher fees (e.g., B, C, and Service share classes) or impose restrictions on who can purchase them (e.g., Institutional share classes). Our approach is to focus on the discontinuity and ranking variables for fund *j*'s largest share class.[15]

The estimated coefficients in the first two columns of Table 3 are qualitatively similar to those in Table 2, with slightly smaller magnitudes because the denominator is now fund assets rather than share class assets. Seven of the ten coefficients are statistically significantly different from zero at conventional levels, with the lack of a discontinuity between one and two stars being the major exception. Importantly, we continue to find little evidence of a discontinuity in current month flows.

*** FIGURE 3A, FIGURE 3B, AND TABLE 4 HERE ***

As in Table 2, the coefficient estimates from the regressions with additional controls are slightly smaller than the baseline estimates, but these differences are never statistically significant. In later tables, we focus primarily on stacked regressions that include the full set of return and flow control variables. Figure 3A provides graphical evidence of the discontinuity in future inflows at each rating threshold.[16] Figure 3B provides graphical evidence of the lack of the discontinuity in current month inflows. Finally, in Table 4, we switch our focus from inflows to other mutual fund characteristics. The fact that we find little evidence of discontinuities in our control variables at Morningstar rating thresholds further increases our confidence in our identification strategy.

IV. Testing for Diseconomies of Scale

Having shown that mutual funds receive incremental flows when they pass across Morningstar rating thresholds, we now use these incremental flows to test for diseconomies of scale. We begin by estimating first stage and reduced form regressions on the full sample of mutual

---

[15] As an alternative, we experimented with taking the highest Morningstar rating and ranking variable across all share classes, on the assumption that this would be the rating marketed to investors, and found very similar results.

[16] Residual flows in Figures 3A and 3B are estimated from the baseline specification in Table 4 that omit the Morningstar within-category percentile ranking and discontinuity dummy variable.

funds over longer investment horizons. Then, because diseconomies of scale may differ across asset classes, we re-estimate first stage and reduced form regressions for different subsamples of mutual funds. Finally, we compare the IV estimates implied by our first stage and reduced form regressions to the diseconomies of scale estimates implied by standard OLS regressions.

A. Evidence from the Full Sample of Mutual Funds

In Table 5, we extend the analysis in Table 4 along two dimensions. First, rather than estimating first stage regressions focused on log net flows in month *t+1*, we estimate first stage regressions focused on cumulative log net flows over different investment horizons. Our goal is to measure the long-term impact of rating thresholds on fund flows. Second, for each first stage regression of log net flows on the discontinuity variable (and full set of controls), we estimate a matching reduced form regression of log net returns on the discontinuity variable (and full set of controls). Given our identification assumption that flows associated with rating thresholds are uncorrelated with manager skill, these flows should only impact fund returns through diseconomies of scale. The reduced form regressions are intended to measure this impact. Again, the sample is restricted to actively managed funds, and all standard errors are clustered on fund.

*** TABLE 5 HERE ***

When we restrict attention to net flows in *t+1* and net returns in month *t+1*, we find little evidence of diseconomies of scale. In both the stacked regressions and the regressions focusing on the discontinuity between four and five stars, the estimated incremental flows and returns are both positive, although the coefficient on the discontinuity variable in the return regression is not statistically different from zero.

When we focus on cumulative log net flows beyond month *t+1*, we continue to find that Morningstar rating thresholds are associated with significant incremental flows. For example, in the stacked regressions, the incremental flows associated with the discontinuity variable (measured in month *t*) are 1.12 log percentage points through month *t+6*, 1.55 log percentage points through month *t+12*, and 2.29 log percentage points through months *t+24.* These estimates imply that while the effect of an extra Morningstar star in the ranking disseminated during month *t+1* is strongest in month *t+1*, the effect of the extra star persists beyond the initial month. There are numerous mechanisms that could produce this effect. Investors may make an initial investment in month *t+1* based on the current-month Morningstar rating, and that initial investor may affect the placement of subsequent investments. Investors may also make investment decisions based on an accumulation of signals received over several months. Regardless of the mechanism, our findings about the timing of investor reactions to Morningstar are consistent with prior findings on the timing of investor reactions to media mentions or advertising (e.g., Reuter and Zitzewitz (2006)).

When we examine returns after month *t+1,* we continue to find little evidence that the extra inflows from the threshold effects affect future returns. The strongest evidence of scale diseconomies for the stacked regressions appears in month *t+21*, where funds above a threshold receive flows totally 2.08 log percentage points of assets and underperform by 9 basis points. The strongest evidence from the regressions include use only the 4/5 star boundary is in month t+15, where incremental flows of 2.29 log percentage points are accompanied by underperformance of 12 basis points. In neither case, however, is the underperformance close to statistically significant at conventional levels. Figures 4A and 4B graph the contemporaneous and cumulative flow and return effects presented in Table 5 as a function of time.

*** FIGURES 4A AND 4B HERE ***

B. Evidence from Different Investment Categories

Although we find limited evidence of diseconomies of scale within the full sample of mutual funds, we might reasonably expect the degree of diseconomies of scale to vary across asset classes. For example, CHHK find their strongest evidence of diseconomies of scale among small-cap equity funds. More generally, we might expect the strongest diseconomies of scale in asset classes with less liquidity or where the inflows experienced by a typical fund can be large relative to the investment options available (e.g., small-cap equity, sector funds, and municipal debt).

In Table 6, we re-estimate the first stage and reduced form regressions in Table 5 for different sets of mutual funds over four different investment horizons. We use the Morningstar category variable to create the following seven non-overlapping subsamples of mutual funds: large-cap equity; mid-cap equity; small-cap equity; sector funds; international equity; taxable bonds; municipal bonds. (We exclude a small set of funds that do not fall into these categories, such as balanced funds, commodities funds, and target date retirement funds.) We also create an "All equity" sample that combines large-cap equity, mid-cap equity, small-cap equity, sector funds, and international equity. We focus on cumulative log flows and log returns through month *t+6*, *t+12*, *t+18*, and *t+24*.

*** TABLE 6 HERE ***

The estimated flow and return effects in the first column of Table 6 are for all funds, and match those reported in Table 5. The other columns focus on different samples of funds. Looking across eight subsamples, we see that the estimated flow effects are almost always positive, but also that the standard errors tend to be much larger than in the full sample. The evidence that Morningstar rating thresholds impact flows is strongest for sector funds, international equity funds, and municipal bond funds. Flows effects are also statistically significant when we focus on the "All equity" sample.

Turning to the reduced form regressions for the seven subsamples, we see that 16 of the 21 estimated coefficients are negative, but only six negative coefficients (and four positive coefficients) are statistically different from zero. Sector funds exhibit the strongest evidence of diseconomies of scale, while municipal bond funds exhibit positive economies of scale. Estimated diseconomies of scale are also negative and statistically significant (at the 10-percent level and below) within the "All equity" sample.

C. A Comparison of IV and OLS Estimates of Diseconomies of Scale

Our regression discontinuity approach allows us to directly estimate the causal impact of rating thresholds on flows and the causal impact of rating thresholds on returns. However, we are ultimately interested in measuring the causal impact of flows on returns. To obtain an IV estimate of the diseconomy of scales for a particular subsample of mutual funds and investment horizon, we scale the estimated coefficient from the reduced form by the estimated coefficient from the first stage.[17] For example, for the full sample of funds through month *t+24*, one log percentage point in incremental flows is associated with incremental returns that are 0.04 log percentage points lower. (The IV estimate of -0.04 equals -0.08 divided by 2.29.)

Table 7 reports IV estimates for different sets of mutual funds and investment horizons alongside the first stage and reduced form estimates (from Table 6). Fifteen of the 25 IV estimates are negative. Among the seven mutually exclusive categories of funds, ten of the 19 IV estimates are negative. However, the standard errors associated with many of the estimates are quite large, particularly at longer time horizons or in categories with smaller inflow effects.

[17] The fact that the estimated flow effects are negative for small-cap equity funds for months t+12 and t+24 prompts us to drop these subsample-horizon combinations from Tables 7 and 8.

*** TABLE 7 HERE ***

In the last several columns of Table 7, we compare our IV estimates to the partial correlation between fund size and fund returns that we estimate within our sample using standard OLS regressions. Specifically, the partial correlation for each asset class and investment horizon is estimated as the coefficient on fund size in a regression of future returns on the variables listed under fund characteristics in Table 1, a control for past-12-month log returns, and a separate fixed effect for each Morningstar category each month. Consistent with Berk and Green's prediction, the IV estimates tend to be much more negative than the OLS estimates. The average IV estimate is -0.056 versus an average OLS estimate of -0.0015. Similarly, the median IV estimate is -0.027 versus a median OLS estimate of -0.0011. However, because of the larger standard errors on the IV estimates, we can only reject the hypothesis that the OLS and IV estimates are equal in the seven cases where the IV estimate is statistically significant from zero (the *p*-values of the Hausman tests range from 0.01 to 0.07). In four of these seven cases, the IV estimate is more negative than the OLS estimate. However, in the other three cases, the IV estimate is more positive than the OLS estimate. We turn now to the final, important, question of whether our estimated diseconomies of scale have an economically significant impact on estimates of performance persistence.

V. Adjusting Performance Persistence for Diseconomies of Scale

Berk and Green (2004) show that the combination of diseconomies of scale with endogenous fund flows will cause researchers to underestimate the true degree of performance persistence in the mutual fund industry. In Table 8, we adjust measures of performance persistence for the causal impact of flows on performance. To begin, we estimate standard OLS regressions that predict future returns from past 12-month returns, as well as the category-by-month fixed effects and control variables included in Tables 2-5. We report the estimated coefficient on the past return measure, and its standard error, in the first set of columns. Within the full sample of funds, assuming a 12-month horizon, the estimated coefficient is 0.100. However, this estimate will be downwardly biased if past returns attract incremental flows and there are diseconomies of scale.

*** TABLE 8 HERE ***

We use a similar set of regressions to predict the impact of past returns on log net flows. And, we report the estimated coefficient on the past return measure, and its standard error, in the second set of columns. For the full sample of funds, again assuming a 12-month horizon, an additional log percentage points in past returns is associated with 0.92 log percentage points in additional flow. The larger the diseconomies of scale associated with these additional flows, the greater the downward bias in the persistence coefficient.

Finally, we use the diseconomies of scale that we estimated in the prior section to adjust the persistence coefficient. Specifically, we estimate the "Corrected persistence coefficient" as the "Persistence coefficient" minus the "Flow coefficient" times the "Causal effect of flows" from Table 7.[18] For the full sample of funds and a 12-month horizon, adjusting for diseconomies of scale increases the persistence correlation from 0.100 to 0.121. If we use the standard errors from Table 7 to construct a 95% confidence interval for the corrected persistence coefficient, we find that it ranges from 0.046 to 0.196. In this case, we cannot reject the hypothesis that the "Persistence coefficient" and "Corrected persistence coefficient" are equal. However, we can reject at the 1-percent level the hypothesis that the corrected persistence coefficient is equal to

[18] The "Persistence coefficient" is the increase in expected next-period log percentage point return associated with a one log percentage point increase in 12-month past returns. The "Flow coefficient" times the "Causal effect of flows" is the expected log percentage point decrease in next period's return based on the expected log percentage point increase in flows times the expected diseconomies of scale associated with the incremental flow. The "Corrected persistence coefficient" removes the impact of the return-induced flows on the expected log percentage point increase in next period's return.

0.42, which is the value implied by Berk and Green's calibration.[19] The same is true for the sample of "All equity" funds, for which the corrected persistence coefficient is 0.193, and the upper bound of the 95% confidence interval is 0.300.

Unfortunately, within many subsamples of funds, the corrected persistence coefficient is imprecisely estimated. As we possess data on virtually every U.S. mutual fund in operation between December 1996 and August 2009, it may prove difficult to significantly increase the statistical power of our tests. Thus, while we conclude that correcting for scale diseconomies would not significantly affect our view of performance persistence for all mutual funds, we cannot rule out the possibility that doing so might affect our views for certain subsamples of funds, particularly sector funds.

VI. Conclusion

The Berk-Green model poses a serious challenge to the commonly held view that mutual fund managers are unskilled and mutual fund investors are unsophisticated. The prediction that more skilled managers will manage larger funds also poses a serious challenge to existing evidence on diseconomies of scale. We use a regression discontinuity approach to determine how important the endogeneity problem implied by the Berk-Green model is in practice. Specifically, we use the discrete changes in flows associated with discrete changes in Morningstar ratings to identify flows that should only impact fund returns through diseconomies of scale. On the one hand, the point estimates of scale diseconomies implied by these plausibly exogenous flows are larger than those implied by cross-sectional comparisons of large and small funds. This is consistent with Berk and Green's prediction that more-skilled managers will manage larger funds.

[19] In Berk and Green's calibration exercise, managers' skill, defined as the annualized alpha they would achieve in the absence of scale diseconomies, is distributed normally with mean 6.5% and standard deviation 6%. Given the within-objective standard deviations of returns of 5.1, 7.4, and 10.0 percent for the 6, 12, and 24 month horizons, respectively, in the absence of scale diseconomies, this distribution of alpha would imply (within-objective) persistence coefficients of 0.21. 0.42, and 0.84 for the 6, 12, and 24-month horizons, respectively.

On the other hand, even these larger, and arguably better identified, estimates of scale diseconomies are not large enough to significantly change our views about the extent of performance persistence.

REFERENCES

Berk, Jonathan, and Richard Green, 2004, "Mutual Fund Flows and Performance in Rational Markets," *Journal of Political Economy* 112, 1269–1295.

Berk, Jonathan, 2005, "Five Myths of Active Portfolio Management," UC-Berkeley mimeo. Available at http://faculty.haas.berkeley.edu/berk/papers/myth.pdf.

Blake, Christopher and Matthew Morey, 2003, "Morningstar Ratings and Mutual Fund Performance," *Journal of Financial and Quantitative Analysis* 35, 451-483.

Blume, Marshall, 1998, "An Anatomy of Morningstar Ratings," *Financial Analysts Journal* 54, 19-27.

Carhart, Mark, 1997, "On Persistence in Mutual Fund Performance," *Journal of Finance* 52, 57–82.

Chevalier, Judith, and Glenn Ellison, 1997, "Risk Taking by Mutual Funds as a Response to Incentives," *Journal of Political Economy* 105, 1167–1200.

Chen, Joseph, Harrison Hong, Ming Huang, and Jeffrey Kubik, 2004, "Does Fund Size Erode Mutual Fund Performance? The Role of Liquidity and Organization," *American Economic Review* 94, 1276–1302.

Chen, Joseph, Harrison Hong, and Jeffrey Kubik, 2008, "Outsourcing Mutual Fund Management: Firm Boundaries, Incentives and Performance," Working Paper.

Del Guercio, Diane, and Paula Tkac, 2008, "Star Power: The Effect of Morningstar Ratings on Mutual Fund Flow," *Journal of Financial and Quantitative Analysis* 43, 907–936.

Evans, Richard, 2010, "Mutual Fund Incubation," *Journal of Finance* (forthcoming).

Iliev, Peter, 2010, "The Effect of SOX Section 404: Costs, Earnings Quality, and Stock Prices," Journal of Finance 65, 1163–1196.

Imbens, Guido, and Thomas Lemieux, 2008, "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics* 142, 615-635.

Jensen, Michael, 1968, "The Performance of Mutual Funds in the Period 1945–1964," *Journal of Finance* 23, 389–416.

Lee, David, and Thomas Lemieux, 2009, "Regression Discontinuity Designs in Economics," NBER Working Paper #14723.

Massimo, Massa, Jonathan Reuter, and Eric Zitzewitz, 2009, “When Should Firms Share Credit with Employees? Evidence from Anonymously Managed Mutual Funds ,” *Journal of Financial Economics* 95, 400–424.

Reuter, Jonathan, and Eric Zitzewitz, 2006, “Do Ads Influence Editors? Advertising and Bias in the Financial Media,” *Quarterly Journal of Economics* 121, 197-227.

Sirri, Erik, and Peter Tufano, 1998, “Costly Search and Mutual Fund Flows,” *Journal of Finance* 53, 1589–1622.

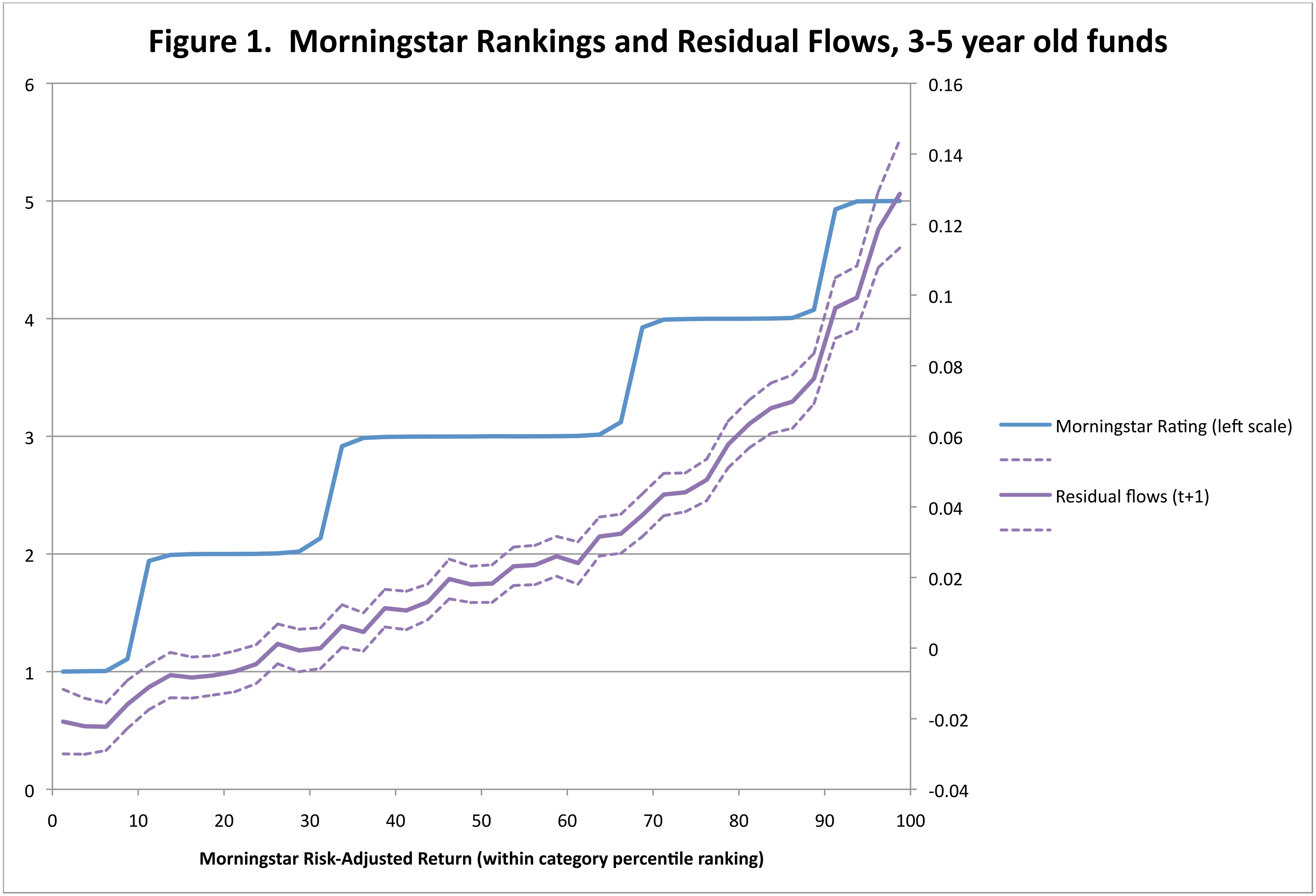


Figure 1. Morningstar Rankings and Residual Flows, 3-5 year old funds
0
1
2
3
4
5
6
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
-0.02
-0.04
0
10
20
30
40
50
60
70
80
90
100
Morningstar Risk-Adjusted Return (within category percentile ranking)
Morningstar Rating (left scale)
Residual flows (t+1)

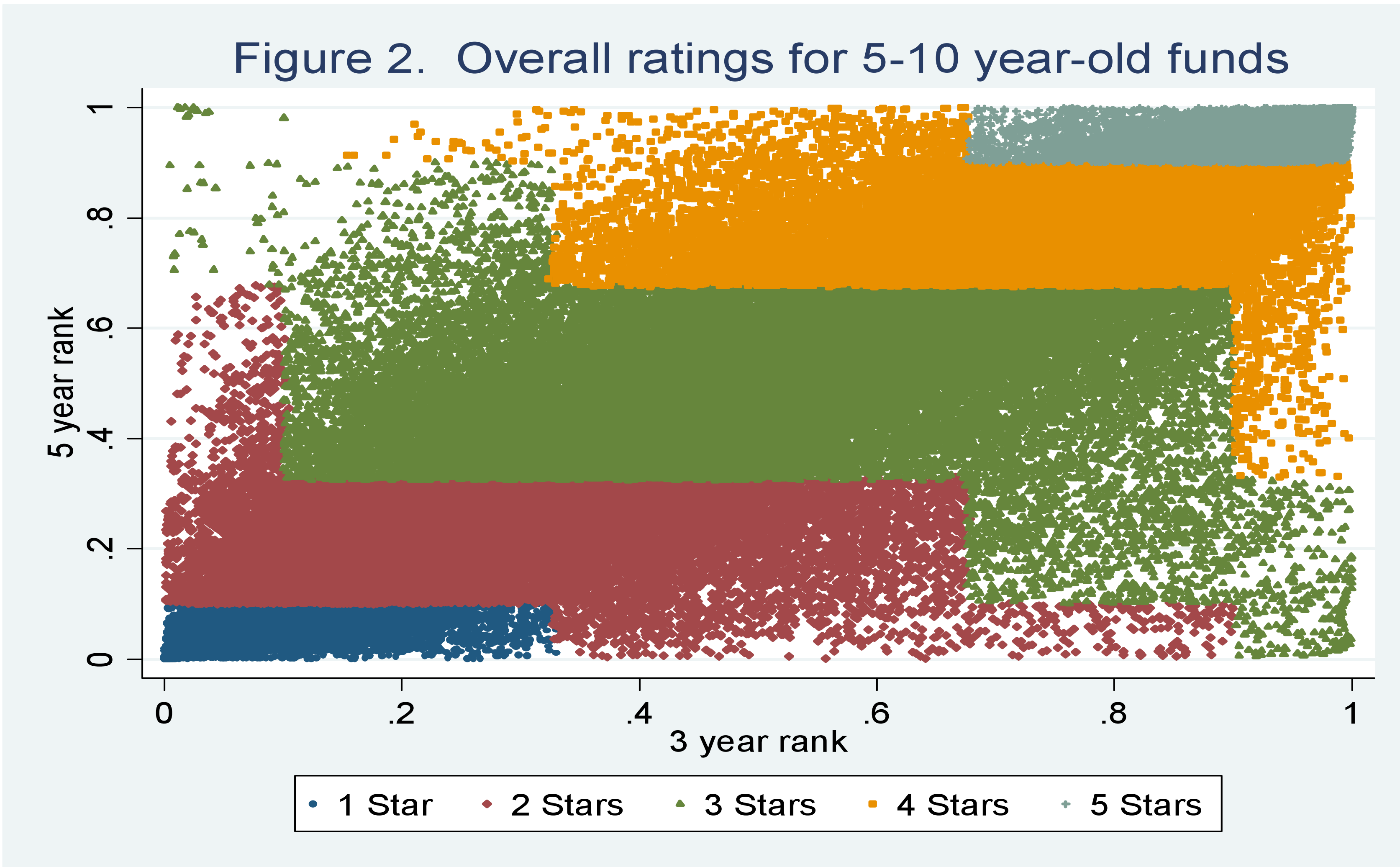


Figure 2. Overall ratings for 5-10 year-old funds
5 year rank
0
.2
.4
.6
.8
1
3 year rank
0
.2
.4
.6
.8
1
1 Star
2 Stars
3 Stars
4 Stars
5 Stars

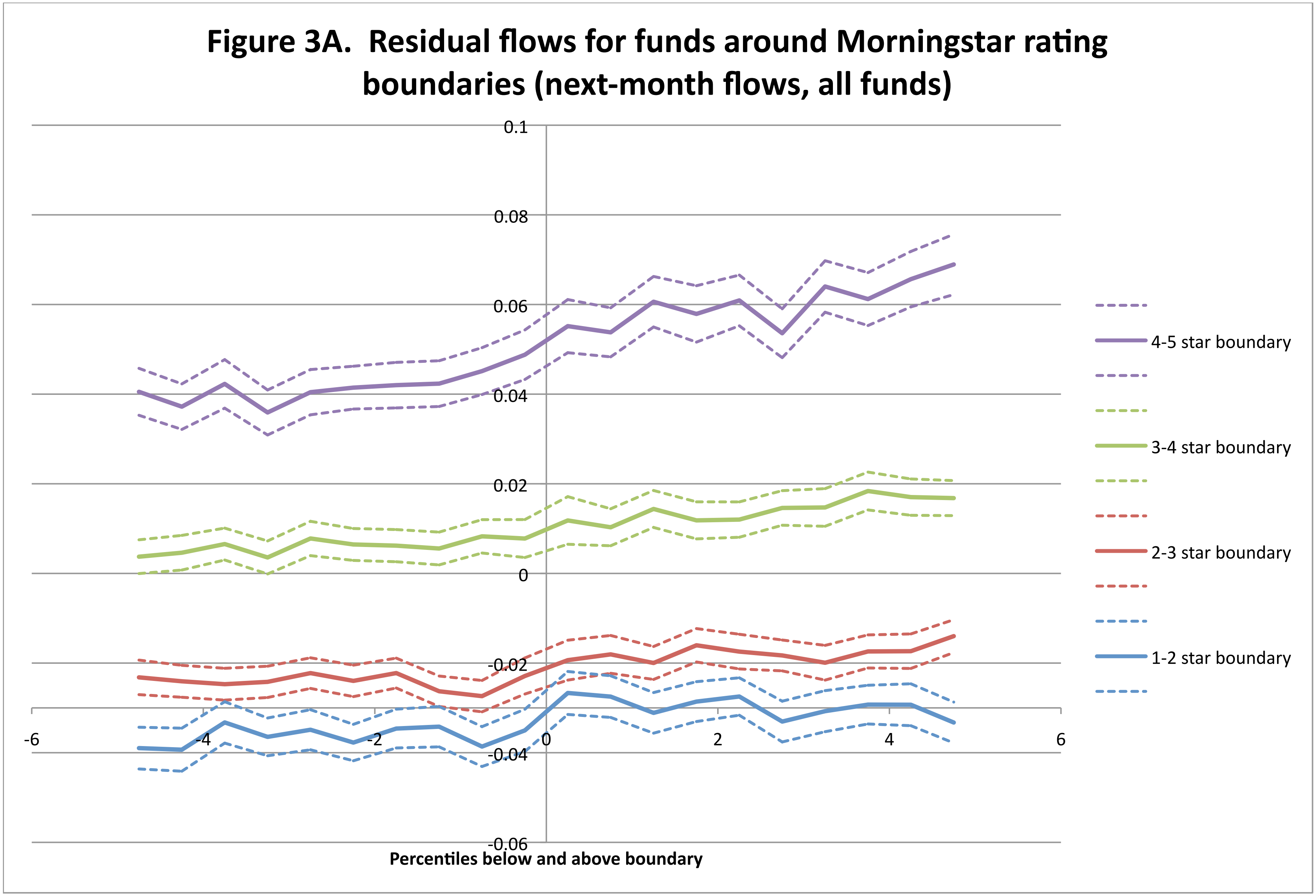
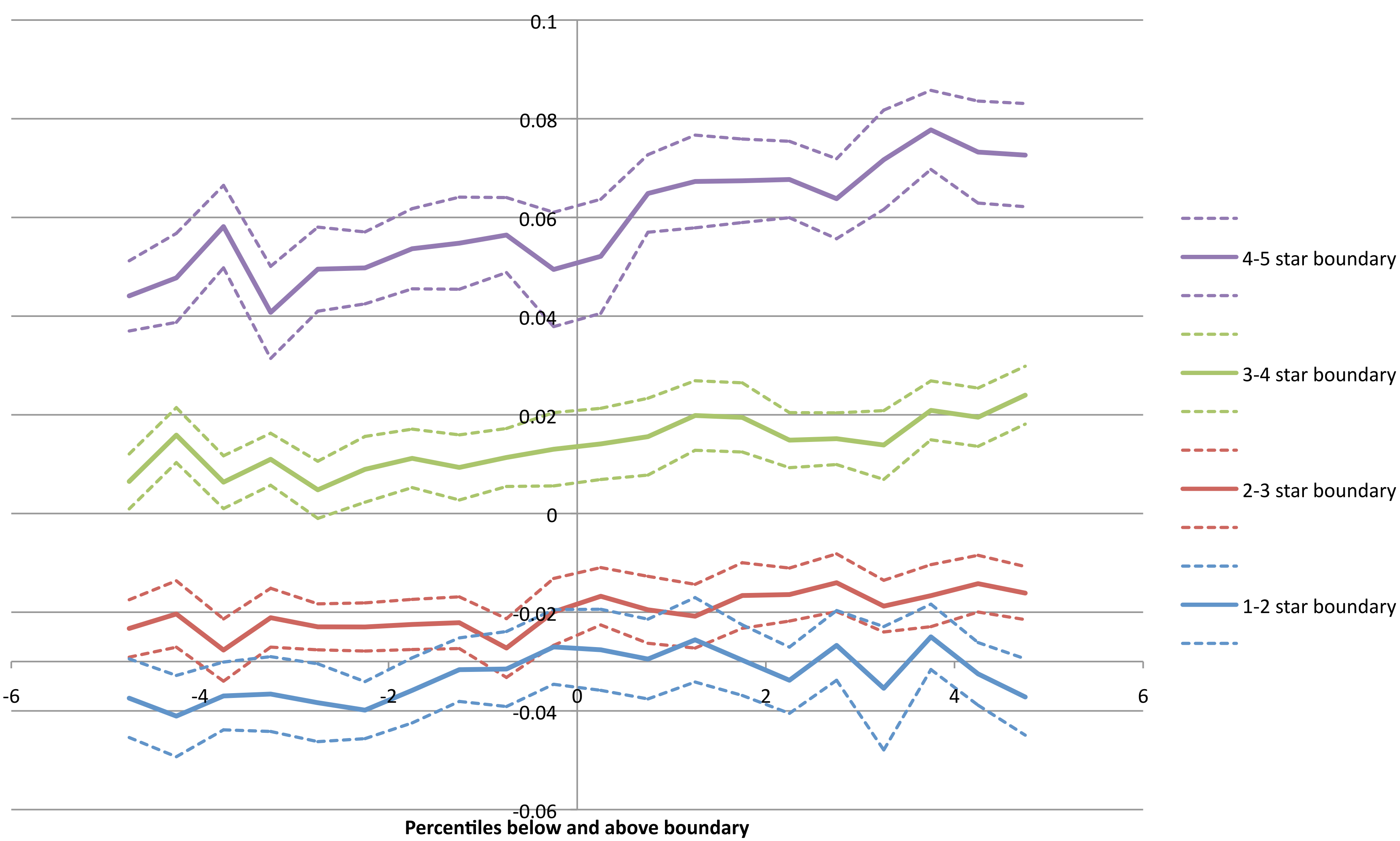


Figure 3A. Residual flows for funds around Morningstar rating boundaries (next-month flows, all funds)
0.1
0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-6
-4
-2
0
2
4
6
Percentiles below and above boundary
4-5 star boundary
3-4 star boundary
2-3 star boundary
1-2 star boundary

Figure 3B. Residual flows for funds around Morningstar rating boundaries (current-month flows, all funds)



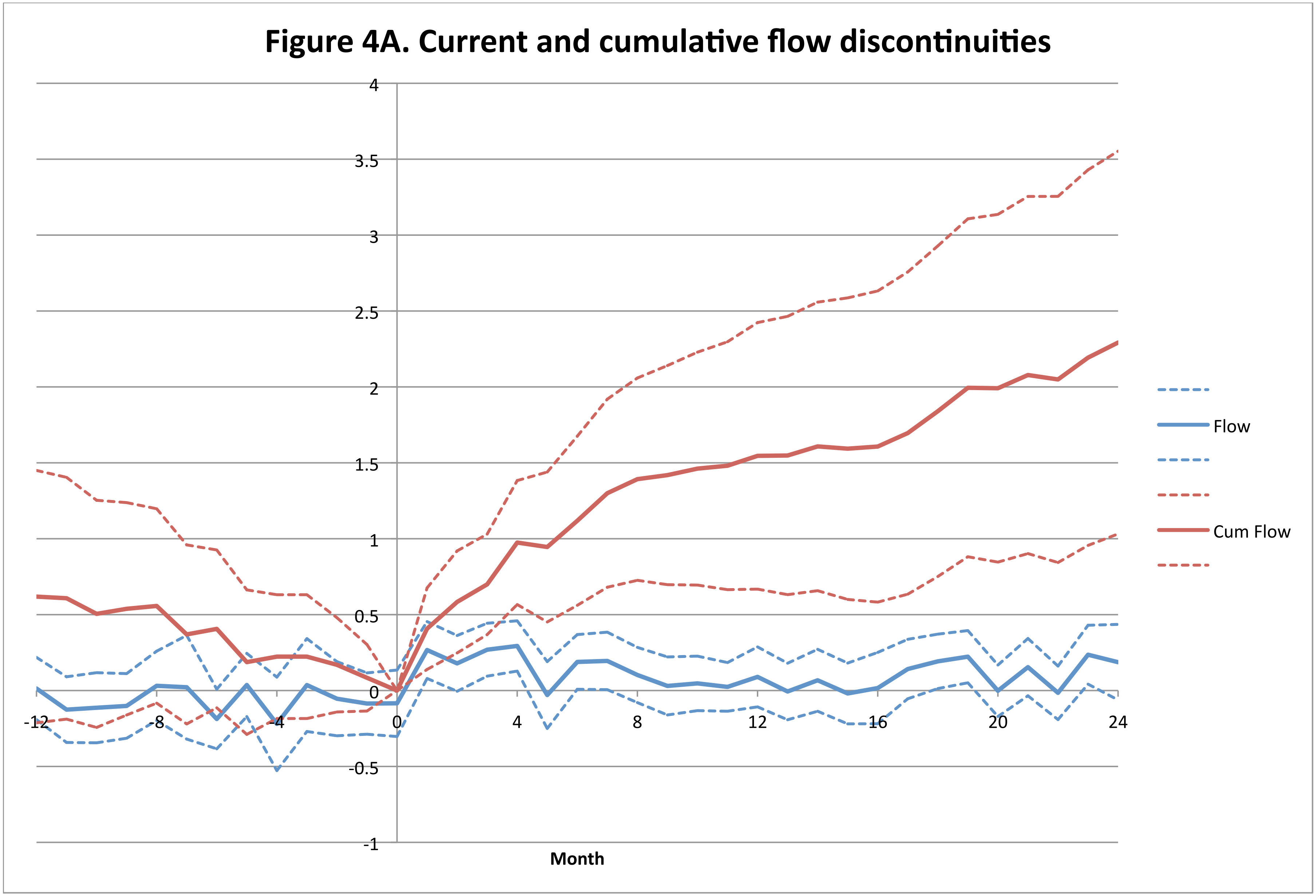


Figure 4A. Current and cumulative flow discontinuities
4
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-12
-8
-4
0
4
8
12
16
20
24
Month
Flow
Cum Flow

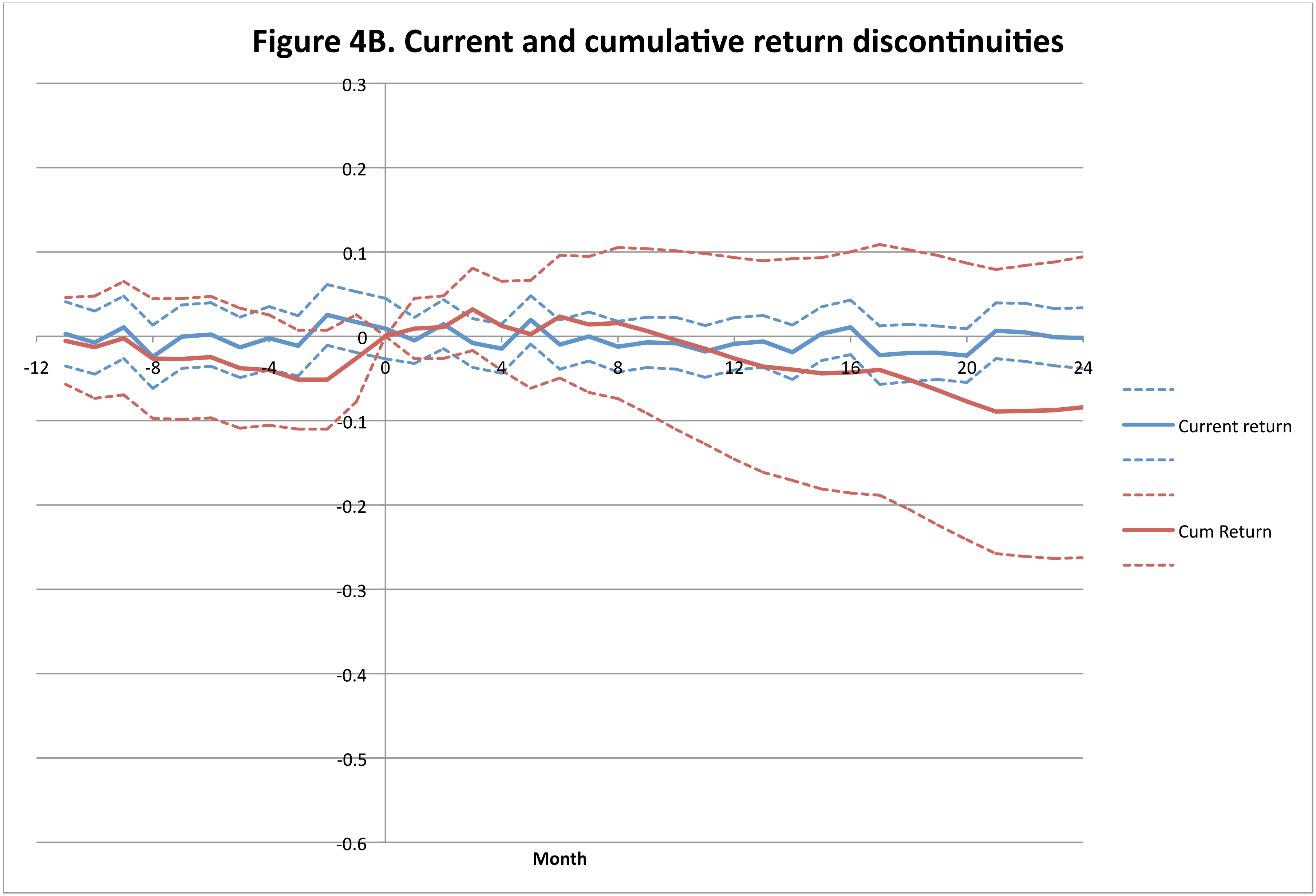


Figure 4B. Current and cumulative return discontinuities
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-12
-8
-4
0
4
8
12
16
20
24
Month
Current return
Cum Return

Table 1. Summary statistics

Summary statistics are reported for portfolio*month combinations in which at least one share class receives a Morningstar rating, which requires at least three years of history. Fund characteristics are the average characteristics of the funds' share classes, weighted by prior-month assets.

	All funds		By (asset-weighted average) Morningstar rating				
	Mean	SD	1 star	2 stars	3 stars	4 stars	5 stars
Number of portfoilo*months	491,863		35,419	101,102	178,293	125,422	51,627
Returns (cumulative log percentage points, adjusted for category mean)							
Percent suriving to t+24 months?	86	35	73	80	86	91	93
Return (t)	0.0	1.8	-0.6	-0.2	0.0	0.2	0.3
Return (t+1)	0.0	1.8	-0.2	0.0	0.0	0.0	0.1
Return (t+1 to t+6)	0.0	5.1	-0.9	-0.2	0.0	0.2	0.4
Return (t+1 to t+12)	0.0	7.4	-1.4	-0.2	0.1	0.3	0.3
Return (t+1 to t+24)	0.0	10.0	-2.1	-0.2	0.1	0.4	0.4
Flows (cumulative log percentage points, adjusted for category mean)							
Flow (t)	0.0	10.6	-1.6	-1.0	-0.2	0.7	2.2
Flow (t+1)	0.0	10.3	-1.4	-0.9	-0.2	0.6	1.9
Flow (t+1 to t+6)	0.0	25.5	-7.1	-4.6	-1.2	3.3	9.8
Flow (t+ 1 to t+12)	0.0	37.8	-12.0	-8.1	-2.0	5.9	16.1
Flow (t+1 to t+24)	0.0	55.2	-17.7	-11.8	-2.9	8.9	23.0
Other fund characteristics							
Ln(Portfolio TNA)	5.2	1.8	4.2	4.8	5.2	5.6	5.9
Ln(Family TNA)	8.9	2.3	8.0	8.8	8.9	9.2	9.2
Expense ratio	1.20	0.72	1.74	1.37	1.15	1.03	1.03
Expense ratio (t+12)	1.19	0.71	1.77	1.38	1.16	1.03	1.02
Percent with any load	69	46	77	77	72	61	55
Portfolio turnover (%)	100	157	150	108	92	87	104
Number of share classes	2.4	1.6	2.3	2.5	2.5	2.3	2.2
Percent of assets in largest share class	84.5	19.4	82.6	82.1	84.5	86.2	86.8
Morningstar ratings							
Percent with same rating for all share classes	73	44	82	70	70	73	83
Average rating (t)	3.10	1.06	1.03	2.01	2.99	3.97	4.97
Average rating (t + 36 months)	3.07	1.02	2.16	2.62	2.99	3.40	3.63
Average risk-adjusted return percentile score							
3 year	51	29	8	25	48	72	91
5 year	52	29	5	22	49	76	93
10 year	51	28	6	23	48	74	91
Percent of portfolios with:							
5 year rating (5-9 year-old funds)	83	37	78	84	85	84	78
10 year rating (10+ year-old funds)	46	50	39	47	49	46	40

Table 2. Estimated future inflow discontinuity at Morningstar rating boundaries -- share class level
Dependent variable: log net flows (in percent)

The unit of observation is share class i in month t. The sample is restricted to share classes of actively managed funds that are at least three years old, because younger share classes are not eligible for a Morningstar rating. The sample for each regression is further restricted to share classes with within-category performance rankings within 5 percentiles on either side of a ranking boundary. The stacked models include all 4 ranking boundaries; the other models include only the indicated boundary. All regressions include a local linear control for within-category percentile ranking and and a discontinuity variable that equals one when the share classes' within-category percentile ranking is above the boundary. Baseline regressions also include controls for lagged log size of the share class, portfolio, and family, expense ratio, portfolio turnover, and loads (front, deferred, and trailing), and category*month fixed effects. "Additional controls" include controls for Morningstar's measure of risk-adjusted returns, lagged log returns from t-12 to t-1, t-24 to t-13, and t-36 to t-25, and lagged log inflows from t-12 to t-1 and t-3 to t-1. Standard errors (in parenthesis) allow for clustering within funds. Statistical significance at the 10-percent, 5-percent, and 1-percent level in two-sided tests is denoted by *, **, and ***.

		Next-month flows				Current-month flows			
Boundary		Baseline		Additional Controls		Baseline		Additional Controls	
Stacked	Discontinuity	0.518	***	0.432	***	-0.006		0.033	
		(0.062)		(0.062)		(0.091)		(0.090)	
	Local linear control	0.007		0.003		0.054	***	0.037	**
		(0.011)		(0.010)		(0.016)		(0.016)	
4/5 stars	Discontinuity	0.946	***	0.880	***	0.299		0.367	
		(0.141)		(0.143)		(0.258)		(0.268)	
	Local linear control	0.046	*	0.022		0.070		0.020	
		(0.025)		(0.026)		(0.046)		(0.048)	
3/4 stars	Discontinuity	0.854	***	0.671	***	-0.196		-0.151	
		(0.130)		(0.132)		(0.165)		(0.161)	
	Local linear control	-0.040	*	-0.036	*	0.072	**	0.053	*
		(0.022)		(0.022)		(0.028)		(0.028)	
2/3 stars	Discontinuity	0.610	***	0.514	***	0.232		0.247	*
		(0.108)		(0.106)		(0.159)		(0.154)	
	Local linear control	-0.027		-0.037	*	0.011		-0.001	
		(0.019)		(0.019)		(0.028)		(0.027)	
1/2 stars	Discontinuity	0.337	**	0.312	**	-0.225		-0.180	
		(0.137)		(0.138)		(0.197)		(0.186)	
	Local linear control	0.009		0.001		0.076	**	0.057	*
		(0.026)		(0.026)		(0.033)		(0.031)	

Table 3. Estimated future inflow discontinuity at Morningstar rating boundaries -- portfolio level
Dependent variable: log net flows (in percent)

This table repeats the analysis in Table 2, but the unit of observation is actively managed fund j in month t. The local linear control is the within-category percentile ranking for the largest share class, and the discontinuity measure indicates whether the portfolio's largest share class is on the positive side of the rating boundary. Other variables are aggregated to the portfolio level, weighted by prior-month assets. The sample is restricted to portfolios for which the largest share classes' percentile ranking is within 5 percentiles of the rating boundary. Standard errors allow for clustering within funds. Statistical significance at the 10-percent, 5-percent, and 1-percent level in two-sided tests is denoted by *, **, and ***.

		Next-month flows				Current-month flows			
		Baseline		Additional Controls		Baseline		Additional Controls	
Boundary									
Stacked	Discontinuity	0.578	***	0.408	***	-0.073		-0.084	
		(0.132)		(0.134)		(0.109)		(0.109)	
	Local linear control	0.066	***	0.049	**	0.087	***	0.070	***
		(0.024)		(0.025)		(0.019)		(0.020)	
4/5 stars	Discontinuity	0.565	**	0.527	*	-0.035		-0.006	
		(0.270)		(0.282)		(0.203)		(0.213)	
	Local linear control	0.218	***	0.151	***	0.147	***	0.112	***
		(0.055)		(0.055)		(0.039)		(0.040)	
3/4 stars	Discontinuity	0.856	***	0.615	**	-0.232		-0.218	
		(0.264)		(0.269)		(0.192)		(0.193)	
	Local linear control	0.009		0.000		0.100	***	0.076	**
		(0.043)		(0.044)		(0.030)		(0.030)	
2/3 stars	Discontinuity	0.520	**	0.392		0.046		0.044	
		(0.250)		(0.262)		(0.207)		(0.217)	
	Local linear control	0.003		0.007		0.001		-0.004	
		(0.043)		(0.045)		(0.035)		(0.037)	
1/2 stars	Discontinuity	0.083		-0.231		-0.055		-0.073	
		(0.372)		(0.385)		(0.328)		(0.335)	
	Local linear control	0.067		0.014		0.110	*	0.093	
		(0.079)		(0.085)		(0.062)		(0.066)	

Table 4. Tests for discontinuities in control variables -- portfolio level

This table conducts tests for discontinuities at Morningstar ranking borders in the control variables used in the portfolio-level regressions in Tables 3 and 5-8. Since these controls are pre-determined at the time of ranking, the Morningstar rating should not have a causal effect on them. The sample is the same as in Table 2. Standard errors allow for clustering within funds. Statistical significance at the 10-percent, 5-percent, and 1-percent level in two-sided tests is denoted by *, **, and ***.

	Discontinuity			Local linear control		
Dependent variable	Coef.		S.E.	Coef.		S.E.
Log Portfolio TNA	0.69		(1.58)	1.10	***	(0.31)
Log Family TNA	-3.70	*	(2.16)	1.16	***	(0.42)
Expense ratio	-0.41		(0.38)	0.02		(0.08)
Expense ratio (t+12)	-0.30		(0.41)	0.00		(0.08)
Has load?	-2.29		(5.45)	-0.55	***	(0.10)
Portfolio turnover	-0.12		(0.15)	-0.42	***	(0.03)
Log return (t-12 to t)	-0.15		(0.14)	0.80	***	(0.03)
Morningstar 3-year risk-adjusted return	-0.22		(0.23)	0.73	***	(0.04)

Table 5. Discontinuities in cumulative flows and returns -- portfolio level
Dependent variable: log net flows or log net returns (in percentage points)

The table repeats the regressions in Table 4 for different time horizons. The dependent variable is the cumulative fund-level log flows or fund-level log returns, calculated either from the ranking month to a future month, or from a past month to the ranking month. As in Table 4, all regressions include the discontinuity and within-category percentile ranking variables for the largest share class. They also include the full set of fund-level controls, including the "additional controls" for past returns and inflows, and fixed effects for category*month combinations. Standard errors allow for clustering within funds. Statistical significance at the 10-percent, 5-percent, and 1-percent level in two-sided tests is denoted by *, **, and ***.

	All boundaries (stacked)						4/5 star boundary					
Month	Flows			Returns			Flows			Returns		
-12	-0.62		(0.44)	0.01		(0.02)	-1.33		(0.92)	-0.04		(0.04)
-9	-0.51		(0.37)	0.00		(0.03)	-1.23		(0.77)	-0.02		(0.07)
-6	-0.37		(0.29)	0.02		(0.04)	-1.19	*	(0.67)	0.03		(0.07)
-3	-0.22		(0.20)	0.05	*	(0.03)	-0.68		(0.44)	0.04		(0.06)
0	-0.08		(0.11)	0.01		(0.02)	-0.01		(0.21)	-0.01		(0.04)
1	0.41	***	(0.13)	0.01		(0.02)	0.53	*	(0.28)	0.01		(0.04)
2	0.58	***	(0.17)	0.03		(0.02)	0.72	*	(0.38)	0.05		(0.05)
3	0.70	***	(0.17)	0.01		(0.03)	1.06	***	(0.38)	0.00		(0.05)
6	1.12	***	(0.28)	0.01		(0.04)	2.06	***	(0.57)	0.06		(0.09)
9	1.42	***	(0.36)	0.00		(0.05)	2.05	***	(0.75)	-0.03		(0.12)
12	1.55	***	(0.44)	-0.04		(0.06)	2.35	**	(0.93)	-0.09		(0.14)
15	1.59	***	(0.50)	-0.04		(0.07)	2.29	**	(1.04)	-0.12		(0.15)
18	1.84	***	(0.54)	-0.06		(0.08)	2.50	**	(1.16)	-0.09		(0.17)
21	2.08	***	(0.59)	-0.09		(0.09)	2.32	*	(1.30)	-0.06		(0.18)
24	2.29	***	(0.63)	-0.08		(0.09)	2.29	*	(1.40)	-0.02		(0.20)

Table 6. Cumulative flow and return effects by asset class

Dependent variable: log net flows or log net returns (in percentage points)

This table repeats the first-stage and reduced-form regressions in Table 5 for different time horizons and subsamples of actively managed mutual funds. "All equity" includes "Large-cap equity", "Mid-cap equity", "Small-cap equity", "Sector funds", and "International equity". "All funds" includes "All equity", "Taxable bonds", "Municipal bonds", and a relatively small number of funds that do not fit into these subsamples, such as balanced funds. Standard errors allow for clustering within funds. Statistical significance at the 10-percent, 5-percent, and 1-percent level in two-sided tests is denoted by *, **, and ***.

	All funds	All equity	Large-cap equity	Mid-cap equity	Small-cap equity	Sector funds	International equity	Taxable bonds	Municipal bonds
Flow effects									
T+6 months	1.118 ***	1.437 ***	0.564	0.981	1.118	2.630 **	3.335 **	1.090 *	0.455
	(0.279)	(0.437)	(0.567)	(0.849)	(0.985)	(1.277)	(1.486)	(0.650)	(0.351)
T+12 months	1.546 ***	1.834 **	1.067	1.901	-0.661	3.500 *	4.415 *	0.826	1.172 **
	(0.439)	(0.713)	(0.906)	(1.262)	(1.989)	(1.909)	(2.377)	(0.970)	(0.510)
T+18 months	1.840 ***	1.784 **	1.167	1.669	-1.882	4.842 *	4.511 *	1.514	1.297 **
	(0.544)	(0.886)	(1.253)	(1.668)	(2.175)	(2.507)	(2.784)	(1.203)	(0.595)
T+24 months	2.292 ***	1.808 *	1.165	2.093	-2.299	5.679 *	4.180	2.270	1.982 ***
	(0.630)	(1.014)	(1.496)	(1.894)	(2.457)	(3.352)	(3.004)	(1.445)	(0.708)
Return effects									
T+6 months	0.014	-0.038	-0.083	0.095	-0.333	-0.229	0.215	-0.018	0.045 **
	(0.040)	(0.072)	(0.089)	(0.199)	(0.236)	(0.266)	(0.180)	(0.046)	(0.021)
T+12 months	-0.036	-0.206 *	-0.141	0.071	-0.638 *	-1.103 **	0.041	-0.027	0.081 ***
	(0.063)	(0.113)	(0.144)	(0.310)	(0.358)	(0.471)	(0.267)	(0.068)	(0.030)
T+18 months	-0.064	-0.322 **	-0.214	-0.005	-0.929 **	-1.431 **	0.016	0.025	0.115 ***
	(0.080)	(0.144)	(0.183)	(0.405)	(0.440)	(0.560)	(0.343)	(0.087)	(0.041)
T+24 months	-0.083	-0.399 **	-0.303	0.236	-1.289 ***	-1.426 **	-0.148	0.063	0.113 **
	(0.091)	(0.162)	(0.204)	(0.470)	(0.481)	(0.649)	(0.377)	(0.101)	(0.049)

Table 7. Size-performance and flow-performance correlations compared with estimated causal effects of flows

This table compares the partial correlation between portfolio size and future performance with the estimated causal effect of inflows on performance. The partial correlation is estimated as the coefficient on portfolio size in a regression of future returns on the variables listed under fund characteristics in Table 1, a control for past-12-month log returns, and category*month fixed effects. The estimated effects of an extra Morningstar star on future inflows and returns are taken from Table 6. The ratio of these estimated effects provides an instrument variables estimate of the causal effect of inflows on (contemporaneous) performance. Hausman tests generally do not reject the null hypothesis that the "Standard OLS" size-performance correlation and "IV" causal effects are equal. Statistical significance at the 10-percent, 5-percent, and 1-percent level in two-sided tests is denoted by *, **, and ***.

Specification:		"First Stage"		"Reduced Form"		"IV"		"Standard OLS"		Hausman test
		Causal effect of star on flows		Causal effect of star on returns		IV estimate of flow effect on returns		Portfolio size coefficient		p-value
Horizon	Asset class	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	
6 months	All funds	1.12	(0.28)	0.01	(0.04)	0.01	(0.04)	-0.0009	(0.0001)	0.71
	All equity	1.44	(0.44)	-0.04	(0.07)	-0.03	(0.05)	-0.0013	(0.0002)	0.61
	Large-cap equity	0.56	(0.57)	-0.08	(0.09)	-0.15	(0.16)	-0.0017	(0.0003)	0.35
	Mid-cap equity	0.98	(0.85)	0.09	(0.20)	0.10	(0.20)	-0.0017	(0.0005)	0.63
	Small-cap equity	1.12	(0.99)	-0.33	(0.24)	-0.30	(0.21)	-0.0018	(0.0008)	0.16
	Sector funds	2.63	(1.28)	-0.23	(0.27)	-0.09	(0.10)	-0.0010	(0.0009)	0.40
	International equity	3.33	(1.49)	0.21	(0.18)	0.06	(0.05)	-0.0002	(0.0005)	0.23
	Taxable bonds	1.09	(0.65)	-0.02	(0.05)	-0.02	(0.04)	0.0000	(0.0001)	0.69
	Munis	0.46	(0.35)	0.05	(0.02)	0.10	(0.05)	0.0000	(0.0001)	**0.03**
12 months	All funds	1.55	(0.44)	-0.04	(0.06)	-0.02	(0.04)	-0.0016	(0.0003)	0.59
	All equity	1.83	(0.71)	-0.21	(0.11)	-0.11	(0.06)	-0.0024	(0.0004)	**0.07**
	Large-cap equity	1.07	(0.91)	-0.14	(0.14)	-0.13	(0.13)	-0.0030	(0.0005)	0.34
	Mid-cap equity	1.90	(1.26)	0.07	(0.31)	0.04	(0.16)	-0.0036	(0.0010)	0.80
	Small-cap equity	-0.66	(1.99)	-0.64	(0.36)	Not meaningful		-0.0038	(0.0015)	
	Sector funds	3.50	(1.91)	-1.10	(0.47)	-0.32	(0.13)	-0.0011	(0.0017)	**0.02**
	International equity	4.42	(2.38)	0.04	(0.27)	0.01	(0.06)	-0.0001	(0.0009)	0.88
	Taxable bonds	0.83	(0.97)	-0.03	(0.07)	-0.03	(0.08)	0.0000	(0.0003)	0.70
	Munis	1.17	(0.51)	0.08	(0.03)	0.07	(0.03)	0.0000	(0.0002)	**0.01**
24 months	All funds	2.29	(0.63)	-0.08	(0.09)	-0.04	(0.04)	-0.0023	(0.0005)	0.39
	All equity	1.81	(1.01)	-0.40	(0.16)	-0.22	(0.09)	-0.0037	(0.0007)	**0.02**
	Large-cap equity	1.16	(1.50)	-0.30	(0.20)	-0.26	(0.17)	-0.0047	(0.0009)	0.14
	Mid-cap equity	2.09	(1.89)	0.24	(0.47)	0.11	(0.22)	-0.0059	(0.0018)	0.60
	Small-cap equity	-2.30	(2.46)	-1.29	(0.48)	Not meaningful		-0.0063	(0.0026)	
	Sector funds	5.68	(3.35)	-1.43	(0.65)	-0.25	(0.11)	0.0005	(0.0030)	**0.03**
	International equity	4.18	(3.00)	-0.15	(0.38)	-0.04	(0.09)	0.0000	(0.0016)	0.70
	Taxable bonds	2.27	(1.44)	0.06	(0.10)	0.03	(0.04)	0.0002	(0.0005)	0.53
	Munis	1.98	(0.71)	0.11	(0.05)	0.06	(0.02)	0.0000	(0.0003)	**0.02**

Table 8. Adjusting performance persistence for causal effect of flows

This table reports coefficients from regressions of future returns and inflows on past 12-month returns. Regressions include category*month fixed effects and also control for the same control variables as in the regressions in Tables 2 and 4-6. The table then reports return persistence coefficients that are corrected for the causal effect of inflows on performance. The corrections are calculated as the product of the flow coefficient (i.e., the extra inflows that accompany a 1% higher return) and the causal effect of an additional 1% of asset inflow on future performance, as reported in Table 7. We calculate the upper and lower bounds of 95% confidence intervals for the corrected persistence coefficient using the standard errors reported in Table 7. Statistical significance at the 10-percent, 5-percent, and 1-percent level in two-sided tests is denoted by *, **, and ***.

		Persistence coefficient		Flow coefficient		Corrected persistence coefficient		
Horizon	Asset class	Coef.	S.E.	Coef.	S.E.	Coef.	Lower CI	Upper CI
6 months	All funds	0.082	(0.006)	0.55	(0.023)	0.075	0.036	0.115
	All equity	0.081	(0.006)	0.52	(0.024)	0.095	0.042	0.147
	Large-cap equity	0.067	(0.015)	0.57	(0.068)	0.151	-0.029	0.332
	Mid-cap equity	0.086	(0.011)	0.49	(0.040)	0.039	-0.159	0.237
	Small-cap equity	0.109	(0.017)	0.54	(0.051)	0.270	0.041	0.498
	Sector funds	0.062	(0.014)	0.42	(0.043)	0.099	0.013	0.185
	International equity	0.069	(0.010)	0.55	(0.039)	0.034	-0.025	0.093
	Taxable bonds	0.093	(0.013)	1.04	(0.114)	0.111	0.022	0.199
	Munis	0.092	(0.024)	0.80	(0.214)	0.012	-0.060	0.085
12 months	All funds	0.100	(0.011)	0.92	(0.040)	0.121	0.046	0.196
	All equity	0.096	(0.012)	0.87	(0.042)	0.193	0.086	0.300
	Large-cap equity	0.088	(0.030)	0.95	(0.116)	0.214	-0.042	0.470
	Mid-cap equity	0.096	(0.023)	0.85	(0.072)	0.064	-0.214	0.343
	Small-cap equity	0.122	(0.027)	0.88	(0.089)		Not meaningful	
	Sector funds	0.076	(0.027)	0.66	(0.078)	0.285	0.106	0.463
	International equity	0.075	(0.019)	0.91	(0.070)	0.066	-0.044	0.177
	Taxable bonds	0.122	(0.023)	1.85	(0.202)	0.181	-0.125	0.487
	Munis	0.140	(0.042)	1.38	(0.406)	0.044	-0.026	0.115
24 months	All funds	0.091	(0.016)	1.39	(0.063)	0.142	0.032	0.251
	All equity	0.085	(0.017)	1.30	(0.065)	0.371	0.139	0.603
	Large-cap equity	0.055	(0.044)	1.37	(0.169)	0.412	-0.068	0.891
	Mid-cap equity	0.050	(0.033)	1.29	(0.113)	-0.096	-0.676	0.484
	Small-cap equity	0.123	(0.041)	1.29	(0.135)		Not meaningful	
	Sector funds	0.043	(0.044)	0.97	(0.143)	0.285	0.064	0.506
	International equity	0.110	(0.027)	1.46	(0.117)	0.162	-0.102	0.425
	Taxable bonds	0.145	(0.032)	3.12	(0.316)	0.058	-0.220	0.336
	Munis	0.206	(0.072)	2.10	(0.658)	0.087	-0.017	0.191