JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 43, No. 4, Dec. 2008, pp. 907–936

Star Power: The Effect of Morningstar Ratings on Mutual Fund Flow

Diane Del Guercio and Paula A. Tkac*

Abstract

We apply an event-study methodology on over 10,000 Morningstar star rating changes and find that Morningstar has substantial *independent* influence on the investment allocation decisions of retail mutual fund investors. It is the discrete change in the star rating itself and not the change in the underlying performance measures that drives flow. We document economically and statistically significant positive abnormal flow following rating upgrades, and negative abnormal flow following rating downgrades. In contrast to the cross-sectional flow performance literature, we find evidence of investor punishment of performance declines, some of which is evident immediately in the month of the rating change.

I. Introduction

> [T]he brand that has emerged as dominant in the 1990s is not Fidelity, Putnam or even Merrill Lynch—but instead is Morningstar.
>
> R. Pozen, *The Mutual Fund Business* (1998), p. 75

Information intermediaries, such as *Consumer Reports* with its product ratings or *U.S. News and World Report* with its college rankings, are commonplace in

*Del Guercio, dianedg@uoregon.edu, Lundquist College of Business, University of Oregon, Eugene, OR 97403; and Tkac, paula.tkac@atl.frb.org, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree St NE, Atlanta, GA 30309. We gratefully acknowledge research support from a grant from the Institute for Quantitative Research in Finance (Q-Group). We thank John Rekenthaler, Director of Research at Morningstar, Inc., for providing data, and Paul Gozali of Morningstar's Research Department for answering numerous data questions. We thank Robert Battalio, Stephen Brown (the editor), John Chalmers, Larry Dann, Harry and Linda DeAngelo, Jerry Dwyer, Lily Fang, Laura Field, David Gallagher, Ro Gutierrez, Rebecca Hann, Woodrow Johnson, Christopher Jones, Aditya Kaul, Matthew Morey (the referee), Elizabeth Odders-White, Cesare Robotti, Rudi Schadt, Robert Tamura, Jonathan Taylor, Tracie Woidtke, and seminar participants at the Western Finance Association meetings, Pacific Northwest Finance Conference, Chicago Quantitative Alliance Conference, Atlanta Fed Brown Bag series, Australian National University, Arizona State, Monash University, University of New South Wales, University of Queensland, University of Southern California, and University of Utah for helpful comments. The views expressed here are the authors' and do not necessarily represent the views of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

markets where consumers face important purchase decisions. In the mutual fund marketplace, where consumers face thousands of choices and often lack the financial expertise to assess product quality, Morningstar Inc. with its one- to five-star fund rating is the undisputed market leader among retail investors. Morningstar's star ratings are arguably the primary inputs to many investors' decisions, and as a consequence they influence investor fund flow, a measure of the aggregate net effect of individual fund investor allocation decisions.[1] While many previous papers have documented an empirical relation between investor fund flow and various performance measures, including Morningstar ratings, a direct *causal* link between any given measure of fund quality and investor behavior has remained elusive.[2] In this study, we provide robust empirical evidence that the particular package of fund quality information embodied in the Morningstar star rating affects investor flow independently of the influence of other common measures of fund performance. Moreover, we estimate the magnitude of this isolated "Morningstar effect" in terms of the additional dollars allocated by investors that can be reliably linked to Morningstar's endorsement.

From the mutual fund investor's point of view, the Morningstar star rating has several desirable attributes. It is a freely available, risk-adjusted performance measure that is updated monthly. While its predictive ability is comparable to more academic risk-adjusted performance measures, its one-to five-star system is easier for the typical mutual fund investor to understand.[3] Because Morningstar's rating is based on ranking all available funds according to its proprietary algorithm, the star rating provides a quick source of information on the full distribution of funds, in that a five-star fund is known to be in the top 10% of the fund universe.[4] The very attributes that appeal to do-it-yourself investors are also attractive to the clients of financial advisors. Thus, a financial advisor or a trustee responsible for choosing the menu of funds for a 401(k) plan may also value the validation or certification a top Morningstar rating provides, especially should their recommendations be subsequently questioned by clients or beneficiaries. In short, the information packed into a Morningstar rating, coming from a reputable

[1] As support for this claim, many publications cite the statistic that equity mutual funds rated with four or five stars by Morningstar have received 80%–100% of all net inflows. In addition, several academic studies motivate their analysis with the argument that Morningstar is the "most popular" and "best-known" ranking service among investors (e.g., Blume (1998), Sharpe (1998), and Blake and Morey (2000)). We provide support for Morningstar's popularity in Section II.A.

[2] For example, Ippolito (1992), Patel, Zeckhauer, and Hendricks (1994), Gruber (1996), Chevalier and Ellison (1997), and Sirri and Tufano (1998) find that several performance measures are simultaneously related to fund flow in the cross section. Del Guercio and Tkac (2002), Bergstresser and Poterba (2002), and Ivkovic and Weisbenner (2006) specifically find that Morningstar ratings are significantly related to fund flow.

[3] Blake and Morey (2000) report that Morningstar ratings are as persistent and useful for forecasting future performance as other fund performance measures, such as Jensen's alpha. Additionally, in unpublished results using 34 months of star rating data, we found that following a trading strategy of investing in only five-star rated funds produces positive risk-adjusted performance out of sample. These findings suggest that relying on low-cost Morningstar ratings when making forward-looking investment decisions is a reasonable approach to fund selection.

[4] In July 2002, after our sample period, Morningstar altered its rating algorithm. It remains the case that the top 10% of funds receive five-star ratings, but the ranking universe is now the fund's more narrow style category rather than all domestic equity funds. We address the interpretation of our results under the new algorithm in Section IV.E.

and unbiased source, plausibly reduces search costs for investors and provides valuable certification for financial professionals, implying a high likelihood of use by the investing public.

From a research design point of view, Morningstar ratings have several unique features that offer measurement advantages over other performance metrics. The star rating's inherent discreteness and wide availability at low cost to investors, along with the fact that rating updates arrive at known monthly intervals, imply that a monthly rating change is an unambiguous and easily observable event for fund investors. Importantly, we provide evidence that these rating changes exhibit low correlations, 0.03 or less, with changes in alternative performance measures, such as Jensen's alpha or raw return. Altogether, these features suggest a new method for analyzing the relation between flow and fund performance, an event study of Morningstar rating changes. In contrast to the previous literature that correlates flow and performance in the cross section of funds, we have the opportunity to directly examine, on a fund-by-fund basis, the investor flow response to changes in fund performance.[5]

Benchmarking fund flow prior to a rating change allows us to isolate a "Morningstar effect" as distinct from the influence of other measures of fund performance, and to establish a strong case for a causal relation between flow and this measure of perceived fund quality. In other words, we find that flow increases for a fund when Morningstar certifies that it has improved, rather than because other measures of its prior long-term performance have changed. To validate this causal claim, Section IV.C contains the results of a battery of tests that highlight the success of our benchmark in filtering out the influence of other performance measures and mutual fund industry factors. For example, we separate changes in Morningstar's star ratings from changes in the underlying continuous percentile ranking that Morningstar uses to assign the stars. We find that it is the discrete change in the star rating itself and not the change in the underlying performance measures that drives mutual fund flow.

A detailed and focused study of Morningstar rating changes offers new insights into mutual fund investor decision-making that are difficult to infer from existing results on flow and performance. For example, most previous studies have focused on the level of performance rather than performance changes and implicitly assume that the flow effect of a performance increase (e.g., from four to five stars) is identical to the effect of a performance decrease of the same magnitude (e.g., from five to four stars). Rather than assuming this symmetry, we test for it directly by estimating the effects of performance improvements and declines separately. As we discuss in Section IV.B, separate measurement of the flow response to ratings declines also allows for an alternative definition of "punishment for poor performance" than that which is commonly used in the flow-performance literature. As an alternative to the traditional cross-sectional approach that asks whether funds that underperform their peers experience outflow, we ask whether

[5]In addition to the studies cited in footnote 2, other research that exclusively uses cross-sectional (panel or Fama-MacBeth) methods to analyze the relation between flow and past performance includes Gallaher, Kaniel, and Starks (2006) and Huang, Wei, and Yan (2007).

a fund loses assets when its own performance declines. Similarly, because most previous studies assume an annual tournament for investor dollars, we know little about whether investors revise their investment allocations and seek performance data at higher frequencies. Our monthly event study can directly address this by analyzing whether there is abnormal flow following monthly changes in star ratings.

We employ event-study methods to over 10,000 Morningstar rating changes over the period November 1996 to October 1999. Our benchmark model for a fund's expected flow filters out the effects of both more recent and longer term prior fund performance, as well as the influence of aggregate flows to a fund's investment style and unobservable fund-specific characteristics, such as reputation. Comparing post-rating-change flow to the predictions of this benchmark, we find statistically significant positive abnormal flow following upgrades and negative abnormal flow following downgrades. Moreover, the flow associated with rating changes is economically significant, suggesting that investors direct their investments in accordance with Morningstar's recommendations.

Not all rating changes are created equal, however, as evidenced by significant asymmetry in investor responses to improving versus deteriorating fund quality. An upgrade from four to five stars has a large effect on flow in the six months following the upgrade, resulting in $32 million in abnormal flow, or 25% above normal. While a downgrade from five to four stars appears to be a nonevent, we do find a very large negative and significant response to a downgrade from four to three stars, on the order of twelve times larger than the normal flow for these funds. These results suggest that investors handsomely reward funds that achieve the elite five-star status but only punish funds when performance falls to a three-star rating (i.e., falls below the 68th percentile of funds). In contrast to the cross-sectional flows literature that finds that funds at the bottom of the performance distribution do not experience significant outflows, we find that funds whose performance deteriorates below four stars experience a significant loss of flow, relative to the flow they would get if they maintained their four-star rating. This result paints a different picture of investor punishment of poor performance than currently offered in the literature.

Most of the observed flow response is significant beginning in event month 0 and remains significant in many cases for as many as seven months after the rating change, indicating that both vigilant and casual investors respond to the rating update. We also find that the flow response to rating changes occurring at year-end, when anecdotally investors reevaluate their portfolios, is not different from that of other months. This lack of an end-of-the-year effect is consistent with the findings of Goetzmann and Massa (2002) for a large sample of individual accounts in an S&P 500 index fund. Overall, our results suggest that the tournament among fund managers is not annual, as implicitly assumed in Brown, Harlow, and Starks (1996) and Chevalier and Ellison (1997). We find it to be much more of an ongoing and high frequency tournament, suggesting a more complicated and dynamic strategic environment for fund managers than assumed in current tests of managerial incentives and behavior. As we describe in Section IV.F, our findings on monthly flow response suggest that more powerful tests of managerial behavior are possible.

II. Data

A. Background on Morningstar's Star Rating

Morningstar first offered its star ratings in 1986 by paid subscription to its print product *Morningstar Mutual Funds,* but they are now available free on several popular Web sites including Morningstar.com, Yahoo! Finance, Fidelity.com, and MSN's Moneycentral. Numerous popular press articles and casual observation of mutual fund advertisements attest to Morningstar's popularity among investors, but star ratings have also been the subject of academic study. For example, Jones and Smythe (2003) report that out of 170 fund advertisements in *Money* magazine in 1999, 80, or 47.1%, contained performance information from an independent research service, such as Morningstar or Lipper. Of these 80 advertisements, 73% specifically mention Morningstar ratings.[6] Fund marketing executives responsible for placing these ads clearly recognize the influence of the Morningstar brand. As Spence Fitzgibbons of American Century Investments put it, Morningstar ratings are "like a Good Housekeeping seal of approval . . . I just think it helps put people at ease."[7] Morningstar's reputation as an unbiased and independent information source, coupled with its simple-to-understand five-star system, likely underlie its appeal. Morningstar has successfully entered overseas markets, providing local fund information to investors in 15 countries.[8] The universality of the star rating suggests that it passes the market test for lowering the search and decision costs of investors, many of whom would otherwise find fund selection to be an intimidating or overwhelming decision.

Press articles also provide anecdotal evidence that Morningstar is the dominant information intermediary among financial advisors, beating competitors such as Lipper and Standard & Poor's. Morningstar's platform for financial advisors is said to have up to an 80% market share among the 250,000 independent financial advisors and planners in the U.S.[9] Part of this dominance occurs precisely because Morningstar is influential with retail investors, many of whom seek professional advice. Advisors report that some clients are only interested in hearing about four- and five-star funds, and have been fired for putting the client in a one-star fund.[10] For this reason, we might expect financial advisors and 401(k) plan trustees to use Morningstar ratings.

---

[6] Jones and Smythe (2003) report only aggregate statistics on ads containing rating information. We are grateful to these authors for providing the breakdown of the percentage of ads containing specific mentions of Morningstar.

[7] "Fund Ratings and Recent Results Diverge," *The Wall Street Journal* (May 3, 2000), p. C27. Also, see "Mutual Fund Ratings Come Under Fire," *The Wall Street Journal* (January 15, 2003), p. D1, and "Insurers Find that the Fate of their Funds is in the Stars," *National Underwriter* (October 18, 1999), p. 7.

[8] See Morningstar Inc.'s 2005 offering prospectus (Form S-1).

[9] "Making a Play for Financial Advisers—Lipper, Thomson, and S&P Have All Created Products Targeting Growing Market," *The Wall Street Journal* (December 4, 2002), p. D13. Market share estimates vary widely depending on whether one counts just subscriptions to the *Advisor Workstation* or *Principia* or also counts those advisors who use Morningstar.com.

[10] "Fund Ratings and Recent Results Diverge," *The Wall Street Journal* (May 3, 2000), p. C27. Also in this article, Keith Hartstein, an executive at John Hancock Funds, stated, "I can't tell you the number of times we as an industry sit in front of people picking managers for their 401(k) plan and hear them start the conversation saying, 'We're only selecting four- and five-star funds.'"

Morningstar covers virtually all funds in existence and calculates star ratings for all funds that are at least three years old. Star ratings are recomputed monthly and are available on Morningstar's Web site approximately three business days after month-end. Below we briefly describe the rating process in place during our sample period, but interested readers should see Blume (1998), Sharpe (1997), (1998), and Morey (2000) for a thorough description. In July 2002, after our sample period ends, Morningstar changed its rating algorithm. We discuss the new rating system in the context of our findings in Section IV.E.

A fund's Morningstar star rating is based on its historical performance with respect to both return and risk relative to its peer group. Specifically, Morningstar uses a rolling 36 months of load-adjusted returns to compute a three-year risk-adjusted performance measure for each fund. During our sample period, Morningstar used four mutually exclusive peer groups (domestic equity, international equity, taxable bond, and municipal bond) and ranked funds according to this three-year risk-adjusted measure. A three-year star rating is assigned for each peer group based on this percentile ranking. Specifically, funds with risk-adjusted ratings in the top 10% of their peer group are assigned five stars, the next 22.5% receive four stars, the next 35% receive three stars, the next 22.5% receive two stars, while the bottom 10% of funds in each peer group receive one star. Morningstar computes a five-year and 10-year star rating in the same way, with the only difference being the number of months of load-adjusted returns used in the calculations. We analyze the overall star rating, which is a weighted average of these ratings, because they are the most widely used and cited among publications geared to investors.[11]

B. Description of Our Sample

We obtained the monthly time-series of Morningstar overall star ratings from November 1996 to October 1999 directly from Morningstar, Inc. These data include star rating histories of all funds in Morningstar's domestic equity star rating category as of October 1999. To be included in this database, the fund must exist and have a star rating as of October 1999. Survivorship bias is a legitimate concern given the high rate of liquidation and mergers over this period. Therefore, through the use of monthly editions of Morningstar's CD-ROM *Principia,* we fill in star ratings for all of the rated funds that disappeared over this period, totaling 4,040 fund-months. We supplement fund ratings with data on returns, total net assets, and other fund characteristics from the *Survivor-Bias-Free Mutual Fund Database* from the Center for Research in Security Prices (CRSP). We perform numerous checks to ensure that the data from the two sources are properly linked.[12] Finally, we delete all fund-months that coincide with a fund

[11]For example, for a fund at least 10 years old, its overall star rating equals 0.2 times its three-year rating, plus 0.3 times its five-year rating, plus 0.5 times its 10-year rating. For funds too young to have a five- and 10-year rating, the overall rating equals the three-year rating. In support of our focus on the overall rating, it is the only Morningstar rating provided on Yahoo! Finance and MSN MoneyCentral.

[12]We are grateful to Morningstar for providing a spreadsheet with fund name and a monthly time-series of total net assets. We confirm that the ratings data we use match those available to investors at that time by comparing our star ratings to those on the original Morningstar CD-ROM at two points in

merger, since the fund flow numbers are likely to be distorted. Our final dataset contains 111,715 fund-months from 3,388 distinct funds.[13]

C. Timing and Measurement of Fund Flows and Summary Statistics

We analyze monthly net dollar flows in or out of a fund, using the standard definition in the literature of the change in total net assets minus appreciation:

$$\text{FLOW}_{it} = \text{TNA}_{it} - \text{TNA}_{it-1}(1 + R_{it}).$$

This definition of fund flow assumes that flow occurs at the end of the month. Updated star ratings reflect the most recent month's performance and are available to investors at the beginning of the following month. Thus, if a rating change occurs using return data through October 1998, we record November 1998 as the month of the rating change and designate the month-end flow in November as the first month possibly affected by the change.

III. Empirical Approach and Methodology

Prior research on fund flow has established the importance of past performance in mutual fund investors' allocation decisions. It is also well known from previous work that this relation is highly convex; funds that perform poorly suffer very modest asset losses while top performers garner a disproportionate amount of flow (Sirri and Tufano (1998), Chevalier and Ellison (1997)). One thing that remains unknown, however, is exactly what type of performance information the typical investor uses in her decision-making. Research to date has employed cross-sectional analysis and found that many past performance measures simultaneously have a strong statistical relation to new money fund flow (raw return, one-factor alpha, four-factor alpha, Sharpe ratio, whether the fund advertised, Morningstar star rating, excess returns). The high degree of correlation among performance variables makes it difficult to get an accurate estimate of the marginal impact of any one given performance measure on investor flow.

A study of the flow response to changes in Morningstar's star rating allows us to overcome this obstacle. The univariate relation between fund flow and Morningstar ratings, shown in Figure 1, displays the familiar convex flow-performance relation that exists for many performance measures and illustrates the need to employ a methodology that can disentangle the independent Morningstar effect from other influences on fund flow. Analyzing the effect of

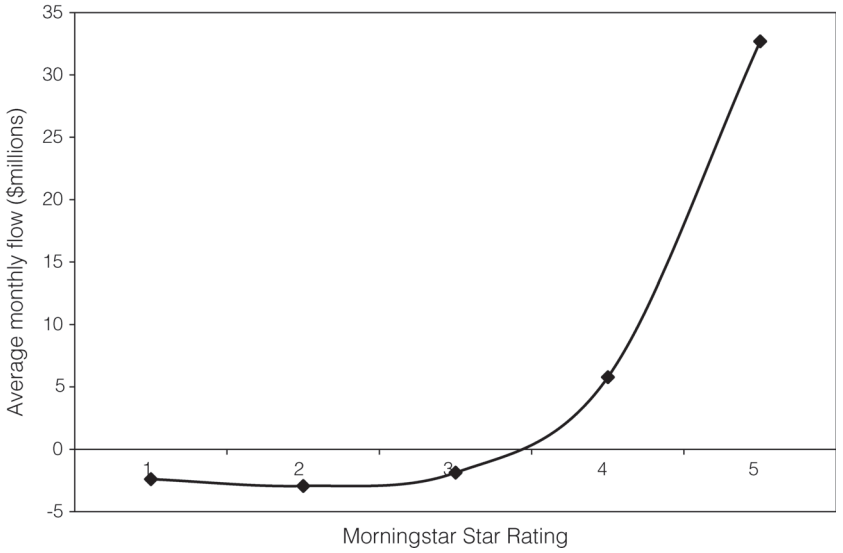
time, July 1997 and October 1999. Only 2.3% of the funds with a star rating for July 1997 in our data could not be found on the original CD-ROM, indicating that backfilling is not a concern. We also find that only 103, or 4.5%, of the funds have different star ratings listed in the two sources in July 1997. In October 1999, we find no discrepancies in star ratings between the two sources.

[13] We consider a share class (e.g., class A) of a fund to be a distinct fund for two reasons. First, Morningstar treats each share class independently for the purpose of assigning a star rating. In fact, we commonly observe that a star rating differs between classes of the same underlying fund. Second, since our main interest is in flow measurement, it is not obvious that flows of different classes are closely related. For this reason, we are not double-counting observations as would happen in a study of fund performance.

FIGURE 1

Average Monthly Dollar Flow in Five Morningstar Star Rating Categories

To create this plot, we average monthly dollar fund flow for funds with a one- through five-star rating from November 1996 through October 1999. The sample includes all funds in Morningstar's domestic equity star rating category. In addition to the diversified equity funds that Morningstar assigns to one of nine equity styles (e.g., large-cap growth), this category also includes sector funds, convertible bond funds, and domestic hybrid funds (e.g., asset allocation funds). The sample contains flow data for a total of 111,715 fund-months and includes data for 3,388 unique funds. For these statistics, and in our empirical tests, we delete 220 fund-months that coincide with a fund merger.



changes in performance on flow, rather than the relation between flow and the *level* of performance, enables us to do just that. Table 1 illustrates the stark differences in correlation between Morningstar's star rating with other performance variables when measured in changes versus levels. Panel A contains the pairwise correlations between Morningstar's star rating and four popular performance measures and shows that funds with four- and five-star ratings are also quite likely to have high Jensen's alphas, Sharpe measures, and three-year cumulative raw returns and return rank. Since all five measures are based on the same underlying return data, correlations that range from 0.46 to 0.65 are not unexpected. In contrast, Panel B shows that changes in these same performance measures have substantially lower correlations with star rating changes (correlation equal to 0.03 or less). Thus, while the same month's return is at the root of each performance change, the star rating algorithm differs enough from the computation of other measures to make the performance changes largely unrelated to each other.

In addition to its low correlation with changes in alternative performance measures, other features inherent in the Morningstar rating imply a cleaner setting in which to measure flow effects. The monthly updates in investor information that we observe, discrete changes in Morningstar ratings, are identical to the signals observed by investors. Thus, isolating a flow response to rating changes amounts to a clean test of whether investors use star ratings, without having to

TABLE 1

Descriptive Statistics

Panel A of Table 1 reports Pearson pairwise correlations across all fund-months from November 1996 through October 1999. The sample includes all funds in Morningstar's domestic equity star rating category. In addition to the diversified equity funds that Morningstar assigns to one of nine equity styles (e.g., large-cap growth), this category also includes sector funds, convertible bond funds, and domestic hybrid funds (e.g., asset allocation funds). The sample size ranges from 81,797 to 105,010 fund-months for 3,388 unique funds depending on data requirements for calculating the performance measure. To calculate Jensen's alpha, the Sharpe measure, three-year cumulative return, and three-year return rank, we use returns over the previous 36 months. Jensen's alpha uses the S&P 500 as a proxy for the market. The Sharpe measure uses the three-month T-bill rate as a proxy for the risk free rate to calculate the mean excess return in the numerator and the standard deviation of the excess return in the denominator. The three-year cumulative return is the compounded monthly raw return, while the rank of this variable uses all funds in the domestic equity universe. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Correlation of Morningstar's Star Rating with Other Performance Measures

	Morningstar Star Rating	Jensen's Alpha	Sharpe Measure	Three-Year Cumulative Return
Jensen's alpha	0.617***			
Sharpe measure	0.650***	0.783***		
Three-year cumulative return	0.459***	0.539***	0.667***	
Three-year return rank	0.622***	0.478***	0.542***	0.730***

Panel B reports Pearson pairwise correlations across monthly changes in the performance measures defined in Panel A. The sample size ranges from 80,218 to 104,482 fund-months for 3,388 unique funds depending on data requirements for calculating changes in the performance measures. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel B. Correlation of Changes in Morningstar's Star Rating with Changes in Other Performance Measures

	Changes in:			
Changes in:	Morningstar Star Rating	Jensen's Alpha	Sharpe Measure	Three-Year Cumulative Return
Jensen's alpha	0.029***			
Sharpe measure	0.016***	0.529***		
Three-year cumulative return	0.021***	0.376***	0.528***	
Three-year return rank	0.028***	0.363***	0.210***	0.437***

Panel C reports the frequency count of fund-months where the fund's Morningstar star rating is different from its previous month's star rating over the period November 1996 to October 1999. These star rating changes are divided into one-star rating upgrades (i.e., a change from one to two stars), and one-star rating downgrades. We also report the frequency of rating changes greater than one star (e.g., a change from one to three stars), but we do not analyze these.

Panel C. Frequency of Fund-Months with Changes in Morningstar Star Rating

Morningstar Star Rating After Change	One-Star Upgrade (*N*)	One-Star Downgrade (*N*)
1	N/A	739
2	653	1,678
3	1,588	2,065
4	2,053	931
5	944	N/A
Subtotal	5,238	5,413
Total number of one-star changes	10,651	
Upgrades greater than one star	40	
Downgrades greater than one star	44	
Total star rating changes	10,735	

make assumptions about how they use them.[14] The discreteness of a rating change naturally lends itself to the use of an event study method, where each fund's flow

[14] Jensen's alpha, in contrast, is a continuous measure that does not lend itself to a well-defined test of investor response to a change in performance. It is not obvious what magnitude of change in alpha is large enough to induce investors to reallocate their assets, nor do we know how often investors seek updated information, implying that we do not know the frequency likely to generate a flow response.

following a rating change is benchmarked to a model of its own normal flow to obtain an estimate of the abnormal flow due to a performance change. The following sections describe the details of how we apply event study methods in the context of mutual fund flow and the diagnostic tests we employ to assess the reliability of the benchmark model.

A. Frequency of Star Rating Change Events

Panel C of Table 1 provides sample summary statistics on the frequency of star rating change events. We observe 10,735 rating changes, which represent close to 10% (10,735/111,715) of total fund-months. The vast majority of these changes are upgrades or downgrades of one star. We do not analyze changes greater than one star, which represent less than 1% of all rating changes. Given the age bias in star ratings in levels noted by Blume (1998) and Morey (2002), we examine whether the distribution of rating changes varies across fund age categories (3–5 years, 5–10 years, and over 10 years) but do not find significant differences.

B. Event Study Method for Calculating Abnormal Flow

Similar to a traditional event study of stock returns, an event study on fund flow aims to parsimoniously purge raw fund flow of the influence of all performance and nonperformance characteristics other than the change in Morningstar rating and thereby isolate the incremental flow due to a rating change. To compute normal flow we estimate a time-series benchmark regression for each individual fund i:

$$(1) \quad F_t^i = \gamma^i + \beta_1^i SF_t^i + \beta_2^i RET_{t-1}^i + \beta_3^i \Delta\alpha_{t-1}^i + \beta_4^i (\Delta\alpha_{t-1}^i)^2 + \beta_5^i F_{t-1}^i + \varepsilon_t^i,$$

where F_t^i is the net dollar flow to fund i at month t, and SF_t^i is the aggregate net flow to all funds in the same style category as fund i at month t, RET_{t-1}^i is fund i's monthly return at $t-1$, $\Delta\alpha_{t-1}^i$ is the change in fund i's Jensen's alpha from month $t-2$ to $t-1$, and F_{t-1}^i is the net flow to fund i at month $t-1$.

Our benchmark regression includes variables found to be important predictors of flow in the literature. Specifically, we include controls for the current popularity of a fund's investment style, recent innovations in a fund's raw and risk-adjusted performance, and a fund's long-term performance record. Because a fund's raw return and risk-adjusted performance in month $t-1$ occur contemporaneously with the change in Morningstar rating and may independently affect flow, it is important to include these variables. We include a squared term on the change in Jensen's alpha to account for potential convexities between flow and recent performance. Including lagged flow as a regressor is equivalent to an infinitely distributed lag model that contains the fund's full performance history and thus is a way to control for the effect of a fund's long-term performance record.

We define the first month an investor would have information available about a new Morningstar rating as event time 0. To estimate the coefficients for the

In short, with traditional performance measures, we do not necessarily observe the same information signal as investors do, which makes it especially difficult to infer causality.

benchmark flow regressions we use 24 months of data, ending three months before time 0 (i.e., event month –26 to –3).[15] We calculate the fund alphas using 36 months of return data and the S&P 500 as the market index but drop funds with less than 24 months of data. We use 19 Morningstar-defined style categories to identify each fund's style and to compute aggregate style flows. Appendix A contains details of our procedure for constructing these style flow benchmarks.

C. Benchmark Diagnostics

If our benchmark is working properly we should observe much higher normal flow for those funds that have performed well in the past. Figure 2, which plots the average time 0 normal flow for all funds according to their Morningstar star rating at time –1, illustrates that the benchmark is performing as expected. Average normal flow is negative for funds below three stars, while the typical five-star fund has large monthly normal flow of $34 million, closely mirroring the pattern in the raw total flow picture of Figure 1. Thus, our specification of normal flow appears to be capturing a fund's longer-term performance track record reasonably well, which in turn gives us confidence in properly labeling the flow response to a rating change as abnormal flow. Appendix B contains evidence that our benchmark model results in a good fit and that our events are not clustered in calendar time, as well as other diagnostic tests that support a well-specified benchmark.

D. Statistical Tests

Taking our estimated benchmark coefficients into the event window, our measure of abnormal flow around a change in Morningstar star rating is therefore:

$$(2)\quad \mathrm{AF}_t^i = \mathrm{F}_t^i - \hat{\gamma}^i - \hat{\beta}_1^i \mathrm{SF}_t^i - \hat{\beta}_2^i \mathrm{RET}_{t-1}^i - \hat{\beta}_3^i \Delta\alpha_{t-1}^i + \hat{\beta}_4^i (\Delta\alpha_{t-1}^i)^2 - \hat{\beta}_5^i \mathrm{F}_{t-1}^i.$$

In words, the abnormal flow to fund i at time t equals the actual flow at time t minus the expected flow due to its aggregate style flow, lagged return, lagged change in alpha, the square of lagged change in alpha, and lagged flow, minus the average abnormal flow to fund i ($\hat{\gamma}^i$). This average abnormal flow captures fund-specific determinants of flow that do not vary over time, such as fund size and unobservable reputation. By grouping funds according to their pre-change star rating, we define events as a change from three to two stars, from four to five stars, etc., and assess the flow response measures to each event for statistical significance.

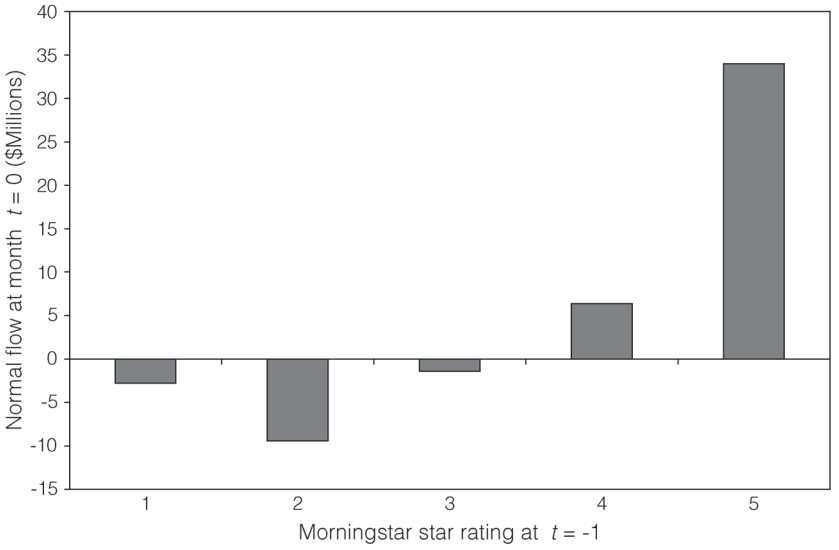
In our main tests we report abnormal flow statistics over the event window month 0 to +6. Unlike an event study of returns where market efficiency implies an immediate stock price reaction to new information, the impact of a rating change on fund flows may persist over many months, and because there is no mispricing to exploit in an open-end mutual fund, a flow response in months 1 through 6

[15] Analysis of benchmark residuals and other diagnostics suggest that a 24-month estimation period provides the best balance between parameter stability and precision. However, the results are qualitatively very similar when we vary the length of the estimation period from 12 to 36 months.

FIGURE 2

Average Month Zero Normal Flow in Five Morningstar Star Rating Categories

To create this plot, we average month 0 normal flow across all funds by Morningstar star rating in month −1. Normal flow is based on a benchmark regression, analogous to a market model regression, whereby a fund's monthly flow is regressed on aggregate flow at time t to funds in its same style group, its own time $t-1$ flow, and its own time $t-1$ return, its change in Jensen's alpha from $t-2$ to $t-1$, and its change in Jensen's alpha from $t-2$ to $t-1$ squared. We use all funds in Morningstar's domestic equity star rating category from November 1996 through October 1999. The number of funds used to compute each average normal flow in the order of one- to five-stars are: 8,016, 18,305, 28,180, 18,176, and 6,815.



is not anomalous. An immediate flow reaction to a rating change is presumably due to vigilant investors who monitor funds on a monthly basis. A delayed flow response is also plausible, either because investors make allocation decisions at different intervals or because more casual investors take time to respond to the new rating.

We apply the event study methods described in Dodd and Warner (1983) to these abnormal flow estimates. Specifically, we focus on standardized abnormal flows, where event period abnormal flows are standardized by the estimated forecast variance (RMSE) of the abnormal flow. Under this method, funds with more precisely measured abnormal flows, and hence lower forecast variance, are implicitly weighted more heavily in our computation of the average abnormal flow across funds in each event month. In addition to these month-by-month measures, we also compute cumulative standardized abnormal flows by summing the standardized abnormal flows for each fund from event time 0 to t, and then dividing by the square root of the number of months used in the cumulation. Following standard practice, at each event date t we compute average standardized abnormal flows ($ASTAF_t$) and average cumulative standardized abnormal flows ($ACSTAF_t$) by averaging across N events (funds).

To assess the statistical significance of event period $ASTAF_t$ and $ACSTAF_t$, we use the standardized cross-sectional test of Boehmer, Musumeci, and Poulsen (1991). Specifically, we divide $ASTAF_t$ and $ACSTAF_t$ by its contemporaneous cross-sectional standard error. Because this measure uses event-period abnormal

flows to calculate the standard errors, it adjusts for any change in the variance of the standardized abnormal flow from the estimation period to the event-window.[16] The tables also contain nonparametric χ^2 test statistics under the null hypothesis that 50% of sample funds have positive standardized abnormal flow.

IV. Results

A. Abnormal Flow Response to Star Rating Changes

This section presents event study results of testing the null hypothesis of zero abnormal flow response for all monthly star rating changes during the 1996 to 1999 period.[17] Panels A and B of Table 2 report the ASTAF for all rating upgrades and downgrades for each event month from 0 to 6, while Panels C and D contain the corresponding ACSTAF estimates. Abnormal flow in this context can be interpreted as a fund's actual flow relative to that expected if it had maintained its pre-change star rating. We report the unconditional Morningstar effect based on the full sample of funds experiencing a rating change, including those that undergo subsequent rating changes within the seven-month event window. This is the most appropriate sample to study because investors cannot know in advance and, hence, condition their investment allocations on, whether a rating will change again.

In most event months, we can reject the null hypothesis of no Morningstar effect, as most ASTAF and ACSTAF are significantly different from zero. In fact, in many cases abnormal flow is significant in event month zero and out to month six, suggesting that some investors respond immediately to a rating change while others respond with a considerable lag. We generally observe a surprisingly consistent flow response across the rating change categories, positive for rating upgrades and negative for rating downgrades, with confirmatory evidence from the nonparametric χ^2 tests. In particular, the attainment of five-star status stands out as an overwhelmingly positive event for the typical fund. In the four- to five-star upgrade category, all test statistics are highly significant in every month in both panels, the magnitude of the mean standardized response is nearly three times that

[16] An alternative approach is the methodology of Brown and Warner (1980), (1985), which instead uses the standard deviation of estimated average abnormal flow (residuals) in the estimation period as the standard error for abnormal flow in the event window. If abnormal flow estimates exhibit an increased variance in the event window, the Boehmer, Musumeci, and Poulsen (1991) standard errors will be larger, and the resulting t-statistics lower, than using the Brown and Warner method and thus will make it more difficult to reject the null hypothesis of no abnormal flow response. Indeed, we do find that Brown and Warner (1985) standard errors are only 33%–70% as large as the standard errors using Boehmer, Musumeci, and Poulsen (1991). Therefore, our choice of standard errors will, if anything, bias against finding a significant Morningstar effect. We also perform diagnostics recommended in Brown and Warner (1985), such as tests for nonnormality and serial correlation in the average abnormal return estimates. We find no evidence that these issues affect our inferences.

[17] To minimize the possibility that extreme observations drive the results, we also delete the top and bottom 1% of standardized abnormal flows at each event date in each rating subsample, and we note that inferences do not change when we include them. We have also examined the ASTAF of rating changes in the pre-event period $(-2, -1)$ and find no evidence of a flow response at time –2, but some evidence of a response at time -1. In all of these cases, the ASTAF at time -1 is the same sign as it is at time 0, indicating that this is most likely a timing issue. We note that if we include up to $t = -1$ in the estimation period, we find similar results in the event window. All unreported diagnostic tests are available from the authors.

for any other rating change category, and the percentage of positive abnormal flow observations ranges from 53% to 61%.

There are a few exceptions to the general finding that upgrades are followed by positive abnormal flow and downgrades are followed by negative abnormal flow. Although the abnormal flow has the correct sign, upgrades from one to two stars and downgrades from two to one star exhibit only weak evidence of a flow response. This might not be surprising because these funds still remain near the bottom of the performance distribution. More surprising is the result that downgrades from five to four stars are associated with significantly *positive* ASTAF and ACSTAF. Given that we find significant abnormal flow for seven event months, we investigate whether this seemingly anomalous result is driven by previous rating changes up to five stars for this group.

Appendix C reports the ACSTAF results of Table 2 when we repeat the analysis for a subsample of funds with no change in star rating in the previous eight months. This sample of "clean" rating changes shows even stronger evidence of positive abnormal flow for rating upgrades and negative abnormal flow following rating downgrades. Furthermore, we find no significant abnormal flow following

TABLE 2

Morningstar Star Rating Changes of Domestic Equity Funds

In Panels A and B, the average standardized abnormal flow ($ASTAF_t$) reported below is averaged across domestic equity funds within the same star rating change group. We define standardized abnormal flow in month t as the actual dollar flow in month t minus the normal, or expected, flow standardized by the estimated forecast variance of the normal flow. Normal flow is based on a market-model regression whereby a fund's monthly flow is regressed on aggregate flow at time t to funds in its same style group, its time $t-1$ flow, its time $t-1$ return, its change in Jensen's alpha from $t-2$ to $t-1$, and its change in Jensen's alpha from $t-2$ to $t-1$ squared. We use nineteen Morningstar-defined style categories to compute aggregate style flows. See the appendix for details on the style categories we use. The estimation period for computing the market-model parameters is months (−26, −3). The standard error is equal to the cross-sectional standard deviation over N sample events of the standardized abnormal flows, divided by the square root of N. Average standardized abnormal flows significantly different from zero at the 5% level or higher in a two-tailed test are indicated in bold. For each event date, we also report the percentage of sample events with positive abnormal flows. We indicate in bold the percentage positive that differs from 50% at the 5% significance level or better, using a χ^2 test with one degree of freedom. The number of observations differs from Panel C of Table 1 because we delete fund-months that coincide with mergers, observations that do not have the data necessary to compute flows, and the top and bottom 1% of standardized abnormal flows at each event date.

Panel A. Average Standardized Abnormal Flow for Rating Upgrades

	From 1 to 2 Stars ($n = 534$)			From 2 to 3 Stars ($n = 1{,}382$)			From 3 to 4 Stars ($n = 1{,}763$)			From 4 to 5 Stars ($n = 805$)		
Event Month	$ASTAF_t$	t-Stat.	$\% > 0$	$ASTAF_t$	t-Stat.	$\% > 0$	$ASTAF_t$	t-Stat.	$\% > 0$	$ASTAF_t$	t-Stat.	$\% > 0$
0	0.01	0.15	51.7	**0.22**	3.46	**54.5**	**0.21**	3.24	49.9	**0.56**	5.40	**55.8**
1	0.02	0.25	52.4	**0.22**	2.94	51.4	**0.19**	3.12	51.8	**0.76**	6.56	**57.8**
2	0.01	0.08	49.4	**0.24**	3.58	**55.5**	**0.35**	4.80	52.3	**0.99**	6.59	**57.3**
3	0.06	0.62	52.6	**0.35**	4.41	**55.7**	**0.54**	6.78	**52.6**	**0.90**	5.33	**55.1**
4	**0.25**	2.44	**55.5**	**0.24**	3.01	**54.1**	**0.55**	6.39	**52.8**	**1.04**	5.84	**55.3**
5	0.19	1.33	52.4	**0.52**	4.98	**57.8**	**0.41**	4.46	**53.8**	**1.00**	5.16	**58.1**
6	**0.26**	2.25	54.0	**0.29**	2.70	**54.4**	**0.45**	4.43	50.5	**1.44**	5.57	52.7

Panel B. Average Standardized Abnormal Flow for Rating Downgrades

	From 2 to 1 Star ($n = 589$)			From 3 to 2 Stars ($n = 1{,}407$)			From 4 to 3 Stars ($n = 1{,}718$)			From 5 to 4 Stars ($n = 770$)		
Event Month	$ASTAF_t$	t-Stat.	$\% > 0$	$ASTAF_t$	t-Stat.	$\% > 0$	$ASTAF_t$	t-Stat.	$\% > 0$	$ASTAF_t$	t-Stat.	$\% > 0$
0	−0.08	−1.16	47.5	**−0.18**	−2.98	48.5	**−0.16**	−2.66	47.7	**0.46**	4.26	**53.9**
1	−0.15	−1.81	48.9	−0.08	−1.43	49.5	**−0.12**	−2.05	49.1	**0.60**	3.95	53.0
2	**−0.23**	−3.09	46.5	**−0.16**	−2.51	47.9	−0.09	−1.49	**46.1**	0.04	0.36	47.6
3	−0.10	−1.22	47.6	**−0.20**	−3.14	**46.7**	**−0.21**	−3.02	**47.0**	**0.26**	2.03	50.7
4	**−0.20**	−1.98	46.9	**−0.15**	−1.98	**47.1**	−0.03	−0.46	**46.9**	**0.82**	4.76	52.4
5	−0.17	−1.81	**44.8**	**−0.27**	−3.66	**44.8**	**−0.21**	−2.91	**45.7**	**0.67**	3.44	49.4
6	**−0.32**	−3.26	**44.4**	−0.18	−1.91	48.1	**−0.15**	−1.79	**45.7**	**0.85**	4.09	49.9

(*continued on next page*)

TABLE 2 (continued)

Morning Star Rating Changes of Domestic Equity Funds

Panel C reports the average cumulative standardized abnormal flow ($ACSTAF_t$) for each date t in the event window. We first compute the cumulative standardized abnormal flow for each fund by summing the standardized abnormal flow from 0 to t and then dividing by the square root of the number of months used in the cumulation. We then average these standardized cumulated abnormal flows over the N sample events in each star rating change group to obtain ACSTAF at each event date t. ACSTAFs significantly different from zero at the 5% level or higher in a two-tailed test are indicated in bold. For each event date, we also report the percentage of sample events with positive cumulative standardized abnormal flows. We indicate in bold the percentage positive that differs from 50% at the 5% significance level or better, using a χ^2 test with one degree of freedom. The number of observations differs from Panel C of Table 1 because we delete fund-months that coincide with mergers, observations that do not have the data necessary to compute flows, and the top and bottom 1% of standardized abnormal flows at each event date.

Panel C. Average Cumulative Standardized Abnormal Flow for Rating Upgrades

Event	From 1 to 2 Stars ($n = 534$)			From 2 to 3 Stars ($n = 1{,}382$)			From 3 to 4 Stars ($n = 1{,}763$)			From 4 to 5 Stars ($n = 805$)		
Month	$ASTAF_t$	*t*-Stat.	% > 0	$ASTAF_t$	*t*-Stat.	% > 0	$ASTAF_t$	*t*-Stat.	% > 0	$ASTAF_t$	*t*-Stat.	% > 0
0	0.01	0.15	51.7	**0.22**	3.46	**54.5**	**0.21**	3.24	49.9	**0.56**	5.40	**55.8**
1	0.02	0.19	48.7	0.25	1.39	**54.2**	**0.31**	2.54	51.7	**0.97**	7.08	**58.1**
2	−0.08	−0.59	49.4	**0.49**	3.97	**54.9**	**0.37**	2.86	52.0	**1.81**	3.90	**59.5**
3	0.08	0.41	48.7	**0.58**	4.49	**55.7**	**0.64**	3.12	**52.8**	**2.17**	4.65	**60.5**
4	0.14	0.88	51.1	**0.77**	4.34	**55.1**	**0.77**	4.22	**52.6**	**2.81**	5.01	**61.6**
5	0.16	0.87	50.4	**0.96**	4.92	**56.5**	**0.88**	4.47	**53.0**	**2.89**	5.16	**61.0**
6	0.39	1.93	51.5	**1.02**	4.84	**57.1**	**0.86**	4.00	**53.3**	**3.59**	5.78	**60.4**

Panel D. Average Cumulative Standardized Abnormal Flow for Rating Downgrades

Event	From 2 to 1 Star ($n = 589$)			From 3 to 2 Stars ($n = 1{,}407$)			From 4 to 3 Star ($n = 1{,}718$)			From 5 to 4 Star ($n = 770$)		
Month	$ASTAF_t$	*t*-Stat.	% > 0	$ASTAF_t$	*t*-Stat.	% > 0	$ASTAF_t$	*t*-Stat.	% > 0	$ASTAF_t$	*t*-Stat.	% > 0
0	−0.08	−1.16	47.5	**−0.18**	−2.98	48.5	**−0.16**	−2.66	47.7	**0.46**	4.26	**53.9**
1	−0.14	−1.24	46.6	**−0.19**	−1.99	47.5	**−0.27**	−2.80	48.2	**0.79**	4.70	53.1
2	−0.09	−0.51	47.2	**−0.25**	−2.63	**46.7**	**−0.29**	−1.98	**46.4**	**0.56**	3.35	52.0
3	**−0.33**	−2.81	**44.2**	**−0.28**	−2.66	**46.0**	**−0.35**	−2.20	**46.7**	**2.00**	2.05	52.5
4	**−0.38**	−2.62	**45.3**	**−0.35**	−2.74	**46.0**	**−0.39**	−2.20	**46.8**	**1.84**	3.36	52.0
5	−0.29	−1.78	**42.6**	**−0.52**	−3.96	**44.3**	**−0.52**	−2.82	**45.2**	**1.78**	3.32	52.2
6	−2.12	−1.26	**43.1**	**−0.53**	−3.62	**44.6**	−0.34	−1.68	**44.8**	**2.99**	2.62	50.9

a downgrade to four stars; we cannot reject that there is zero flow response to a rating change from five to four stars. Thus, the positive response in the full sample reflects the abnormal flow following prior upgrades, and the downgrade from five to four stars can be viewed as the only nonevent among the rating change groups.

B. The Dollar Value of a Morningstar Rating Change and a New Measure of Punishment

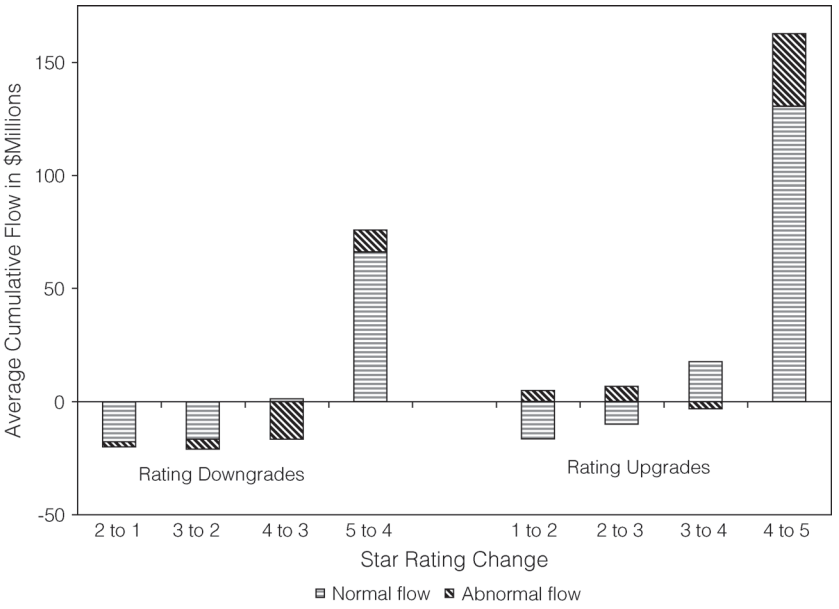
While statistical significance indicates whether star ratings reliably influence investors, the economic significance determines whether Morningstar's influence is likely to alter the behavior of investment companies and fund managers. One obvious question of interest is how many millions of dollars a five-star rating is worth to the average fund. To answer such questions, we analyze nonstandardized abnormal dollar flow, rather than the standardized version used in the statistical tests.

Figure 3 graphs the nonstandardized average normal and average abnormal flows cumulated through the six event months following a rating change in each of the upgrade and downgrade categories. The pattern of normal flow in the figure suggests that the benchmark reliably captures the flow resulting from the *recent* performance that led to an upgrade or downgrade. For example, funds that were downgraded from four to three stars have a normal flow of $1.4 million, while

FIGURE 3

Average Seven-Month Cumulative Abnormal Flow for Morningstar Star Rating Changes

To create this plot, we averaged seven-month (0,6) cumulative abnormal flow and seven-month cumulative normal flow across domestic equity funds within the same star rating change group. Normal flow is based on a benchmark regression, analogous to a market model regression, whereby a fund's monthly flow is regressed on aggregate flow at time t to funds in its same style group, its own time $t - 1$ flow, and its own time $t - 1$ return, its change in Jensen's alpha from $t - 2$ to $t - 1$, and its change in Jensen's alpha from $t - 2$ to $t - 1$ squared. Abnormal flow in month t is the actual dollar flow in month t minus normal flow. Note that unlike the numbers reported in Table 2, this figure contains measures of abnormal flow that are *not* standardized by the forecast variance.



funds that were upgraded from four to five stars have a normal flow of $131 million. Thus, normal flow is starkly different for these two sets of funds despite the fact that they were both at the same four-star level before the rating change.

Many of the rating changes are economically significant events for the funds, resulting in flows that are much different than the expected flow. We find that an upgrade from four to five stars results in an average of $32 million in abnormal flow, or 25% above average expected flow, over the seven months following the upgrade. Rating changes also affect funds with negative normal flow, such as those at the low end of the performance spectrum. A downgrade from three to two stars, for example, increases the outflow in the average fund by 27%. The average fund downgraded from four to three stars experiences outflow of $16.7 million, nearly 12 times larger than the expected normal flow of only $1.4 million.

Our method of benchmarking normal flow for each fund allows us to compute an alternative measure of investor "punishment for poor performance." While the mutual fund literature has noted that the total outflows of poor performers are quite small, the abnormal flow response to a rating change might provide a better measure of punishment since it captures the incremental loss of flow, or opportunity cost, due to the performance decline. For example, a technology fund at the peak of the high-tech boom might have both high total flow and high total flow

relative to other funds but experience negative abnormal flow when its performance declines. We find that the negative abnormal flow following rating downgrades from four to three stars is economically meaningful (12 times larger than normal flow), suggesting that investors punish performance declines more than previously believed.

C. Isolating the Star Rating Effect

Central to our interpretation of a "Morningstar effect" is the confidence that we have isolated a flow response to a Morningstar rating change, that the abnormal flow we report is not due to other factors or is not a response to other performance measure changes. In this section, we report results of a variety of tests designed to address this issue. In all tests, the results indicate a strong case for the isolation of a Morningstar effect.

1. No Abnormal Flow in "Nonevent" Months

If the average abnormal flow we report is due to a response to a Morningstar rating change, then we should expect to see zero average abnormal flow in "nonevent" fund-months in which no rating change occurs. To conduct this test, we identify funds where the star rating has not changed over the eight months prior to time 0 (i.e., the last six months of the estimation period and pre-"nonevent" window). Because we find that a rating change response lasts up to seven months, it is necessary to eliminate funds with a recent rating change. For this test, we group funds according to their constant-over-time star rating and compute ASTAF in the same way that we do for star rating change events. We find that in all but one case, the ASTAF are not statistically different from zero. Only the ASTAF for three-star funds at time 0 is statistically significant ($n = 9{,}894$) and equal to -0.09. Most notably, for four- and five-star funds where we report the largest flow response to a rating change, we find zero flow response in these nonevent months.

2. Is it Morningstar or Morningstar's Algorithm?

At some level, identifying a flow response to a Morningstar rating change cannot be distinguished from a flow response to the change in performance as defined by the Morningstar rating. That is, it is conceptually impossible to prove that investors are following the Morningstar rating per se, rather than recreating the Morningstar algorithm (or something correlated with it) and using this measure in their allocation decisions. While it is hard to imagine why an investor would incur the cost in time and effort to recreate the freely available Morningstar rating, the next test demonstrates that it not the underlying Morningstar algorithm assessing fund performance that appears to drive investor flow, but rather the discrete implementation into a star rating.

Recall that Morningstar star ratings are assigned according to predetermined cutoffs in the distribution of the underlying Morningstar performance measure (e.g., funds in the top 10% of the distribution receive five stars). By replicating Morningstar's algorithm and computing their continuous percentile performance measure, we can identify funds that experience a change in percentile ranking but experience no change in star rating. For example, a fund that moves from the 72nd

to the 85th percentile has improved its Morningstar performance measure but not enough to generate a discrete change in star rating. If investors use a performance measure in their decision-making that is highly correlated with the base Morningstar performance measure but are *not* using the Morningstar star rating, then we would expect to see a flow response to funds experiencing a change in this measure. More specifically, if the underlying performance is really driving flow, then funds with a Morningstar percentile increase should experience positive abnormal flow and funds with a percentile decrease should experience negative abnormal flow, even if the star rating does not change. A finding of no abnormal flow, in contrast, would suggest that it is the star rating itself (and hence, Morningstar's endorsement) that drives flow rather than a coincidental correlation with another measure of performance used by investors.

We run the same test as in the previous analysis of nonevents on a sample of funds that experience no Morningstar star rating change but do experience a change in their underlying Morningstar performance percentile.[18] We split funds into groups based on their Morningstar rating and the sign of the change in the Morningstar percentile, positive or negative. The *t*-tests and nonparametric χ^2 tests are shown in Table 3. Overwhelmingly the results indicate that abnormal flows are insignificantly different from 0. The exceptions to this finding are the *t*-test for one-star funds experiencing an increase in percentile performance (though not supported by the χ^2 test), and the significantly higher percentage of three-star funds with a percentile decrease that experience positive abnormal flow (though not supported by the *t*-test). This test adds, perhaps, the most pointed evidence that it is the Morningstar star rating itself that is generating the abnormal flow seen in the data.

3. Other Tests

An interesting special case of a rating change event is a rating initiation, whereby Morningstar rates a fund for the very first time exactly 36 months after the fund's inception. For this event, all funds share the common status of being unrated and thereby enter the pool of funds considered by investors who rely on Morningstar's recommendation. Appendix D contains the event study results of 1,400 rating initiations using the same event study method as before, but modifying the specification to account for the shorter performance history common to these funds. Consistent with a "Morningstar effect," we find that initial four- and five-star ratings are significantly related to abnormal flow. In particular, a five-star initiation results in an average of $19.9 million in abnormal flow over the following seven months, or 34% above average normal flow for these funds.

As stated earlier, an advantage of our fund-specific benchmark is its ability to control for the influence of all variables included in the benchmark, including

[18]We use data from Morningstar CDs for the period April to December 1997 for this test. We use this period because we have lost access to other CDs in the overall sample period since compiling the original data for the study. For details on calculating the Morningstar percentile performance measure, see Blume (1998). Using this data, we verify that the funds with a change in percentile ranking exceeding the cutoff do indeed change star ratings (e.g., a fund going from the 85th to 92nd percentile will change from four to five stars), confirming our replication of Morningstar's percentiles.

TABLE 3

Average Standardized Abnormal Flow for Funds with a Change in Morningstar Percentile Rank but No Change in Morningstar Star Rating

We use information on Morningstar CDs to compute the Morningstar percentile performance measure in each month from April to December 1997. We then use the 5,385 fund-months in this period where there was no change in the Morningstar star rating over the eight months prior to time 0 and group these observations by star rating in month $t-1$ and by whether the Morningstar percentile performance measure increased or decreased from $t-1$ to 0. For each category we compute the average standardized abnormal flow ($ASTAF_0$) in month 0 in the same manner as in Table 2. The standard error is equal to the cross-sectional standard deviation over N sample events of the standardized abnormal flows, divided by the square root of N. Average standardized abnormal flows significantly different from zero at the 5% level or higher in a two-tailed test are indicated in bold. We also report the percentage of sample events with positive abnormal flows. We indicate in bold the percentage positive that differs from 50% at the 5% significance level or better, using a χ^2 test with one degree of freedom. Similar to Table 2, we delete fund-months that coincide with mergers, observations that do not have the data necessary to compute flows, and the top and bottom 1% of standardized abnormal flows.

Morningstar Star Rating at $t-1$	Change in Morningstar Percentile from $t-1$ to 0	Number of Observations	$ASTAF_0$	t-Stat.	% > 0
1	Negative	181	−1.44	−1.39	47.0
1	Positive	213	**0.61**	1.97	54.2
2	Negative	546	−0.38	−0.55	53.3
2	Positive	661	1.20	1.72	51.2
3	Negative	1,065	0.34	1.13	**53.9**
3	Positive	1,029	−0.43	−0.73	51.8
4	Negative	672	0.14	0.92	50.2
4	Positive	645	1.22	1.39	51.9
5	Negative	209	0.37	1.67	49.8
5	Positive	164	−3.27	−0.92	51.8
Total		5,385			

a fund's performance history. Appendix E presents evidence that our measure of abnormal flow, which we attribute to a star rating change, is indeed net of the effect of alternative performance measures. Specifically, Appendix E contains tests of the ability of changes in Sharpe ratio and changes in Jensen's alpha to explain the variation in abnormal flow relative to Morningstar rating change groups. We do not find support for these alternative contemporaneous measures, either in changes or in levels, being able to explain variation in abnormal flow, conditional on star rating changes. In an unreported robustness check, we also test whether event-window performance of these alternative measures is related to abnormal flow after a rating change but find no support for this. In sum, these tests reinforce the interpretation of an isolated Morningstar effect.

D. Interpretations of Abnormal Flow Patterns and Morningstar's Influence on Investor Decision-Making

We have shown that the Morningstar effect is economically meaningful and statistically robust, implying that a substantial number of investors use star ratings in their fund allocation decision, and that for these investors, Morningstar appears to add value as an information intermediary. Two investor types most likely to value Morningstar's ratings are novice investors who lack the financial expertise to use more sophisticated measures and financial professionals who value the external certification. In this section, we interpret our main findings in the context of what we learn about this set of investors and their decision-making.

Even though we generally observe a negative response to rating downgrades, we distinctly find no response to a downgrade from five to four stars. One explanation is that fund companies aggressively promote the perception that four-star funds are also high quality. In support of this, we find that of the 86 distinct funds advertising their star rating in June 1999, 36 (42%) are four-star funds and 32 (37%) are five-star funds.[19] In addition, an Internet search reveals that fund companies issue press releases when their funds are upgraded to four stars as well as when upgraded to five stars. Another explanation is that this effect largely reflects the behavior of financial advisor clients if, for example, five-star funds added to 401(k) plans or advisor "recommended lists" are only removed if performance lags either more dramatically (e.g., a fall to three stars) or more permanently. Alternatively, the tax lock-in effect discussed in Bergstresser and Poterba (2002) might explain the lack of response exclusive to five- to four-star downgrades, as investors in these funds might be reluctant to realize the capital gains of these recent good performers.

Finally, our event study also provides interesting new evidence on the timeliness of investor response to fund quality information. We find a significant flow response beginning in event month 0, indicating that investor response is detectable at the same monthly frequency that Morningstar releases new ratings. While somewhat surprising, the quick response by some investors is consistent with anecdotal evidence. We have learned from an official at Morningstar that traffic on their Web site spikes on the days surrounding the release of new ratings. Consistent with this apparent investor interest, MSN's MoneyCentral Web site allows investors to screen on funds that Morningstar upgraded since last month. Morningstar's Web site lists rating changes as "news alert" items on each fund's information page, and a registered user can elect to be sent an e-mail when a rating changes. Our results show that this vigilant investor group is indeed large enough to measurably affect flow.

Our results also imply the existence of a substantial group of investors who respond to rating changes with a considerable lag. We report that abnormal flow is significant out to month six and confirm in unreported month-by-month results that this also holds for a subsample of funds that do not experience any further rating changes. This slow response might be driven in part by investors who have different schedules for evaluating funds, or who rely on the somewhat stale information in fund advertising or magazine summary lists. For example, of 37 mutual fund advertisements touting star ratings in June 1999, 12 featured star ratings as of April 1999, 20 as of March, one as of February, and only four reflected ratings as of May 1999.[20] In addition, one of Morningstar's popular products is a quarterly subscription to its CD-ROM database. Subscribers relying on this source would necessarily get the rating change with a lag. Thus, news of a rating change, or rating status, may take several months to reach some investors.

[19]To investigate this we sampled all mutual fund ads featuring domestic equity funds appearing in *Money*, *Barron's,* and *The Wall Street Journal's* Money & Investing Section during June 1999.

[20]The June issues of *Barron's*, *Money* magazine, or *The Wall Street Journal* contained 187 mutual fund advertisements, 43 of which featured Morningstar star ratings. Of these, 37 are in our sample of domestic equity funds.

E. Morningstar's July 2002 Rating Algorithm Change

In July 2002, Morningstar substantially changed its star rating algorithm, most notably by ranking mutual funds according to 48 style categories (e.g., small-cap value) rather than pooling all domestic equity funds into one ranking universe.[21] Thus, during our sample period, star ratings are more likely to reflect the performance of the style category rather than the skill of the active manager. A practical implication of the algorithm change is that funds that were in outperforming styles in July 2002 (i.e., growth) but underperformed other funds in their same style category immediately lost stars, while funds in underperforming styles (i.e., value) tended to gain stars.[22]

We expect that our main results would be qualitatively similar in a more recent period under the new rating methodology. Specifically, given that the core users of Morningstar's rating, novice investors and financial advisors, are likely to continue to value the same ease of use and certification attributes of Morningstar ratings as before, we fully expect star ratings to remain an independent driver of investor flow. Indeed, many financial advisors have applauded the algorithm change, citing their new ability to recommend funds they believe to belong in their client's portfolio (e.g., natural resource funds) but that were perennially below three stars.[23] Thus, a likely result of the change in Morningstar's rating algorithm for core users is that allocating toward five-star funds is no longer synonymous with allocating toward the styles that have recently done well, as now five-star ratings are evenly distributed across styles by construction. This should lead to greater style diversity in the portfolios of Morningstar's core users, as well as a weaker overall relation between their fund allocations and total return performance. Thus, in aggregate, the weakened relation between total return and star rating should also weaken the relation between fund flow and total return performance levels. The extent to which Morningstar star ratings have the power to move flow in opposition to total return ranking is an interesting subject for future research.

F. Applications of Fund-Specific Benchmarking

Our fund-specific benchmarking method is applicable to other research questions involving mutual funds. Straightforward applications include studies of the investor response to well-defined mutual fund events, such as mergers, manager changes, or SEC investigations. In each case, a model of normal flow, or a fund-specific benchmark estimated in the time-series, provides an estimate of the flow that the fund would be expected to gain or lose had the event not occurred. As such, it provides the best estimate of a flow response to the event. More generally,

[21] Morningstar also changed to a risk-adjustment methodology based on certainty equivalents rather than the previously employed downside risk measure. However, this change likely had a lesser effect on fund ratings than the move to style-based rankings.

[22] See Morey and Gottesman (2006) for a description of the algorithm change and evidence of the enhanced predictive power of star ratings after July 2002.

[23] See "Morningstar Revises Star Ratings," *Financial Planning* (June 2002), p. 28.

a fund-specific measure of performance sensitivity can be useful in any study with cross-sectional predictions regarding variation in investors' performance monitoring across funds. Qian (2005) and Chen, Goldstein, and Jiang (2008) are examples of mutual fund governance studies that relate the degree of investor vigilance (i.e., performance sensitivity) to the need for monitoring by the board of directors.

One widely studied application of the flow-performance relation that might benefit from a fund-specific measure of performance sensitivity is the study of managerial incentives and behavior. Brown, Harlow, and Starks (1996) and Chevalier and Ellison (1997) were the first to investigate the incentives facing fund managers implicit in the relation between flow and performance and to test whether managers respond to these incentives. Subsequent empirical studies have revisited the manager response to these incentives with mixed support.[24] Because most of these tests are based on a cross-sectional analysis, we suspect that the substantial heterogeneity in the time-series relation between flow and performance is partly responsible for the mixed evidence. Including all funds in these tests without taking account of the performance sensitivities of their investor clienteles adds noise and reduces the power of such tests. In particular, funds with relatively performance-insensitive clienteles have much less incentive to take actions to increase performance because portfolio performance improvements do not lead to increased assets for these managers. Instead, by estimating the flow-performance sensitivity for each fund, our approach suggests a simple but more powerful method for identifying subsamples of managers with stronger implicit incentives.[25]

The features of the Morningstar rating system that allow us to measure its influence apart from other performance metrics are the same features that may make it a powerful incentive for fund managers.[26] Our findings suggest specific points in the performance distribution where the incentive to improve a fund's Morningstar rating are strongest. Rather than separating funds by performance quartile, or relative to the median fund, as in Brown, Harlow, and Starks (1996), our findings indicate that incentives are stronger along the boundaries between

[24]Brown, Harlow, and Starks (1996) and Chevalier and Ellison (1997) find evidence that mutual fund managers respond to the implicit incentives in the convex relation between flow and performance and systematically alter the riskiness of their portfolios during the last part of the year. Busse (2001) and Koski and Pontiff (1999) argue that the empirical relation between performance and risk is driven by methodology or mechanically by flows and is not the result of incentives. Chen and Pennacchi (2009) provide an alternative test based on tracking error.

[25]There is reason to believe that performance sensitivity does indeed vary substantially both across funds and over time for a given fund. For example, Mitchell, Mottola, Utkus, and Yamaguchi (2006) report that 80% of 401(k) investors in their sample did not make a single account trade over a two-year holding period, implying that funds with a large percentage of 401(k) investors would have much fewer performance-sensitive investors relative to other funds. In contrast, Goetzmann and Massa (2002) report that the median investor in a large S&P 500 index fund made one trade per year. Similarly, Christoffersen, and Musto (2002) argue that a fund's performance sensitivity should fall after a performance decline because the performance-sensitive investors withdraw their assets, leaving only the relatively insensitive investors in the fund.

[26]The more dominant a rating system is among the other information inputs decision-makers use, and the easier the rating's effect is to measure, the greater the incentive for fund managers to focus their energies on improving their rating. Crowley (2001), in an article geared to practitioners, describes a trading strategy for improving one's Morningstar rating.

star ratings. In particular, the finding that upgrades to five stars are handsomely rewarded, but downgrades to four stars often do not elicit an outflow response suggests that the relation between flow and performance is especially convex at the boundary between four and five stars. Managers of four-star funds that are close to the five-star cutoff might have the greatest incentive to take actions that have a chance of improving their Morningstar ratings, while managers attaining five-star status have little incentive to change their behavior. The large penalty that we find for dropping to three stars also suggests that managers of four-star funds close to this precarious cutoff have a strong incentive to guard their position relative to other funds.

Our evidence that some investors respond immediately to *monthly* rating changes and that the response continues for seven months implies that the tournament for investor flow more likely operates on an ongoing basis, rather than at the annual frequencies assumed in previous studies. Moreover, in an unreported test we find no economically significant difference in the flow response for performance changes that occur at year-end. These findings suggest that more work in the theoretical modeling of managerial behavior that better captures the dynamic and repeated game aspects of the tournament may be fruitful. A theoretical model of managerial incentives in such an environment could point the way toward new empirical tests of strategic managerial behavior.

V. Conclusion

In the market for colleges, cars, and mutual funds, the choice set is large and the set of relevant attributes that translate into quality are difficult for the typical decision-maker to assess. Information intermediaries, such as *U.S. News and World Report* and Morningstar, can add value in this environment by providing low-cost summary product quality information from a reputable and unbiased source. Using discrete informational events occurring at regular monthly intervals, we provide evidence that Morningstar's repackaging of fund quality information in the form of a simple one- to five-star rating has a large independent influence on mutual fund flow, and therefore on investor allocation decisions. Overall, we conclude that investors view Morningstar ratings as informative quality measures and confirm Morningstar's reputation as an influential player in the market for mutual funds.

Through an event study of fund flow, we uncover patterns that are difficult to infer from cross-sectional analysis, namely, the timeliness and magnitude of the investor response to a change in a fund's own performance. In contrast to prior studies, we find evidence that investors significantly punish funds whose performance drops below the top one-third of funds to a three-star rating, and that some investors respond immediately to this change in performance. Similar to other studies, we find a disproportionate positive response to an upgrade to the highest five-star rating. Together this implies that fund managers with the same four-star rating might have very different incentives to increase risk depending on whether they are closer to the five- or three-star cutoff. Direct tests of this effect on managerial incentives await future research.

APPENDIX A

Morningstar Style Categories and the Construction of Style Benchmarks

We use Morningstar's style category as a style identifier for sample funds. Morningstar divides the domestic equity star rating category peer group into 19 mutually exclusive categories. We list the 19 categories in the table below. For each fund in our sample, we assign it to the category Morningstar places it in as of October 1999. (For dead funds we use the Morningstar category reported in the month before its disappearance.) This procedure should result in an accurate categorization, since according to Morningstar's definition, the style category assignment is based on the fund's average investment style over the previous three years. At each point in time, Morningstar assigns a fund to a style category based on the style and size of the stocks the fund owns.

The style flow we use in our benchmarking procedure is an in-sample measure, using all funds in the sample within the same Morningstar category. Thus, large-cap growth funds are benchmarked using their sensitivity to the overall flow to large-cap growth funds. Because our sample is constructed to include all domestic equity funds that were rated at any time during the sample period, the only funds missing from the aggregate style flow measure are funds coming into existence after November 1996 (and thus not rated by Morningstar throughout our sample period). Below we report the number of funds in each category that we use to compute aggregate style flows.

Morningstar Category	Number of Funds
Large-cap growth	376
Large-cap blend	552
Large-cap value	406
Medium-cap growth	279
Medium-cap blend	168
Medium-cap value	147
Small-cap growth	246
Small-cap blend	132
Small-cap value	123
Specialty-health	30
Specialty-technology	51
Specialty-utilities	83
Specialty-communications	16
Specialty-financials	25
Specialty-real estate	59
Specialty-natural resources	43
Specialty-precious metals	40
Convertible bond	47
Domestic hybrid	564

APPENDIX B

Estimation Period Statistics of the Event-Study Methodology

Appendix B reports diagnostic statistics on the event-study benchmark model used to calculate abnormal flow for star rating change events. The benchmark model, analogous to a market model regression, regresses a fund's monthly flow on aggregate flow at time t to funds in its same style group, its own time $t-1$ flow, its own time $t-1$ return, its change in Jensen's alpha from $t-2$ to $t-1$, and its change in Jensen's alpha from $t-2$ to $t-1$ squared. We separately fit this benchmark for each of the 10,651 rating change events, and define the estimation period as months $(-26, -3)$. The pairwise cross-sectional correlations reported in the table are computed by pooling observations over all funds and months in the estimation period. Panel A contains the sample distribution of the R^2 of the benchmark regressions, and suggests that our relatively parsimonious model results in an overall good fit. We also report in the table that the number of rating change events per month is reasonably dispersed throughout the sample period, indicating that calendar time clustering is not a concern. Panel B contains cross-sectional correlations, pooled over all funds and over all months in the estimation period, of both predicted (normal) and residual flow with the benchmark regressors, and with fund size. Del Guercio and Tkac (2002), among others, find a positive cross-sectional relation between fund flow and size. Because a fund's size relative to its peers remains roughly constant over time, in a time-series benchmark regression, the effect of fund size is likely to be absorbed into the intercept. The strong positive correlation between predicted flows and fund size, along with the lack of a correlation between the benchmark residuals and fund size, implies that larger funds have higher normal flow but not higher estimated levels of abnormal flow. This is evidence that the benchmark captures the established positive relation between flow and fund size. As expected from a well-specified benchmark, the cross-sectional correlation between the estimation residuals and all of the regressors are not significantly different from zero. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Sample Distribution of the R^2 of the Benchmark Regressions

	Estimated Benchmark Regressions		
	25th Percentile	Median	75th Percentile
R^2	0.25	0.43	0.66
No. of events per month	251	293	363

Panel B. Cross-Sectional Correlations Pooled Over All Funds and Over All Months

Pairwise Correlations	Style Flow	Lagged Return	Lagged Flow	Change in Alpha	Change in Alpha Squared	Fund Size
Standardized residuals	–0.000	0.0002	0.0004	0.0001	–0.0002	0.003
Normal flow	0.096***	0.020***	0.030***	0.033***	0.004***	0.042***

APPENDIX C

Morningstar Star Rating Changes of Domestic Equity Funds

Panel A of Appendix C reports the average cumulative standardized abnormal flow ($ACSTAF_t$) for each date t in the event window for the subsample of funds with no change in star rating in the eight months prior to event month 0. We first compute the cumulative standardized abnormal flow for each fund by summing the standardized abnormal flow from 0 to t, and then dividing by the square root of the number of months used in the cumulation. We then average these standardized cumulated abnormal flows over the N sample events in each star rating change group to obtain ACSTAF at each event date t. ACSTAF significantly different from zero at the 5% level or higher in a two-tailed test are indicated in bold. For each event date, we also report the percentage of sample events with positive cumulative standardized abnormal flows. We indicate in bold the percentage positive that differs from 50% at the 5% significance level or better, using a χ^2 test with one degree of freedom. Comparisons of the sample size in Table 2 Panels C and D indicates the number of funds that are deleted due to a previous rating change.

Panel A. Average Cumulative Standardized Abnormal Flow for the Subsample of Rating Upgrades with No Rating Change in the Previous Eight Months

Event	From 1 to 2 Stars ($n = 173$)			From 2 to 3 Stars ($n = 446$)			From 3 to 4 Stars ($n = 484$)			From 4 to 5 Stars ($n = 234$)		
Month	$ACSTAF_t$	*t*-Stat.	$\% > 0$	$ACSTAF_t$	*t*-Stat.	$\% > 0$	$ACSTAF_t$	*t*-Stat.	$\% > 0$	$ACSTAF_t$	*t*-Stat.	$\% > 0$
0	**0.35**	2.01	**60.1**	**0.51**	3.36	**58.3**	**0.53**	2.86	52.7	**0.65**	3.96	**57.3**
1	**0.66**	2.53	53.9	**1.07**	2.91	**59.5**	**0.82**	2.83	**55.5**	**1.06**	5.01	**65.1**
2	0.60	1.84	57.4	**1.30**	4.53	**60.8**	**0.90**	3.36	54.2	**1.58**	5.12	**66.7**
3	0.73	1.80	56.3	**1.35**	4.76	**60.7**	**1.80**	3.41	**55.9**	**1.86**	4.75	**61.2**
4	**1.02**	2.05	**58.3**	**1.88**	4.34	**64.4**	**1.79**	3.99	**59.4**	**2.69**	2.54	**61.5**
5	**1.52**	2.22	54.6	**2.24**	5.12	**64.0**	**1.93**	5.02	**57.8**	**3.44**	3.02	**62.5**
6	**1.96**	2.12	**58.8**	**2.47**	5.42	**65.8**	**2.05**	5.49	**60.2**	**3.61**	3.25	**61.8**

Panel B. Average Cumulative Standardized Abnormal Flow for the Subsample of Rating Downgrades with No Rating Change in the Previous Eight Months

Event	From 2 to 1 Star ($n = 216$)			From 3 to 2 Stars ($n = 410$)			From 4 to 3 Stars ($n = 405$)			From 5 to 4 Stars ($n = 198$)		
Month	$ACSTAF_t$	*t*-Stat.	$\% > 0$	$ACSTAF_t$	*t*-Stat.	$\% > 0$	$ACSTAF_t$	*t*-Stat.	$\% > 0$	$ACSTAF_t$	*t*-Stat.	$\% > 0$
0	–0.15	–1.25	48.1	**–0.31**	–2.36	**44.9**	**–0.65**	–3.95	**38.0**	0.10	0.46	49.0
1	**–0.31**	–1.99	44.0	–0.25	–1.23	**45.0**	**–0.80**	–3.55	**42.9**	0.29	0.96	46.2
2	**–0.46**	–2.79	**42.9**	**–0.57**	–2.80	**41.0**	**–0.86**	–2.59	**38.1**	–0.13	–0.43	45.5
3	**–0.52**	–2.50	**38.9**	**–0.74**	–2.83	**40.2**	**–1.07**	–2.92	**37.5**	–0.12	–0.33	**42.7**
4	**–0.84**	–3.48	**39.8**	**–0.97**	–3.45	**39.8**	**–0.90**	–2.08	**39.3**	0.58	0.85	45.9
5	**–0.90**	–3.58	**33.7**	**–1.26**	–4.41	**38.6**	**–1.16**	–2.76	**36.7**	0.27	0.49	42.8
6	**–1.11**	–3.84	**32.5**	**–1.21**	–3.95	**40.4**	–0.83	–1.76	**37.7**	0.63	0.98	46.0

APPENDIX D

Initial Morningstar Star Ratings of Domestic Equity Funds

Average standardized abnormal flow ($ASTAF_t$) reported below is averaged across domestic equity funds within the same initial star rating group. Morningstar assigns an initial star rating in the 36th month of a fund's existence. We define standardized abnormal flow in month t as the actual dollar flow in month t minus the normal, or expected, flow standardized by the estimated forecast variance of the normal flow. Normal flow is based on a market-model regression whereby a fund's monthly flow is regressed on aggregate flow at time t to funds in its same style group, its time $t-1$ flow, its time $t-1$ return, and its time $t-1$ return squared. The estimation period for computing the market-model parameters is months $(-26,-3)$. The standard error is equal to the cross-sectional standard deviation over N sample events of the standardized abnormal flows, divided by the square root of N. $ASTAF_t$ significantly different from zero at the 5% level or higher in a two-tailed test are indicated in bold. For each event date, we also report the percentage of sample events with positive standardized abnormal flows. We indicate in bold the percentage positive that differs from 50% at the 5% significance level or better, using a χ^2 test with one degree of freedom. The number of observations differs from Panel C of Table 1 because we delete fund-months that coincide with mergers, observations that do not have the data necessary to compute flows, and the top and bottom 1% of standardized abnormal flows at each event date.

Panel A. Average Standardized Abnormal Flow

Event	1 Star ($n=141$)			2 Stars ($n=344$)			3 Stars ($n=527$)			4 Stars ($n=277$)			5 Stars ($n=154$)		
Month	$ASTAF_t$	*t*-Stat.	$\%>0$	$ASTAF_t$	*t*-Stat.	$\%>0$	$ASTAF_t$	*t*-Stat.	$\%>0$	$ASTAF_t$	*t*-Stat.	$\%>0$	$ASTAF_t$	*t*-Stat.	$\%>0$
0	0.12	0.84	49.6	–0.15	–1.40	**41.6**	0.01	0.08	**45.2**	**0.42**	2.65	49.8	**0.54**	2.13	55.8
1	–0.16	–0.83	42.9	–0.16	–1.49	**44.0**	0.17	1.84	47.2	**0.38**	2.50	51.6	**1.38**	3.81	**66.2**
2	–0.06	–0.37	48.5	–0.09	–0.78	**42.2**	0.12	1.20	46.7	0.03	0.21	48.9	**0.79**	2.48	56.6
3	0.18	0.90	44.4	–0.08	–0.85	**41.5**	0.12	1.24	**44.8**	**0.37**	2.18	51.9	0.29	0.99	51.2
4	0.01	0.07	47.3	0.05	0.47	46.5	–0.07	–0.72	**44.7**	0.23	1.14	48.2	0.64	1.65	54.0
5	0.27	0.96	50.8	0.04	0.37	47.4	–0.03	–0.30	**41.4**	0.31	1.49	51.0	0.26	1.15	57.6
6	0.35	1.38	55.3	0.29	1.68	46.1	–0.16	–1.35	**44.7**	**0.51**	2.46	49.8	**0.68**	1.94	52.2

Panel B reports the average cumulative standardized abnormal flow ($ACSTAF_t$) for each date t in the event window. We first compute the cumulative standardized abnormal flow for each fund by summing the standardized abnormal flow from 0 to t, and then dividing by the square root of the number of months used in the cumulation. We then average these over the N sample events to obtain ACSTAF at each event date t. ACSTAF significantly different from zero in a two-tailed test, and the percentage positive different from 50% using a X^2 test with one degree of freedom, both at the 5% significance level or better, are indicated in bold.

Panel B. Average Cumulative Standardized Abnormal Flow

Event	1 Star ($n=141$)			2 Stars ($n=344$)			3 Stars ($n=527$)			4 Stars ($n=277$)			5 Stars ($n=154$)		
Month	$ASTAF_t$	*t*-Stat.	$\%>0$	$ASTAF_t$	*t*-Stat.	$\%>0$	$ASTAF_t$	*t*-Stat.	$\%>0$	$ASTAF_t$	*t*-Stat.	$\%>0$	$ASTAF_t$	*t*-Stat.	$\%>0$
0	0.12	0.84	49.6	–0.15	–1.40	**41.6**	0.01	0.08	**45.2**	**0.42**	2.65	49.8	**0.54**	2.13	55.8
1	0.02	0.07	45.7	**–0.34**	–2.07	**39.5**	0.24	1.22	47.2	**0.56**	2.74	52.3	**1.28**	3.74	**61.6**
2	–0.21	–1.41	43.4	–0.22	–1.61	**37.5**	0.10	0.82	45.5	**0.66**	2.72	53.8	**1.30**	3.44	**62.5**
3	–0.04	–0.18	43.6	–0.19	–1.36	**38.9**	0.15	1.00	46.9	**0.86**	3.28	53.5	**1.51**	2.90	**64.3**
4	–0.03	–0.13	48.1	–0.14	–0.86	**42.1**	0.15	0.85	46.5	**0.80**	2.94	53.0	**1.44**	2.96	**60.3**
5	0.25	0.80	49.2	–0.22	–0.74	**43.6**	–0.06	–0.37	**45.3**	**1.08**	2.98	50.6	**1.45**	2.89	55.9
6	0.10	0.38	48.0	–0.01	–0.03	45.8	–0.08	–0.43	**44.4**	**1.45**	3.18	50.7	**1.59**	2.81	53.9

APPENDIX E

Determinants of the Six-Month Cumulative Standardized Abnormal Flow Response to a Change in Star Rating: Contemporaneous Change in Jensen's Alpha and Sharpe Ratio versus Morningstar Star Rating

Appendix E reports the results of a two-way analysis of variance of cumulative standardized abnormal flows from month 0 to 6 ($CSTAF_6$) for the full sample of funds experiencing a Morningstar star rating change ($N = 7{,}060$). We separately compare the explanatory power of a fund's Morningstar rating change at $t = 0$ to either a change in a fund's Jensen's alpha group or a change in a fund's Sharpe ratio group, measured at the same time as the Morningstar rating change. We present one way of defining Jensen's alpha or Sharpe ratio groups: "Morningstar quintiles," where the top 10% of alphas (Sharpe ratios) are in quintile 5, the next 22.5% in quintile 4, the next 35% in quintile 3, the next 22.5% in quintile 2, and the bottom 10% in quintile 1. We also define alpha and Sharpe groups as equal-sized quartiles and equal-sized quintiles, with very similar results (not reported). For each comparison below labeled (1) and (2), we report the *F*-statistic and the corresponding *p*-value for the test of the null hypothesis that the sample means are identical across groups in the one rating system, conditional on the other rating system. The *F* test has $(n_1 - 1, n - p)$ degrees of freedom, where n_1 equals the number of change groups in the rating system being tested for significance, n equals the number of observations, and p equals the number of change groups in the both rating systems. *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Categorizations	Type II Sum of Squares *F*-Statistic	Prob > *F*
(1)	Jensen's alpha change ("Morningstar quintiles")	0.83	0.617
	Morningstar star rating change groups	9.26***	0.000
(2)	Sharpe ratio change ("Morningstar quintiles")	0.97	0.470
	Morningstar star rating change groups	10.22***	0.000

This table presents analysis of variance tests that compare the ability of Morningstar star rating changes to explain variation in abnormal flow against the ability of alternative contemporaneous performance measures. Specifically, the table reports the results of two-way analysis of variance (Type II sum of squares) tests on seven-month cumulative standardized abnormal flows ($CSTAF_6$) following a fund's Morningstar rating change. To put alpha and Sharpe measures on equal footing with Morningstar's star rating in terms of its discreteness, we examine various groupings of funds based on rankings by these two measures. We report results using a "Morningstar-type" grouping where the top 10% of alphas (Sharpe measures) are in quintile 5, the next 22.5% in quintile 4, the next 35% in quintile 3, the next 22.5% in quintile 2, and the bottom 10% in quintile 1, although the results are similar if we use an equal-sized quartile or quintile grouping instead. These groupings might be thought of as a simulated rating system, with each representing a different method investors might use to translate the continuous alpha or Sharpe measures onto a coarser grid of fund quality.

Given the low correlation between changes in rating and changes in these performance measures reported in Table 1, and the fact that our benchmark specification includes a monthly change in alpha as a control variable, we expect abnormal flow to be unrelated to changes in Jensen's alpha. Consistent with this prediction, the *F*-statistics indicate that conditional on the star rating change, changes in an alpha rating cannot explain variation in abnormal flow. Although not explicitly controlled for in the benchmark, changes in the Sharpe measure rating also cannot explain variation in abnormal flow. The star rating change, in contrast, remains significant even after controlling for these contemporaneous changes in simulated alpha and Sharpe ratings. Thus, while all three performance measures are based on the same underlying set of returns data, and consequently may be changing at the same time, these tests indicate that our benchmark has successfully purged any flow effect of a change in alpha or Sharpe measure.

We perform similar tests on alpha and Sharpe measures in levels rather than changes (not reported). We cannot reject the null hypothesis that the conditional average abnormal flows are equivalent across alpha and Sharpe measure groups. At the same time, Morningstar's star rating remains highly significant conditional on any of the alpha or Sharpe categorizations; we can reject the null hypothesis that Morningstar star rating change does

not matter (i.e., that average abnormal flows are equivalent across all star ratings) at the 1% confidence level. Therefore, the abnormal flow around star rating changes of Table 2 are not driven by the level of fund alphas and Sharpe measures. In particular, the large positive abnormal flow following an upgrade from four to five stars is not driven by high alpha or high Sharpe measure funds. These results are consistent with our earlier analysis suggesting that our benchmark successfully incorporates a fund's long-term performance history. Similar tests on the sample of rating initiations yield the same results. The flow response to an initial rating is not explained by either the level or the change in fund alphas and Sharpe measures.

References

Bergstresser, D., and J. Poterba. "Do After-Tax Returns Affect Mutual Fund Inflows?" *Journal of Financial Economics,* 63 (2002), 381–414.

Blake, C. R., and M. R. Morey. "Morningstar Ratings and Mutual Fund Performance." *Journal of Financial and Quantitative Analysis,* 35 (2000), 451–483.

Blume, M. "An Anatomy of Morningstar Ratings." *Financial Analysts Journal*, 54 (1998), 1–9.

Boehmer, E.; J. Musumeci; and A. Poulsen. "Event-Study Methodology under Conditions of Event-Induced Variance." *Journal of Financial Economics,* 30 (1991), 253–272.

Brown, K.; W. V. Harlow; and L. Starks. "Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry." *Journal of Finance,* 51 (1996), 85–110.

Brown, S. J., and J. B. Warner. "Measuring Security Price Performance." *Journal of Financial Economics,* 8 (1980), 205–258.

Brown, S. J., and J. B. Warner. "Using Daily Stock Returns: The Case of Event Studies." *Journal of Financial Economics,* 14 (1985), 3–31.

Busse, J. "Another Look at Mutual Fund Tournaments." *Journal of Financial and Quantitative Analysis*, 36 (2001), 53–73.

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

Chen, H., and G. Pennacchi. "Does Prior Performance Affect a Mutual Fund's Choice of Risk? Theory and Further Empirical Evidence." *Journal of Financial and Quantitative Analysis*, forthcoming (2009).

Chen, Q.; I. Goldstein; and W. Jiang. "Directors' Ownership in the U.S. Mutual Fund Industry." *Journal of Finance*, forthcoming (2008).

Christoffersen, S., and D. Musto. "Demand Curves and the Pricing of Money Management." *Review of Financial Studies,* 15 (2002), 1499–1524.

Crowley, P. "Improve Your Morningstar Rating Using Options." *Journal of Investing*, 10 (2001), 73–86.

Del Guercio, D., and P. A. Tkac. "The Determinants of the Flow of Funds of Managed Portfolios: Mutual Funds versus Pension Funds." *Journal of Financial and Quantitative Analysis,* 37 (2002), 523–557.

Dodd, P., and J. B. Warner. "On Corporate Governance: A Study of Proxy Contests." *Journal of Financial Economics,* 11 (1983), 401–438.

Gallaher, S.; R. Kaniel; and L. Starks. "Madison Avenue Meets Wall Street: Mutual Fund Families, Competition, and Advertising." Working Paper, Duke University (2006).

Goetzmann, W., and M. Massa. "Daily Momentum and Contrarian Behavior of Index Fund Investors." *Journal of Financial and Quantitative Analysis,* 37 (2002), 375–389.

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

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

Ippolito, R. "Consumer Reaction to Measures of Poor Quality." *Journal of Law and Economics,* 35 (1992), 45–70.

Ivkovic, Z., and S. Weisbenner. "'Old' Money Matters: The Sensitivity of Mutual Fund Redemption Decisions to Past Performance." Working Paper, University of Illinois (2006).

Jones, M., and T. Smythe. "The Information Content of Mutual Fund Print Advertising." *Journal of Consumer Affairs,* 37 (2003), 22–41.

Koski, J., and J. Pontiff. "How are Derivatives Used? Evidence from the Mutual Fund Industry." *Journal of Finance,* 54 (1999), 791–816.

Morey, M. R. "Mutual Fund Age and Morningstar Ratings." *Financial Analysts Journal,* 58 (2002), 56–63.

Morey, M. R., and A. A. Gottesman. "Morningstar Mutual Fund Ratings Redux." Working Paper, Pace University (2006).

Patel, J.; R. Zeckhauser; and D. Hendricks. "Investment Flows and Performance: Evidence from Mutual Funds, Cross-Border Investments and New Issues." In *Japan, Europe and the International Financial Markets: Analytical and Empirical Perspectives,* R. Satl, R. Levitch, and R. Ramachandran, eds. New York, NY: Cambridge University Press (1994).

Pozen, R. C. *The Mutual Fund Business*. Cambridge, MA: MIT Press (1998).

Qian, M. "Whom Can You Trust? A Study of Mutual Fund Governance." Working Paper, Boston College (2005).

Sharpe, W. "Morningstar's Performance Measures." Working Paper, Stanford University (1997).

Sharpe, W. "Morningstar's Risk-Adjusted Ratings." *Financial Analysts Journal*, 54 (1998), 21–33.

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