NBER WORKING PAPER SERIES

RATINGS-DRIVEN DEMAND AND SYSTEMATIC PRICE FLUCTUATIONS

Itzhak Ben-David
Jiacui Li
Andrea Rossi
Yang Song

Working Paper 28103
http://www.nber.org/papers/w28103

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2020, Revised October 2021

We thank Sylvester Flood (Morningstar), Paul Kaplan (Morningstar), Nick Barberis, John Campbell, Alex Chinco, Darrell Duffie, Bing Han, Charles Lee, Martin Lettau, Juhani Linnainmaa, Andrei Shleifer, Xin Wang, and Motohiro Yogo for helpful comments. We thank seminar participants at The Ohio State University, the University of Utah, the University of Washington, Arrowstreet Capital, and Hong Kong University of Science and Technology as well as participants at the NBER Summer Institute (Asset Pricing), the NBER Behavior Finance Workshop, the Pacific Northwest Finance Conference, and the 2021 AFA Annual Meeting for comments. Ben-David is a cofounder and a partner in an investment advisor. Send correspondence to Itzhak Ben-David, ben-david.1@osu.edu. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Ratings-Driven Demand and Systematic Price Fluctuations
Itzhak Ben-David, Jiacui Li, Andrea Rossi, and Yang Song
NBER Working Paper No. 28103
November 2020, Revised October 2021
JEL No. G11,G24,G41

ABSTRACT

We show that mutual fund ratings generate correlated demand that creates systematic price fluctuations. Mutual fund investors chase fund performance via Morningstar ratings. Until June 2002, funds pursuing the same investment style had highly correlated ratings. Therefore, rating-chasing investors directed capital into winning styles, generating style-level price pressures, which reverted over time. In June 2002, Morningstar reformed its methodology of equalizing ratings across styles. Style-level correlated demand via mutual funds immediately became muted, significantly altering the time-series and cross-sectional variation in style returns.

Itzhak Ben-David
The Ohio State University
Fisher College of Business
606A Fisher Hall
Columbus, OH 43210-1144
and NBER
ben-david.1@osu.edu

Jiacui Li
David Eccles School of Business
University of Utah
SFEBB 8123, 1655 Campus Center Dr
Salt Lake City, UT 84112
jiacui.li@eccles.utah.edu

Andrea Rossi
University of Arizona
Eller College of Management
Department of Finance
1130 E. Helen St.
Tucson, AZ 85721
rossi2@email.arizona.edu

Yang Song
Foster School of Business
University of Washington
4273 E Stevens Way NE
Seattle, WA 98195
songy18@uw.edu

Stock ownership by retail-owned mutual funds in the United States has been steadily rising since the 1980s, reaching about 25% of the entire market capitalization in the early 2000s. With such a high ownership rate, the financial advice that guides households' capital investments could play a central role in driving flows and shaping financial markets. Indeed, previous studies have found that mutual fund flows respond to past performance in the form of external ratings (Del Guercio and Tkac, 2008; Kaniel and Parham, 2017; Evans and Sun, 2021; Reuter and Zitzewitz, 2021) and that mutual fund flows can generate large price pressure in the underlying stocks (Coval and Stafford, 2007; Lou, 2012). Hence, it is important to test how the advice consumed by households filters into security prices and, more importantly, whether this advice translates into systematic price patterns.

Among the different types of financial advice that U.S. mutual fund investors follow, Morningstar star ratings are perhaps the most prominent. Soon after introducing the ratings in 1985, Morningstar became the industry leader in rating mutual funds, with millions of subscribed investors who obtain ratings either directly or through investment advisors. Moreover, investment platforms and fund families prominently feature Morningstar ratings as part of the information they display to investors. While ratings are a "backward-looking measure of a fund's past performance"[1] and therefore are not necessarily useful as forward-looking guidance, they are important inputs in households' investment decisions. Discrete changes in Morningstar star ratings drive mutual fund flows (Del Guercio and Tkac, 2008; Reuter and Zitzewitz, 2021), and Morningstar ratings are the prime determinant of flows in the cross-section of equity mutual funds (Ben-David, Li, Rossi, and Song, 2021b).

In this study, we test whether rating-driven demand causes *systematic* price fluctuations in the stock market. Investors heavily rely on Morningstar ratings throughout our sample period (1991–2018). As a consequence, stocks experience price pressures based on the rating changes of the mutual funds that own them. Because Morningstar reformed its rating methodology in June 2002, we can trace changes in systematic price patterns back

[1] From Morningstar's website (`https://www.morningstar.com/company/morningstar-ratings-faq`), retrieved December 20, 2020.

to rating-driven mutual fund flows. Our analysis shows that the 2002 Morningstar reform dramatically changed the allocation of investors' capital across styles, which in turn significantly altered the time-series and cross-sectional variation of style returns. Overall, our results show that Morningstar ratings, a form of investment advice provided to households, can cause systematic and persistent price fluctuations in the stock market.

The key mechanism by which Morningstar influences capital flows is best explained through the lens of its rating methodology. Prior to June 2002, Morningstar ratings were broadly aligned with mutual funds' past performance. In that period, Morningstar rated all mutual funds—regardless of their style tilts—based on their performance ranking across the *entire universe* of U.S. equity funds, with minor adjustments for loads and volatility. Because a significant fraction of fund performance is determined by style exposure (e.g., small-cap or growth-oriented), funds that pursued similar investment style mandates had highly correlated ratings. Under Morningstar's pre-June 2002 methodology, investing in high-rated funds was broadly equivalent to chasing funds' past returns.

In June 2002, Morningstar implemented a simple yet impactful revision of its rating methodology. Instead of ranking all equity funds against each other, Morningstar began benchmarking funds against peer funds *within* their style. The style-peer groups are based on the well-known Morningstar three-by-three "style box" (value/blend/growth $\times$ small/mid/large cap). By design, the revised methodology removes the style-performance component from the fund ranking, and therefore fund ratings immediately became balanced across styles. Panel A of Figure 1 shows how the dispersion of average ratings across styles suddenly collapsed in June 2002 as a consequence of the reform.

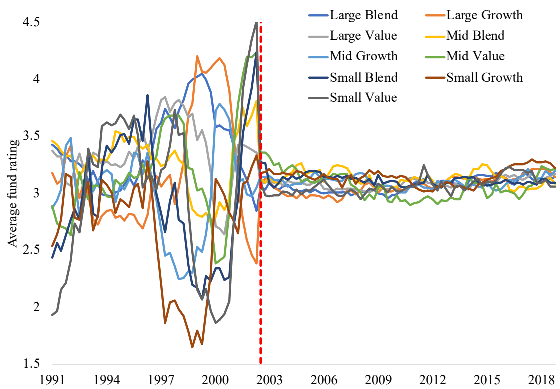
Our paper analyzes the systematic impact of this reform on the stock market. The empirical analysis has four parts.

In the first part, we demonstrate that the key elements of the mechanism exist in the data: (a) investor flows respond to ratings, and (b) flow-induced trades create price pressures.

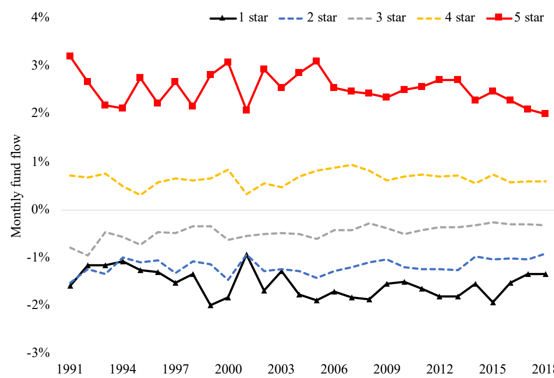
Figure 1. Morningstar rating methodology change and style price pressures

This figure highlights the main results in the paper. Panel A plots the average mutual fund rating by the 3×3 size-value Morningstar styles. The vertical dashed line represents the June 2002 methodology change event. Panel B plots the average monthly fund flow by one- to five-star Morningstar ratings. In panels C and D, we sort the 3×3 style portfolios by their lagged rating changes (ExpSum(ΔRating), defined in Section 3.2). We then plot the cumulative differences in flows and returns between the top and bottom styles for the subsequent 3 years. Following the existing fund flow literature, flows are defined as a fraction of mutual fund assets under management. The shaded areas represent 95% bootstrapped confidence intervals.

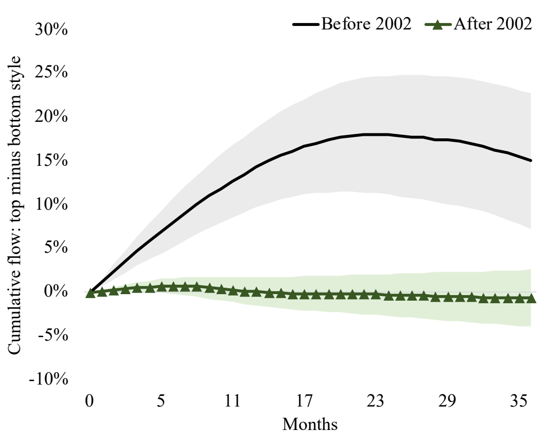
(a) Morningstar fund rating by style



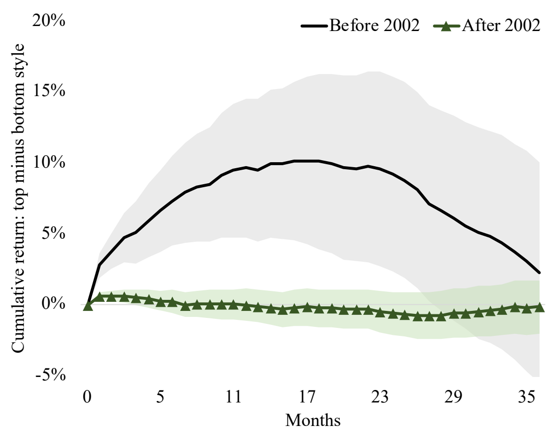
(b) Fund flow by Morningstar rating



(c) Style fund flows



(d) Style returns



First, we document that investors chase ratings *regardless* of the rating methodology.[2] This finding is evident in Panel B of Figure 1: monthly fund flows to mutual funds strongly depend on Morningstar ratings, and the dependence magnitude is stable over the 28 years

[2]Ben-David et al. (2021b) and Evans and Sun (2021) make similar observations. These findings are most consistent with the explanation that investors view Morningstar's ratings as a recommendation about the best funds from a trusted advisor (e.g., as in the "money doctors" model proposed by Gennaioli, Shleifer, and Vishny, 2015).

in our sample.[3] Second, we use an impulse-response analysis to show that a rating increase results in a surge in mutual fund flows, and that flow increases lead to contemporaneous stock price appreciation and subsequent reversals.

We also formally analyze the magnitude of this rating-driven effect in the cross-section of monthly stock returns. To do so, we regress stock returns on the average lagged rating changes of the funds holding the stock as of the previous month, while controlling for those funds' past return history, for the stock's past returns, and for prominent stock return predictors. In this case, lagged rating changes proxy for expected rating-induced mutual fund flows. This specification allows us to measure the *marginal* impact of rating-induced price pressure on stock returns. We find that the effect of ratings on stock returns is economically large and statistically significant. In fact, the marginal effect of rating-induced flows on stock return predictability is even larger than that of many prominent return predictors, such as value, momentum, and profitability.

The fact that expected flow-induced trading exerts price pressure on stocks is not a novel contribution of our paper (see, e.g., Lou, 2012). This evidence demonstrates that changes in ratings have an independent causal effect on stock returns, which serves as a premise for our main contribution on the influence of rating-driven flows on systematic price fluctuations.

In the second part of our analysis, we examine the impact of rating-driven demand on systematic return patterns. Morningstar's reform affected styles; hence, our analysis is at the style level. Before the reform, ratings clustered at the style level, leading to correlated style-level flows. Consequently, inflows hit a small subset of winning styles, forming concentrated price pressures in those styles. Indeed, before June 2002, the most upgraded styles (i.e., styles with the highest recent changes in ratings) drew large fund inflows, and their returns exhibited momentum and subsequent reversals.[4] Opposite patterns can be observed for

[3]Over a typical 2-year period, five-star funds receive inflows equal to more than 75% of their initial assets under management (AUM), while one-star funds shrink by over one-third due to outflows.

[4]As we will discuss in Section 3, we focus on rating changes (upgrading and downgrading) rather than rating levels. This specification allows us to identify *changes* in flows and hence changes in *prices*, avoiding reversals due to earlier price pressures caused by earlier flows.

the most downgraded styles. After June 2002, ratings became evenly distributed across styles, and therefore rating-chasing flows were distributed across the entire spectrum of styles. Because rating-chasing investors unknowingly stopped applying price pressure on a small subset of winning styles, the rating-induced style-level momentum and reversal effects became muted.

Panels C and D of Figure 1 illustrate these results. Panel C shows that, before 2002, the most upgraded style attracted approximately 15% more fund flows (as a fraction of mutual fund assets) than the most downgraded style over the subsequent 12 months. Once the reform was enacted and ratings became evenly distributed across styles, the spread in flows to the most upgraded versus downgraded styles disappeared, demonstrating the power of Morningstar ratings in driving flows. Panel D shows that the return difference between upgraded and downgraded styles mirrors the pattern observed for rating-induced flows. In fact, the rating-driven style momentum strategy (long/short) generated a considerable return of about 90 to 100 basis points per month before June 2002 and became unprofitable afterward. We show later that these price pressure effects are stronger in stocks with higher mutual fund ownership, consistent with our fund flow channel.

In addition, the 2002 reform altered the dispersion of style flows and style returns. As ratings became more homogeneous across styles due to the reform, so did flows and returns. The average monthly style-level flow spread (top flow minus bottom flow) dropped from 3.3% before June 2002 to 1.4% after June 2002, and the return spread dropped from 5.5% to 3.0%. The general finding that return and flow dispersion collapsed after the rating reform is robust to using alternative dispersion measures or shorter time windows around the reform event.

In the third part of our analysis, we focus on a short period around June 2002. Because we are concerned that other factors (e.g., stock market decimalization) may have caused the effects we find, focusing on the months surrounding the event allows for a sharper identification of the reform-induced effects. In the months leading up to June 2002, funds in top-rated styles gathered inflows and the underlying stocks performed well. By the same token, funds in bottom-rated styles experienced outflows and the underlying stocks performed poorly. The methodology reform caused rating dispersion across styles to sharply collapse, and so did flow dispersion. As predicted, the style return patterns also immediately halted and even slightly reversed. We also carry out a battery of tests to tackle possible alternative explanations for the event study results. The results are all supportive of our interpretation that the rating reform causally affected style-level flows and returns.

To summarize, we document that a seemingly innocuous reform implemented by a single rating firm created a long-lasting impact on the allocation of investors' capital across styles. This reallocation of capital flows altered the time-series and cross-sectional variation of style returns. These findings highlight the importance of nonfundamental demand in shaping systematic returns.

Our paper is related to the literature on demand effects in asset prices. Unlike traditional asset pricing, which assumes that price movements are only explained by cash flow and discount rate variation (Cochrane, 2011), studies in this field have also found price effects of demand in index additions and deletions (Harris and Gurel, 1986; Shleifer, 1986; Wurgler and Zhuravskaya, 2002; Chang, Hong, and Liskovich, 2015), mutual fund flows (Coval and Stafford, 2007; Lou, 2012; Huang, Song, and Xiang, 2020; Li, 2021), exchange-traded fund flows (Ben-David, Franzoni, and Moussawi, 2018; Brown, Davies, and Ringgenberg, 2021), and other sources of institutional investor demand (Koijen and Yogo, 2019; Ben-David, Franzoni, Moussawi, and Sedunov, 2021a; Parker, Schoar, and Sun, 2020). Consistent with the model in Barberis and Shleifer (2003), the pioneering work of Teo and Woo (2004) and Froot and Teo (2008) shows evidence that institutional demand can drive style-level returns. Most recently, Gabaix and Koijen (2021b) show that the demand-induced price impact coefficient at the aggregate level is larger than that at the idiosyncratic level, a finding that is corroborated in our study.

1 Data and Variable Construction

In this section, we describe the data set and explain how we construct the rating and flow variables.

1.1 Mutual fund sample

Mutual funds are one of the largest classes of equity investors in the United States and a prime investment vehicle for retail investors. When our sample begins in 1991, U.S. equity mutual funds held a total AUM of $326 billion, which was 8.9% of the entire market capitalization. Their ownership fraction grew steadily, reaching about 30% in 2005 and has remained steady since then. By the end of our sample period in 2018, equity mutual funds owned $10,849 billion, which represented 29.3% of the entire market capitalization (panel A of Figure 2).

We obtain monthly fund returns and total net assets (TNA) from the Center for Research in Security Prices (CRSP) survivorship-bias-free mutual fund data set. We use all U.S. domestic equity mutual funds. While funds are often marketed to different clients through different share classes, they invest in the same portfolio and typically only differ in the fee structure. Therefore, we aggregate all share classes at the fund level using Russ Wermers' MFLINKS (Wermers, 2000). We also obtain quarterly fund holdings from Thomson Reuters' S12 data. We augment the holdings data with stock returns and characteristics from the CRSP/Compustat merged database.

Our monthly sample starts in January 1991 and ends in December 2018. The starting date is based on data availability: monthly AUM (which is required to calculate monthly flows) from CRSP starts in 1990, and some measures require 1 year of lagged data to construct. Following the mutual fund literature (e.g., Coval and Stafford, 2007), the fund flow for fund

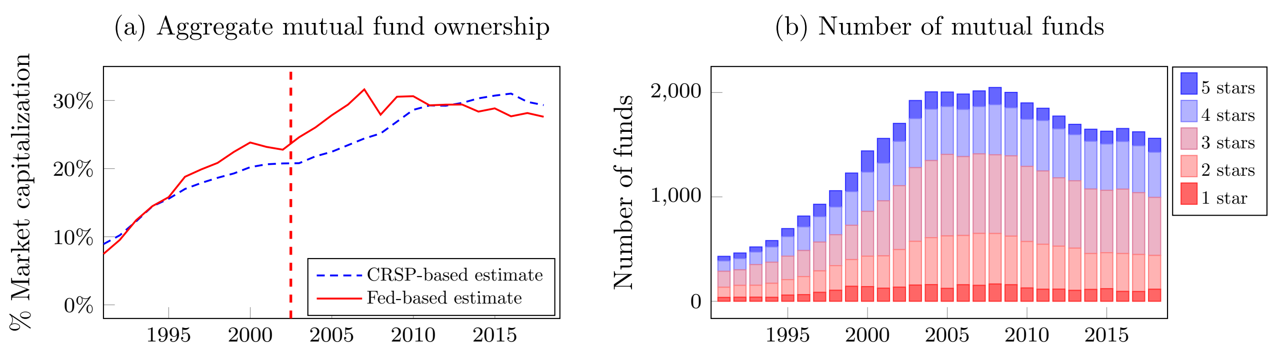
j in month t is defined as the net flow into the fund divided by lagged TNA:

$$\mathrm{Flow}_{j,t} = \frac{\mathrm{TNA}_{j,t}}{\mathrm{TNA}_{j,t-1}} - (1 + \mathrm{Ret}_{j,t}). \tag{1}$$

We obtain Morningstar ratings and style categories from Morningstar Direct and merge them with the CRSP mutual fund data using the matching table from Pástor, Stambaugh, and Taylor (2020).[5] Morningstar assigns ratings at the share class level. We follow Barber, Huang, and Odean (2016) and aggregate them at the fund level by TNA-weighting different share classes. We restrict our analysis to mutual funds with at least \$1 million TNA and winsorize fund flows at the 0.5% and 99.5% levels within each month. We require the existence of 12 lags of monthly flows, returns, and ratings. The resultant sample comprises a total of 3,305 funds with 454,787 fund-month observations.

Figure 2. Summary statistics of mutual funds

Panel A shows the aggregate domestic mutual fund holding of U.S. stocks as a fraction of the overall market. The blue line is based on the CRSP mutual fund database, and the red line is based on Federal Reserve Board flow of fund reports (L.223). Panel B shows the number of funds in each Morningstar star rating classification during our sample period of 1991–2018. Table A.1 in the appendix further summarizes the mutual fund sample used in this paper.



Panel B of Figure 2 summarizes the time series of the number of funds per Morningstar rating. The number of funds quadrupled from 1991 to 2005, and then plateaued before slightly declining from 2009 onward. Additional summary statistics are provided in Appendix A.1.

[5] We thank the authors for kindly providing the matching table.

1.2 Stock- and style-level ratings

As the main focus of this study is the effects of rating-induced demand on stocks and style portfolios, we summarize ratings at both the stock and style levels.

We define the level of and change in the Morningstar rating of stock i in month t as the holdings-weighted average rating of all funds J that hold the stock i as of the end of the prior month:[6]

$$\text{Rating}^{\text{stock}}_{i,t} = \frac{\sum_{\text{fund } j \in J} \text{SharesHeld}_{i,j,t-1} \cdot \text{Rating}_{j,t}}{\sum_{\text{fund } j \in J} \text{SharesHeld}_{i,j,t-1}}, \tag{2}$$

$$\Delta\text{Rating}^{\text{stock}}_{i,t} = \frac{\sum_{\text{fund } j \in J} \text{SharesHeld}_{i,j,t-1} \cdot (\text{Rating}_{j,t} - \text{Rating}_{j,t-1})}{\sum_{\text{fund } j \in J} \text{SharesHeld}_{i,j,t-1}}. \tag{3}$$

We now define ratings and rating changes at the style level. For a given style π, we aggregate up the stock-level ratings:

$$\text{Rating}^{\text{style}}_{\pi,t} = \sum_{i \in \text{style } \pi} w^{\pi}_{i,t-1} \cdot \text{Rating}^{\text{stock}}_{i,t}, \tag{4}$$

$$\Delta\text{Rating}^{\text{style}}_{\pi,t} = \sum_{i \in \text{style } \pi} w^{\pi}_{i,t-1} \cdot \Delta\text{Rating}^{\text{stock}}_{i,t}, \tag{5}$$

where $w^{\pi}_{i,t-1}$ is the portfolio weight of stock i in the corresponding style, based on the aggregate holdings of mutual funds that are classified by Morningstar as investing in the style π.[7]

To measure style-level flows, we compute the TNA-weighted average flows to all funds in that style. We later drop the superscripts "stock" and "style" when unambiguous. Table A.2 in the appendix presents summary statistics of ratings, flows, and returns for styles.

[6] We use the latest holdings available in the past 3 months. We normalize by total shares held by mutual funds to be consistent with the specification in Lou (2012).

[7] Specifically, for each stock i, its weight in style π is given by

$$w^{\pi}_{i,t-1} = \frac{\sum_{\text{fund } j \in \text{ style } \pi} \text{Price}_{i,t-1} \cdot \text{SharesHeld}_{i,j,t-1}}{\sum_{\text{fund } j \in \text{ style } \pi} \text{TNA}_{j,t-1}}.$$

Because $\text{TNA}_{j,t-1} = \sum_{\text{stock } i} \text{Price}_{i,t-1} \cdot \text{SharesHeld}_{i,j,t-1}$, weights in each style sum to one ($\sum_{\text{stock } i} w^{\pi}_{i,t-1} = 1$). The use of Morningstar-based style weights will be discussed in more detail in Section 4.

2 Background of the Morningstar Ratings and the 2002 Reform

In this section, we describe the simple, yet radical, methodological reform in the popular Morningstar star rating system that took place in June 2002. In later sections, we demonstrate that this exogenous reform had a far-reaching impact on style return dynamics.

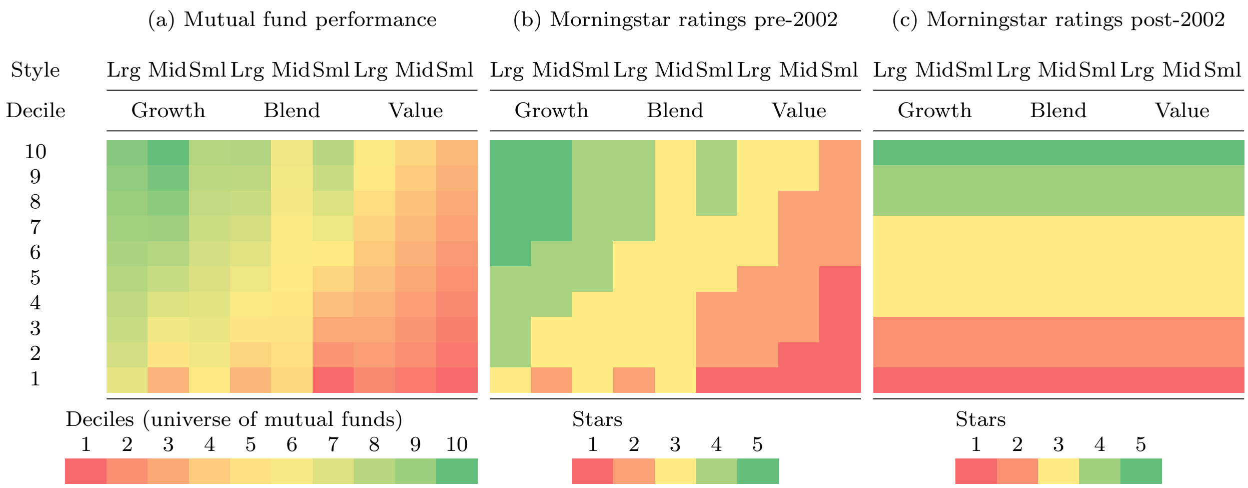
After launching its mutual fund rating system in 1985, Morningstar quickly became the industry leader in guiding investors' mutual fund selection. Since its early days, Morningstar's methodology has been transparent and publicly available. To assign ratings, Morningstar first summarizes the recent past returns of funds and conducts minor adjustments for return volatility and expenses. Depending on a fund's age, the lookback horizon for past performance can be 3, 5, or 10 years, and more weight is applied to the most recent 3 years of returns. Then, Morningstar ranks funds by their performance and assigns a rating of one to five stars with fixed proportions (10%, 22.5%, 35%, 22.5%, and 10%).

Morningstar's rating methodology changed abruptly in June 2002. The reason behind the change is related to the fact that many funds pursue specific investment styles (e.g., large-cap growth) by mandate.[8] Since its inception and until June 2002, Morningstar ranked all U.S. equity funds against each other. Because style performance is a significant part of fund performance, fund ratings were highly dependent on style performance. Following the dot-com crash, many fund managers specializing in large-cap growth stocks complained that their fund ratings dropped sharply. These managers argued that ratings barely reflected their own contributions and mostly echoed style-level returns outside of their control. As a result, the research team at Morningstar, spearheaded by the economist Dr. Paul Kaplan,

[8]Historically, mutual funds have followed different investment philosophies in identifying investment opportunities ("styles"). For example, managers following Graham and Dodd (1934) look for undervalued firms ("value"), while those following Fisher (1958) search for firms with substantial unrealized growth potential ("growth"). Funds therefore are often classified as value or growth. Funds that invest in the same style have more overlapping holdings.

Figure 3. Illustration of the Morningstar methodology pre- and post-June 2002

The figure presents a hypothetical example of the mapping of mutual fund performance into Morningstar ratings pre- and post-June 2002. The columns represent different investment styles (large-growth, midcap-growth, small-growth, large-blend, midcap-blend, small-blend, large-value, midcap-value, small-value). In panel A, the rows represent the 3-year performance deciles of funds *within* each style. The colors represent the performance decile across the *entire* mutual fund universe: green indicates top-ranked performance, and red indicates bottom-ranked performance across the entire mutual fund universe. Panel B shows ratings by Morningstar based on the pre-2002 methodology. Panel C shows ratings by Morningstar based on the post-June 2002 methodology.



redesigned the rating system.[9]

The main modification in the post-June 2002 methodology was that funds were ranked *within* style categories,[10] as opposed to across the entire universe. Morningstar classifies diversified U.S. equity funds into the well-known 3×3 style matrix based on funds' holdings: combinations of small/midcap/large and value/blend/growth. Since June 2002, the distribution of star ratings has been the same for funds within each of the 3×3 styles for diversified funds. The modified methodology was announced as early as April 2002 and was implemented at the end of June 2002. Appendix B provides additional technical details about the rating methodology.

Figure 3 schematically illustrates the relation between past fund performance and Morn-

[9] We learned this information during phone conversations with Morningstar management. Making ratings more balanced across styles was also one of the stated objectives for the methodology reform. For instance, in a *New York Times* interview, Don Phillips, a managing director of Morningstar, said, "Two years ago, every growth fund looked wonderful. . . Now, none does" (Norris, 2002).

[10] The modified methodology also ranked sector funds within their industrial sectors (e.g., financial, utilities). For simplicity, our analysis focuses on ratings and flows of diversified U.S. equity funds, which constitute 87% of all equity mutual funds in our sample period.

ingstar ratings. Before June 2002, Morningstar's mutual fund ratings closely mapped past overall fund performance into star ratings. Panel A shows a snapshot of mutual funds' past hypothetical performance (colors) for funds within styles. The columns represent the different styles, and the rows represent past fund performance deciles within each style. Funds in the top rows performed the best within their styles.

The pre-reform rating methodology largely mapped past performance into star ratings. As panel A shows, in this hypothetical example, large-growth funds had the best performance. Panel B shows that the best-performing funds were ranked the highest by Morningstar. In other words, before 2002, Morningstar ratings were highly correlated with raw past returns.

Since June 2002, Morningstar has ranked funds *within* style; hence, rankings are independent of style performance (panel C). The demand from rating-chasing investors, therefore, became more evenly spread over all styles.

2.1 Discussion: Timing of the methodology reform

While the reform was prompted by the dot-com crash and therefore did not occur on a random date, its *exact timing* is exogenous, a fact that we will exploit in Section 5. While the dot-com peak was in March 2000, the designer of the reform, Dr. Paul Kaplan, was only appointed as the head of Morningstar research in February 2001 (Morningstar, 2001). While we do not observe the decision process within Morningstar, it likely took significant work and deliberation before the reform was finalized and approved, as Morningstar rarely changes its methodology and this reform is arguably the most significant change to date.

Furthermore, as will be shown in Section 3, investors' rating-chasing behavior did not change around the dot-com bust or the 2002 reform. Therefore, even though the reform timing is not entirely exogenous, it appears unrelated to the specific channel of rating-induced flows and price pressures that we are interested in.

3 Rating-Chasing Behavior and Price Impact

In this section, we examine the mechanism that links ratings to flows and then to returns. First, we present evidence that investors rely strongly on ratings throughout the sample period. Second, we use an impulse-response analysis to investigate the impact of rating changes on flows and returns. Finally, we show that rating changes lead to robust cross-sectional predictability in stock returns. The effect of rating-induced demand on stock return predictability is even stronger than that of some prominent predictors, such as value, momentum, and profitability.

3.1 Investors chase ratings regardless of rating methodology

Because our identification relies on the reform in the Morningstar rating methodology, it is important to examine whether mutual fund investors continued to rely on Morningstar ratings after June 2002.

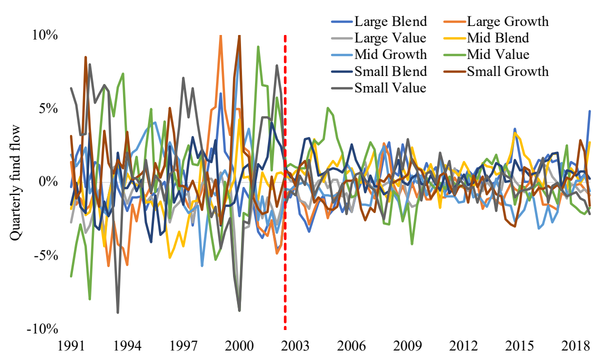
We begin by examining simple summary statistics. Panel B of Figure 1 plots the average flows to mutual funds with different Morningstar ratings. Throughout our sample period, five-star funds receive flows that amount to +2% to +3% of their AUM per month on average. This is economically large as it implies that the AUM of five-star funds increases by about 25% to 40% over 1 year. In contrast, one-star funds experience outflows of −1.5% to −2% of their AUM per month on average. Importantly, these patterns do not change after June 2002.

More formally, we estimate the response of fund flows to lagged fund ratings using 3-year rolling-window TNA-weighted Fama-MacBeth regressions (Fama and MacBeth, 1973), after controlling for 36 lags of monthly fund returns. Panel B of Figure 4 plots the results. The estimated coefficient varies only slightly over the sample without a material drop around or following the 2002 reform. For example, the average flow-to-rating response was 0.37%

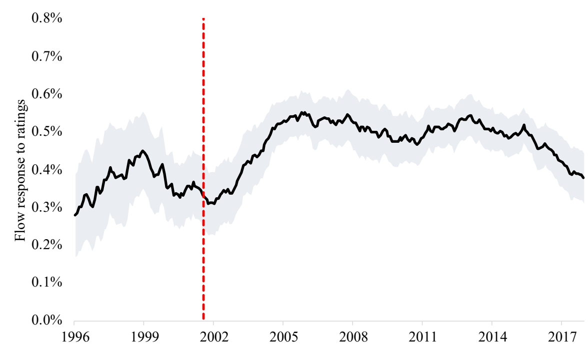
Figure 4. The June 2002 Morningstar methodology reform

This figure examines the relation between Morningstar ratings and fund flows over the sample period. The vertical dashed red lines represent the June 2002 methodology change event. Panel A plots the TNA-weighted average quarterly fund flow by Morningstar 3×3 styles. Flows are demeaned cross-sectionally to focus on the dispersion. Panel B explores the stability of the relation between ratings and flows at the fund level. Specifically, it plots the regression coefficient of fund flows on lagged ratings estimated using 3-year rolling windows. Because the regression controls for 36 lags of monthly fund returns and because it takes 3 years to compute a rolling average, the graph starts in 1996. (Raw monthly data are available from 1991.) The shaded area indicates the two standard error bands.

(a) Fund flow, by style



(b) Regression coefficient of fund flows on ratings



before June 2002 and was 0.48% after June 2002.[11]

In summary, these results indicate that mutual fund investor capital allocation chased Morningstar ratings regardless of the rating methodology. However, the rating reform led to significant changes in style-level fund flows. Because ratings are constructed within styles after June 2002, style-level fund flow dispersion dropped significantly after the methodology reform, as is easily visible in panel A of Figure 4. As we see later in Section 4.1, the style-level correlated demand due to rating-chasing behavior mostly disappeared after the reform.

3.2 Stock-level rating-induced price pressures

Next, we confirm that Morningstar ratings substantially affect stock prices through flow-induced trading. This step is necessary before we explore the influence of rating-induced style demand on style returns in the subsequent analysis.

[11]See further analysis indicating that investors did not change their rating-chasing behavior in Evans and Sun (2021) and Ben-David et al. (2021b).

We assess the price impact of ratings on stock returns by first separately estimating the two chained effects: (a) the response of fund flows to Morningstar rating changes, and (b) the response of stock returns to flow-induced trading.

First, we estimate the fund flow response to lagged fund rating changes:

$$\text{Flow}_{j,t} = a + b_1 \cdot \Delta\text{Rating}_{j,t-1} + \ldots + b_{36} \cdot \Delta\text{Rating}_{j,t-36} + \gamma X_{j,t} + u_{j,t}, \tag{6}$$

where $\Delta\text{Rating}_{j,t}$ is the month t rating change of fund j, and controls $X_{j,t}$ include 36 monthly lags of fund flows, as well as decile indicators of the previous 3-year cumulative fund return and benchmark-adjusted returns. The benchmark-adjusted returns are defined as fund returns in excess of the AUM-weighted average returns of funds in the same Morningstar category (e.g., small cap-growth). We use decile indicators to allow for the nonlinear response of fund flows to past performance (Chevalier and Ellison, 1997).

Panel A of Figure 5 plots the cumulative response coefficients ($b_1, b_1 + b_2, \ldots$). In response to a one-star change in rating, funds experience an average of 6%–7% additional flows over the next 24 months. This result is consistent with prior research showing that, when controlling for past fund performance, discrete changes in ratings cause sizeable differences in fund flows (Del Guercio and Tkac, 2008; Reuter and Zitzewitz, 2021).

Second, we estimate the response of stock returns to stock-level flow-induced trading. To measure the amount of stock-level trading caused by fund flows, we follow Lou (2012) to calculate flow-induced trading (FIT) for each stock i in each month t:[12]

$$\text{FIT}_{i,t} = \frac{\sum_{\text{fund } j \in J} \text{SharesHeld}_{i,j,t-1} \cdot \text{Flow}_{j,t}}{\sum_{\text{fund } j \in J} \text{SharesHeld}_{i,j,t-1}}. \tag{7}$$

In short, FIT is the amount of mutual fund trading in stock i that is mechanically caused by fund flows. As explained in Lou (2012), whereas discretionary trading can reflect managers' information about fundamentals, FIT isolates the nondiscretionary trading that is

[12]Lou (2012) also applies different scaling factors to inflows and outflows. We omit this scaling for simplicity, but our results are robust to using his scaling factors.

only attributable to fund flows and thus likely does not contain fundamental information.[13]

We then estimate the response of stock returns to FIT,

$$\text{Ret}_{i,t} = a + c_0 \cdot \text{FIT}_{i,t} + c_1 \cdot \text{FIT}_{i,t-1} + \ldots + c_{36} \cdot \text{FIT}_{i,t-36} + u_{i,t}, \tag{8}$$

and plot the cumulative response $(c_0, c_0 + c_1, \ldots)$ in panel B of Figure 5. An increase of 1% in mutual fund ownership through FIT (i.e., expected trading due to flows) leads to immediate price pressure of approximately 0.6% in the contemporaneous month, and a complete reversion in the subsequent 1 to 2 years. This result is consistent with the findings related to FIT in Lou (2012).

Combining these two effects, we predict that rating changes (especially recent changes) should affect stock returns. We expect the impact to come from rating changes rather than rating levels. This is because while a higher rating level in the more distant past also generates flows (Figure 5, panel A), the price pressures created by their initial impact are already embedded in the later part of the "price pressure cycle" and are already reverting (Figure 5, panel B). For this reason, we use rating changes in the rest of our analysis.[14]

To facilitate our later analysis of rating-induced price impact, it is convenient to summarize recent rating changes into a weighted average sum such that the weights correspond to how much each lag affects returns. We obtain such a weighting scheme by directly estimating the response of stock returns on the past 24 lags of stock-level rating changes (defined in Equation (3)) with the same controls in Equation (10) discussed later.[15] We plot the coefficients in panel C of Figure 5. As expected, more recent rating changes are more impactful,

[13]Consistent with this interpretation, Lou finds that FIT leads to price pressures that revert over time. Wardlaw (2020) recently argued that some flow measures, such as that in Edmans, Goldstein, and Jiang (2012), inadvertently include contemporaneous stock returns. This does not apply to our flow measure, which is constructed following Lou (2012) and does not use price information.

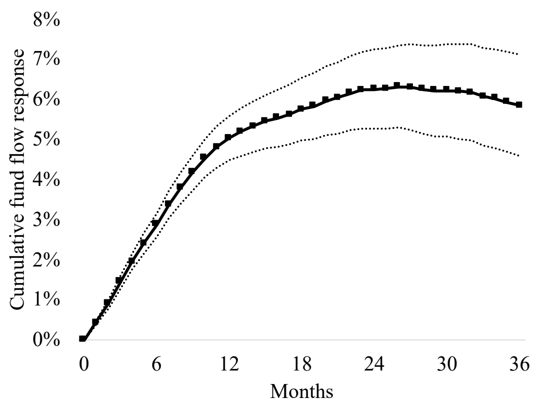
[14]Consistent with the patterns discussed here, in unreported results we find that rating levels also tend to positively predict future stock returns. However, the effect is statistically significant only if we also control for rating levels lagged by several months. This result implies that the effect is better specified by using rating changes.

[15]Requiring the existence of 36 lags of ratings and controls leads to a significantly smaller sample size, so we use 24 lags here. We obtained nearly identical results when using 36 lags, despite the smaller sample.

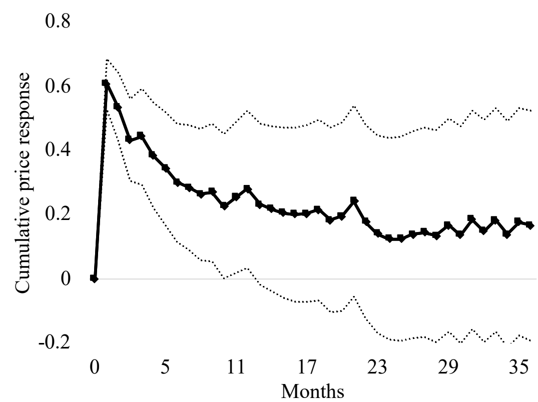
Figure 5. Price impact of ratings and flows

Panel A shows the cumulative response of fund flows to changes in fund ratings. Panel B shows the cumulative response of stock returns to flow-induced trading (FIT), defined as the nondiscretionary trading induced by mutual fund managers proportionally adjusting existing portfolio holdings in response to fund flows. Panel C shows the *noncumulative* response of stock returns to changes in ratings as well as the fitted exponential response (green line). Panel D plots the cumulative value-weighted price path of stocks with top and bottom deciles of the lagged exponential sum of rating changes ($\text{ExpSum}(\Delta\text{Rating})_{i,t-1}$). The sorting decile breakpoints are based on NYSE stocks. In panels A, B, and D, the dashed lines represent two standard errors bands.

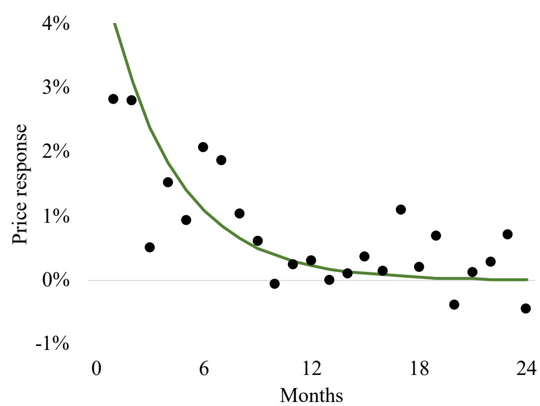
(a) Rating change → Fund flow



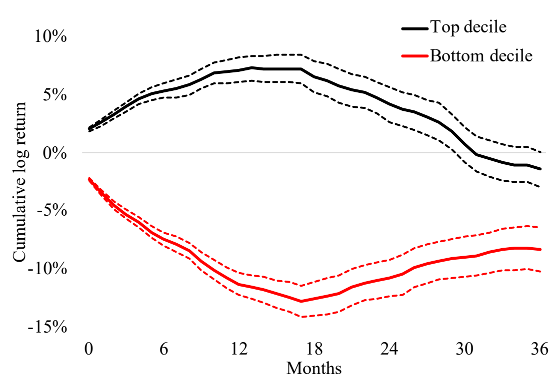
(b) Flow-induced trading → Stock return



(c) Rating change → Stock return



(d) Price pressure and reversion



and the coefficients on more distant rating changes converge toward zero.

Because the impact primarily takes place over the first 12 months, we summarize past rating changes using the following weighted sum:

$$\text{ExpSum}(\Delta\text{Rating})_{i,t-1} = \sum_{k=1}^{12} \tau_k \cdot \Delta\text{Rating}_{i,t-k}, \tag{9}$$

where $\tau_k = \frac{12\cdot(1-\delta)}{1-\delta^{12}} \cdot \delta^{k-1}$ and $\sum_{k=1}^{12} \tau_k = 12$. The weights decay with factor $\delta = 0.76$, which is estimated from a least-squares fit to the response (panel C of Figure 5).[16] Because the weights sum to 12 (months), in terms of units, ExpSum(ΔRating) should be interpreted as the rating change over 1 year. The estimated decay factor $\delta = 0.76$ implies a half-life of $-\ln(2)/\ln(\delta) \approx 2.58$ months. As we show in the next subsection, our results are not sensitive to reasonable variations in the choice of rating change horizon or weighting scheme.

The results presented so far indicate that recent rating changes create price pressures. To further validate the price pressure interpretation, we examine whether the price movements revert. In panel D of Figure 5, for each month, we sort stocks into NYSE breakpoint-based decile portfolios based on ExpSum$(\Delta\text{Rating})_{i,t-1}$ and track the performance over the following 3 years. Stocks in the top decile of past rating changes outperform stocks in the bottom decile by about 20% over the subsequent 12 to 18 months. Importantly, the cumulative return difference between the two groups of stocks indeed reverts over the 36-month horizon.[17]

3.3 Return predictability in the cross-section of stock returns

The results presented above suggest that rating-induced flows create price pressure at the stock level. To put the magnitude of rating-induced stock return predictability into perspective, in this section, we compare rating changes against some of the most prominent return predictors.

[16]The estimate of δ is robust to the regression methodology in estimating the dependence of stock returns on rating changes. The estimate in our paper is based on a panel regression with market cap-weighted observations ($\delta = 0.764$). An equal-weighted panel regression yields $\delta = 0.814$. If we use a Fama-MacBeth regression, δ becomes 0.793 if market-cap weighted and 0.815 if equal weighted. All of our subsequent results are robust to much larger variation in δ.

[17]The bootstrapped standard errors are obtained via randomly permuting stocks in each year. That is, in each year and for each stock, we assign the decile ranking of another randomly chosen stock to it. We repeat this procedure 1,000 times to measure the variation of the subsequent price paths of these randomly sorted stocks. Our approach amounts to a randomization test to sample from the null hypothesis that $\text{ExpSum}_{i,t-1}$ does not affect stock returns.

Specifically, we estimate the following return predictability model:

$$\text{Return}_{i,t} = d_1 \Delta \text{Rating}_{i,t-1-h \to t-1} + \gamma^s X^s_{i,t} + \gamma^f X^f_{i,t} + u_{i,t}, \quad (10)$$

where the dependent variable is the return of stock i in month t; $\Delta\text{Rating}_{i,t-1-h\to t-1}$ is the share-weighted average change in ratings from month $t-1-h$ to month $t-1$ for the funds that hold stock i as of the end of month $t-1$; $X^s_{i,t}$ is a vector of stock-based controls that include known predictors of stock returns: the lagged one-month return, momentum (i.e., the stock return from month $t-12$ to month $t-2$ as in Jegadeesh and Titman, 1993), long-term reversal (i.e., the stock return from month $t-36$ to month $t-13$ as in De Bondt and Thaler, 1985), size (Banz, 1981), value (Fama and French, 1993), profitability (i.e., gross profitability as in Novy-Marx, 2013), and investment (i.e., asset growth as in Cooper, Gulen, and Schill, 2008); and $X^f_{i,t}$ is a vector of fund-stock-based controls that include the fraction of stock i's outstanding shares held by funds and the share-weighted average 3-year fund return *and* benchmark-adjusted returns of the funds that hold stock i as of the end of month $t-1$. The benchmark-adjusted returns are computed as fund returns in excess of the AUM-weighted returns of funds in the same Morningstar category.

To allow for an easy comparison of rating changes and other return predictors, we standardize the within-month mean and standard deviation of the right-hand-side variables to zero and one, respectively. (The approach is similar to that of Green, Hand, and Zhang, 2017, and other studies that present coefficients as z-scores.) Only stocks held by at least one fund as of the end of the previous month are included in the analysis. Following literature standards, the regression is estimated via the Fama-MacBeth procedure with the Newey-West standard error correction of 12 lags.

Note that this analysis implicitly builds on the discontinuity approach of Del Guercio and Tkac (2008) and Reuter and Zitzewitz (2021). A key difference is that our analyses are carried out at the stock level rather than at the fund level.[18] In our setting, stock-level ratings are weighted averages of individual fund ratings and are therefore not a discrete variable. For this reason, unlike in Del Guercio and Tkac (2008) and Reuter and Zitzewitz (2021), it is not possible to perform an exact discontinuity-design regression here.

Panel A of Table 1 presents baseline results with rating changes summarized using the exponential decay function described in Section 3.2. That is, lagged rating changes are measured as ExpSum(ΔRating) over the prior 12 months. The first column shows how stock returns are related to characteristics, without including rating changes or any fund-level controls. Overall, the results are broadly consistent with those reported in recent literature (Fama and French, 2015; Hou, Xue, and Zhang, 2015). That is, there is no significant size premium in the sample period, and investment and profitability are robust predictors of stock returns.

We estimate Equation (10) using three different samples: all stocks, stocks held by at least three mutual funds as of the end of the previous month,[19] and all stocks excluding microcaps.[20] The results are reported in columns 2, 4, and 6, respectively. Across the three specifications, the effect of rating changes is stronger than the effect of size, value, profitability, momentum, and long-term reversal in terms of both magnitude and statistical significance.[21] When restricting to nonmicrocap stocks, ratings can predict returns better

[18]This is, in fact, a major difference and explains why we find larger effects than do Reuter and Zitzewitz (2021). Note that we aggregate the rating changes of *all* funds holding each stock. Because most stocks are held diversely by many funds, the change of a single fund's rating, which Reuter and Zitzewitz (2021) exploit, is unlikely to significantly affect stock returns. This can be seen in a back-of-the-envelope exercise. The median stock is held by 132 funds, with each fund holding an average of 0.4% of the stock's market cap. Even the largest holder only accounts for 3.1% on average. If the largest-holding fund's rating changed, that would cause monthly fund flows to change by up to 2% per month (Figure 1, panel B). Assuming a large price impact coefficient of 5 only translates to a stock return impact of $3.1\% \times 2\% \times 5 = 0.31\%$. Note that this value is calculated using the most generous assumptions in each step of the calculation. If we consider rating change for an average fund, the effect drops to 0.04%. Therefore, changes in *individual* fund ratings are unlikely to have a large impact on stock returns.

[19]Dropping stocks held by fewer than three funds from the sample is intended to reduce the noise in the relation between rating changes and expected rating-induced trading. Lou (2012) impose the same data requirement in his table 9.

[20]Microcaps are defined as stocks with lagged market capitalization below the 20^{th} percentile of NYSE market capitalization.

[21]The signs of stock characteristics are "flipped" when necessary so that they positively predict returns based on the original studies documenting the return predictability.

than all other predictors.[22]

In columns 3, 5, and 7, we modify Equation (10) by multiplying rating changes and lagged 3-year fund returns by the fraction of a stock's market cap held by mutual funds as of the end of the previous month. The key results from the analysis are robust to this specification change.

[22] While momentum is a strong return predictor before 1991, this is not the case in our sample (see, e.g., Lewellen, 2015).

Table 1. Rating-induced price pressures in the cross-section of stocks

Panel A presents coefficient estimates for Equation (10), which are estimated via the Fama-MacBeth (Fama and MacBeth, 1973) procedure. Panel B provides robustness tests with respect to the length of the rating change measurement horizon, weighting method of observations, nonlinear fund return controls, and the heterogeneity in the flow response to ratings. Fund-level controls include the lagged fraction of each stock held by mutual funds as well as the lagged share-weighted average past 3-year fund returns and benchmark-adjusted fund returns. In panel B, *nonlinear control* refers to specifications in which the past fund returns and benchmark-adjusted fund returns are controlled for by using decile indicators instead of a continuous variable. *Heterogeneous response* refers to specifications in which rating changes are weighted by the relative magnitude of their threshold effect (e.g., four to five star upgrades may produce different effects relative to one to two star upgrades). Independent variables are transformed into z-scores (mean = 0, standard deviation = 1) within each cross-section. The t-statistics, presented in parentheses, are computed based on standard errors with Newey-West corrections of 12 lags. *$p < .1$; **$p < .05$; ***$p < .01$.

A. Return predictability from ratings and stock characteristics

	All stocks			Min. 3 funds		Ex. microcaps	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ExpSum(ΔRating)		0.17***		0.19***		0.26***	
		(3.99)		(4.21)		(3.59)	
ExpSum(ΔRating) ×%Held			0.16***		0.16***		0.17***
			(3.78)		(3.84)		(3.49)
Size	0.06	0.04	0.04	0.04	0.04	0.03	0.03
	(1.62)	(1.13)	(1.02)	(1.16)	(1.06)	(1.11)	(1.07)
Value	0.15**	0.14**	0.14**	0.12*	0.12*	0.06	0.07
	(2.09)	(2.21)	(2.22)	(1.92)	(1.94)	(0.86)	(0.90)
Profitability	0.11	0.11*	0.11*	0.12*	0.11*	0.09	0.09
	(1.50)	(1.74)	(1.73)	(1.79)	(1.78)	(1.28)	(1.24)
Investment	0.22***	0.21***	0.22***	0.20***	0.20***	0.14***	0.15***
	(4.62)	(4.88)	(4.94)	(4.75)	(4.80)	(3.77)	(3.84)
Momentum	0.15	0.14	0.13	0.15	0.15	0.20	0.19
	(1.07)	(0.99)	(0.94)	(1.07)	(1.03)	(1.24)	(1.22)
Reversal	0.06	0.05	0.05	0.05	0.05	0.04	0.04
	(1.36)	(1.21)	(1.24)	(1.20)	(1.24)	(1.01)	(1.08)
Fund-level controls	No	Yes	Yes	Yes	Yes	Yes	Yes
No. observations	1,270,055	1,270,055	1,270,055	1,204,473	1,204,473	616,636	616,636
Average R^2	.030	.039	.039	.040	.040	.058	.058

Table 1. Continued

B. Robustness with respect to specification						
	ΔRating, linear control			ΔRating × %Held, linear control		
	All stocks	Min 3 funds	Ex. microcaps	All stocks	Min. 3 funds	Ex. microcaps
	(1)	(2)	(3)	(4)	(5)	(6)
ExpSum(ΔRating)	0.17***	0.19***	0.26***	0.16***	0.16***	0.17***
	(3.99)	(4.21)	(3.59)	(3.78)	(3.84)	(3.49)
3-month ΔRating	0.11***	0.13***	0.22***	0.11***	0.11***	0.13***
	(2.75)	(3.01)	(3.41)	(3.39)	(3.47)	(3.46)
6-month ΔRating	0.14***	0.16***	0.23***	0.12***	0.12***	0.14***
	(3.23)	(3.43)	(3.16)	(2.89)	(2.89)	(2.83)
9-month ΔRating	0.13***	0.14***	0.22***	0.13***	0.13***	0.14***
	(2.79)	(2.82)	(2.87)	(2.66)	(2.65)	(2.72)
12-month ΔRating	0.12**	0.13**	0.23***	0.13***	0.13***	0.16***
	(2.29)	(2.50)	(2.86)	(2.69)	(2.75)	(2.92)
	ΔRating, nonlinear control			ΔRating × %Held, nonlinear control		
	All stocks	Min 3 funds	Ex. microcaps	All stocks	Min. 3 funds	Ex. microcaps
ExpSum(ΔRating)	0.18***	0.19***	0.25***	0.16***	0.16***	0.16***
	(4.34)	(4.48)	(3.82)	(4.01)	(4.05)	(3.74)
3-month ΔRating	0.11***	0.12***	0.21***	0.11***	0.11***	0.12***
	(2.94)	(3.12)	(3.50)	(3.63)	(3.64)	(3.54)
6-month ΔRating	0.14***	0.16***	0.22***	0.12***	0.12***	0.13***
	(3.42)	(3.55)	(3.31)	(3.06)	(3.07)	(3.05)
9-month ΔRating	0.12***	0.14***	0.20***	0.13***	0.12***	0.14***
	(2.83)	(2.87)	(3.04)	(2.82)	(2.80)	(2.93)
12-month ΔRating	0.12**	0.13***	0.22***	0.13***	0.13***	0.15***
	(2.39)	(2.58)	(3.03)	(2.94)	(3.00)	(3.24)
	ΔRating, heterogeneous response			ΔRating × %Held, heterogeneous response		
	All stocks	Min 3 funds	Ex. microcaps	All stocks	Min 3 funds	Ex. microcaps
ExpSum(ΔRating)	0.15***	0.16***	0.21***	0.14***	0.14***	0.15***
	(3.46)	(3.52)	(3.33)	(3.42)	(3.45)	(3.38)
3-month ΔRating	0.11***	0.12***	0.18***	0.11***	0.11***	0.12***
	(2.85)	(3.05)	(3.32)	(3.29)	(3.36)	(3.47)
6-month ΔRating	0.13***	0.14***	0.19***	0.12***	0.11***	0.13***
	(3.08)	(3.17)	(3.01)	(2.83)	(2.81)	(2.89)
9-month ΔRating	0.13***	0.13***	0.18***	0.13***	0.12***	0.14***
	(2.88)	(2.86)	(2.80)	(2.77)	(2.75)	(2.78)
12-month ΔRating	0.13***	0.14***	0.21***	0.13***	0.13***	0.15***
	(2.64)	(2.69)	(2.87)	(2.75)	(2.78)	(2.95)

In panel B of Table 1, we show that rating-induced return predictability is not sensitive to reasonable variations in the choice of rating change horizon or weighting scheme. Specifically, we report estimates for the coefficient on $\Delta\text{Rating}_{i,t-1-h\to t-1}$ with h equal to 3, 6, 9, and 12 months and where each lagged rating change is weighted equally. The upper part of the panel presents the results. Consistent with the patterns presented in Figure 5, panel C, the effect is particularly strong over the first 6 months and tends to weaken after the initial 9 months.

In the middle part of panel B, we also verify that the results are robust to controlling for potential nonlinearity in the relation between past fund returns and fund flows.[23] In most studies, this relation is found to be convex (e.g., Chevalier and Ellison, 1997). However, Spiegel and Zhang (2013) argue that the relation may actually be linear, especially when studying size-weighted flows. We take an agnostic approach and simply modify Equation (10) to allow for a nonlinear effect. Specifically, we change the fund–stock-based control for the lagged average 3-year fund return from a continuous variable into 10 indicator variables representing lagged fund return deciles. All the results are robust to this specification.

Finally, in the bottom part of panel B, we verify the robustness of the results after taking into account differential fund flow responses to different rating change thresholds. As documented in Del Guercio and Tkac (2008) and Reuter and Zitzewitz (2021), at the fund level, the fund flow response is larger around the four/five and three/four star rating thresholds relative to the two/three and one/two star thresholds. We confirm this empirical fact in our data. Because the vast majority of stocks are held by multiple funds, whether this heterogeneity would matter at the stock level is unclear, but we again take an agnostic approach and simply modify the specification to allow for heterogeneous rating change effects. We estimate the following marginal effects: 0.120, 0.340, 0.620, and 0.623 for the one-/two-, two-/three-, three-/four-, and four-/five-star rating thresholds, respectively.[24] We then use these estimates as weights when aggregating fund-level rating changes at the stock level. The results are virtually unchanged.

Notably, the patterns described here are obtained while controlling for the past returns of the funds that hold the stock and the past returns of the stock itself. Thus, the effect

[23] We thank Motohiro Yogo for suggesting this test.

[24] Similar to Reuter and Zitzewitz (2021), we estimate these threshold effects using four threshold regressions of fund flows on lagged rating changes and past fund return controls.

we document should be interpreted as evidence that discrete changes in Morningstar ratings cause flow-induced trading that has a significant *marginal* impact on the cross-section of stock returns. This evidence serves as a premise for our main contribution in the next few sections because it demonstrates that changes in ratings have an independent causal effect on stock returns. We now show that rating-driven demand strongly influences the time-series and cross-sectional variation of style returns.

4 Impact of Rating-Chasing Demand on Style Performance

So far, we have presented evidence that ratings affect returns at the stock level. In this section, we move up a level and examine the impact of the Morningstar reform on style portfolios. To be better aligned with the definition of Morningstar ratings—the key driver of results—we use Morningstar's fund style classifications to define styles portfolios. For instance, the large-cap growth style portfolio is defined by the aggregate holdings of all funds in the Morningstar Large-Cap Growth category.[25]

In the subsections that follow, we document our main results. We start by showing the existence of robust style-level rating-induced price pressures as well as style-level momentum and reversal patterns before the June 2002 reform. Consistent with the idea that these patterns were due to correlated rating-induced trading at the style level, the effects dissipated after the reform. Moreover, style return dispersion declined dramatically following the rating methodology change. To provide sharper identification, in Section 5, we conduct an event study using a short window to focus exclusively on the reform-induced rating movements in

[25] We use this classification because it is the basis for the style-level adjustments in Morningstar ratings. Lettau, Ludvigson, and Manoel (2019) document that fund-based style classification in the financial industry does not exactly map to the size and value definitions used by academics, which are based on market capitalization and book-to-market ratios (Fama and French, 1993). Appendix A.3 shows that the industry classification is a "smoothed" version of the academic style definitions. In Appendix A.6, we present results repeating the main analyses for styles based on the academic definitions. The results generally extend to the academic-based styles qualitatively, though with weaker magnitudes, as expected.

June 2002.

4.1 Style-level rating-induced price pressures

We start by examining the effects of the Morningstar reform on style-level demand and return dynamics. To this end, we first calculate the style-level changes in Morningstar ratings. We aggregate the stock-level rating changes for each style portfolio π:

$$\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1} = \sum_{\text{stock } i \in \text{ style } \pi} w^{\pi}_{i,t-1} \cdot \text{ExpSum}(\Delta\text{Rating})_{i,t-1}, \tag{11}$$

where the stock-level lagged 12-month rating change, $\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$, is as defined in Equation (9), and $w^{\pi}_{i,t-1}$ is the portfolio weight of stock i in style π.

To examine the effects of rating changes on style flows and returns, we rank the nine style portfolios by $\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$ within each month and track average cumulative flows and returns over the following months. The results, which are also presented graphically in panels C and D of Figure 1, are tabulated in Table 2. Panel A shows statistics for the top- versus bottom-ranked styles over horizons of up to 36 months, and panel B repeats the analysis for the three top-ranked versus three bottom-ranked styles. The standard errors are bootstrapped by randomly permuting the style portfolios in each year.[26]

Before 2002, the top style experienced approximately 1% higher flows per month relative to the bottom style over the next 12 months. As expected, the spread in flows between the top-three and the bottom-three styles is smaller, at about 0.7%, but the difference is still statistically significant. Following the rating reform of June 2002, which mechanically shrunk the dispersion of ratings across styles, the spread in flows over the 12 months after sorting styles became about an order of magnitude smaller.

Table 2 also shows that the patterns observed in style flows are mirrored in style returns. Before June 2002, the top style outperformed the bottom style by about 10% in total over

[26] In other words, we conduct a randomization test to sample from the null hypothesis that ratings do not affect style returns at all.

Table 2. Rating-induced price pressures in style portfolios

We sort style portfolios using the lagged exponential sum of rating changes (ExpSum(ΔRating)$_{\pi,t-1}$) and tabulate their average monthly fund flow and return over the subsequent 36 months. Panel A shows the difference between the top and bottom styles. Panel B shows the difference between the averages of the top-three and the bottom-three styles. Bootstrapped standard errors are reported in parentheses. $^{*}p < .1$; $^{**}p < .05$; $^{***}p < .01$.

	A. Top 1 minus bottom 1				
	Months:	1–6	7–12	13–24	25–36
Monthly flow (%)	Before June 2002	1.14***	0.92***	0.38*	−0.25
		(0.33)	(0.28)	(0.23)	(0.19)
	After June 2002	0.09	−0.09*	−0.04	−0.02
		(0.07)	(0.05)	(0.05)	(0.05)
	Before − After	1.05***	1.01***	0.42*	−0.22
		(0.34)	(0.29)	(0.23)	(0.19)
Monthly return (%)	Before June 2002	0.76**	0.39	−0.04	−0.58***
		(0.31)	(0.35)	(0.22)	(0.22)
	After June 2002	−0.07*	−0.04	−0.05	0.04
		(0.04)	(0.06)	(0.05)	(0.04)
	Before − After	0.83***	0.43	0.02	−0.62***
		(0.32)	(0.36)	(0.23)	(0.23)
	B. Top 3 minus bottom 3				
	Months:	1–6	7–12	13–24	25–36
Monthly flow (%)	Before June 2002	0.81***	0.66***	0.14	−0.14
		(0.22)	(0.19)	(0.16)	(0.09)
	After June 2002	0.10**	−0.08**	−0.04	−0.05**
		(0.04)	(0.03)	(0.02)	(0.02)
	Before − After	0.71***	0.74***	0.17	−0.09
		(0.23)	(0.20)	(0.16)	(0.10)
Monthly return (%)	Before June 2002	0.47**	0.28	−0.10	−0.39***
		(0.21)	(0.22)	(0.17)	(0.13)
	After June 2002	−0.08***	−0.04	−0.05	0.03
		(0.03)	(0.03)	(0.03)	(0.03)
	Before − After	0.55**	0.31	−0.05	−0.42***
		(0.22)	(0.22)	(0.17)	(0.13)

the next 12 to 18 months, and the return spread reverted subsequently. Strikingly, the return spread is effectively zero after June 2002. Again, we find similar patterns when comparing the top-three and bottom-three styles (panel B). Overall, these results are consistent with style-level ratings creating flow-induced price pressures and subsequent reversals before June 2002, but not after that date as flows spread out across styles after the reform.[27]

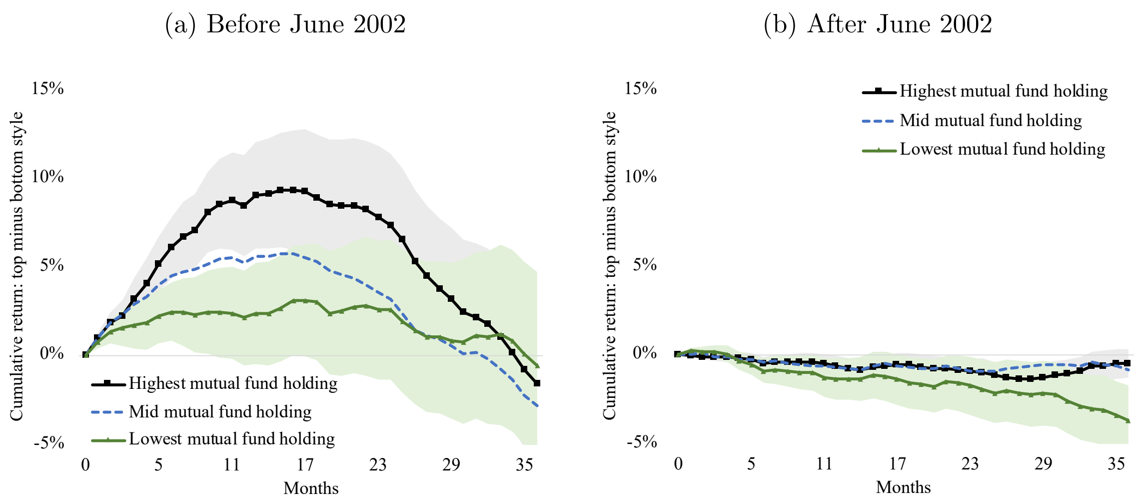
[27] In Figures A.3 and A.4 in the appendix, we verify that our results are robust to using alternative measures of past rating changes instead of ExpSum(ΔRating).

4.1.1 Heterogeneous exposure to Morningstar ratings within styles

To further sharpen the test, we examine whether stocks more heavily held by mutual funds within a given style portfolio experience larger rating-induced price pressures. In each style portfolio, we further sort stocks into three equal-stock-count terciles based on the lagged fraction of shares held by all mutual funds. On average, mutual funds hold 30.6% of the stocks in the top tercile, followed by 18.8% and 11.6% for the next two terciles.

Figure 6. Rating-induced style returns: Splitting on mutual fund ownership

As in panel D in Figure 1, we sort the 3×3 style portfolios by their lagged rating changes (ExpSum(ΔRating)) and plot the cumulative differences in returns between the top and bottom styles for the subsequent 36 months. Both bottom- and top-ranked styles are split into three subsets of stocks, based on the fraction of total shares held by mutual funds. The shaded areas represent 95% confidence intervals based on bootstrapped standard errors.



We then repeat the same exercise—sorting styles using $\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$ and examining the difference between the top and bottom styles for each of the three mutual-fund-holding terciles. Figure 6 plots the results. As predicted, the price effect before June 2002 is stronger in the style portfolios consisting of stocks with higher mutual fund ownership than in those consisting of stocks with low mutual fund ownership.[28] There is no effect in any of the terciles after June 2002.

4.2 Profitability of the rating-driven style momentum strategy

These price pressure results suggest that a rating-based style momentum strategy would be profitable before June 2002, but not afterward. This is indeed the case. In Table 3, we examine the monthly returns of style portfolios sorted by the lagged exponential sum of rating changes ($\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$). Panels A and B show that, prior to June 2002, styles with high past rating changes generally have better performance than other styles and that that difference shrank significantly after June 2002. As a consequence, a trading strategy that goes long on the top-ranked style and short on the bottom-ranked style was profitable, with about a 1% monthly return and capital asset pricing model (CAPM) alpha before June 2002. As we predicted, the strategy became unprofitable after June 2002.

4.3 Cross-sectional dispersion in style returns

As a result of the reform, the dispersion in average Morningstar ratings across fund styles declined sharply after June 2002 (Panel A of Figure 1). Therefore, if our hypothesis—that ratings drive flows and then lead to price impact— is correct, we should observe a decline in the dispersion in style flows and returns after the reform.

To test this prediction, we use two definitions of dispersion: the spread between the styles with highest and lowest realizations, and the standard deviation across all styles. We calculate style-level dispersion in ratings, flows, and returns. We then regress these dispersion measures on an indicator that equals one after June 2002. In addition to using

[28] The style portfolios with higher mutual-fund-holding stocks in fact have *more* overlap in their portfolio constituents. Note that a higher overlap mechanically leads to less return dispersion (100% overlap means zero return dispersion), and therefore could bias against finding a significant result. To examine this possibility, we compute the pairwise overlap of each pair of portfolios j and k, defined as $\sum_{\text{stock } i} |\min(w_{i,\text{portfolio } j}, w_{i,\text{portfolio } k})|$, where $w_{i,\text{portfolio } l}$ is the weight of portfolio l in stock i. We then take an average across all pairs of style portfolios within each tercile. The average overlap for the highest, mid, and lowest mutual fund holding terciles are 24.2%, 17.0%, and 11.9%, respectively.

Table 3. Rating-induced style momentum strategy before and after June 2002

This table shows monthly flows and returns of 3 × 3 Morningstar style portfolios sorted each month by the lagged exponential sum of rating changes (ExpSum(ΔRating)$_{\pi,t-1}$). Panels A and B, respectively, show the average monthly return and CAPM alpha as percentages. The last column is the difference between the top- and bottom-ranked styles. In panel A, style returns are demeaned in each month to focus on the cross-sectional difference. The standard errors are reported in parentheses. *$p < .1$; **$p < .05$; ***$p < .01$.

	Bot.	2	3	4	5	6	7	8	Top	Top − Bot.
A. Return (demeaned)										
Before 2002	−0.42*	−0.45**	−0.25	0.00	−0.08	0.21	−0.06	0.49**	0.54**	0.96**
	(0.22)	(0.22)	(0.18)	(0.17)	(0.11)	(0.15)	(0.16)	(0.23)	(0.24)	(0.44)
After 2002	−0.02	0.08	−0.06	0.08	−0.07	0.04	−0.07	0.04	−0.01	0.01
	(0.08)	(0.07)	(0.08)	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)	(0.09)	(0.15)
B. CAPM alpha										
	Bot.	2	3	4	5	6	7	8	Top	Top − Bot.
Before 2002	−0.24	−0.29	−0.11	0.23	0.13	0.46**	0.20	0.71***	0.82***	1.06***
	(0.23)	(0.23)	(0.18)	(0.18)	(0.16)	(0.19)	(0.23)	(0.27)	(0.29)	(0.37)
After 2002	−0.01	0.09	−0.02	0.12	−0.05	0.05	−0.06	0.08	0.04	0.05
	(0.11)	(0.11)	(0.10)	(0.09)	(0.11)	(0.10)	(0.10)	(0.10)	(0.10)	(0.15)

the full sample, to account for the impact of the dot-com bust, we also use a 4-year window centered on the methodology change event, as well as a full sample window that excludes the 4 years surrounding the event. Standard errors are adjusted using the Newey-West procedure.

Table 4. Dispersion of style ratings, flows, and returns

We regress dispersion measures of monthly ratings, flows, and returns of style portfolios on a dummy that equals one after June 2002. We report the coefficient on the dummy variable in this table. In columns 1, 3, and 5, we measure dispersion using the spread between the styles with the highest and lowest realizations. In columns 2, 4, and 6, we measure dispersion using the standard deviation of those variables. Across the different rows, we vary the sample size used in the regressions. Newey-West standard errors are reported in parentheses. *$p < .1$; **$p < .05$; ***$p < .01$.

Regression coefficient on the post-June 2002 dummy						
Dependent variables:	Rating		Flow (%)		Return (%)	
	Spread	SD	Spread	SD	Spread	SD
	(1)	(2)	(3)	(4)	(5)	(6)
Full sample	−0.61***	−0.22**	−1.88***	−0.60***	−2.54***	−0.90***
	(0.22)	(0.11)	(0.23)	(0.08)	(0.68)	(0.25)
2000Q3–2004Q2	−0.53***	−0.20***	−1.74***	−0.63***	−4.45***	−1.53***
	(0.19)	(0.06)	(0.45)	(0.17)	(0.85)	(0.31)
Exclude 2000Q3–2004Q2	−0.62**	−0.22*	−1.91***	−0.59***	−2.11***	−0.76***
	(0.26)	(0.13)	(0.27)	(0.09)	(0.73)	(0.25)

Table 4 shows the regression coefficients on the post-June-2002 dummy variable. As predicted, regardless of the dispersion measures used, we find that ratings, flows, and returns of styles became less dispersed after June 2002. Columns 1 to 4 show that the dispersion in ratings and flows declined dramatically regardless of the time window.

Columns 5 and 6 show that style return dispersion also dropped precipitously after June 2002. Over the entire sample, the monthly return spread between the top- and bottom-ranked styles dropped by 2.54% after June 2002 (from 5.5% to 2.9%).When excluding the sample period from 2000Q3 to 2004Q2 to alleviate the concern about the dot-com bust, the monthly return spread between the top- and bottom-ranked styles dropped by 2.1% (from 5.0% to 2.9%). The results are qualitatively similar when dispersion is measured using the standard deviation of returns.[29]

Overall, the results in this section suggest that Morningstar ratings have had a noticeable impact on style flow and return dynamics over the sample period studied. We now conduct an event study around the 2002 shock to further zoom in on the methodology reform event.

5 Event Study around the Morningstar Reform

The style-level price pressure results documented so far build on the evidence that rating-induced fund flows have a causal effect on stock returns (see Section 3 and especially Table 1).

In this section, we provide an additional and independent test of rating-induced demand effects on style returns. To do so, we conduct an event study using a 1-year window (January to December 2002) around the reform implementation date. By focusing on a short window and by relying on the degree of exposure of the various styles to the Morningstar reform, we can ensure that the rating changes are primarily caused by the methodology change (as opposed to by managerial skill, for instance). Focusing on a narrow window also reduces the chance that our findings are confounded by other events, such as NYSE decimalization

[29] Moreover, in untabulated tests, we find that the results presented in Table 4 are robust to the inclusion of a time trend control, indicating that they are not driven by a general decline in style-level dispersion.

in early 2001 and the introduction of NYSE autoquoting in 2003 (Hendershott, Jones, and Menkveld, 2011). In addition, we examine other variables around that date to verify that the effects we document do not stem from shocks to the fundamentals of the stocks forming the style portfolios or from the trading behavior of market participants other than mutual funds.

5.1 Performance of styles, by predicted rating impact

Our analysis tracks style ratings, flows, and returns in 2002; the styles are sorted by their exposure to Morningstar's methodology reform. The reform caused the style ratings to converge to three stars (the average rating). Thus, styles that had ratings greater than three stars as of May 2002 experienced a drop in their ratings due to the methodology reform. In contrast, styles that had ratings lower than three stars experienced an increase. The objective of our analysis is to compare the ratings, flows, and returns of the styles that experience the largest changes in ratings due to the reform.

Our ranking of the exposure of styles to the Morningstar reform relies on prewindow information.[30] Specifically, since the change in ratings between May and June includes components that are related both to the methodology reform and to style performance, we rank styles by the *predicted* rating change due to the reform, computed using December 2001 data. We calculate the predicted rating changes in the following fashion. For each fund j, we compute

$$\widehat{\Delta\mathrm{Rating}}_j = \mathrm{Rating}^{\mathrm{counterfactual}}_{j,\mathrm{Dec\ 2001}} - \mathrm{Rating}^{\mathrm{actual}}_{j,\mathrm{Dec\ 2001}}, \tag{12}$$

where $\mathrm{Rating}^{\mathrm{counterfactual}}_{j,\mathrm{Dec\ 2001}}$ is our estimate of what its December 2001 rating would have been under the post-2002 methodology.[31] $\widehat{\Delta\mathrm{Rating}}_j$ thus measures how the fund's rating would

[30]Ranking on prewindow information alleviates concerns about mean-reversion in returns, which would be an issue, for instance, if we rank styles based on their average performance during January to June 2002.

[31]Appendix A.8 provides more details about how the counterfactual ratings are computed in a bottom-up fashion using past fund returns.

have changed in December 2001 had the reform happened then. We then aggregate up these fund-level predictions at the style level.

When we sort the nine styles by the predicted rating change, we find that this procedure correctly predicts which style portfolios experienced the largest changes in June 2002. Specifically, the small-value style enjoyed the highest rating in December 2001 and is therefore predicted to experience the largest reform-driven decline; the large-growth style had the lowest rating in December 2001 and is predicted to experience the largest increase. The reason we can successfully predict the exposure of different styles to the reform using data from 6 months before the reform is that ratings are slow moving.[32]

In Figure 7, we present the evolution of style ratings, flows, and returns in 2002. Panel A, which plots average style ratings (demeaned cross-sectionally), shows a sharp methodology-induced rating collapse exactly at the event. The style most negatively affected by the reform suffered a drop of about 0.4 stars, while the most positively affected style experienced an increase of about 0.4 stars. Similar patterns can be observed when comparing the flows to the second-most positively and negatively affected styles.

Next, we look at flows. Panel C shows that in the months leading up to the event, the top-rated style (expected to be most negatively affected by the reform) experienced approximately 23% additional flows relative to the bottom-rated style. As a result of the reform, the average star rating across styles became approximately 3 at the end of June 2002. Thus, as expected, the divergence in cumulative flows between styles disappeared.

Finally, we examine cumulative style returns in panel E. Pre-event style returns lined up with pre-event ratings and flows. Following the June 2002 event, return differences leveled off with a slight reversal. The most-negatively affected style had returns of 2.6% per month during the pre-event period and reverted to -0.5% after the event. The most-positively affected style had a -1.9% monthly return before the event and 0.3% after.

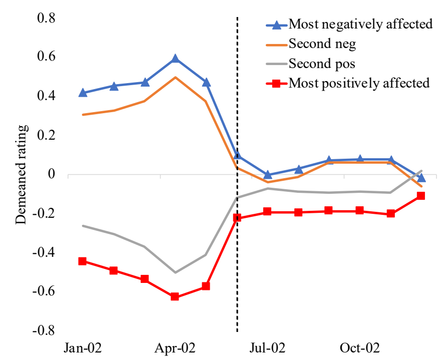
Similar to this style-level exercise, we also perform this event study using individual stocks

[32] As explained in Appendix B, ratings are slow moving because they are based on 3, 5, and 10 years of past fund returns.

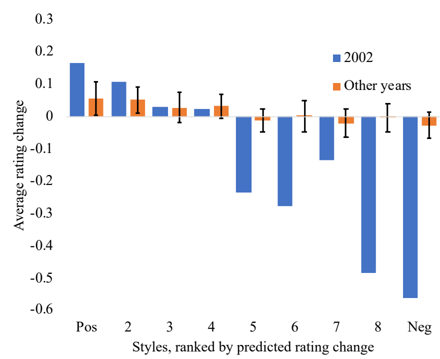
Figure 7. Event study around June 2002

We perform event studies on the 3×3 size-value Morningstar style portfolios during the 6 months before and after the June 2002 methodology change. In the left panels, we sort styles by their *predicted* rating change at the June 2002 event using December 2001 data, and then plot the evolution of their ratings in panel A, cumulative flows in panel C, and cumulative returns in panel E. The dashed vertical line represents the June 2002 event. The right panels conduct the same exercises in years other than 2002 as a placebo test. The blue bars plot the average rating, flow, and return changes after June 2002 (average of July to December 2002 minus average of January to June 2002), while the orange bars plot the corresponding results for years other than 2002. The whiskers represent two-standard-error bands. All variables—ratings, returns, and flows—are demeaned cross-sectionally to focus on cross-sectional dispersion.

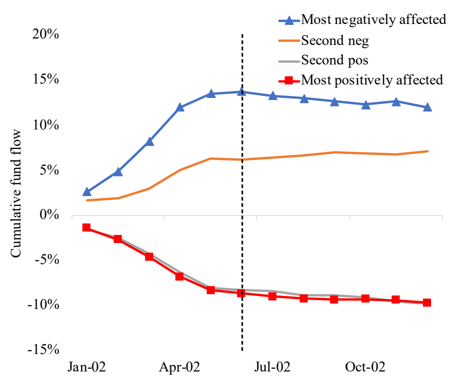
(a) Ratings, 2002



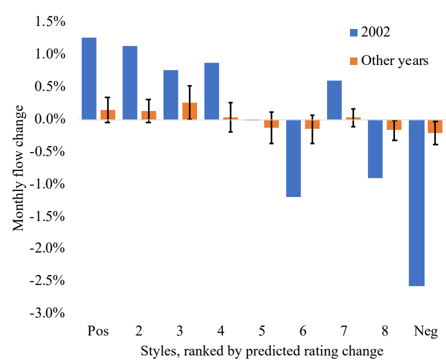
(b) Ratings change: 2002 vs. placebo



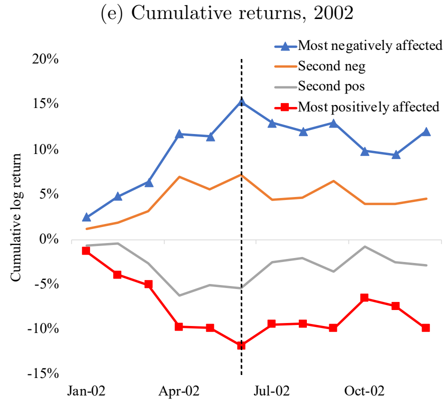
(c) Cumulative flows, 2002



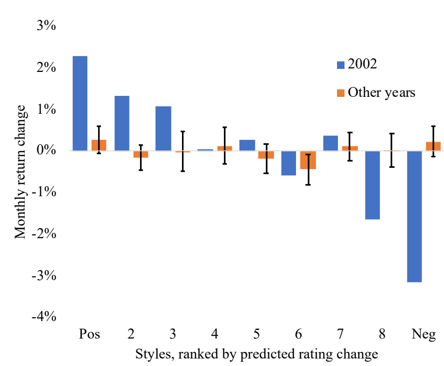
(d) Flow change: 2002 vs. placebo



(e) Cumulative returns, 2002



(f) Return change: 2002 vs. placebo



and obtain similar results. The benefit of repeating the exercise using stocks is that doing so allows us to study a cross-section of thousands of observations (stocks), each of which has a different degree of ex ante "exposure" to the Morningstar reform. Specifically, we follow the same procedure described above to estimate the predicted stock-level methodology-induced rating change using data as of December 2001 and sort stocks into quintiles. Figure A.8 in the appendix plots the evolution of the ratings and cumulative returns of these stocks. The results are consistent with the style-level exercise.

To put the size of the price impact coefficient (reciprocal of demand elasticity) into perspective, a back-of-the-envelope exercise shows that the style-level price impact coefficient in our results is approximately 5.3. That is, buying 1% of the market cap outstanding creates a price impact of 5.3%.[33] This magnitude is consistent with the existing literature. Because the 2002 shock generated less diversifiable style-level flows, we expect our estimate to be higher than estimates based on stock-level shocks but smaller than those based on market-level shocks. Using a demand system approach, Koijen and Yogo (2019) estimate the coefficient to be between 2 to 4 at the stock level. Using the granular instrument approach of Gabaix and Koijen (2021a), Gabaix and Koijen (2021b) estimate the market-level coefficient to be between 5.28 and 7.08. Our estimate lies between these two.[34]

5.2 Testing for alternative explanations

The fact that the reform happened exactly in June 2002 alleviates the concern about other contaminating events. For instance, one may worry about the introduction of NYSE decimalization and autoquoting, both of which increased market liquidity. However, the former happened in early 2001 and the latter in 2003 (Hendershott et al., 2011). The acceleration of 10K filings by the Securities and Exchange Commission (SEC) became effective

[33] In our exercise, the cumulative fund flow difference between the top- and bottom-ranked styles is 22.4% in the 6 months leading up to the event (panel C of Figure 7). The return difference is 27.2%. In 2002, mutual funds held approximately 22.8% of the U.S. stock market in 2002; for this period, we calculate a style-level price impact coefficient of $\frac{27.2\%}{22.5\%} \times \frac{1}{22.8\%} \approx 5.3$.

[34] Consistent with our finding, Li (2021) estimates the size- and value-factor level price impact parameter to be 5.7 using a structural approach.

in November 2002 and can only have had an impact when companies filed their 10Ks, that is, not before early 2003 (Securities and Exchange Commission, 2002). The Sarbanes-Oxley Act was passed in July 2002, but it is unlikely to have driven our results for two reasons. First, the Sarbanes-Oxley Act is concerned with regulating the financial reporting of firms and thus should not have had large impacts on mutual fund flows. Second, while some studies have found that the Sarbanes-Oxley Act event had differential price impacts on different stocks, the findings are unrelated to style differences across firms and are much smaller in magnitude than the effects we document. (See, for instance, results in Jain and Rezaee, 2006; Li, Pincus, and Rego, 2008).[35]

In addition, we conduct three further tests around the June 2002 event window to examine alternative explanations. The results do not support any of the alternative hypotheses.

5.2.1 Placebo test: Other years

First, to alleviate the concern that the style flow and return patterns occur mechanically due to regression to the mean, we conduct a placebo test by rerunning an identical exercise in all years other than 2002. Panels B, D, and F of Figure 7 show that the patterns observed in 2002 did not take place in other years.[36] The orange bars represent the same exercise in other years together with two-standard-error bands. The sharp changes in style ratings, flows, and returns are unique to 2002. In Table A.4 in the appendix, we perform a formal test of the results shown here using panel regressions, and the results support the patterns presented in panel F of Figure 7.

[35] This discussion is motivated by the conjectures put forth in Green et al. (2017) about the differences before and after 2002.

[36] For each placebo year T other than 2002, we sort styles by $\widehat{\text{Rating}}_{\pi}^{\text{pre 2002 methodology}} - \widehat{\text{Rating}}_{\pi}^{\text{post 2002 methodology}}$, which is computed using data in December of year $T-1$.

5.2.2 Other factors that may have affected style returns around 2002

Our event study methodology assumes that no other sudden style-level shocks occurred around June 2002 that could have caused the patterns we observe. Such shocks would need to affect flows and returns of styles in a manner that happens to be aligned with the impact of the Morningstar reform. While we are not aware of similar shocks around June 2002, the inexistence of such shocks is a key assumption that merits further validation.

For this purpose, in Figure 8 we look for sudden changes in a number of other variables. Theoretically speaking, asset prices can move due to changes in fundamentals or due to trading behaviors,[37] and thus we investigate these two possibilities separately.

To investigate changes in fundamentals, we compute return on assets (ROA) and return on equity (ROE) using quarterly Compustat data and plot their evolution at the style-level (value-weighted) in panels A and B of Figure 8. While the fundamentals of different styles do differ and fluctuate, there is no discernible sudden change around June 2002.

To investigate trading behaviors of institutions, we examine trades by 13F institutions at the style level. We obtain quarterly 13F holdings data from Thomson Reuters.[38] We then plot the cumulative trading by different types of institutions as a fraction of market capitalization in panels C to E.

Panel C plots the combined trading by investment companies and independent investor advisors, a category that includes most mutual funds. Consistent with our flow-induced price pressure mechanism, these institutions indeed trade in a manner aligned with the flow patterns depicted in Figure 7, panel A: they traded into (out of) styles with high (low) pre-2002 ratings, before halting suddenly right after June 2002. Panel D plots the trading of other 13F institutions including banks, insurance companies, pension funds, endowment funds, and other institutions. Panel E plots the trading of 13F institutions that represent hedge funds

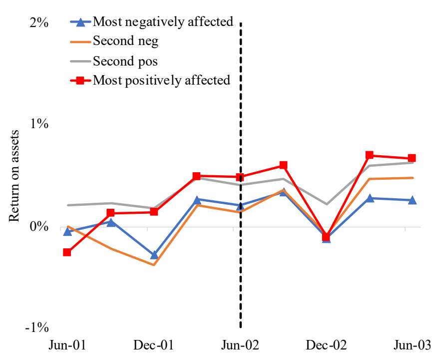
[37] It is unlikely that a rationally determined discount rate can vary so much in a short period of time, so we do not investigate this possibility. Note that 2002 is not a recession period; the U.S. economy was already out of the dot-com-related recession by November 2001, according to the National Bureau of Economic Research.

[38] For panels C and D, we separate 13F institutions by their legal types, which are obtained from Brian Bushee's website. See `https://accounting-faculty.wharton.upenn.edu/bushee/` for details.

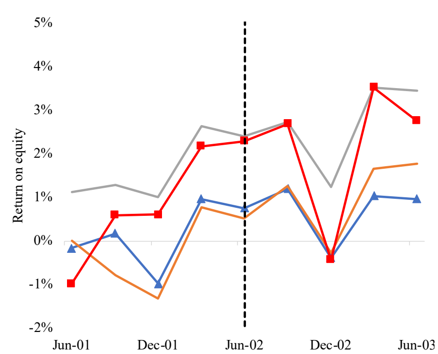
Figure 8. Event study around June 2002: Alternative explanations

We perform event studies on the 3×3 Morningstar style portfolios around the June 2002 methodology change. As in Figure 7, we sort styles by their *predicted* rating change at the June 2002 event using December 2001 data, and then plot the evolution of various variables. To assess changes in fundamentals, panels A and B plot the evolution of return on assets and return on equity, respectively. Panels C to E plot the cumulative trading (as a fraction of overall market capitalization) by different types of 13F institutions. Panel C plots trading by investment companies and advisors, which include most mutual funds. Panel D plots trading by other 13F institutions; panel E plots trading by hedge funds that do not offer mutual funds (i.e., pure-play 13F hedge funds). The trading measures in panels C to E are demeaned by period to focus on cross-sectional dispersion. Panel F plots the evolution of aggregate short interest. The vertical dashed line represents the June 2002 methodology change event.

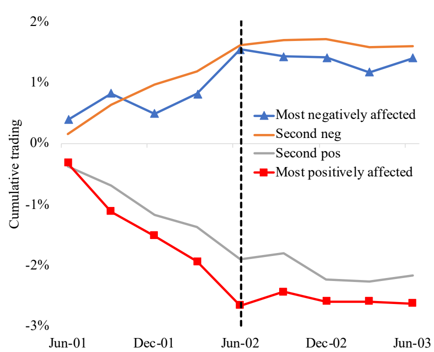
(a) Return on assets (ROA)



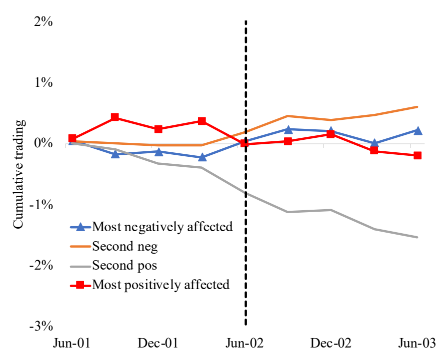
(b) Return on equity (ROE)



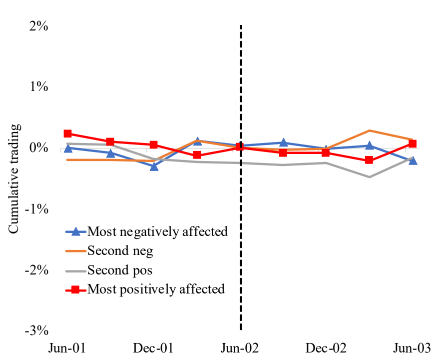
(c) Trading: Investment companies and advisors



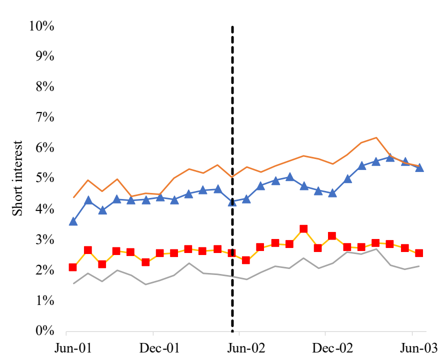
(d) Trading: Other 13F institutions



(e) Trading: Hedge funds



(f) Short interest



that do not offer mutual funds.[39] Neither panel D nor panel E show any discernible "kink" in the trading of any of those non-mutual-fund institutions. Finally, because 13F data only record long positions, we also examine the evolution of aggregate short interest in panel F. We see a general slow rise in short interest across all styles over the window but no clear change around the event.

5.2.2.1 Controlling for stock characteristics. The previous two tests show that there were no sudden changes in fundamentals or trading behaviors of non-mutual-fund institutions around 2002. However, one might still argue that our results could be driven by sudden characteristics-related return changes that happened for other reasons. For example, one might hypothesize that investor sentiment suddenly decreased for small-value stocks after June 2002, causing returns to decrease for those stocks.

To further alleviate this concern, we show that our results—that predicted rating changes explain return changes—also take place at the stock level after controlling for size and book-to-market ratio characteristics. Specifically, for each stock i, we define

$$\text{Rating}_{i,t}^{\text{idiosyncratic}} = \text{Rating}_{i,t} - \text{Rating}_{\text{size-book/market portfolio } p,t}, \tag{13}$$

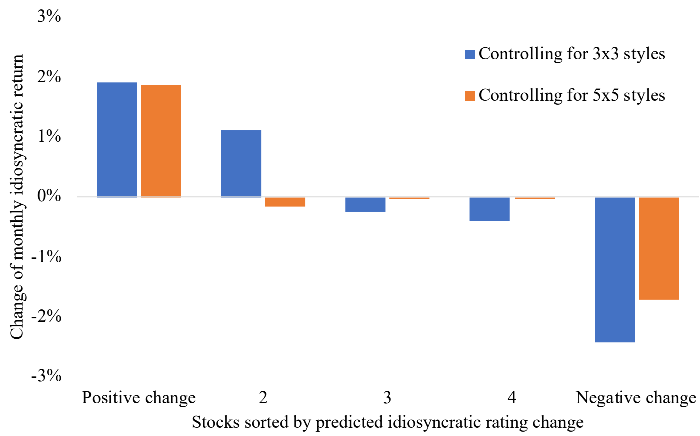
$$\text{Ret}_{i,t}^{\text{idiosyncratic}} = \text{Ret}_{i,t} - \text{Ret}_{\text{size-book/market portfolio } p,t}, \tag{14}$$

where p is the 3×3 or 5×5 size and book-to-market ratio sorted portfolio (constructed using NYSE cutoffs) to which stock i belongs. We compute the rating and return of those portfolios by aggregating from the underlying stocks using market cap weights. We choose size and book-to-market ratio because they are the most commonly used characteristics in the literature and because they are the most parsimonious way to "mimic" the size-value style box used by Morningstar. Thus, $\text{Rating}_{i,t}^{\text{idiosyncratic}}$ and $\text{Ret}_{i,t}^{\text{idiosyncratic}}$ measure the components of ratings and returns that are *not* explained by realized characteristics-related returns.

[39] These hedge funds are identified in Ben-David, Franzoni, and Moussawi (2012). Please see that paper for details.

Figure 9. Return change around June 2002: Controlling for characteristics

We first subtract value-weighted average ratings and returns of 3×3 or 5×5 size–book-to-market ratio portfolios to obtain *idiosyncratic* ratings and returns for each stock (see Equations (13) and (14)). Then, we sort stocks into quintiles based on their *predicted* idiosyncratic rating changes at the June 2002 event using December 2001 data. The top (bottom) quintile comprises stocks expected to experience the largest decline (increase) of idiosyncratic ratings. The figure plots the difference between the average 6-month idiosyncratic return after the event (July to December 2002) and the average 6-month idiosyncratic return before the event (January to June 2002).



We then sort on the predicted idiosyncratic rating change (using December 2001 data)[40] and examine the change in idiosyncratic stock returns after the event in Figure 9. Note that our mechanism—that rating changes affect returns—works both at the characteristic-spanned and and the characteristic-orthogonal (idiosyncratic) levels. Therefore, even after controlling for characteristics, we should still expect to see an effect. Consistent with this expectation, stocks predicted to experience large upgrades (downgrades) in *idiosyncratic* rating (Equation (13)) experience positive (negative) changes in *idiosyncratic* returns (Equation (14)). The quantitative relation between rating changes and return changes is also comparable to that found at the style level (panel F in Figure 7). These results further suggest that our style-level findings around June 2002 are unlikely to be driven by unspecified characteristics-level return movements.

[40]Specifically, we first compute counterfactual ratings for each fund in December 2001 under the post-reform Morningstar methodology. We then aggregate those at the stock and portfolio levels and subtract the latter from the former to get the counterfactual *idiosyncratic* rating at the stock level. The difference between this counterfactual idiosyncratic rating and the actual idiosyncratic rating in December 2001 is our prediction of the change.

6 Conclusion

In recent years, evidence has mounted that investor demand can exert significant pressure on asset prices. However, it is difficult to identify demand that is nonfundamental, that is large enough to affect a broad set of securities, and that can also plausibly cause systematic price fluctuations. Our study presents evidence that Morningstar rating-driven household demand for mutual funds contributes to economically significant price fluctuations at the style level.

In our empirical setting, a reform that Morningstar enacted in June 2002 serves to shift demand between two regimes. The reform equalized ratings across styles, causing capital flows to be spread more evenly across styles. Throughout the sample period, investors directed capital in accordance with Morningstar ratings, and such rating-driven flows generated price impact. Prior to the reform, style-level rating imbalances created style-level price pressures and increased return dispersion across styles. After the reform, these patterns became much weaker, consistent with the removal of style-level rating-induced price pressures. Using an event study around the exact reform date, we find evidence that style-level price pressure from fund flows ceased at the event date.

Our results focus on one specific source of nonfundamental correlated demand—Morningstar ratings provided to mutual fund investors—and show its impact on the equity market. The overall role of correlated demand in determining asset prices is likely greater than what is documented here. Correlated demand also can arise from other sources such as demand for certain styles driven by institutional frictions (Froot and Teo, 2008; Koijen and Yogo, 2019) and performance chasing in index-linked products (Broman, 2016; Dannhauser and Pontiff, 2019). Taken together, these findings should alter the way economists interpret systematic price movements: instead of solely reflecting fundamental risks, they also may be determined by nonfundamental demand.

References

Abarbanell, Jeffery S, and Brian J Bushee, 1998, Abnormal returns to a fundamental analysis strategy, *Accounting Review* 73, 19–45.

Amihud, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Arnott, Robert D., Mark Clements, Vitali Kalesnik, and Juhani T. Linnainmaa, 2021, Factor momentum, Working Paper, Dartmouth College.

Asness, Clifford, Andrea Frazzini, Ronen Israel, Tobias J Moskowitz, and Lasse H Pedersen, 2018, Size matters, if you control your junk, *Journal of Financial Economics* 129, 479–509.

Bali, Turan G, Nusret Cakici, and Robert F Whitelaw, 2011, Maxing out: Stocks as lotteries and the cross-section of expected returns, *Journal of Financial Economics* 99, 427–446.

Ball, Ray, Joseph Gerakos, Juhani T Linnainmaa, and Valeri Nikolaev, 2016, Accruals, cash flows, and operating profitability in the cross section of stock returns, *Journal of Financial Economics* 121, 28–45.

Banz, Rolf W., 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9, 3–18.

Barbee Jr, William C, Sandip Mukherji, and Gary A Raines, 1996, Do sales–price and debt–equity explain stock returns better than book–market and firm size?, *Financial Analysts Journal* 52, 56–60.

Barber, Brad M., Xing Huang, and Terrance Odean, 2016, Which factors matter to investors? Evidence from mutual fund flows, *Review of Financial Studies* 29, 2600–2642.

Barberis, Nicholas, and Andrei Shleifer, 2003, Style investing, *Journal of Financial Economics* 68, 161–199.

Barry, Christopher B, and Stephen J Brown, 1984, Differential information and the small firm effect, *Journal of Financial Economics* 13, 283–294.

Basu, Sanjoy, 1977, Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis, *Journal of Finance* 32, 663–682.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, 2012, Hedge fund stock trading in the financial crisis of 2007–2009, *Review of Financial Studies* 25, 1–54.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, 2018, Do ETFs increase volatility?, *Journal of Finance* 73, 2471–2535.

Ben-David, Itzhak, Francesco Franzoni, Rabih Moussawi, and John Sedunov, 2021a, The granular nature of large institutional investors, *Management Science* Advanced access published April 30, 2021, 10.1287/mnsc.2020.3808.

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, 2021b, What do mutual fund investors really care about?, *Review of Financial Studies* Advance Access published April 30, 2021, 10.1287/mnsc.2020.3808.

Blume, Marshall E., 1998, An anatomy of Morningstar ratings, *Financial Analysts Journal* 54, 19–27.

Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R Roberts, 2007, On the importance of measuring payout yield: Implications for empirical asset pricing, *Journal of Finance* 62, 877–915.

Bradshaw, Mark T, Scott A Richardson, and Richard G Sloan, 2006, The relation between corporate financing activities, analysts' forecasts and stock returns, *Journal of Accounting and Economics* 42, 53–85.

Broman, Markus S., 2016, Liquidity, style investing and excess comovement of exchange-traded fund returns, *Journal of Financial Markets* 30, 27–53.

Brown, David C., Shaun Davies, and Matthew Ringgenberg, 2021, ETF arbitrage and return predictability, *Review of Finance* 25, 937–972.

Campbell, John Y, Jens Hilscher, and Jan Szilagyi, 2008, In search of distress risk, *Journal of Finance* 63, 2899–2939.

Chan, Louis KC, Josef Lakonishok, and Theodore Sougiannis, 2001, The stock market valuation of research and development expenditures, *Journal of Finance* 56, 2431–2456.

Chang, Yen-Cheng, Harrison Hong, and Inessa Liskovich, 2015, Regression discontinuity and the price effects of stock market indexing, *Review of Financial Studies* 28, 212–246.

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

Cochrane, John H., 2011, Presidential address: Discount rates, *Journal of Finance* 66, 1047–1108.

Cooper, Michael J., Huseyin Gulen, and Michael J. Schill, 2008, Asset growth and the cross-section of stock returns, *Journal of Finance* 63, 1609–1651.

Coval, Joshua, and Eric Stafford, 2007, Asset fire sales (and purchases) in equity markets, *Journal of Financial Economics* 86, 479–512.

Daniel, Kent, and Sheridan Titman, 2006, Market reactions to tangible and intangible information, *Journal of Finance* 61, 1605–1643.

Dannhauser, Caitlin D., and Jeffrey Pontiff, 2019, Flow, Working Paper, Boston University.

De Bondt, Werner F.M., and Richard Thaler, 1985, Does the stock market overreact?, *Journal of Finance* 40, 793–805.

Del Guercio, Diane, and Paula A. Tkac, 2008, Star power: The effect of Morningstar ratings on mutual fund flow, *Journal of Financial and Quantitative Analysis* 43, 907–936.

Dichev, Ilia D, 1998, Is the risk of bankruptcy a systematic risk?, *Journal of Finance* 53, 1131–1147.

Edmans, Alex, Itay Goldstein, and Wei Jiang, 2012, The real effects of financial markets: The impact of prices on takeovers, *Journal of Finance* 67, 933–971.

Evans, Richard B., and Yang Sun, 2021, Models or stars: The role of asset pricing models and heuristics in investor risk adjustment, *Review of Financial Studies* 34, 67–107.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Fisher, Philip A., 1958, *Common stocks and uncommon profits* (Harper).

Francis, Jennifer, Ryan LaFond, Per M Olsson, and Katherine Schipper, 2004, Costs of equity and earnings attributes, *Accounting Review* 79, 967–1010.

Froot, Kenneth, and Melvyn Teo, 2008, Style investing and institutional investors, *Journal of Financial and Quantitative Analysis* 43, 883–906.

Gabaix, Xavier, and Ralph S.J. Koijen, 2021a, Granular instrumental variables, Working Paper, Harvard University.

Gabaix, Xavier, and Ralph S.J. Koijen, 2021b, In search of the origins of financial fluctuations: The inelastic markets hypothesis, Working Paper, Harvard University.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, 2015, Money doctors, *Journal of Finance* 70, 91–114.

George, Thomas J., and Chuan-Yang Hwang, 2004, The 52-week high and momentum investing, *Journal of Finance* 59, 2145–2176.

Graham, Benjamin, and David L. Dodd, 1934, *Security analysis*, 7th edition (Whittlesey House, McGraw-Hill, New York).

Green, Jeremiah, John R.M. Hand, and X. Frank Zhang, 2017, The characteristics that provide independent information about average US monthly stock returns, *Review of Financial Studies* 30, 4389–4436.

Griffin, John M, and Michael L Lemmon, 2002, Book-to-market equity, distress risk, and stock returns, *Journal of Finance* 57, 2317–2336.

Harris, Lawrence, and Eitan Gurel, 1986, Price and volume effects associated with changes in the S&P 500 list: New evidence for the existence of price pressures, *Journal of Finance* 41, 815–829.

Haugen, Robert A, and Nardin L Baker, 1996, Commonality in the determinants of expected stock returns, *Journal of Financial Economics* 41, 401–439.

Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld, 2011, Does algorithmic trading improve liquidity?, *Journal of Finance* 66, 1–33.

Hirshleifer, David, Kewei Hou, Siew Hong Teoh, and Yinglei Zhang, 2004, Do investors overvalue firms with bloated balance sheets?, *Journal of Accounting and Economics* 38, 297–331.

Hou, Kewei, and David T Robinson, 2006, Industry concentration and average stock returns, *Journal of Finance* 61, 1927–1956.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 650–705.

Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating anomalies, *Review of Financial Studies* 33, 2019–2133.

Huang, Shiyang, Yang Song, and Hong Xiang, 2020, Noise trading and asset pricing factors, Working paper, University of Washington.

Jain, Pankaj K., and Zabihollah Rezaee, 2006, The Sarbanes-Oxley Act of 2002 and capital-market behavior: Early evidence, *Contemporary Accounting Research* 23, 629–654.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Kaniel, Ron, and Robert Parham, 2017, WSJ category kings: The impact of media attention on consumer and mutual fund investment decisions, *Journal of Financial Economics* 123, 337–356.

Koijen, Ralph S.J., and Motohiro Yogo, 2019, A demand system approach to asset pricing, *Journal of Political Economy* 127, 1475–1515.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, Contrarian investment, extrapolation, and risk, *Journal of Finance* 49, 1541–1578.

Lettau, Martin, Sydney C. Ludvigson, and Paulo Manoel, 2019, Characteristics of mutual fund portfolios: Where are the value funds?, Working Paper, University of California at Berkeley.

Lewellen, Jonathan, 2015, The cross-section of expected stock returns, *Critical Finance Review* 4, 1–44.

Li, Haidan, Morton Pincus, and Sonja Olhoft Rego, 2008, Market reaction to events surrounding the Sarbanes-Oxley Act of 2002 and earnings management, *Journal of Law and Economics* 51, 111–134.

Li, Jiacui, 2021, What drives the size and value factors?, Working Paper, University of Utah.

Lockwood, Larry, and Wikrom Prombutr, 2010, Sustainable growth and stock returns, *Journal of Financial Research* 33, 519–538.

Lou, Dong, 2012, A flow-based explanation for return predictability, *Review of Financial Studies* 25, 3457–3489.

Loughran, Tim, and Jay W Wellman, 2011, New evidence on the relation between the enterprise multiple and average stock returns, *Journal of Financial and Quantitative Analysis* 46, 1629–1650.

Morningstar, 2001, Morningstar appoints paul kaplan, ph.d., cfa, as director of research, vahid fathi named director of stock research., Press release, Morningstar.

Moskowitz, Tobias J., and Mark Grinblatt, 1999, Do industries explain momentum?, *Journal of Finance* 54, 1249–1290.

Norris, Floyd, 2002, Morningstar to grade on a curve, *New York Times* April 23.

Novy-Marx, Robert, 2011, Operating leverage, *Review of Finance* 15, 103–134.

Novy-Marx, Robert, 2012, Is momentum really momentum?, *Journal of Financial Economics* 103, 429–453.

Novy-Marx, Robert, 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108, 1–28.

Parker, Jonathan, Antoinette Schoar, and Yang Sun, 2020, Retail financial innovation and stock market dynamics: The case of target date funds, Working Paper, Massachusetts Institute of Technology.

Pástor, Ľuboš, Robert F. Stambaugh, and Lucian A. Taylor, 2020, Fund tradeoffs, *Journal of Financial Economics* 614–634.

Piotroski, Joseph D, 2000, Value investing: The use of historical financial statement information to separate winners from losers, *Journal of Accounting Research* 1–41.

Pontiff, Jeffrey, and Artemiza Woodgate, 2008, Share issuance and cross-sectional returns, *Journal of Finance* 63, 921–945.

Reuter, Jonathan, and Eric Zitzewitz, 2021, How much does size erode mutual fund performance? A regression discontinuity approach, *Review of Finance* Advance Access published May 28, 2021, 10.1093/rof/rfab016.

Securities and Exchange Commission, 2002, Final rule: Acceleration of periodic report filing dates and disclosure concerning website access to reports, Release Nos. 33-8128, 34-46464, Washington, DC, Securities and Exchange Commission.

Shleifer, Andrei, 1986, Do demand curves for stocks slope down?, *Journal of Finance* 41, 579–590.

Sloan, Richard G, 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings?, *Accounting Review* 71, 289–315.

Soliman, Mark T, 2008, The use of dupont analysis by market participants, *Accounting Review* 83, 823–853.

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

Teo, Melvyn, and Sung-Jun Woo, 2004, Style effects in the cross-section of stock returns, *Journal of Financial Economics* 74, 367–398.

Thomas, Jacob K, and Huai Zhang, 2002, Inventory changes and future returns, *Review of Accounting Studies* 7, 163–187.

Titman, Sheridan, KC John Wei, and Feixue Xie, 2004, Capital investments and stock returns, *Journal of financial and Quantitative Analysis* 39, 677–700.

Wardlaw, Malcolm, 2020, Measuring mutual fund flow pressure as shock to stock returns, *Journal of Finance* 75, 3221–3243.

Wermers, Russ, 2000, Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses, *Journal of Finance* 55, 1655–1695.

Wurgler, Jeffrey, and Ekaterina Zhuravskaya, 2002, Does arbitrage flatten demand curves for stocks?, *Journal of Business* 75, 583–608.

Xing, Yuhang, 2008, Interpreting the value effect through the q-theory: An empirical investigation, *Review of Financial Studies* 21, 1767–1795.

Appendix A Additional Results

In this section, we provide additional results and robustness checks for the analysis presented in the main body of the paper.

A.1 Sample Statistics

Table A.1 shows detailed statistics of our mutual fund sample from 1991 to 2018. Our sample includes 433 mutual funds at the beginning of our sample period, and the number peaked in 2008 at 2,062. Since then, the number of funds decreased slightly, whereas total assets under management increased from 2009 onward, reaching about $4 trillion by the end of our sample period in 2018. Columns 4 to 8 report the distribution of funds in each rating category; column 9 shows the fraction of sector funds; columns 10 to 13 report the fraction of funds in different styles. Table A.2 shows the summary statistics of the nine Morningstar fund-based styles.

Table A.1. Summary statistics of mutual funds, by year

Columns 1 to 3 show the year, the number of mutual funds, and their aggregate AUM. Columns 4 to 8 indicate the fraction of funds assigned to each Morningstar star rating. Note that these fractions can differ from (10%, 22.5%, 35%, 22.5%, 10%) because Morningstar assigns those fixed fractions of ratings at the share-class level, but we follow Barber et al. (2016) in aggregating ratings at the fund level by value-weighting different share classes and rounding to the nearest integer. Column 9 indicates the fraction that are sector funds. The other U.S. domestic equity funds that are considered diversified are classified into the 3×3 style box categories, and columns 10 to 13 indicate the fraction of funds that fall within the different styles.

Year	Number funds	AUM ($ billion)	Fraction by rating					Sector funds	Diversified fund style			
			1 star	2	3	4	5 star		Large	Small	Growth	Value
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1991	433	458.2	9%	23%	36%	23%	10%	19%	51%	16%	30%	28%
1992	466	600.8	9%	25%	32%	23%	11%	18%	52%	17%	29%	29%
1993	525	733.6	8%	22%	38%	23%	9%	17%	54%	16%	27%	30%
1994	587	748.8	7%	23%	34%	25%	10%	16%	54%	17%	27%	30%
1995	702	971.8	9%	22%	32%	27%	11%	15%	53%	17%	28%	28%
1996	826	1,177.6	8%	21%	31%	28%	13%	15%	51%	20%	30%	28%
1997	942	1,416.0	9%	22%	30%	26%	13%	14%	53%	20%	30%	29%
1998	1,069	1,524.5	10%	22%	28%	25%	14%	14%	55%	20%	33%	28%
1999	1,238	1,721.4	12%	21%	27%	26%	14%	14%	55%	22%	37%	27%
2000	1,454	1,510.0	10%	20%	30%	26%	14%	14%	57%	23%	37%	28%
2001	1,595	1,238.7	9%	20%	34%	23%	15%	15%	57%	22%	38%	27%
2002	1,731	964.3	8%	21%	36%	25%	10%	15%	57%	22%	41%	23%
2003	1,948	1,072.5	8%	22%	36%	24%	9%	16%	56%	22%	43%	22%
2004	2,020	1,224.5	8%	22%	37%	24%	8%	16%	56%	22%	43%	22%
2005	2,021	1,366.5	6%	25%	39%	23%	7%	15%	56%	22%	42%	23%
2006	1,997	1,567.6	8%	24%	38%	23%	7%	15%	56%	22%	41%	23%
2007	2,019	1,681.6	8%	25%	38%	22%	7%	15%	56%	23%	41%	23%
2008	2,062	946.6	8%	24%	37%	23%	8%	15%	55%	23%	41%	23%
2009	2,019	1,249.2	8%	23%	38%	23%	7%	14%	54%	23%	42%	23%
2010	1,912	1,472.2	7%	23%	38%	24%	8%	14%	55%	23%	41%	23%
2011	1,853	1,574.3	6%	23%	38%	26%	6%	14%	56%	23%	40%	23%
2012	1,778	1,819.9	7%	23%	37%	26%	7%	14%	56%	23%	41%	22%
2013	1,700	2,503.1	6%	24%	38%	26%	6%	15%	56%	23%	42%	23%
2014	1,651	2,924.7	7%	21%	38%	28%	7%	15%	56%	24%	41%	24%
2015	1,635	2,969.2	8%	21%	37%	27%	8%	15%	55%	24%	40%	25%
2016	1,666	3,046.6	6%	22%	37%	27%	7%	16%	55%	24%	40%	25%
2017	1,633	3,723.2	6%	22%	37%	28%	8%	16%	54%	25%	38%	25%
2018	1,563	3,820.4	7%	21%	36%	28%	9%	16%	54%	26%	38%	26%

Table A.2. Summary statistics of styles, by year

A total of nine stock styles (small/mid/large cap × value/blend/growth) are used in this study. Columns 2 and 3 summarize the mean and standard deviation of Morningstar ratings for each style. Columns 4 and 5 summarize the lagged 12-month rating changes (ExpSum(ΔRating)). Columns 6 and 7 summarize monthly fund flows in the styles, and columns 8 and 9 summarize monthly style returns.

Year	Rating		ExpSum(ΔRating)		Monthly fund flow		Monthly return	
	Average	SD	Average	SD	Average	SD	Average	SD
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1991	3.17	0.67	−0.03	0.27	0.59%	1.10%	3.20%	5.15%
1992	3.33	0.45	−0.05	0.34	1.24%	1.69%	1.10%	3.25%
1993	3.51	0.50	0.06	0.24	1.08%	1.67%	1.41%	2.74%
1994	3.68	0.44	0.04	0.17	0.98%	0.89%	0.04%	3.29%
1995	3.84	0.43	0.07	0.19	1.18%	0.82%	2.65%	2.50%
1996	3.75	0.40	−0.21	0.28	1.24%	1.06%	1.75%	3.81%
1997	3.58	0.63	−0.18	0.40	1.08%	1.04%	2.16%	4.72%
1998	3.39	0.71	−0.12	0.41	0.12%	1.08%	1.33%	7.47%
1999	3.23	0.82	−0.01	0.38	−0.49%	1.51%	2.12%	5.19%
2000	3.38	0.67	0.01	0.47	0.15%	1.37%	0.38%	7.47%
2001	3.67	0.51	−0.03	0.52	0.84%	1.23%	−0.18%	6.83%
2002	3.69	0.46	−0.16	0.35	0.46%	1.50%	−1.70%	5.86%
2003	3.58	0.23	−0.08	0.12	0.76%	0.76%	2.88%	3.91%
2004	3.57	0.19	−0.09	0.09	0.53%	0.79%	1.37%	3.17%
2005	3.58	0.20	−0.09	0.11	0.13%	0.60%	0.81%	3.23%
2006	3.65	0.20	−0.03	0.09	0.01%	0.55%	1.19%	2.85%
2007	3.64	0.23	−0.13	0.14	−0.17%	0.69%	0.51%	3.11%
2008	3.50	0.24	−0.21	0.16	−0.39%	0.73%	−3.79%	7.52%
2009	3.42	0.25	−0.14	0.12	−0.06%	0.74%	2.76%	6.95%
2010	3.41	0.19	−0.04	0.10	−0.19%	0.55%	1.92%	6.04%
2011	3.49	0.15	0.04	0.14	−0.28%	0.58%	0.08%	5.67%
2012	3.59	0.12	0.04	0.10	−0.33%	0.36%	1.41%	3.30%
2013	3.66	0.14	−0.01	0.07	0.23%	0.45%	2.76%	2.72%
2014	3.71	0.15	0.01	0.07	−0.13%	0.71%	0.84%	3.27%
2015	3.74	0.11	−0.04	0.10	−0.26%	0.48%	−0.02%	3.74%
2016	3.77	0.17	−0.01	0.16	−0.37%	0.54%	1.32%	4.01%
2017	3.83	0.15	−0.02	0.14	−0.22%	0.55%	1.57%	1.62%
2018	3.87	0.15	0.04	0.16	−0.19%	0.42%	1.43%	2.93%

A.2 Horse Race against Other Stock Return Predictors

Table 1 in the main paper shows that stock-level lagged rating changes have significant stock price impact after controlling for common return predictors. In this section, in addition to the standard Fama-French-Carhart predictors, we also examine how our rating-based predictor compares with a relatively comprehensive list of 43 additional return predictors, which are chosen by following Arnott, Clements, Kalesnik, and Linnainmaa (2021).[41] Listed in Table A.3, these 43 predictors cover all six predictor categories in Hou, Xue, and Zhang (2020): intangibles, investment, momentum, profitability, trading frictions, and value/growth.

How does the rating-based return predictability compare with the other predictors? We run a simple horse race. In Figure A.1, we compare the return prediction coefficient of $\text{ExpSum}(\Delta\text{Rating})_{i,t-1}$ against all the other predictors. To make the predictive coefficients comparable, we transform all predictors into z-scores. To alleviate the concern that the result is sensitive to the exponential sum specification ($\text{ExpSum}(\Delta\text{Rating})$), we also include the alternative specifications of lagged 3-, 6-, 9-, or 12-month rating changes. We plot the predictive coefficients of Fama-MacBeth regressions in Figure A.1.

Panels A and B plot the coefficients in univariate regressions in which we only control for decile indicators of past-36-month fund returns and benchmark-adjusted fund returns, both of which are also included in Table 1. Panels C and D plot multivariate results after controlling for the six commonly used characteristics of size, value, profitability, investment, momentum, and reversal. For visualization, the regression coefficient for $\text{ExpSum}(\Delta\text{Rating})_{i,t-1}$ is shown in green; the coefficients for alternative rating change specifications are displayed in red; and the other 43 predictors are colored blue.

On a univariate basis, while $\text{ExpSum}(\Delta\text{Rating})_{i,t-1}$ only ranks seventh when all stocks are used (see panel A), panel B shows that it is the strongest predictor when microcap stocks are excluded. Further, this finding holds true in the alternative rating specifications. After controlling for the six commonly used characteristics, when excluding micro-

[41] We restrict our attention to those that can be constructed using CRSP and Compustat data.

Table A.3. Stock return predictors used in horse race analysis

The table lists the stock characteristics that prior studies have documented as return predictors. Following Hou et al. (2020), we classify the stock characteristics into six categories: intangibles, investment, momentum, profitability, trading frictions, and value/growth.

Category	Stock characteristic	Publication
Intangibles (6)	Industry concentration	Hou and Robinson (2006)
	Operating leverage	Novy-Marx (2011)
	Firm age	Barry and Brown (1984)
	Advertising expense	Chan, Lakonishok, and Sougiannis (2001)
	R&D expense	Chan et al. (2001)
	Earnings persistence	Francis, LaFond, Olsson, and Schipper (2004)
Investment (12)	Abnormal capital investment	Titman, Wei, and Xie (2004)
	Accruals	Sloan (1996)
	5-year share issuance	Daniel and Titman (2006)
	Growth in inventory	Thomas and Zhang (2002)
	Industry-adjusted CAPEX growth	Abarbanell and Bushee (1998)
	Investment growth	Xing (2008)
	Investment-to-assets	Hou et al. (2015)
	Investment-to-capital	Xing (2008)
	Net operating assets	Hirshleifer, Hou, Teoh, and Zhang (2004)
	Net working capital changes	Soliman (2008)
	One-year share issuance	Pontiff and Woodgate (2008)
	Total external financing	Bradshaw, Richardson, and Sloan (2006)
Momentum (4)	Fifty-two-week high	George and Hwang (2004)
	Intermediate momentum ($t-7,\ t-12$)	Novy-Marx (2012)
	Industry momentum	Moskowitz and Grinblatt (1999)
	Momentum ($t-2,\ t-6$)	Jegadeesh and Titman (1993)
Profitability (13)	Cash-based profitability	Ball, Gerakos, Linnainmaa, and Nikolaev (2016)
	Change in asset turnover	Soliman (2008)
	Distress risk	Campbell, Hilscher, and Szilagyi (2008)
	Gross profitability	Novy-Marx (2013)
	Ohlson's O-score	Griffin and Lemmon (2002)
	Piotroski's F-score	Piotroski (2000)
	Profit margin	Soliman (2008)
	QMJ profitability	Asness, Frazzini, Israel, Moskowitz, and Pedersen (2018)
	Return on assets	Haugen and Baker (1996)
	Return on equity	Haugen and Baker (1996)
	Sales-minus-inventory growth	Abarbanell and Bushee (1998)
	Sustainable growth	Lockwood and Prombutr (2010)
	Altman's z-score	Dichev (1998)
Trading frictions (2)	Amihud illiquidity	Amihud (2002)
	Maximum daily return	Bali, Cakici, and Whitelaw (2011)
Value/growth (6)	Cash flow-to-price	Lakonishok, Shleifer, and Vishny (1994)
	Earnings-to-price	Basu (1977)
	Enterprise multiple	Loughran and Wellman (2011)
	Sales growth	Lakonishok et al. (1994)
	Sales-to-price	Barbee Jr, Mukherji, and Raines (1996)
	Net payout yield	Boudoukh, Michaely, Richardson, and Roberts (2007)

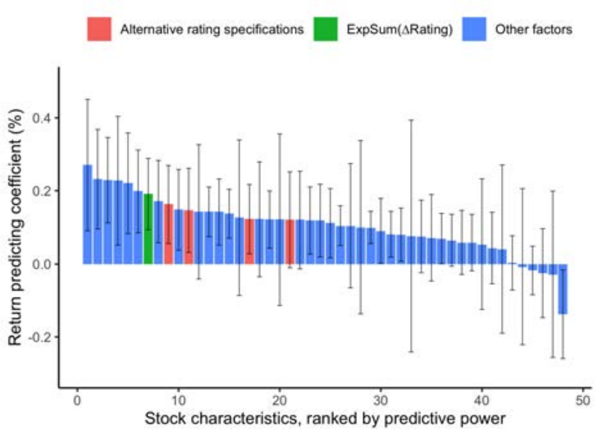
caps, $\text{ExpSum}(\Delta\text{Rating})_{i,t-1}$ is the second-most powerful predictor following the maximum daily return characteristic in Bali et al. (2011), and the result is robust to using alternative rating specifications.

These results indicate that the rating-induced return predictability is one of the strongest

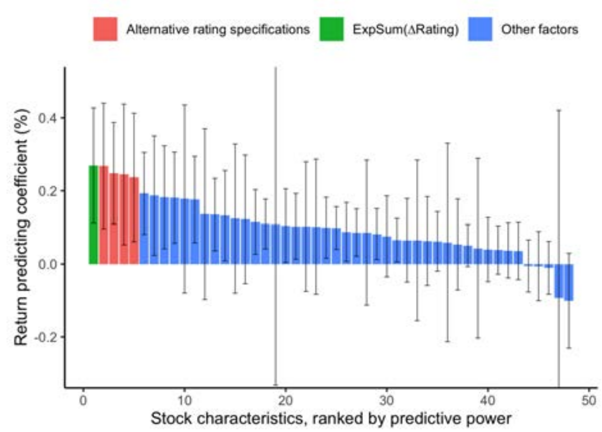
Figure A.1. Horse race between rating and other stock return predictors

We compare the stock return predictability of ExpSum(ΔRating)$_{i,t-1}$ against the 43 other return predictor characteristics listed in Table A.3. We estimate return predictability at the stock level using Fama-MacBeth regressions and plot the coefficients. To make coefficients comparable, we transform all predictors into z-scores. We switch the signs, if necessary, so that the predictors are expected to predict returns positively based on the original studies. To isolate the marginal predictive power of ratings, all regressions control for decile indicators of the past-36-month cumulative raw fund returns and benchmark-adjusted fund returns aggregated at the stock level. Panels A and B show results from regressions without controls, and panels C and D show results from multivariate regressions that also control for the six commonly used characteristics of size, book-to-market ratio, profitability, investment, momentum, and reversal. The left panels use all stocks, and the right panels use stocks that are above the 20% percentile of the NYSE stock market cap. Green bars represent the predictive coefficient of ExpSum(ΔRating)$_{i,t-1}$. Red bars represent alternative rating-based variables of lagged 3-, 6-, 9-, and 12-month rating changes. Blue bars represent the other 43 characteristics. The whiskers represent two-standard-error bands; the standard errors are calculated using the Newey-West procedure with 12 lags.

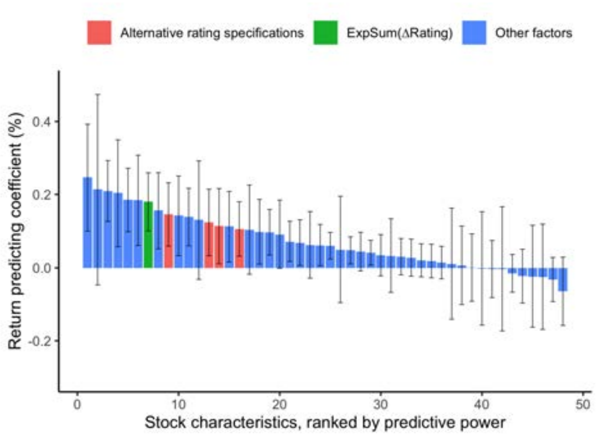
(a) Univariate, all stocks



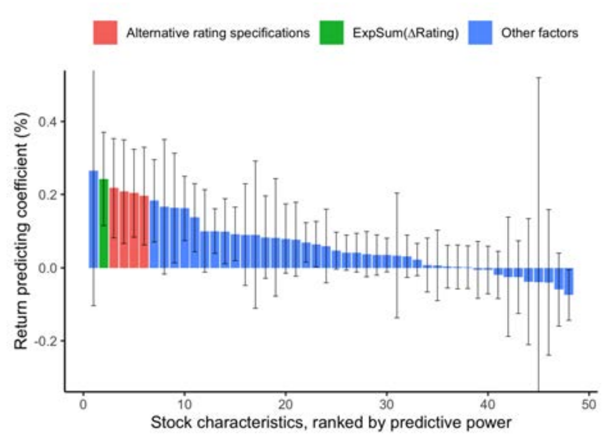
(b) Univariate, no microcap



(c) With controls, all stocks



(d) With controls, no microcap



among all major stock predictors so far discovered. Further, unlike many other predictors, it is more powerful in large-cap stocks, which is consistent with the fact that mutual funds tend to hold large-cap stocks. Even if we include all 43 additional predictors as controls, the predictive coefficient of ExpSum(ΔRating)$_{i,t-1}$ on nonmicrocaps (all stocks) is still highly significant, with values of 0.138 (0.106) and t-statistics of 3.57 (3.77). Therefore, the predictive power of rating changes is not spanned by any of these known return predictors.

A.3 Morningstar versus Academic Style Classification

In the main text, we used Morningstar funds to define 3×3 size-value stock portfolios. These definitions are related to, but different from, the academic style definitions. For instance, Lettau et al. (2019) point out that "value funds" in the industry hold few stocks with high book-to-market ratios—the value stocks defined by academia. This section explores the difference between the Morningstar and the academic style definitions.

In Figure A.2, we sort stocks by market capitalization and book-to-market ratio into 10×10 portfolios using NYSE breakpoints. The heat maps in panel A show the academic style definitions, which are strictly based on stock characteristics. By construction, the stocks in those style portfolios are concentrated in "rectangular regions." Panel B presents the distribution of stocks in Morningstar-based styles, which turn out to be "smoothed" versions of the academic styles. For instance, while the academic large-cap growth style only holds stocks with large market capitalization and low book-to-market ratios, the Morningstar-based style also can hold some, albeit fewer, stocks with other characteristics.

Figure A.2. Comparison of Morningstar and academic stock style definitions

We sort stocks into 10 × 10 portfolios based on NYSE size and book-to-market ratio break points. Panel A plots the distribution of holdings in academic style definitions. Panel B plots the distribution of holdings by funds in different styles. The heat map colors indicate the distribution of these style portfolios in each bin, with red indicating high weights and green indicating low weights.

(a) Academic style definitions

Academic style: large-cap

Academic style: growth

	Growth	2	3	4	5	6	7	8	9	Value
Large	11.1%	11.1%	11.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	11.1%	11.1%	11.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	11.1%	11.1%	11.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
4	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
5	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
6	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
7	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
8	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
9	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Small	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Academic style: blend

	Growth	2	3	4	5	6	7	8	9	Value
Large	0.0%	0.0%	0.0%	8.3%	8.3%	8.3%	8.3%	0.0%	0.0%	0.0%
2	0.0%	0.0%	0.0%	8.3%	8.3%	8.3%	8.3%	0.0%	0.0%	0.0%
3	0.0%	0.0%	0.0%	8.3%	8.3%	8.3%	8.3%	0.0%	0.0%	0.0%
4	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
5	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
6	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
7	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
8	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
9	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Small	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Academic style: value

	Growth	2	3	4	5	6	7	8	9	Value
Large	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	11.1%	11.1%	11.1%
2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	11.1%	11.1%	11.1%
3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	11.1%	11.1%	11.1%
4	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
5	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
6	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
7	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
8	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
9	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Small	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Academic style: mid-cap

	Growth	2	3	4	5	6	7	8	9	Value
Large	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
4	8.3%	8.3%	8.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
5	8.3%	8.3%	8.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
6	8.3%	8.3%	8.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
7	8.3%	8.3%	8.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
8	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
9	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Small	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

	Growth	2	3	4	5	6	7	8	9	Value
Large	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
4	0.0%	0.0%	0.0%	6.3%	6.3%	6.3%	6.3%	0.0%	0.0%	0.0%
5	0.0%	0.0%	0.0%	6.3%	6.3%	6.3%	6.3%	0.0%	0.0%	0.0%
6	0.0%	0.0%	0.0%	6.3%	6.3%	6.3%	6.3%	0.0%	0.0%	0.0%
7	0.0%	0.0%	0.0%	6.3%	6.3%	6.3%	6.3%	0.0%	0.0%	0.0%
8	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
9	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Small	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

	Growth	2	3	4	5	6	7	8	9	Value
Large	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
4	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	8.3%	8.3%	8.3%
5	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	8.3%	8.3%	8.3%
6	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	8.3%	8.3%	8.3%
7	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	8.3%	8.3%	8.3%
8	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
9	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Small	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Academic style: small-cap

	Growth	2	3	4	5	6	7	8	9	Value
Large	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
4	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
5	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
6	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
7	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
8	11.1%	11.1%	11.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
9	11.1%	11.1%	11.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Small	11.1%	11.1%	11.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

	Growth	2	3	4	5	6	7	8	9	Value
Large	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
4	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
5	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
6	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
7	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
8	0.0%	0.0%	0.0%	8.3%	8.3%	8.3%	8.3%	0.0%	0.0%	0.0%
9	0.0%	0.0%	0.0%	8.3%	8.3%	8.3%	8.3%	0.0%	0.0%	0.0%
Small	0.0%	0.0%	0.0%	8.3%	8.3%	8.3%	8.3%	0.0%	0.0%	0.0%

	Growth	2	3	4	5	6	7	8	9	Value
Large	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
4	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
5	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
6	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
7	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
8	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	11.1%	11.1%	11.1%
9	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	11.1%	11.1%	11.1%
Small	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	11.1%	11.1%	11.1%

(b) Morningstar fund-based style definitions

Fund style: large-cap

Fund style: growth

	Growth	2	3	4	5	6	7	8	9	Value
Large	17.2%	10.9%	9.1%	8.3%	5.4%	4.5%	3.8%	2.6%	2.4%	2.8%
2	4.7%	3.2%	2.2%	1.6%	1.4%	1.2%	1.0%	0.7%	0.5%	0.5%
3	2.4%	1.4%	0.9%	0.8%	0.8%	0.6%	0.4%	0.3%	0.3%	0.3%
4	1.0%	0.6%	0.4%	0.3%	0.3%	0.2%	0.2%	0.2%	0.1%	0.1%
5	0.6%	0.3%	0.2%	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
6	0.4%	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%	0.1%
7	0.2%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
8	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
9	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Small	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Fund style: blend

	Growth	2	3	4	5	6	7	8	9	Value
Large	10.6%	9.0%	8.4%	8.4%	7.7%	6.8%	6.4%	5.5%	4.5%	5.1%
2	2.1%	1.7%	1.6%	1.5%	1.5%	1.4%	1.4%	1.3%	1.2%	1.3%
3	0.8%	0.7%	0.7%	0.6%	0.7%	0.6%	0.5%	0.5%	0.5%	0.5%
4	0.4%	0.3%	0.3%	0.3%	0.3%	0.3%	0.2%	0.2%	0.2%	0.3%
5	0.2%	0.2%	0.2%	0.1%	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%
6	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
7	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%
8	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
9	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Small	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Fund style: value

	Growth	2	3	4	5	6	7	8	9	Value
Large	5.0%	5.3%	6.3%	6.8%	7.8%	7.8%	8.0%	8.3%	7.3%	7.9%
2	0.9%	0.9%	1.2%	1.5%	1.8%	1.8%	1.9%	2.2%	2.4%	2.6%
3	0.5%	0.4%	0.6%	0.5%	0.6%	0.7%	0.7%	0.8%	0.8%	1.0%
4	0.2%	0.2%	0.2%	0.2%	0.3%	0.3%	0.3%	0.4%	0.3%	0.4%
5	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.2%	0.2%	0.2%	0.2%
6	0.0%	0.0%	0.0%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
7	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%
8	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
9	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Small	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Fund style: mid-cap

	Growth	2	3	4	5	6	7	8	9	Value
Large	4.2%	1.9%	1.5%	1.4%	1.3%	1.0%	0.8%	0.7%	0.7%	0.7%
2	6.4%	3.4%	2.5%	2.2%	1.5%	1.2%	1.0%	0.8%	0.5%	0.5%
3	6.2%	3.6%	2.5%	2.4%	2.1%	1.4%	1.1%	0.8%	0.8%	0.5%
4	4.7%	3.1%	2.0%	1.5%	1.4%	0.9%	0.8%	0.7%	0.5%	0.4%
5	3.3%	2.0%	1.5%	1.0%	0.9%	0.6%	0.5%	0.4%	0.4%	0.3%
6	2.2%	1.4%	0.9%	0.8%	0.6%	0.4%	0.4%	0.3%	0.3%	0.2%
7	1.8%	0.9%	0.6%	0.4%	0.3%	0.3%	0.2%	0.2%	0.2%	0.2%
8	1.1%	0.5%	0.3%	0.2%	0.2%	0.2%	0.1%	0.1%	0.1%	0.1%
9	0.6%	0.3%	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.1%
Small	0.2%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

	Growth	2	3	4	5	6	7	8	9	Value
Large	2.3%	1.9%	2.6%	2.7%	2.7%	2.1%	2.1%	1.8%	1.6%	1.8%
2	2.5%	1.9%	1.8%	2.0%	1.7%	1.7%	1.9%	1.4%	1.3%	1.4%
3	2.9%	2.2%	2.0%	2.0%	2.1%	1.9%	1.7%	1.4%	1.5%	1.6%
4	1.9%	1.8%	1.6%	1.4%	1.5%	1.3%	1.2%	1.2%	1.1%	1.2%
5	1.2%	1.1%	1.2%	1.0%	1.1%	0.9%	0.8%	0.8%	0.7%	0.8%
6	0.9%	0.8%	0.8%	0.8%	0.7%	0.6%	0.6%	0.5%	0.6%	0.6%
7	0.7%	0.6%	0.5%	0.5%	0.4%	0.4%	0.4%	0.4%	0.4%	0.3%
8	0.6%	0.4%	0.3%	0.3%	0.2%	0.3%	0.3%	0.3%	0.2%	0.2%
9	0.4%	0.3%	0.2%	0.2%	0.2%	0.2%	0.2%	0.1%	0.1%	0.1%
Small	0.3%	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.0%

	Growth	2	3	4	5	6	7	8	9	Value
Large	0.6%	0.8%	1.3%	1.4%	1.6%	2.0%	2.0%	2.4%	2.2%	2.6%
2	1.6%	1.3%	2.0%	2.7%	2.7%	2.7%	3.1%	3.3%	3.1%	3.1%
3	1.3%	1.6%	1.8%	2.1%	2.3%	2.3%	2.2%	2.3%	2.4%	3.1%
4	0.7%	1.0%	1.3%	1.5%	1.6%	1.6%	1.5%	1.8%	1.7%	2.3%
5	0.6%	0.7%	0.8%	0.9%	1.0%	1.1%	1.2%	1.1%	1.0%	1.4%
6	0.3%	0.4%	0.4%	0.5%	0.5%	0.5%	0.5%	0.6%	0.6%	0.7%
7	0.2%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.4%	0.4%
8	0.1%	0.1%	0.1%	0.1%	0.1%	0.2%	0.2%	0.2%	0.2%	0.3%
9	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%
Small	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Fund style: small-cap

	Growth	2	3	4	5	6	7	8	9	Value
Large	0.3%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	1.1%	0.3%	0.3%	0.3%	0.2%	0.1%	0.0%	0.0%	0.0%	0.0%
3	2.5%	1.1%	0.7%	0.7%	0.5%	0.3%	0.4%	0.2%	0.2%	0.1%
4	4.0%	2.6%	1.8%	1.3%	1.2%	0.8%	0.7%	0.6%	0.4%	0.3%
5	5.3%	3.4%	2.5%	1.7%	1.4%	1.0%	0.8%	0.7%	0.5%	0.4%
6	5.2%	3.5%	2.3%	2.0%	1.4%	1.1%	0.9%	0.7%	0.7%	0.5%
7	5.2%	3.1%	2.2%	1.6%	1.1%	0.9%	0.7%	0.7%	0.6%	0.3%
8	4.5%	2.2%	1.6%	1.2%	0.8%	0.7%	0.6%	0.5%	0.4%	0.3%
9	3.0%	1.5%	1.0%	0.7%	0.5%	0.4%	0.4%	0.3%	0.3%	0.2%
Small	1.4%	0.6%	0.4%	0.2%	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%

	Growth	2	3	4	5	6	7	8	9	Value
Large	0.1%	0.1%	0.0%	0.0%	0.1%	0.1%	0.1%	0.0%	0.1%	0.1%
2	0.2%	0.1%	0.2%	0.2%	0.2%	0.2%	0.1%	0.1%	0.1%	0.1%
3	0.6%	0.4%	0.5%	0.5%	0.5%	0.4%	0.5%	0.3%	0.3%	0.3%
4	1.4%	1.3%	1.3%	1.1%	1.1%	1.0%	0.9%	0.9%	0.6%	0.6%
5	2.3%	2.2%	1.9%	1.7%	1.5%	1.4%	1.3%	1.3%	1.2%	1.0%
6	2.5%	2.5%	2.2%	2.1%	1.9%	1.8%	1.5%	1.5%	1.5%	1.0%
7	2.6%	2.3%	2.2%	2.0%	1.8%	1.5%	1.5%	1.4%	1.3%	0.8%
8	2.4%	1.9%	1.7%	1.7%	1.3%	1.3%	1.2%	1.2%	0.9%	0.7%
9	2.1%	1.5%	1.4%	1.1%	1.1%	1.0%	0.9%	0.8%	0.7%	0.5%
Small	1.4%	1.0%	0.9%	0.8%	0.6%	0.5%	0.4%	0.4%	0.3%	0.2%

	Growth	2	3	4	5	6	7	8	9	Value
Large	0.1%	0.1%	0.1%	0.2%	0.2%	0.2%	0.2%	0.2%	0.1%	0.2%
2	0.1%	0.1%	0.1%	0.1%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%
3	0.3%	0.3%	0.3%	0.3%	0.3%	0.4%	0.5%	0.4%	0.4%	0.4%
4	0.4%	0.5%	0.7%	0.7%	0.8%	0.9%	1.0%	1.0%	0.8%	1.0%
5	0.6%	0.9%	1.0%	1.1%	1.3%	1.3%	1.6%	1.7%	1.7%	1.5%
6	0.8%	1.2%	1.3%	1.6%	1.6%	1.8%	2.0%	2.2%	2.4%	1.7%
7	1.0%	1.3%	1.6%	1.6%	1.8%	1.8%	1.9%	2.2%	2.4%	1.7%
8	1.1%	1.2%	1.3%	1.7%	1.6%	1.7%	1.8%	2.1%	1.7%	1.3%
9	1.2%	1.3%	1.4%	1.2%	1.5%	1.6%	1.6%	1.4%	1.2%	1.1%
Small	1.0%	1.1%	1.3%	1.2%	1.1%	1.0%	1.0%	0.9%	0.6%	0.5%

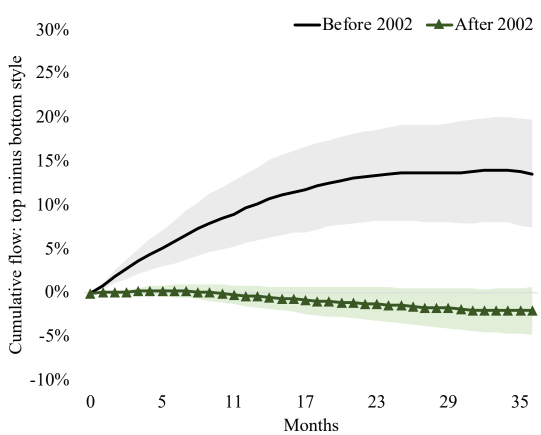
A.4 Robustness Checks

Figures A.3 and A.4 provide robustness checks for the results presented in panels C and D of Figure 1, respectively. Instead of sorting the 3×3 style portfolios by ExpSum(ΔRating), we sort them by the past 3-, 6-, 9-, and 12-month rating changes, and then plot the subsequent flow and return spreads between the top and bottom style portfolios. The results are qualitatively similar to the results in the main paper.

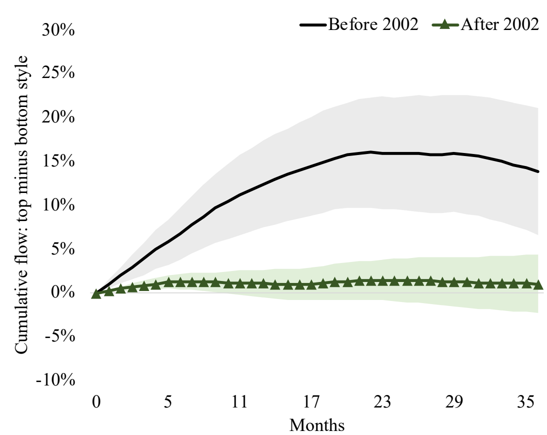
Figure A.3. Style fund flows with alternative rating specifications

This figure provides robustness checks to panel C of Figure 1. We sort the 3×3 style portfolios by their lagged 3-, 6-, 9-, or 12-month rating changes and then plot the cumulative difference in fund flows between the top and bottom styles for the subsequent 3 years. The shaded areas represent 95% bootstrapped confidence intervals.

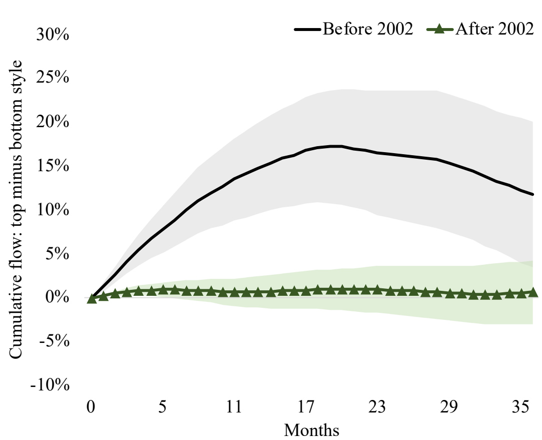
(a) 3-month rating change



(b) 6-month rating change



(c) 9-month rating change



(d) 12-month rating change

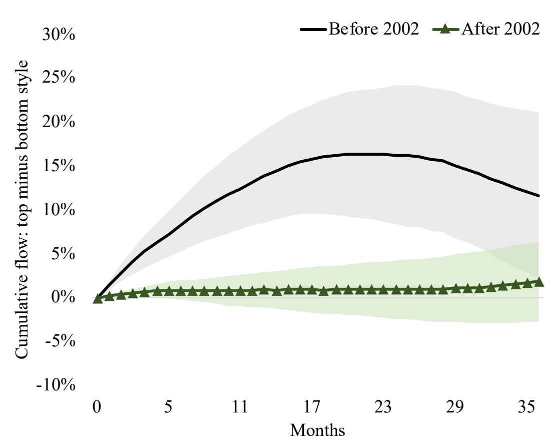
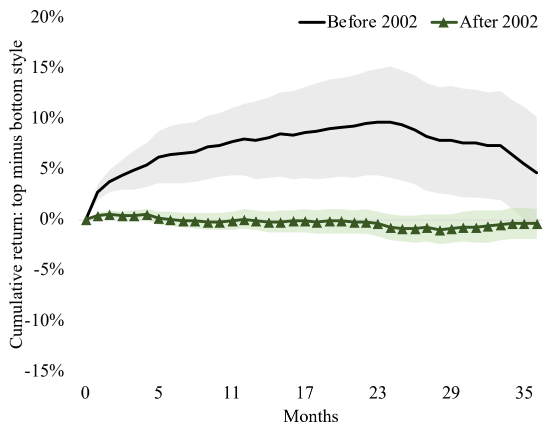


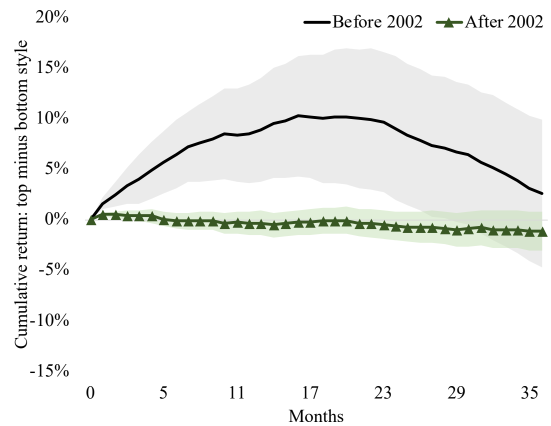
Figure A.4. Style fund return with alternative rating specifications

This figure provides robustness checks to panel D of Figure 1. We sort the 3×3 style portfolios by their lagged 3-, 6-, 9-, or 12-month rating changes and then plot the cumulative difference in returns between the top and bottom styles for the subsequent 3 years. The shaded areas represent 95% bootstrapped confidence intervals.

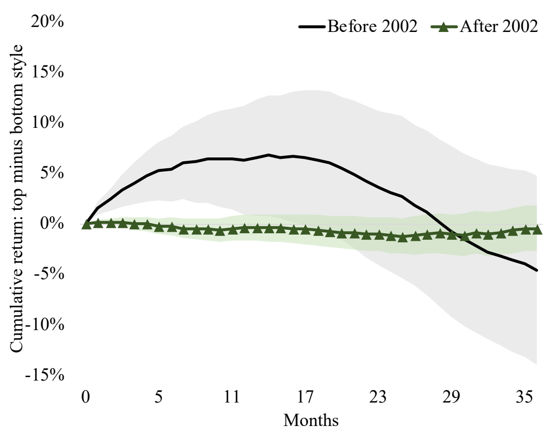
(a) 3-month rating change



(b) 6-month rating change



(c) 9-month rating change



(d) 12-month rating change

In panel F of Figure 7, we show that styles that are predicted to suffer rating decreases experienced declines in returns after June 2002 and that the same did not happen in other years. We now present a more formal test of that result. Specifically, we estimate panel regressions of monthly style returns:

$$\mathrm{Ret}_{\pi,t} = \alpha + \beta\, \mathrm{Post} \times \mathrm{StyleRank}_{\pi} + \kappa\, \mathrm{Post} + \mu\, \mathrm{StyleRank}_{\pi} + \gamma\, X_{\pi,t} + \epsilon_{\pi,t},$$

where the controls $X_{\pi,t}$ include month and style fixed effects, and we cluster standard errors by month. Post is an indicator that equals one after June, and StyleRank_π is the ranking (1 to 9) of the style based on data in December of the previous year.[42] In the event year of 2002, styles ranked high are expected to experience large declines in ratings, as shown in panel B in Figure 7.

Table A.4. Style return change around 2002 and other years

This table provides a panel regression version of panel F of Figure 7. In each year T, we rank styles into 1 to 9 by $\widehat{\text{Rating}}_\pi^{\text{pre 2002 methodology}} - \widehat{\text{Rating}}_\pi^{\text{post 2002 methodology}}$, which is computed using data in December of year $T-1$. We then estimate panel regressions of monthly style returns on the interaction of an indicator that equals one after the month of June ("Post") and another variable related to style ranking. The latter variable is the numerical style rank of 1 to 9 in columns 1 and 4, an indicator of whether a style ranks in the top half (6^{th} to 9^{th}) in columns 2 and 5, and an indicator of whether the style is the top-ranked style in columns 3 and 6. All regressions control for month and style fixed effects, and standard errors are clustered by month. Columns 1 to 3 are estimated on the event year of 2002, and columns 4 to 6 are estimated on the other years as placebo tests. Column 7 shows the difference between estimates in 2002 versus the other years. $^*p < .1$; $^{**}p < .05$; $^{***}p < .01$.

	Dependent variable: Style return $\text{Ret}_{\pi,t}$ (%)						
	2002			Other years			Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Post × Style rank	-0.544^{**}			0.000			-0.543^{**}
	(0.266)			(0.042)			(0.269)
Post × Top-half styles		-2.336^{**}			-0.047		-2.289^{**}
		(1.046)			(0.194)		(1.064)
Post × Top style			-3.542^{***}			0.226	-3.768^{***}
			(1.275)			(0.253)	(1.300)
Month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	
Style fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	
No. observations	108	108	108	2,916	2,916	2,916	
Adjusted R^2	.915	.909	.909	.868	.868	.868	

Columns 1 and 4 in Table A.4 report the regression results for the event year of 2002 and the other years, respectively. While higher style ranks are associated with lower post-June returns in 2002, as predicted, the same is not true in the other placebo years. As shown in column 7, which reports the difference, these two estimates are statistically significantly

[42] Specifically, for each year T, we use data in December of year $T-1$ to compute $\widehat{\text{Rating}}_\pi^{\text{pre 2002 methodology}} - \widehat{\text{Rating}}_\pi^{\text{post 2002 methodology}}$ and rank styles based on that.

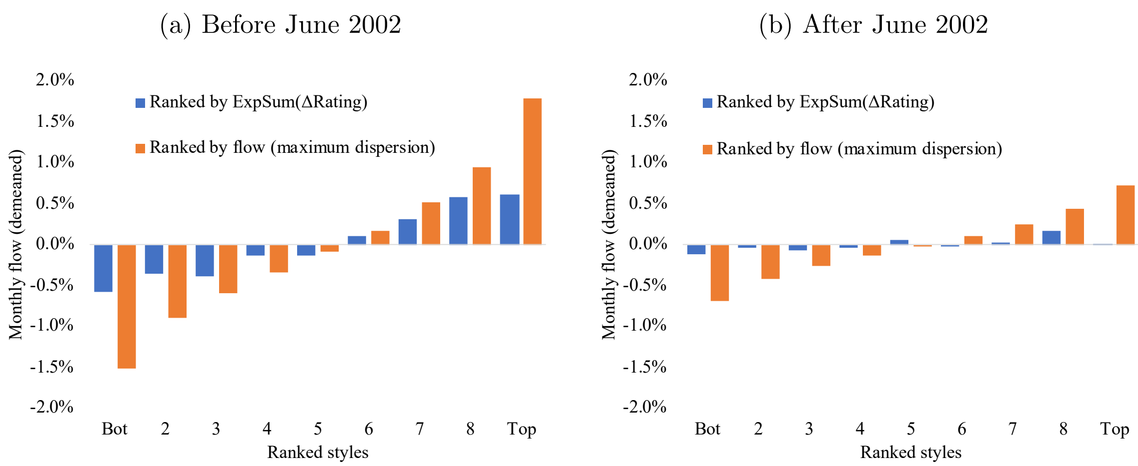
different. In columns 2 and 5, we replace the numerical style rank with an indicator that equals one for styles that are ranked in the top half (6^{th} to 9^{th}). In columns 3 and 6, we use an indicator that equals one for only the top-ranked style. All these results yield the same conclusion: the post-June return changes are unique to 2002.

A.5 Quantifying Rating-Induced Style-Level Fund Flows

To get a sense of the magnitude of style-level flows spanned by ratings, in Figure A.5, we sort the 3×3 styles by ExpSum(ΔRating) and plot their average monthly flows in blue bars. To compare with the *overall* style-level flow dispersion, we also sort by *realized* flows in orange. Because style-level rating variation is small after the Morningstar reform, we split the results by before and after June 2002 and plot them in two separate panels. Flows are demeaned by month to focus on the cross-sectional dispersion.

Figure A.5. Style-level fund flows by lagged rating changes

We sort the 3×3 Morningstar styles by the exponentially weighted sum of lagged ratings ($\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$) and plot the average monthly fund flows in blue. To visualize the overall flow dispersion, we also sort by *realized* flow and plot the results in orange. Panel A plots results for before the Morningstar methodology change in June 2002, and panel B plots the results for after June 2002. Flows are demeaned by month to focus on the cross-sectional dispersion.



Before June 2002, styles with higher $\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$ have higher flows, and the top-minus-bottom monthly flow spread is 1.19%. This is slightly higher than one-third of the maximum flow spread of 3.31%. As expected, because style-level rating variation becomes muted after the Morningstar reform, ratings no longer explain flow variation after June 2002.

A.6 Empirical Results Using Academic Styles

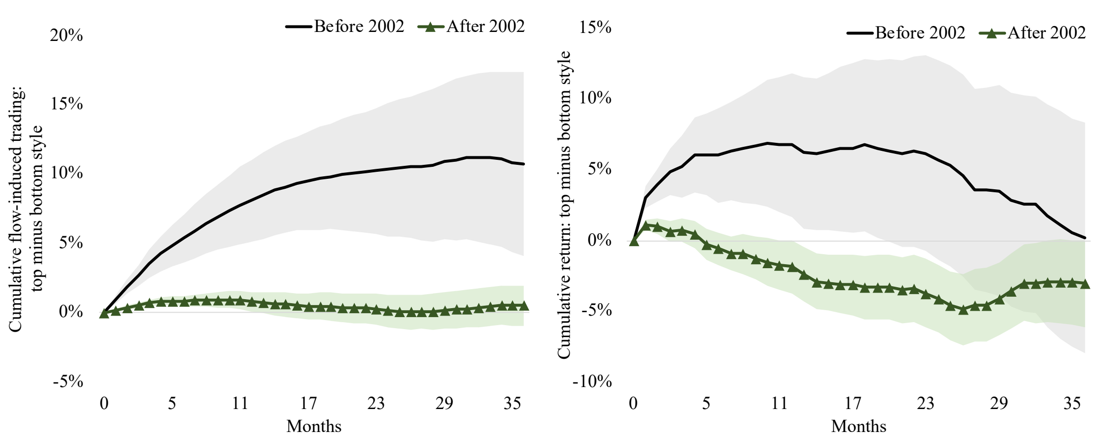
This section shows that the key results based on Morningstar-defined styles also extend to the academic-defined styles using size and book-to-market ratio characteristics.

Figure A.6 reproduces panels C and D in Figure 1 using academic-defined styles. The general patterns are similar.[43]

Figure A.6. Price pressure in academic style portfolios

This figure is similar to panels C and D in Figure 1 but is performed using style portfolios defined using stock characteristics. Stocks are sorted into 3×3 size-value styles using NYSE breakpoints of market capitalization and book-to-market ratios. In each month, we rank styles by their lagged ExpSum(ΔRating) and plot the subsequent cumulative flow-induced trading (panel A) and returns (panel B). We create separate estimates for the sample period before June 2002 and after June 2002. The shaded areas represent 95% bootstrapped confidence intervals.

(a) Flow-induced trading (b) Return



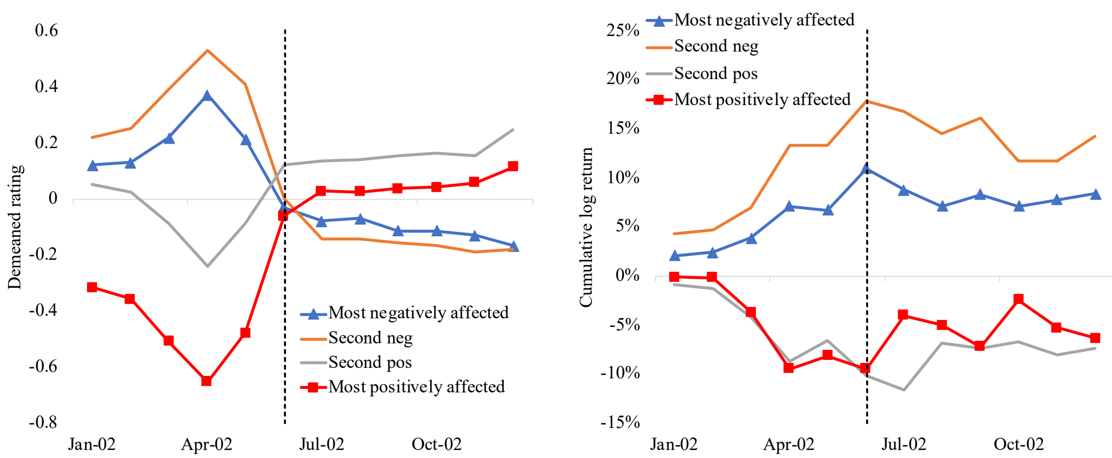
We also conduct an event study based on the academic-defined styles as in Section 5. Figure A.7 illustrates the ratings and returns of the academic styles within this one-year window when sorted on the predicted rating changes. The patterns are similar to those depicted in Figure 7, where style portfolios are instead based on Morningstar-defined styles.

[43] Results based on Morningstar-defined styles are slightly sharper, consistent with the fact that ratings—the source of change around 2002—are computed using Morningstar style definitions.

Figure A.7. Behavior of academic styles around the June 2002 event

We perform event studies on the 3 × 3 size-value academic style portfolios during the 6 months before and after the June 2002 methodology change. The styles are sorted by their predicted rating change at the June 2002 event using December 2001 data. These style portfolios use the standard academic definition by sorting on size and value stock characteristics (Fama and French, 1993). All variables are demeaned cross-sectionally to focus on cross-sectional dispersion.

(a) Ratings (b) Cumulative returns



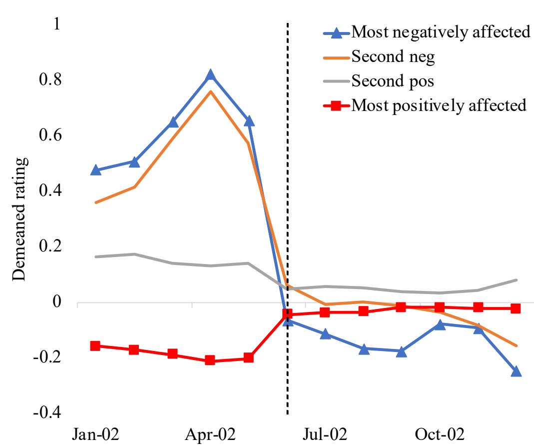
A.7 2002 Event Study at the Stock Level

As mentioned in Section 5.1, we also perform an event study in 2002 at the stock level. Specifically, we follow the same methodology as used in Section 5.1 to estimate the *predicted* stock-level methodology-induced rating change using December 2001 data. We then sort stocks into quintiles using the predicted values. Figure A.8 plots the evolution of ratings and cumulative returns of these stocks. Panel A shows that the prediction is useful: those predicted to be positively (negatively) affected indeed experience large upward (downward) ratings revisions in June 2002. Panel B shows that the behavior of returns is consistent with ratings having a price impact. Stocks that are predicted to be positively (negatively) affected have low (high) ratings and returns before the event, and then reverse right after it.

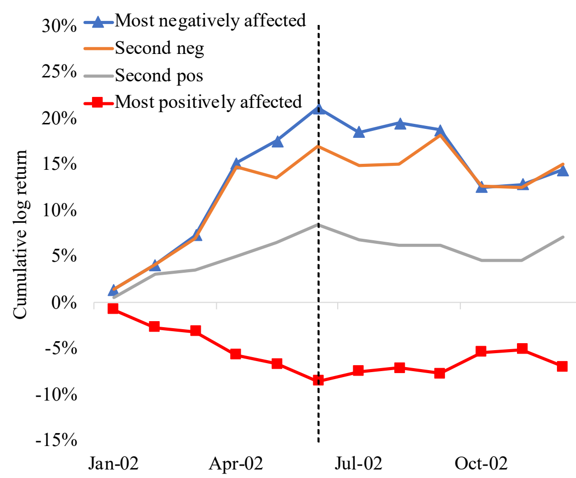
Figure A.8. Event study around June 2002: Stock-level exercise

This figure is similar to Figure 7, but here we examine individual stocks. We perform event studies on all stocks held by mutual funds during the 6 months before and 6 months after the June 2002 methodology change. We sort stocks into quintiles by their *predicted* rating change at the June 2002 event using December 2001 data, and then plot the evolution of their ratings in panel A and cumulative returns in panel B. The dashed vertical line represents the June 2002 event. Both ratings and flows are demeaned cross-sectionally to focus on cross-sectional dispersion.

(a) Ratings



(b) Cumulative returns



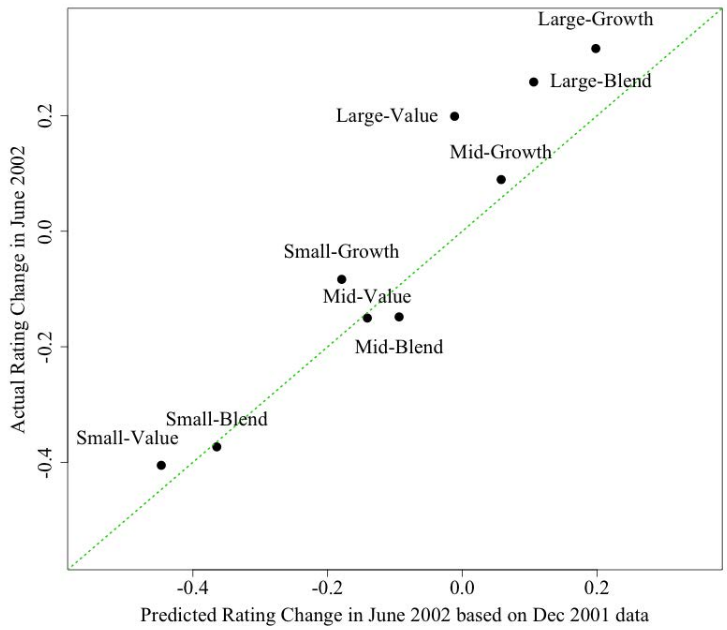
A.8 Predicting Rating Changes

We follow Morningstar's methodology in Appendix B to estimate fund ratings using fund returns and style categories data. In other words, we redo all the calculations done by Morningstar using the data we have.

Because we do not have access to the exact data set used historically by Morningstar, we cannot exactly reproduce all fund ratings. However, the computations are good enough for our purposes because when aggregated at the style-level, we can predict rating changes fairly accurately. Figure A.9 plots the actual style-level rating change in June 2002 against our predictions (computed following the method described in Section 5.1). As shown in the figure, we correctly predict that small-value is the most negatively affected style and small-blend is the second-most negatively affected, while large-growth is the most positively affected and large-blend is the second-most positively affected. The predictions are also reasonably accurate in magnitudes.

Figure A.9. Predicting June 2002 style rating change using end-of-2001 data

We follow Morningstar's rating methodology to calculate what fund ratings would have changed to in December 2001 under the new methodology. The fund ratings are aggregated at the style level. We use the difference between this counterfactual rating and the actual rating as our prediction, and then plot the actual style-level rating changes against the predicted changes. The style portfolios are labeled, and the dashed diagonal line indicates a perfect match.



Appendix B Morningstar Methodology

In this section, we explain the construction of Morningstar ratings and the June 2002 methodology change in detail.

Morningstar ratings are updated every month. Morningstar's rating calculation comprises two steps:

1. For each fund with sufficient data, calculate performance measures using past returns, with some adjustments based on return volatility and fund loads.

2. Rank funds by the performance measure and assign ratings.

In June 2002, Morningstar changed both steps of the methodology. The steps are consecutive, though independent. Our analysis shows that the change to the second step (described in Section B.2) made the biggest difference to the issues of interest in the study.

B.1 Step One: Calculate Performance Measures

The pre-2002 methodology is described in detail in Blume (1998), and we summarize it here. First, the cumulative return is computed over three horizons (36, 60, 120 months):

$$R_i^T = \prod_{t=1}^{T}(1 + r_{i,t}) - 1, \qquad T \in \{36, 60, 120\}, \tag{B.1}$$

where the monthly fund returns $r_{i,t}$ are net of management fees but unadjusted for loads. Then, the cumulative return is adjusted for loads to get a load-adjusted return over the risk-free rate:

$$\text{LoadRet}_i^T = R_i^T L_i - R_f^T, \tag{B.2}$$

where the load adjustment L_i equals 1 minus the sum of the front- and back-end loads. R_f^T is defined as the cumulative risk-free rate return for horizon T using 3-month T-bills. The measure is standardized to

$$\text{MnLoadRet}_i^T = \frac{\text{LoadRet}_i^T}{\max(R_f, \text{AvgLoadRet}^T)}, \tag{B.3}$$

where AvgLoadRet^T is the average of LoadRate_i^T across all funds in the same investment class (equity, corporate bonds, etc.).

Second, Morningstar derives the final performance measure by subtracting a risk-adjustment term:

$$\text{Performance}_{i,t} = \text{MnLoadRet}_{i,t}^T - \text{MnRisk}_{i,t}^T. \tag{B.4}$$

The risk-adjustment term is defined as a normalized average downward return deviation. Concretely, Morningstar calculates

$$\text{Risk}_i^T = \frac{\sum_{t=1}^T - \min(r_{i,t} - r_t^f, 0)}{T}. \tag{B.5}$$

Then, the measure is normalized by the relevant average risk:

$$\text{MnRisk}_t^T = \frac{\text{Risk}_i^T}{\text{AvgRisk}^T}. \tag{B.6}$$

After June 2002, Morningstar changed the way it adjusts for risk.[44] Morningstar summarizes a fund's past performance using the so-called Morningstar risk-adjusted return (MRAR):

$$\text{MRAR}_i^T(\gamma) = \left[\frac{1}{T}\sum_{t=1}^{T}(1 + r_{i,t} - r_t^f)^{-\gamma}\right]^{-\frac{12}{\gamma}} - 1, \tag{B.7}$$

[44] Morningstar explains its post-June 2002 rating methodology in a publicly available manual, available at `https://corporate.morningstar.com/US/documents/MethodologyDocuments/FactSheets/MorningstarRatingForFunds_FactSheet.pdf`. See also Blume (1998).

where $r_{i,t} - r_t^f$ is the geometric return in excess of the risk-free rate after adjusting for loads,[45] and $\gamma = 2$ is the risk aversion coefficient.

The formula penalizes funds with higher return volatility. To see this, notice that when γ converges to 0, $\text{MRAR}^{\text{T}}(0)$ is equal to the annualized geometric mean of excess returns.[46] When γ is set to be greater than 0, holding the geometric mean return constant, the formula yields a lower MRAR value for funds whose monthly returns deviate more from their mean. Specifically, the risk adjustment can be expressed as $\text{MRAR}^{\text{T}}(0) - \text{MRAR}^{\text{T}}(2)$.

B.2 Step Two: Rank Funds and Assign Ratings

Given rankings of funds, Morningstar calculates 3-year, 5-year, and 10-year ratings for funds with the necessary number of historical returns at those horizons, and then takes a weighted average of them (rounded to the nearest integer) to form an overall rating—the rating most commonly reported and used. For funds with more than 3 years but less than 5 years of data, the overall rating is just the 3-year rating. For funds with more than 5 years but less than 10 years of data, the overall rating assigns 60% and 40% weights on the 5-year and 3-year ratings.[47] For those with 10 years of data, 50%, 30%, and 20% weights are assigned on the 10-year, 5-year, and 3-year ratings, respectively.

The ratings are based on rankings of funds. Before June 2002, Morningstar ranked the past performance of all equity funds together and assigned them ratings with fixed proportions: 10%, 22.5%, 35%, 22.5%, and 10%. Since June 2002, Morningstar has ranked funds within each style ("Morningstar category") and assigned ratings based on the within-

[45] For funds with loads, Morningstar uses the load-adjusted return r_t, defined as $r_t = a \cdot (1 + r_t^{\text{raw}}) - 1$. The adjustment factor a is defined as $a = \left(\frac{V_{\text{adj}}}{V_{\text{unadj}}}\right)^{1/T}$, where V_{adj} (V_{unadj}) is the load-adjusted (unadjusted) cumulative fund return over the past T months. For details, see "The Morningstar Rating Methodology," June 2006.

[46] Morningstar motivates the MRAR formula using expected utility theory. Specifically, consider an investor with a power utility and relative risk aversion of $\gamma + 1$. A standard feature of the power utility is that when risk aversion decreases to 1 ($\gamma = 0$), it becomes log utility. Therefore, MRAR(0) simply calculates the geometric mean return.

[47] Because the 5-year history contains the 3-year history, the 3 most recent years are effectively given more weight than more distant history.

style ranking. Styles include the standard 3×3 size-value categories in the Morningstar style box and also a number of specialized sector categories (e.g., financial, technology).