NBER WORKING PAPER SERIES

NON-FUNDAMENTAL DEMAND AND STYLE RETURNS

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

We thank Sylvester Flood (Morningstar), Paul Kaplan (Morningstar), Nick Barberis, John Campbell, Alex Chinco, Darrell Duffie, Bing Han, 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, University of Washington, Arrowstreet Capital, Hong Kong University of Science and Technology, as well as participants in the Pacific Northwest Finance Conference, the NBER Summer Institute (Asset Pricing), and the NBER Behavior Finance Workshop for comments. Ben-David is with The Ohio State University and the NBER; Li is with University of Utah; Rossi is with the University of Arizona; and Song is with the University of Washington. Ben-David is a co-founder and a partner in an investment advisor. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Non-Fundamental Demand and Style Returns
Itzhak Ben-David, Jiacui Li, Andrea Rossi, and Yang Song
NBER Working Paper No. 28103
November 2020
JEL No. G11,G24,G41

ABSTRACT

We present causal evidence that non-fundamental correlated demand exerts a first-order impact on style returns. Mutual fund investors chase fund performance via Morningstar ratings, regardless of the rating methodology. Until June 2002, ratings depended on fund returns without any style adjustment, and thus mutual funds with the same investment style had highly correlated ratings. This methodology led rating chasing investors to direct capital into winning styles, exacerbating return chasing behavior. Capital flows exerted non-fundamental price pressure on the underlying stocks, creating style-momentum that reverted over time. In June 2002, Morningstar reformed its rating methodology so that ratings became equalized across styles. The reform demonstrates the causal impact of rating chasing: once the reform was implemented, style-level price pressures via the mutual fund channel immediately became muted. Furthermore, the dispersion in style performance declined sharply, and style momentum and reversal disappeared. We estimate that Morningstar rating chasing explains a substantial part of the size and value factors' time-series variation.

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
Seattle, WA 98195
songy18@uw.edu

1 Introduction

Economists agree that asset prices reflect investor demand. What is subject to debate is the extent to which security prices are elastic and the nature of investor demand. On one extreme, elastic prices and rational expectations about future cash flows yield an efficient market. On the other extreme, inelastic prices and non-fundamental demand result in mispriced assets. Furthermore, if the non-fundamental demand is systematic in nature, e.g., correlated at the style level (such as value or growth), then systematic mispricings can occur (Barberis and Shleifer, 2003). In the words of the economist Paul Samuelson, markets can become *"macro-inefficient"* (Samuelson, 1998).

Existing evidence suggests that capital flows impact security prices. Some studies show localized price pressure due to institutional reasons such as portfolio rebalancing.[1] Other studies find that flows to investment vehicles create price pressure.[2] Yet, there is limited understanding of the drivers of these flows and thus it is difficult to differentiate non-fundamental demand from preference- or information-driven trading.

In this study, we present *causal* evidence that non-fundamental correlated demand—driven by mutual fund investors' tendency to chase Morningstar ratings—creates systematic style-level price patterns. We identify the link between capital flows and style returns by exploiting a reform in Morningstar's rating methodology that took place in June 2002. By balancing the distribution of ratings across mutual fund investment styles, the reform altered the allocation of immense capital flows across styles. We show that this shock to style-level demand reshaped systematic properties of style portfolios in the stock market.

A priori, the mutual fund sector is sufficiently large to plausibly generate systematic non-fundamental price patterns. In recent decades, mutual funds have been the primary conduit for equity investment by U.S. households. In 1991, when our sample begins, equity mutual funds held about 9% of the U.S. stock market capitalization. Their ownership had

[1] See, for example, rebalancing due to index inclusion/deletion (Shleifer, 1986; Harris and Gurel, 1986; Wurgler and Zhuravskaya, 2002; Chang, Hong, and Liskovich, 2015).

[2] For instance, mutual funds (Teo and Woo, 2004; Coval and Stafford, 2007; Lou, 2012; Huang, Song, and Xiang, 2020), institutional investors (Koijen and Yogo, 2019; Gabaix and Koijen, 2020b; Ben-David, Franzoni, Moussawi, and Sedunov, 2020), and exchange-traded funds (Ben-David, Franzoni, and Moussawi, 2018; Brown, Davies, and Ringgenberg, 2020) trade and generate price pressure as they balance their portfolios and respond to fund flows.

tripled to nearly 30% by 2005 and has remained steady since then. Furthermore, because mutual funds are mostly owned by retail investors, fund flows are likely unsophisticated and not motivated by fundamental value.[3] In turn, mutual fund managers trade in response to capital flows and thus channel the investors' demand, whether informed or not, into the stock market. Indeed, prior studies have documented that flow-induced trades create stock-level price pressures.[4]

The Morningstar rating reform provides an ideal setting to evaluate the effect of non-fundamental demand. While mutual fund investors have relied heavily on Morningstar ratings to guide their flows throughout the sample period,[5] ratings have a different economic meaning before and after June 2002 due to a methodology reform. Hence the reform dramatically rerouted flows across mutual funds and, therefore, to the underlying securities.

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 its 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, chasing ratings was broadly equivalent to chasing funds' past returns.

In June 2002, Morningstar revised its rating methodology. Instead of simply 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 × small/mid/large cap). By design, the revised methodology removes the style-performance component from the fund ranking, and

[3]For example, mutual fund investors allocate money to funds that subsequently underperform (Frazzini and Lamont, 2008; Song, 2020), invest in high-fee funds (Barber, Odean, and Zheng, 2005; Choi and Robertson, 2020), time the market poorly (Akbas, Armstrong, Sorescu, and Subrahmanyam, 2015; Friesen and Nguyen, 2018), and rely on salient and simple signals (Kaniel and Parham, 2017; Hartzmark and Sussman, 2019).

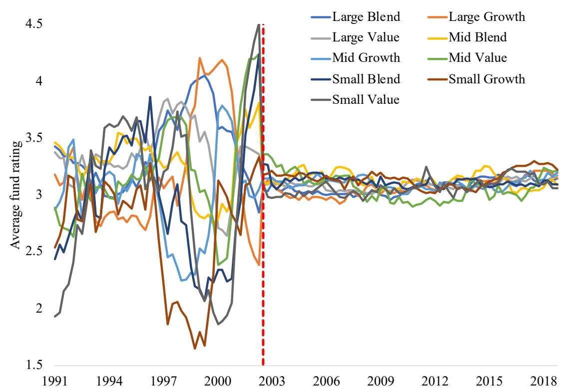
[4]For example, fire sales by mutual funds depress stock prices (Coval and Stafford, 2007), and flow-induced trades can partially explain momentum returns (Lou, 2012).

[5]In fact, Ben-David, Li, Rossi, and Song (2019) find that Morningstar ratings are the most important driver of fund flows among all factors hitherto studied.

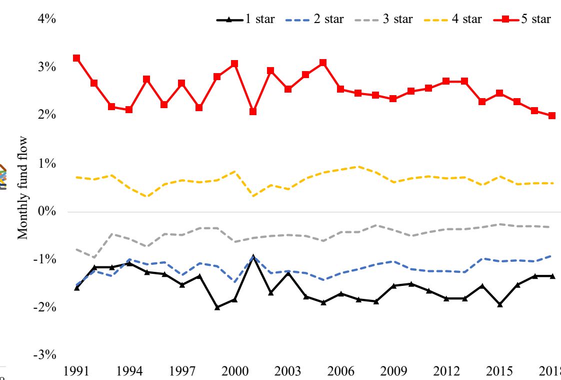
Figure 1. Morningstar Rating Methodology Change and Style Price Pressures

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 marks 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 4.2). Then we plot the cumulative differences in flows and returns between the top and bottom styles for the subsequent three years. The shaded areas are 95% bootstrapped confidence intervals.

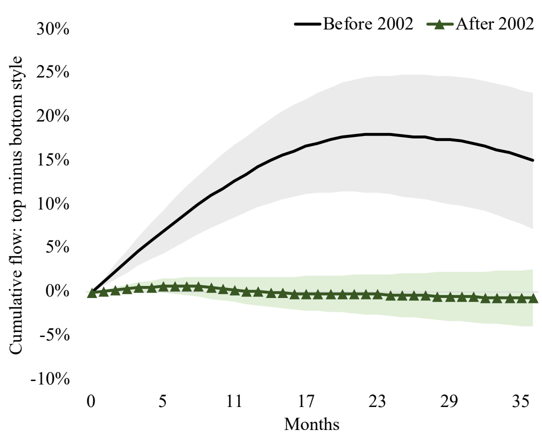
(a) Morningstar fund rating by style



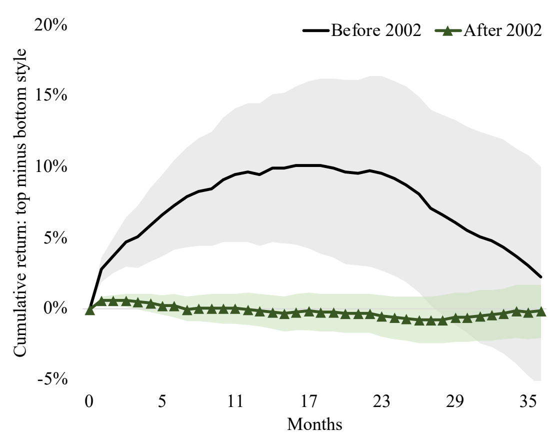
(b) Fund flow by Morningstar rating



(c) Style fund flows



(d) Style returns



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 trace the key mechanisms: investor flows chase ratings and flow-induced trades create price pressures. First, we document that investors chase ratings *regardless* of the rating methodology.[6] 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 in our sample.[7] Second, we use impulse responses to show that a rating upgrade results in a surge in mutual fund flows, and that flow increases lead to contemporaneous stock price appreciation and subsequent reversals.

In the second part, we turn to examine the impact of the reform on style performance. Before the reform, ratings were concentrated at the style level and thus led to correlated style-level flows. Hence, flows hit a small subset of winning styles and price pressures formed at the style level. Indeed, before June 2002, the most upgraded styles (i.e., styles with highest recent changes in ratings) drew large fund inflows, and their returns exhibited momentum and subsequent reversals. Opposite patterns were observed for 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. Since ratings-chasing investors stopped applying price pressure on a small subset of winning styles, style-level momentum and reversal largely disappeared.

These results are illustrated in Panels (c) and (d) of Figure 1. Panel (c) shows that before 2002, the most upgraded style attracted approximately 15% more flows 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. We show later that these price pressure effects are stronger in stocks with higher mutual fund ownership, consistent with our fund flow-based channel.

Morningstar's rating reform also led to a decrease in the profitability of style momentum

[6] Ben-David et al. (2019) and Evans and Sun (2020) 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).

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

strategies after June 2002. Because Morningstar fund ratings depend on past returns, the pre-2002 rating methodology exacerbated the positive feedback at the style level. For instance, high past performance of a style led to rating upgrades, which attracted inflows to funds in that style and led to further price appreciation due to buying pressure. This style-level positive feedback loop was severed in June 2002, and style momentum became weaker after that date: long/short style momentum strategies generated a considerable return of 70 to 90 basis points per month before June 2002 and became entirely unprofitable afterwards.

In addition, the 2002 reform altered the dispersion of style flows and style returns. As ratings became more homogeneous across styles due to Morningstar's reform, so did flows and returns. The average monthly style-level flow spread (top minus bottom) 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, we provide sharper identification for the effects of the reform on the style-level return properties by zooming into a 12-month window around June 2002. Using a shorter window reduces the chance that our results arise from forces other than the reform. In the months leading to June 2002, funds in top-rated styles gathered flows and the underlying stocks performed well. Accordingly, 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 before the reform also immediately halted and even slightly reversed.[8]

We carry out a battery of tests to tackle possible alternative explanations for the event study results. First, using all other years as placebo tests, we confirm that the observed sharp changes in style flows and returns are unique to 2002.[9] Second, we show that other fundamental factors, such as measures of stock fundamentals and trading by non-mutual fund institutions, did not change materially around June 2002. Thus, it is unlikely that our findings are driven by unobserved shocks fortuitously happening at the same time. Third, to

[8]This exercise is not subject to concerns about mean-reversion in returns because we sort styles using *predicted* reform-induced rating changes computed using data *before* the start of the event window.

[9]This test also alleviates the concern that the 2002 result is simply driven by mean-reversion.

further ease the concern about characteristics-related returns driven by unspecified reasons, we show that rating changes orthogonal to size and book/market characteristics also had a similar impact on characteristics-controlled stock returns in the event window. Finally, our estimate of the price impact coefficient is similar to existing estimates in the literature. Taken together, these results are all supportive of our interpretation that the rating reform causally impacted style-level flows and returns.

In the final part of our analysis, we study the explanatory power of ratings-induced demand on fluctuations in the common risk factors related to size and value (Fama and French, 1993). These factors are defined as long-short portfolios along the two style dimensions, size and book/market, and are considered by many financial economists as reflecting fundamental risk. To quantify the impact of rating-induced demand, we first use the Morningstar reform as an instrument to estimate the price pressure coefficient of style ratings on future style returns over a short window around June 2002. Focusing on this window allows us to cleanly isolate variation in ratings that is caused by the methodology change, therefore mitigating endogeneity concerns. We then apply the estimated parameter to quantify the effect of style ratings on style returns for the entire sample period of 1991–2018. Admittedly, this is a crude estimate, but it is informative about the size of the potential impact of rating-induced demand on factors. Our analysis shows that rating-induced price pressure can explain on average 10% to 30% of the variation in monthly factor returns before 2002. As expected, the explanatory power dropped precipitously after June 2002.[10]

To summarize, we find that a sizeable fraction of the common variation in stock returns can be empirically linked to non-fundamental correlated demand. 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. In turn, the reallocation of capital flows altered the cross-sectional variation of style returns, style-level momentum, and widely-used return factors. While style-level return chasing may have taken place on a smaller scale before our sample period, the introduction of fund ratings—which coincided with the exponential growth of the retail mutual fund sector—made it easier for investors to chase

[10]Because ratings are persistent, in untabulated results, we find that the explanatory power of rating-induced price pressure for the size and value factor returns before 2002 rises to an average of 40% at the quarterly frequency.

style-level returns and thus potentially amplified flow-induced feedback trading at the style level. The fact that style return dynamics changed dramatically after the reform highlights the importance of non-fundamental demand in shaping systematic returns.

The rest of the paper is organized as follows. Section 2 introduces the data set. Section 3 describes the Morningstar rating methodology change in June 2002. Section 4 shows that investors chase Morningstar ratings to a similar extent after June 2002 and rating-induced flows to funds indeed exert a large price impact on the underlying stocks. Section 5 demonstrates that style return dynamics changed dramatically since 2002. Section 6 studies ratings, flows, and returns around June 2002 through an event study approach. Section 7 quantifies the influence of correlated demand on the size and value factors. Section 8 concludes. Additional results and robustness checks are provided in the appendices.

2 Data and Variable Construction

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

2.1 Mutual Fund Sample

Mutual funds are one of the largest classes of equity investors in the U.S. 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).

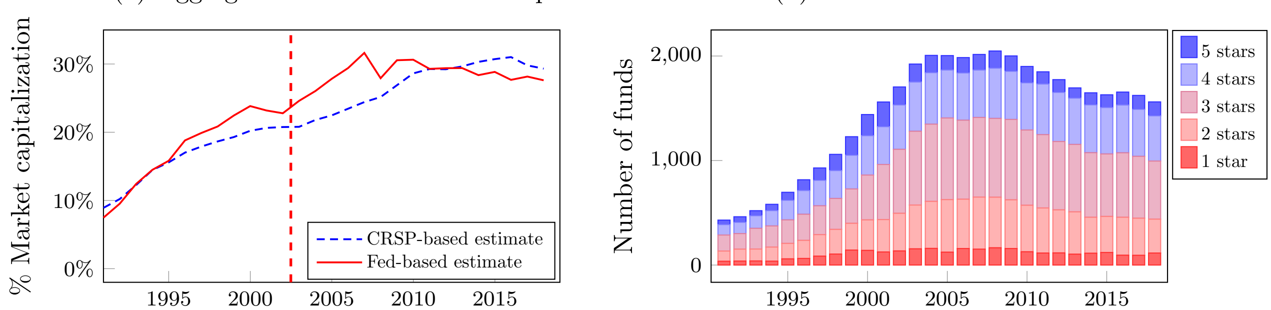
We obtain monthly fund return and total net assets (TNA) from the CRSP survivorship-bias-free mutual fund data set. We use all U.S. domestic equity mutual funds. While funds are often marketed to different clients through different share classes, they invest in the same portfolio and typically only differ in the fee structure. Therefore, we aggregate all share classes at the fund level using Russ Wermer's 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 the Center for Research in Security Prices (CRSP) starts in 1990, and some measures require one year of lagged data to construct.

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. Appendix Table A.1 further summarizes the mutual fund sample used in this paper.

(a) Aggregate mutual fund ownership (b) Number of mutual funds



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:

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

We obtain Morningstar ratings and style categories from Morningstar Direct and merge them with the CRSP mutual fund data using the matching table from Pastor, Stambaugh, and Taylor (2020).[11] 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 resulting sample comprises a total of 3,305 funds with 454,787 fund-month observations.

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

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.

2.2 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 and changes in ratings both at stock and style levels.

We define the level of and change in 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:[12]

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

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

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

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

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

where $w_{i,t-1}^{\pi}$ 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.

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. Appendix Table A.2 presents summary statistics of ratings, flows, and returns for styles.

[12]We use the latest holdings available in the past 3 months.

3 Morningstar Ratings: Background and 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 three, five, or 10 years, and more weight is applied to the most recent three 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.[13] 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, redesigned the rating system.[14]

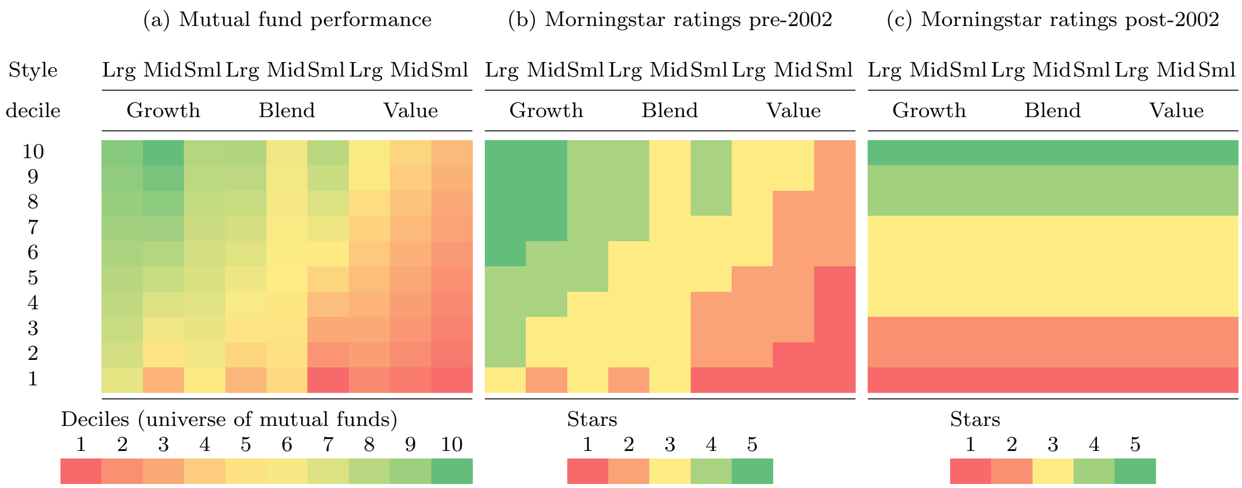
The main modification in the post-June 2002 methodology was that funds were ranked

[13] 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 would have more overlapping holdings.

[14] 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." See Floyd Norris, Morningstar to Grade on a Curve, *New York Times*, April 23, 2002.

Figure 3. Illustration of 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 three-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.



within style categories,[15] 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[16] and was implemented at the end of June 2002. Appendix B provides additional technical details on the rating methodology.

Figure 3 illustrates the relation between fund performance and Morningstar 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

[15]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 constituted 87% of all equity mutual funds.

[16]See `http://news.morningstar.com/pdfs/FactSheet_StyleBox_Final.pdf`.

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.

Discussion: Timing of 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 6. 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.[17] 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 largest change to date.

Furthermore, as shown in Section 4, 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.

4 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 chase ratings throughout the sample period. Second, we use impulse-response analysis to investigate the impact of rating changes on flows and returns.

[17] See "Morningstar Appoints Paul Kaplan, Ph.D., CFA, as Director of Research, Vahid Fathi Named Director of Stock Research" from Morningstar news archive.

4.1 Investors Chase Ratings Regardless of Ratings Methodology

Because our identification relies on the reform in the Morningstar rating methodology, it is important to examine whether mutual fund investors continued to chase 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 one 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 appear to differ before and after June 2002.

More formally, we then estimate the response of fund flows to lagged fund ratings using three-year rolling-window TNA-weighted Fama-MacBeth regressions (Fama and MacBeth, 1973) that controls for 12 lags of monthly fund returns. The results are plotted in Panel (b) of Figure 4. The coefficient estimate only varies slightly over the sample, and there is no material drop around or following the 2002 reform. For example, the average flow-to-rating response was 0.52% before June 2002 and 0.59% after June 2002.[18]

In summary, these results indicate that mutual fund investors chase 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 5.1, the style-level correlated demand due to rating-chasing behavior mostly disappeared after the reform.

4.2 Stock-Level Rating-Induced Price Pressures

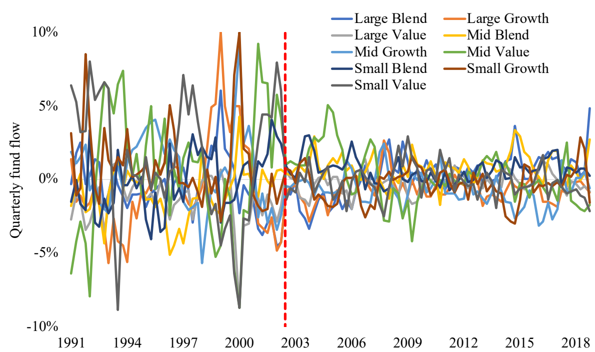
Next, we confirm that Morningstar ratings can have a large price impact on stocks through flow-induced trading. This step is necessary before we explore the influence of rating-induced style demand on style returns in the subsequent analysis.

[18] See further analysis indicating that investors did not change their rating-chasing behavior in Ben-David et al. (2019) and Evans and Sun (2020).

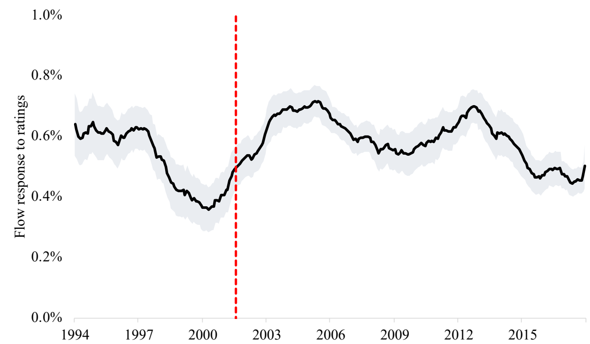
Figure 4. The June 2002 Morningstar Methodology Reform

This figure examines the relationship between Morningstar rating and fund flow over the full sample. The vertical dashed red lines mark 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 relationship between ratings and flows at the fund-level. Specifically, it plots the regression coefficient of fund flows on lagged ratings estimated using three-year rolling windows. The regression controls for 12 lags of monthly fund returns. The shaded area is the two standard error band.

(a) Fund flow, by style



(b) Regression coefficient of fund flows on ratings



We assess the price impact of ratings on stock returns by first separately estimating the two chained effects: (i) the response of fund flows to Morningstar rating changes, and (ii) the response of stock returns to flow-induced trading. We use Fama-MacBeth regressions that are value-weighted by lagged fund TNA in fund-level regressions or by lagged stock market value in stock-level regressions.

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

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

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 and returns. The cumulative response coefficients $(b_1, b_1+b_2, \ldots)$ are plotted in Panel (a) of Figure 5. In response to a one-star change in rating, funds experience an average of 6% additional flows, most of which take place over the first 24 months. This result is consistent with prior research showing that, when controlling for past fund performance, discrete changes in ratings create sizeable differences in fund flows that last for several months (Del Guercio and Tkac, 2008; Reuter and Zitzewitz, 2015).

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:[19]

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

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 non-discretionary trading that is only attributable to fund flows and thus likely does not contain fundamental information.[20]

We then estimate the response of stocks 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 $(c_0, c_0 + c_1, ...)$ 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 one to two 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 impact 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.[21]

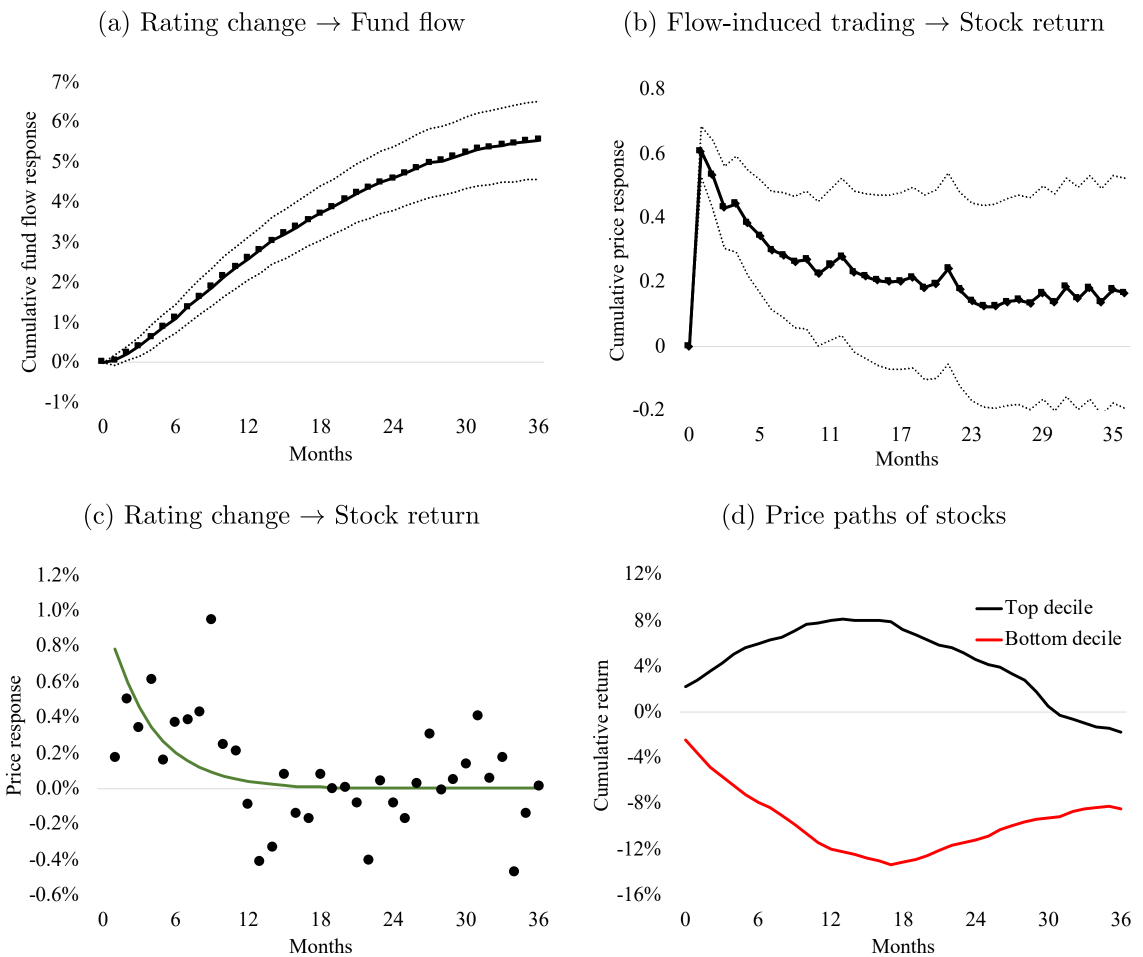
[19] Lou (2012) also applies different scaling factors to inflows and outflows. We omit this scaling for simplicity, but our results are robust to using his scaling factors.

[20] 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, whose construction follows Lou (2012) and does not use price information.

[21] 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—which implies that the effect is better specified by using rating changes.

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. In these two panels, the dashed lines show two standard errors bands. 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 decile breakpoints are based on NYSE stocks.



To facilitate our later analysis of rating-induced price impacts, it is convenient to summarize recent rating changes into a weighted average sum such that the weights correspond to how much each lag impacts returns. We obtain such a weighting scheme by directly estimating the response of stock returns on the past 36 lags of stock-level rating changes (defined in Equation (3)) and plot the coefficients in Panel (c) of Figure 5. As expected, more recent rating changes are more impactful, 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 $\delta = 0.764$, which is estimated from a least-squares fit to the response (Panel (c) of Figure 5). Because the weights sum to 12 (months), in terms of units, ExpSum(ΔRating) should be interpreted as the rating change over one year. The estimated decay factor $\delta = 0.764$ implies a half-life of $-\ln(2)/\ln(\delta) \approx 2.58$ months. Our results are not sensitive to reasonable variations in the parameter δ.

The results presented so far indicate that recent rating changes cause price pressures. To further validate the price pressure interpretation, we examine whether the price movements revert. In Panel (d) of Figure 5, each month we sort stocks into decile portfolios based on $\text{ExpSum}(\Delta\text{Rating})_{i,t-1}$ and track the performance over the following three 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.

In short, the results in this section indicate that rating-induced flows generate substantial price pressure at the stock level.

5 Impact of Rating-Chasing Demand on Style Performance

So far, we have presented evidence that ratings impact 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 style classifications to define stock styles. For instance, the large-cap growth style portfolio is defined by the aggregate holdings of all funds in that category.[22] Specifically, for each stock i, its weight in style π is given by[23]

$$w_{i,t-1}^{\pi} = \frac{\sum_{\text{fund } j \in \text{ style } \pi} \text{Price}_{i,t-1} \cdot \text{SharesHeld}_{i,j,t-1}}{\sum_{\text{fund } j \in \text{ style } \pi} \text{TNA}_{j,t-1}}. \tag{10}$$

We first show that style-level rating-induced price pressures, as well as style-level momentum and reversal, became muted after June 2002. Then, we document that style return dispersion declined dramatically following Morningstar's rating reform. To provide sharper identification, in Section 6, we conduct an event study using a short window to focus exclusively on the reform-induced rating movements in June 2002.

5.1 Style-Level Rating-Induced Price Pressures

To start, we examine 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 up the stock-level rating changes for each style portfolio π:

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

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

To examine the effects of rating changes, we rank styles 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 were presented graphically in Panels (c) and (d) of Figure 1, and are tabulated here in Table 1. 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

[22] We use this classification because it is the basis of the style-level adjustments in Morningstar ratings. Lettau, Ludvigson, and Manoel (2019) document that fund-based style classification in the financial industry does not map exactly to the size and value definitions used by academics that are based on market capitalization and book-to-market ratios (Fama and French, 1993). Appendix A.2 shows that the industry classification is a "smoothed" version of the academic style definitions. In Appendix A.3, we present results repeating the main analyses for styles based on the academic definition. The results generally extend to the academic-based styles qualitatively, though with weaker magnitudes, as expected.

[23] Because $\text{TNA}_{j,t-1} = \sum_{\text{stock } i} \text{Price}_{i,t-1} \cdot \text{SharesHeld}_{i,j,t-1}$, we have $\sum_{\text{stock } i} w_{i,t-1}^{\pi} = 1$.

Table 1. Rating-Induced Price Pressures in Style Portfolios

We sort style portfolios using the lagged exponential sum of rating changes ($\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$) 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.

Panel 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)
Panel 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)

$^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$

bottom-ranked styles. The standard errors are bootstrapped by randomly permuting the style portfolios in each year.

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 for the top three and bottom three styles is smaller, at about 0.7%. After the rating reform in 2002 eliminated style-level rating differences, the spread in flows disappeared.

Table 1 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 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.

Heterogeneous exposure to Morningstar ratings within styles. To further sharpen the test, we make use of the fact that stocks more heavily held by mutual funds should 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 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.

We then repeat the same exercise—sorting styles using ExpSum_{t-1} and examining the difference between the top and bottom styles—for each of the three mutual-fund-holding terciles. As predicted, the price effect before June 2002 is stronger in the style portfolios constructed by stocks with higher mutual fund holdings and relatively muted in those constructed by stocks with low mutual fund holdings.[24] There is no effect in any of the terciles after June 2002.

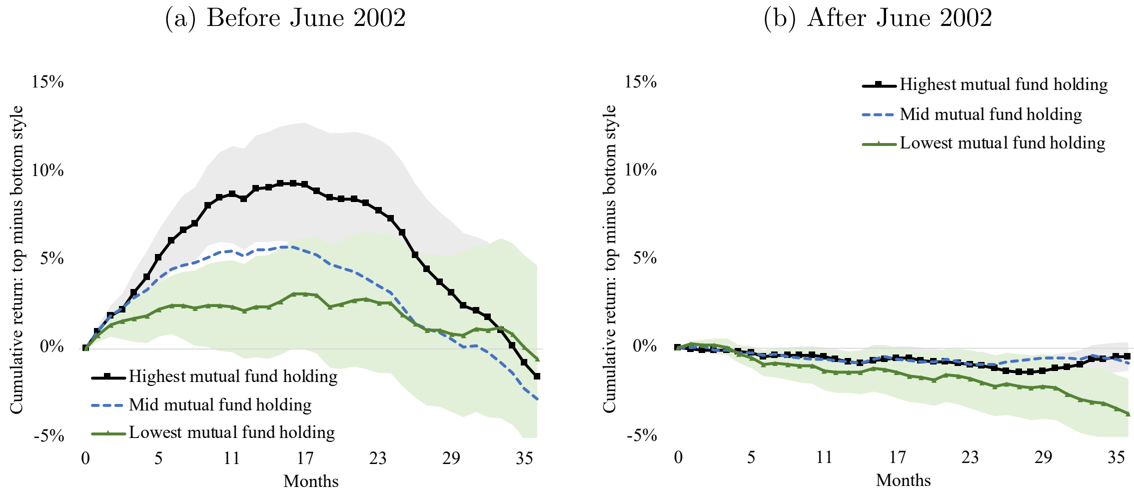
5.2 Rating Chasing and Style Momentum Strategy

These price pressure results suggest the existence of a rating-induced style-level positive feedback trading loop before June 2002. That is, high past performance of a particular style leads to rating upgrade, which in turn attracts inflows to mutual funds in that style. As those funds buy more stocks, that further pushes up the price of the underlying stocks. Due

[24] The style portfolios with higher mutual-fund-holding stocks in fact have *more* overlap in their portfolio constituents. As higher overlap mechanically leads to less return dispersion (100% overlap means zero return dispersion), this goes against us finding a result. To examine this, we compute the pairwise overlap of each two 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.

Figure 6. Rating-Induced Style Returns: Splitting on Mutual Fund Holding

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 to three subsets of stocks, based on the fraction of total shares held by mutual funds. The shaded areas are based on bootstrapped two-standard-error bands.



to the Morningstar methodology reform, this positive feedback loop via the mutual fund channel should be severed in June 2002.

Thus, style momentum should be stronger prior to June 2002 and become weaker after that date. This is indeed the case. In Figure 7, we plot monthly flows and returns of style portfolios in a long-short style-momentum strategy based on lagged returns or the lagged exponential sum of rating changes ($\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$). Panels (a) and (c) show that, prior to June 2002, flows were highly correlated with style sorts based on past performance and rating changes. After June 2002, the flow spread shrank significantly. Panels (b) and (d) plot style returns and show that the style momentum strategy (top minus bottom style) was profitable before June 2002 with a large positive return of about 70 to 90 bps per month. However, after June 2002, the style momentum strategy became entirely unprofitable.

5.3 Cross-Sectional Dispersion in Style Returns

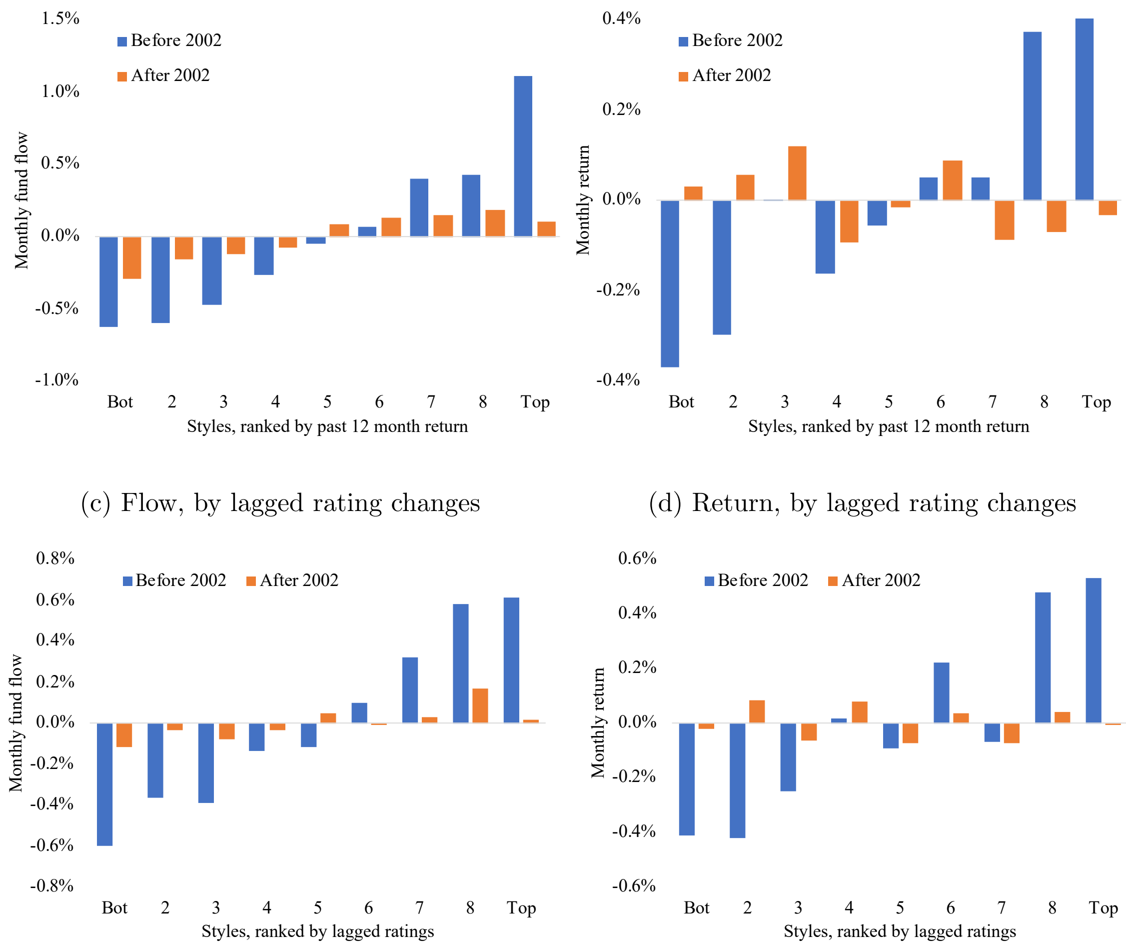
The sharp decline in the style rating dispersion after June 2002 (Panel (a) of Figure 1) also implies that the dispersion in style flows and returns should decrease.

Figure 7. Style Momentum Before and After June 2002

This figure shows monthly flows and returns for a momentum strategy based on styles. We sort the 3×3 styles each month using past-12-month style returns (Panels (a) and (b)) and the lagged exponential sum of rating changes ($\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$) (Panels (c) and (d)). Panels (a) and (c) show the average monthly flow and Panels (b) and (d) show the average monthly return. The bars plot the average monthly flows and returns of those style portfolios before and after June 2002. All variables are demeaned to focus on the cross-sectional difference.

(a) Flow, by lagged return

(b) Return, by lagged return



(c) Flow, by lagged rating changes

(d) Return, by lagged rating changes

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

Table 2. 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.

Regression coefficient on the post-June 2002 dummy						
Dependent variables:	Rating		Flow (%)		Return (%)	
	Spread	Std Dev	Spread	Std Dev	Spread	Std Dev
	(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)

$^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$

The regression coefficients on the post-June-2002 dummy variable are shown in Table 2. 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 top and bottom 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 styles dropped by 2.1% (from 5.0% to 2.9%). The result are qualitatively similar when dispersion is measured using the standard deviation of returns.

Overall, the results in this section indicate that the Morningstar reform has significantly changed style flow and return dynamics after June 2002. We now conduct an event study around the 2002 shock to further zoom in on the event.

6 Event Study Around the Morningstar Reform

To provide a sharper test of rating-induced demand effects on style returns, we conduct an event study using a one-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 different 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). 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.

6.1 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 shrink towards three stars. 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 *pre-window information*. 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:

$$\widehat{\Delta \text{Rating}}_j = \text{Rating}^{\text{counterfactual}}_{j,\text{Dec 2001}} - \text{Rating}^{\text{actual}}_{j,\text{Dec 2001}}, \quad (12)$$

where $\text{Rating}^{\text{counterfactual}}_{j,\text{Dec 2001}}$ is our estimate of what its December 2001 rating would have been under the post-2002 methodology.[25] $\widehat{\Delta \text{Rating}}_j$ thus measures how the fund's rating would 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 enjoys the highest rating in December 2001 and is predicted to experience the largest decline; the large-growth style has the lowest rating in December 2001 and is predicted to experience the largest increase. The success of this prediction using data six months earlier is due to the slow-moving nature of ratings.[26]

In Figure 8, we plot 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 positively and negatively affected style.

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. After June 2002, flows equalized across styles.

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.

To put the size of the price impact into perspective, a back-of-envelope exercise shows that the style-level price impact coefficient in our results is slightly smaller than the estimate

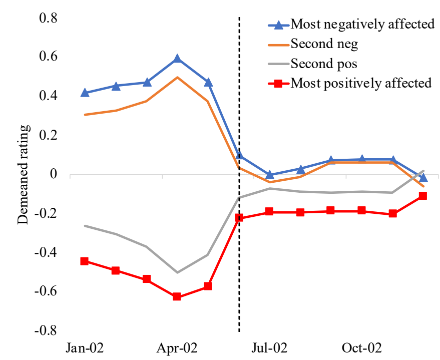
[25] Appendix A.5 provides more details about how the counterfactual ratings are computed in a bottom-up fashion using past fund returns.

[26] As explained in Appendix B, ratings are slow-moving because they are based on three, five, and 10 years of past fund returns.

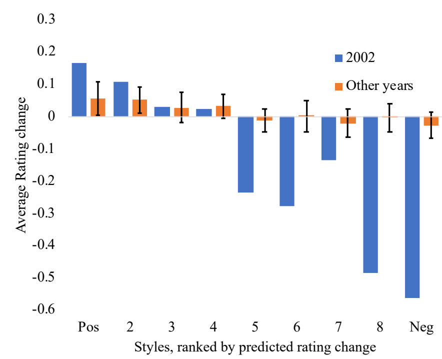
Figure 8. Event Study Around June 2002

We perform event studies on the 3 × 3 size-value Morningstar style portfolios during the six 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 is 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. To focus on cross-sectional dispersion, all variables—ratings, returns, and flows—are demeaned cross-sectionally.

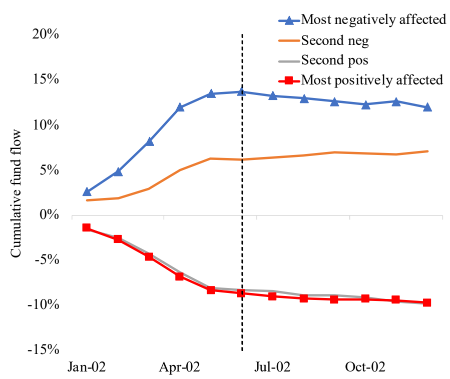
(a) Ratings, 2002



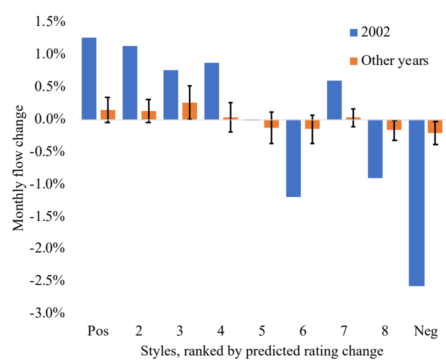
(b) Ratings change: 2002 vs. placebo



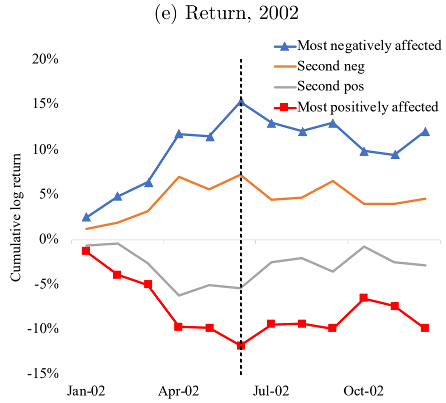
(c) Flow, 2002



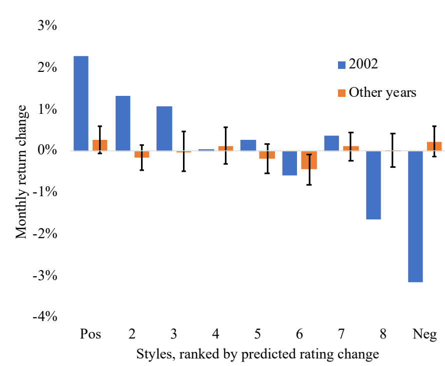
(d) Flow change: 2002 vs. placebo



(e) Return, 2002



(f) Return change: 2002 vs. placebo



obtained by Gabaix and Koijen (2020b).[27] In sum, the patterns in Panels (a), (c), and (e) of Figure 8 are consistent with style-level rating and flow changes having a causal impact on returns.

6.2 Testing Alternative Explanations

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.

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 8 show that the patterns observed in 2002 did not take place in other years. The orange bars show the same exercise in other years together with two-standard-error bands. Clearly, the sharp changes in style ratings, flows, and returns are unique to 2002.

Other measures that may impact 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 impact flows and returns of styles in a manner that is aligned with how Morningstar reform impacted ratings across styles. While we are not aware of similar shocks around June 2002, this is a key assumption that merits further validation.

For this purpose, in Figure 9 we examine whether there are sudden changes in a number of other variables. Theoretically speaking, asset prices can move due to changes in fundamentals or due to trading behaviors,[28] and thus we investigate these two possibilities separately.

[27] Gabaix and Koijen (2020b) use the granular instrument approach of Gabaix and Koijen (2020a) to estimate the price impact coefficient (reciprocal of demand elasticity) at the market level. Their estimate is in the range of 5.28 to 7.08. In our exercise, the cumulative fund flow difference between the top and bottom-ranked styles is 22.4% in the six months leading up to the event (Panel (c) of Figure 8). The return difference is 27.2%. As mutual fund holds approximately 22.8% of the U.S. stock market in 2002, this implies a style-level price impact coefficient of $\frac{27.2\%}{22.5\%} \times \frac{1}{22.8\%} \approx 5.3$.

[28] It is unlikely that 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 is already out of the dot-com-related recession by November 2001, according to the National Bureau of Economic Research.

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). While the fundamentals of different styles do differ and fluctuate, there is no discernible sudden change around June 2002 that is correlated with the style-level return movements.

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 and aggregate by the legal types of institution obtained from Brian Bushee's website. 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 trading by investment companies and independent investor advisors—two categories that include the mutual fund sector. Consistent with our mechanism, these institutions indeed trade in a matter aligned with the flow patterns depicted in Figure 8(a): they traded into (out of) styles with high (low) pre-2002 ratings until a sudden halt right after June 2002. Panel (d) plots the trading of banks, insurance companies, and pension funds. Panel (e) plots the trading of endowment funds and other institutions. There is no discernible "kink" in the trading of any of those non-mutual fund institutions. Finally, because 13F data only records long positions, we also examine the evolution of aggregate short interest in Panel (f). While there is a general slow rise of short interest over the window, there is no clear change around the event.

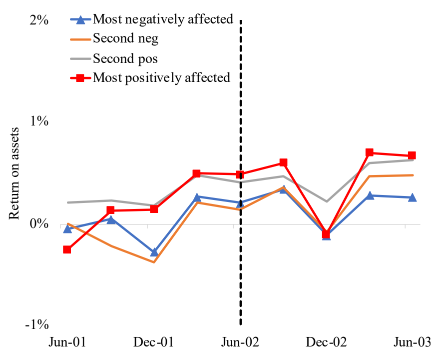
Controlling for stock characteristics. The previous two tests show that there were no sudden changes in fundamentals or trading behaviors of non-mutual-fund institutions around 2002. However, one might still argue that our results could be driven by sudden characteristics-related return changes that happen for other reasons. For example, one might hypothesize that investor sentiment suddenly decreased for small-value stocks after June 2002, and this can cause 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/market, the most commonly-used characteristics. Specifically, for each stock i, we

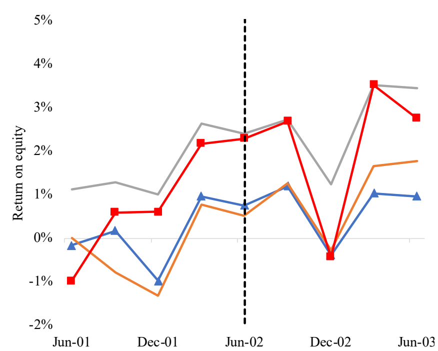
Figure 9. Event Study Around June 2002: Alternative Explanations

We perform event studies on the 3×3 Morningstar style portfolios around the June 2002 methodology change. As in Figure 8, 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. 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 the mutual fund sector. Panel (d) plots trading by banks, insurance companies, and pension funds; Panel (e) plots trading by endowment funds and other institutions. To focus on cross-sectional dispersion, the trading measures in Panels (c) to (e) are demeaned by period. Panel (f) plots the evolution of aggregate short interest. The vertical dashed line indicates the June 2002 methodology change event.

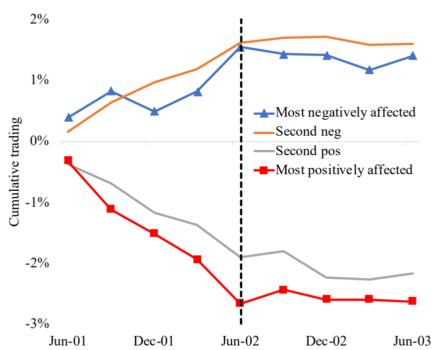
(a) Return on assets (ROA)



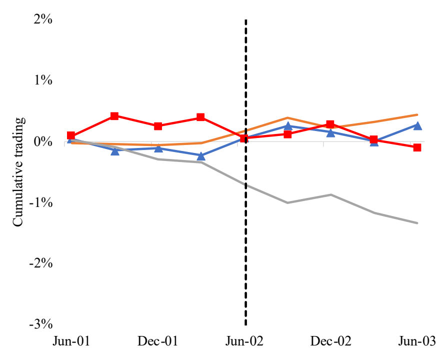
(b) Return on equity (ROE)



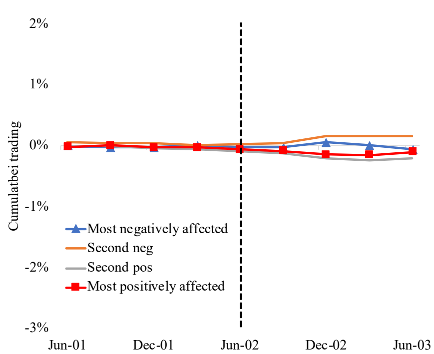
(c) Trading: Investment companies and advisors



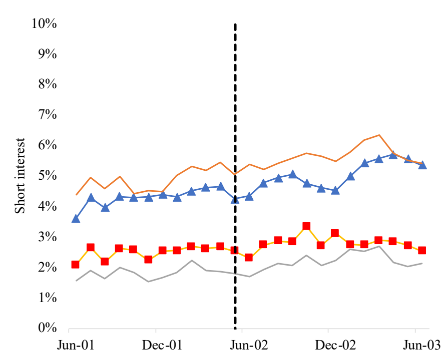
(d) Trading: Banks, insurance, pension funds



(e) Trading: Endowment funds and others



(f) Short interest



define:

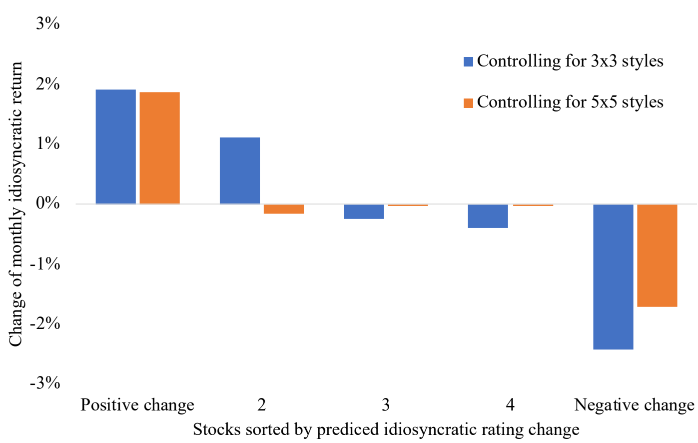
$$\text{Rating}_{i,t}^{\text{idiosyncratic}} = \text{Rating}_{i,t} - \text{Rating}_{\text{size-book/market portfolio } p,t}, \tag{13}$$

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

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

Figure 10. 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/market 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 are stocks that are 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 data in December 2001)[29] and examine the change in idiosyncratic stock returns after the event in Figure 10. Note

[29]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.

that our mechanism—that rating changes impact returns—works both at the characteristic-spanned and characteristic-orthogonal (idiosyncratic) levels. Therefore, even after controlling for characteristics, we should still expect to see an effect. Consistent with this prediction, stocks predicted to experience large upgrade (downgrade) in *idiosyncratic* rating (Equation (13)) experience positive (negative) changes in *idiosyncratic* return (Equation (14)). The quantitative relationship between rating changes and return changes is also comparable to that found at the style level (Panel (f) in Figure 8). These results further suggests that our style-level findings around June 2002 are unlikely to be driven by unspecified characteristics-level return movements.

7 Demand Effects Shaping the Asset Pricing Factors

So far, we have seen that rating-induced demand creates large price pressures on style returns. In this section, we explore the relation between ratings-chasing demand and the size and value factors, which are defined as long-short portfolios on styles. Here we ask: to what extent the returns of the size and value factors can be explained by ratings-induced correlated demand? To be consistent with the earlier sections, we present findings based on Morningstar style but the results based on styles in academic studies (following Fama and French, 1993) are similar (Appendix A.3).

To answer this question, we first use the 12-month short window around June 2002 to estimate the price impact coefficient of Morningstar ratings on style returns—the building blocks of the size and value factors. We choose a short window for two reasons. First, we want to ensure that rating changes primarily come from the methodology change.[30] Second, we want to avoid the impact of other market-level changes, such as the dot-com bubble burst in early 2000, the decimalization event in 2001 (Chordia, Subrahmanyam, and Tong, 2014), or the "momentum crash" period in 2008 (Daniel and Moskowitz, 2016).

Formally, we estimate a forecasting panel regression using the same twelve months around

[30]This is not true if we use a longer sample as ratings are (albeit complex) functions of past returns. Recent studies find that return factors can exhibit momentum (Arnott, Clements, Kalesnik, and Linnainmaa, 2019; Gupta and Kelly, 2019); earlier papers found that returns tend to exhibit long-term reversion (De Bondt and Thaler, 1985). Thus, running a regression of style returns on lagged ratings may be picking up both momentum and reversal effects.

the event (January 2001 to December 2001):

$$\text{Ret}_{\pi,t} = \lambda \cdot \text{ExpSum}(\Delta\text{Rating})_{\pi,t-1} + \text{Controls}_{\pi,t-1} + \epsilon_{\pi,t}, \tag{15}$$

where the controls include style returns over $t-1$, $t-2$ to $t-6$, and $t-7$ to $t-12$ months as well as style and time fixed effects. By controlling for past returns, we can better capture the marginal effect of rating changes in addition to possible style momentum effects (Ehsani and Linnainmaa, 2019; Gupta and Kelly, 2019). Standard errors are clustered by month.

Table 3. Estimating the Price Impact of Ratings (λ) Around June 2002

We estimate the rating price impact coefficient λ through a forecasting panel regression of monthly returns of the 3×3 styles on lagged ratings changes ($\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$). We also control for past style returns over $t-1$, $t-2$ to $t-6$, and $t-7$ to $t-12$ months, as well as style and time fixed effects in some specifications. The sample spans the six months before to six months after the methodology change. Columns (1) and (2) are estimated in a panel regression and standard errors are clustered by month. Columns (3) and (4) are estimated using feasible general least squares (FGLS) where we use the empirically estimated covariance matrix of style returns to adjust for cross-sectional correlations. Standard errors are shown in parentheses.

Dependent variable:	Monthly style return $\text{Ret}_{\pi,t}(\%)$			
	Panel regression		FGLS	
	(1)	(2)	(3)	(4)
$\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$	2.89^{**}	2.63^{**}	2.49^{***}	1.76^{***}
	(1.16)	(1.17)	(0.71)	(0.67)
Past Return Controls	Yes	Yes	Yes	Yes
Style FE	Yes	No	Yes	No
Time FE	Yes	Yes	Yes	Yes
Observations	108	108	108	108
Adjusted R^2	93.0%	93.2%	80.6%	78.9%

$^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$

The estimation results are shown in Table 3. For each star rating change, the style-level price impact is 2.89% per month with a t-statistic of 2.50 (Column (1)). To account for cross-sectional style return correlation, in Columns (3) and (4), we also adjust the standard errors using a feasible generalized least squares (FGLS) approach.[31] When using FGLS, the estimated price impact is 2.49% ($t = 3.50$). The estimates are both statistically and economically significant. Our estimates do not change materially if we use shorter or longer event-time windows, and we present those robustness checks in Appendix Table A.3.

We now use the estimated $\lambda = 2.89\%$ in Column (1) to quantify the influence of ratings on style and factor returns. Specifically, we use the following price-impact specifications:

$$\text{Ret}_{\text{SMB},t} = \underbrace{\lambda \cdot \text{ExpSum}(\Delta\text{Rating})_{\text{SMB},t-1}}_{\text{Rating-induced price pressure}} + \text{Ret}^{\text{counterfactual}}_{\text{SMB},t} \tag{16}$$

$$\text{Ret}_{\text{HML},t} = \underbrace{\lambda \cdot \text{ExpSum}(\Delta\text{Rating})_{\text{HML},t-1}}_{\text{Rating-induced price pressure}} + \text{Ret}^{\text{counterfactual}}_{\text{HML},t}, \tag{17}$$

where

$$\text{ExpSum}(\Delta\text{Rating})_{\text{SMB},t-1} \equiv \left(\sum_{\pi\in\mathcal{S}} - \sum_{\pi\in\mathcal{B}}\right) \text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$$

$$\text{ExpSum}(\Delta\text{Rating})_{\text{HML},t-1} \equiv \left(\sum_{\pi\in\mathcal{V}} - \sum_{\pi\in\mathcal{G}}\right) \text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}.$$

In the equations above, we use $\mathcal{S}$, $\mathcal{B}$, $\mathcal{V}$, and $\mathcal{G}$ to denote the three small-cap styles, the three large-cap styles, the three value styles, and the three growth styles, respectively. For

[31] We use the full sample of style returns to estimate the covariance matrix of style returns and incorporate it into the estimation. Specifically, let y be the vector of style returns stacked together so that the first nine entries are the first month, the next nine entries are the second month, and so forth. Then, we estimate the covariance matrix of y to be

$$\hat{\Omega} = \begin{pmatrix} \hat{C} & 0 & \dots & 0 \\ 0 & \hat{C} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{C} \end{pmatrix},$$

where $\hat{C}$ is the estimated contemporaneous return variance-covariance matrix of the nine styles. Let X denote the matrix of independent variables. Then, we estimate the regression coefficients and variance-covariance matrix using

$$\hat{b} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y,$$
$$\widehat{Var}(\hat{b}) = (X'\hat{\Omega}^{-1}X)^{-1}.$$

example, $\mathcal{S}$ combines the value-small, blend-small, and growth-small portfolios.

To visualize the influence of rating-induced price pressure on the factors, in Panels (a) and (b) of Figure 11, we plot the cumulative returns of factors against the cumulative rating-induced returns ($\lambda \cdot \text{ExpSum}(\Delta\text{Rating})_{t-1}$). To capture value and size premia in one single strategy, in Panel (c), we also plot the returns of the "diagonal" portfolio (SVMBG) that is long the small-value style and short the large-growth style. The plots suggest that that rating-induced price pressures can explain a large portion of factor return variation before 2002. After June 2002, rating-induced demand largely loses explanatory power, as expected.

To quantify the explanatory power on factor return variation, we compute the modified "R-squared" of rating-induced price pressures by using the cleanly identified λ estimate in Equation (15):

$$\text{R-squared}^f = \frac{Var(\lambda \cdot \text{ExpSum}(\Delta\text{Rating})_{f,t-1})}{Var(\text{Ret}_{f,t})},$$
$$\text{where } f \in \{\text{SMB}, \text{HML}, \text{SVMBG}\}.$$

We compute this "R-squared" measure for both before and after June 2002. Before 2002, we find that $\text{R-squared}^{\text{SMB}} = 34.7\%$, $\text{R-squared}^{\text{HML}} = 11.1\%$, and $\text{R-squared}^{\text{SVMBG}} = 27.0\%$. After June 2002, these figures drop to 9.5%, 3.3%, and 9.2%, respectively.[32] [33]

While this exercise admittedly delivers a crude estimate, the results suggest that correlated demand of styles can explain a sizeable fraction of factor return variation.

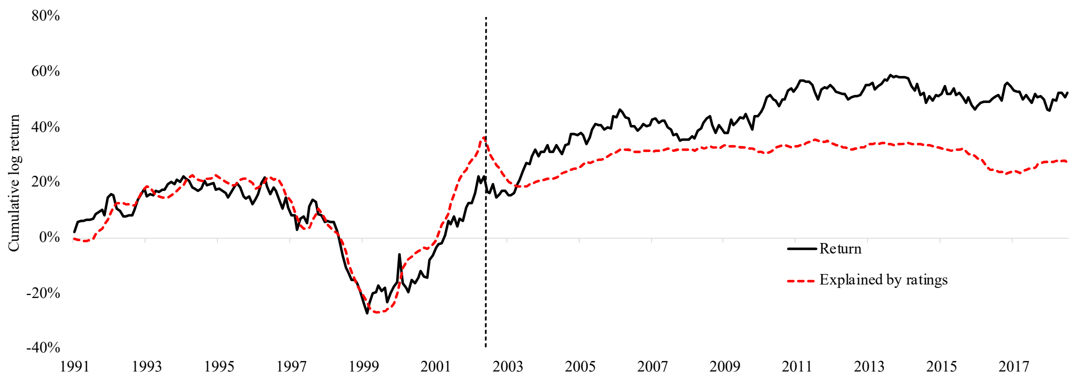
[32] If we only use the residual of $\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$ after partialling out the various past return controls in regression (15), then the "R-squared" for SMB, HML, and SVMBG becomes 29.0%, 7.7%, and 20.5% before June 2002 and 9.2%, 4.0%, and 9.8% after June 2002.

[33] While the estimate of λ based on the 2002 shock is better identified, one may be worried that the price impact could vary over time. Therefore, we also repeat the exercise using λ estimated using five-year rolling window regressions (Figure A.5). As expected, the estimated λ is larger before 2002 and becomes statistically insignificant after June 2002. When using this rolling-window λ as a conservative estimate, for the period before 2002, the "R-squared" for SMB, HML, and SVMBG becomes 12.4%, 3.5%, and 9.5%, respectively. After June 2002, it declines to 2.3%, 0.7%, and 1.9%, respectively.

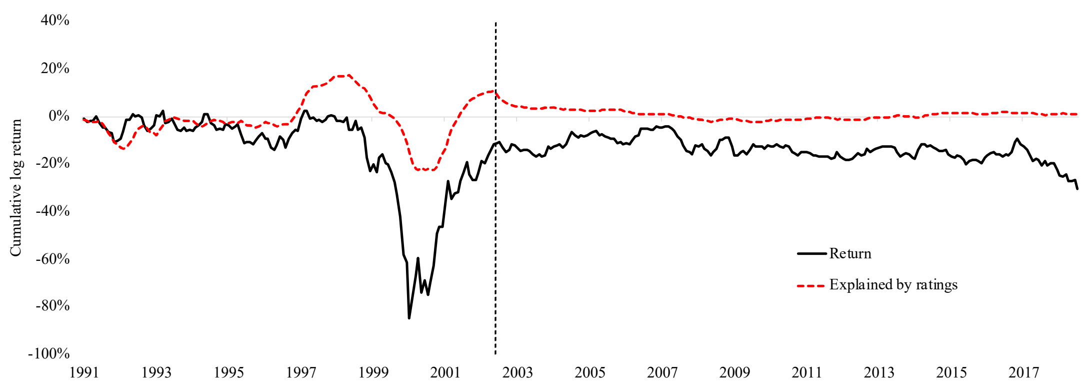
Figure 11. Explanatory Power of Ratings on Size and Value Factors

We quantify the explanatory power of rating pressures on long-short factor portfolios based on the 3 × 3 styles. Panel (a) plots the average returns of the three small capitalization styles minus the three large capitalization styles ("small-minus-big"). Panel (b) plots the average of the three value styles minus the average of the three growth styles ("high-minus-low"), while Panel (c) plots the small-value style minus the large-growth style. The black solid lines are the actual cumulative log returns while the red dashed lines are the returns explained by ratings ($\lambda \cdot \text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$) where λ is estimated in Column (1) of Table 3.

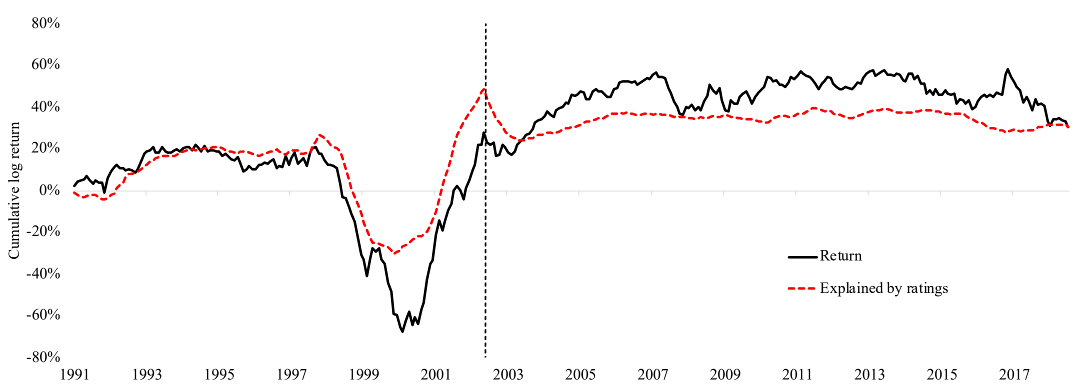
(a) Small Cap − Large Cap



(b) Value − Growth



(c) Small-Value − Large-Growth



8 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 non-fundamental, large enough to impact a broad set of securities, and that also clearly causes systematic price fluctuations. Our study presents evidence that demand fueled by mutual fund investors' rating-chasing behavior can impact large-scale price patterns, including style returns, style momentum, and style dispersion.

The identification of the effects comes from Morningstar's June 2002 reform to its methodology for rating mutual funds. The reform equalized ratings across styles, causing capital flows to spread more evenly across styles. Prior to the reform, style-level rating imbalances created style-level price pressures and increased return dispersion across styles. Moreover, because ratings depend on past returns, rating chasing caused a positive feedback loop involving style returns and fund flows; thus, style portfolios displayed momentum and reversal in both flows and returns. After the reform, these patterns disappeared, consistent with the removal of style-level rating-induced price pressures. Using an event study around the exact reform date, we achieve sharper identification of the effects of rating-induced demand on style returns. Finally, we estimate that rating-induced price pressures can explain a sizeable fraction of the time-series variation in the size and value factors.

Our results show that one specific source of non-fundamental demand—mutual fund investors naively chasing ratings irrespective of rating definition—can drive significant fluctuations in style-level performance. The overall role of correlated demand in determining asset prices is likely greater than what is documented here. Correlated demand can also 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 reflecting fundamental risk, they may also be determined by non-fundamental demand.

References

Akbas, Ferhat, Will J. Armstrong, Sorin Sorescu, and Avanidhar Subrahmanyam, 2015, Smart money, dumb money, and capital market anomalies, *Journal of Financial Economics* 118, 355–382.

Arnott, Robert D., Mark Clements, Vitali Kalesnik, and Juhani T. Linnainmaa, 2019, Factor momentum, Working paper, Dartmouth College.

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.

Barber, Brad M., Terrance Odean, and Lu Zheng, 2005, Out of sight, out of mind: The effects of expenses on mutual fund flows, *Journal of Business* 78, 2095–2120.

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

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, 2020, The granular nature of large institutional investors, *Management Science* forthcoming.

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, 2019, What do investors really care about?, Working paper, The Ohio State University.

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

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, 2020, ETF arbitrage and return predictability, *Review of Finance* forthcoming.

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.

Choi, James J., and Adriana Z. Robertson, 2020, What matters to individual investors? Evidence from the horse's mouth, *Journal of Finance* 75, 1965–2020.

Chordia, Tarun, Avanidhar Subrahmanyam, and Qing Tong, 2014, Have capital market anomalies attenuated in the recent era of high liquidity and trading activity?, *Journal of Accounting and Economics* 58, 41–58.

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

Daniel, Kent D., and Tobias J. Moskowitz, 2016, Momentum crashes, *Journal of Financial Economics* 122, 221–247.

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

De Bondt, Werner FM, 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.

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.

Ehsani, Sina, and Juhani T. Linnainmaa, 2019, Factor momentum and the momentum factor, Working paper, National Bureau of Economic Research.

Evans, Richard B., and Yang Sun, 2020, Models or stars: The role of asset pricing models and heuristics in investor risk adjustment, *Review of Financial Studies* forthcoming.

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 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).

Frazzini, Andrea, and Owen A Lamont, 2008, Dumb money: Mutual fund flows and the cross-section of stock returns, *Journal of Financial Economics* 88, 299–322.

Friesen, Geoffrey C., and Viet Nguyen, 2018, The economic impact of mutual fund investor behaviors, Working paper, University of Nebraska-Lincoln.

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, 2020a, Granular instrumental variables, Working paper, Harvard University.

Gabaix, Xavier, and Ralph S.J. Koijen, 2020b, 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.

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

Gupta, Tarun, and Bryan Kelly, 2019, Factor momentum everywhere, *Journal of Portfolio Management* 45, 13–36.

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.

Hartzmark, Samuel M., and Abigail Sussman, 2019, Do investors value sustainability? A natural experiment examining ranking and fund flows, *Journal of Finance* 76, 2789–2837.

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

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.

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.

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

Pastor, Lubos, Robert F Stambaugh, and Lucian A Taylor, 2020, Fund tradeoffs, *Journal of Financial Economics* forthcoming.

Reuter, Jonathan, and Eric Zitzewitz, 2015, How much does size erode mutual fund performance? A regression discontinuity approach, Working paper, Boston College.

Samuelson, Paul A, 1998, Summing up on business cycles: Opening address, in *Conference Series-Federal Reserve Bank of Boston*, volume 42, 33–36, Federal Reserve Bank of Boston.

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

Song, Yang, 2020, The mismatch between mutual fund scale and skill, *Journal of Finance* 75, 2555–2589.

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

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.

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.

A.2 Morningstar vs. 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/market ratios—the value stocks defined by academia. This section explores the difference between the Morningstar and the academic style definitions.

In Figure A.1, we sort stocks by market capitalization and book/market 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 a "rectangular region." 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 can also hold some, albeit fewer, stocks with other characteristics.

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

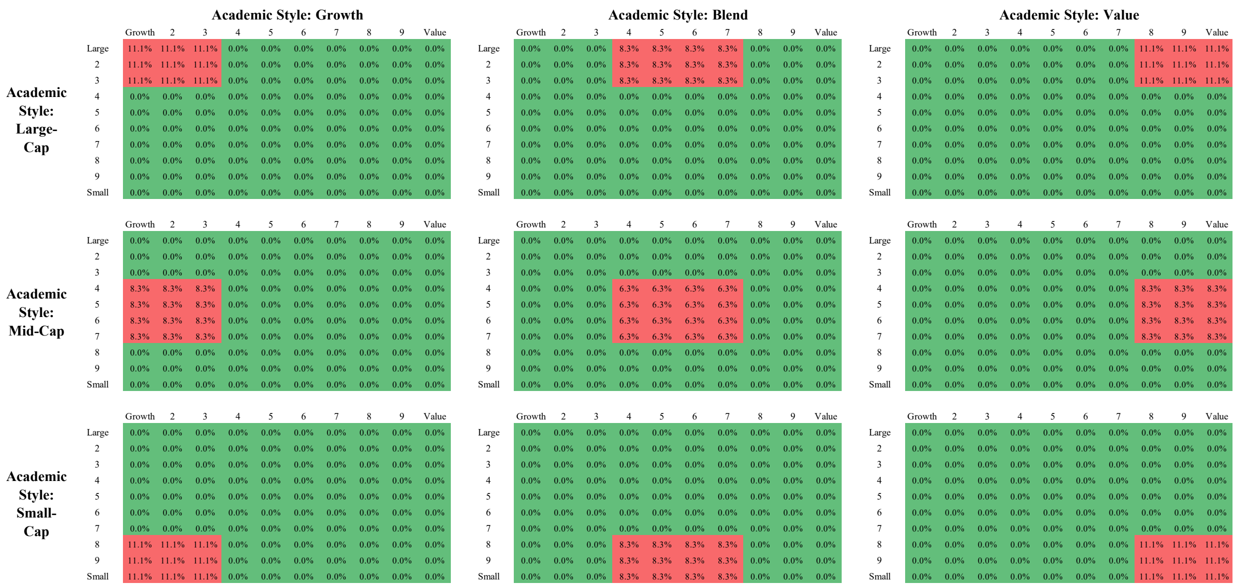
There are a total of nine stock styles (small/mid/large cap × value/blend/growth) used in this study. Columns (2) and (3) summarize the mean and standard deviation of Morningstar ratings for each style. Columns (4) and (5) summarize the lagged-12-month rating changes (ExpSum(ΔRating)). Columns (6) and (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	Std Dev	Average	Std Dev	Average	Std Dev	Average	Std Dev
(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%

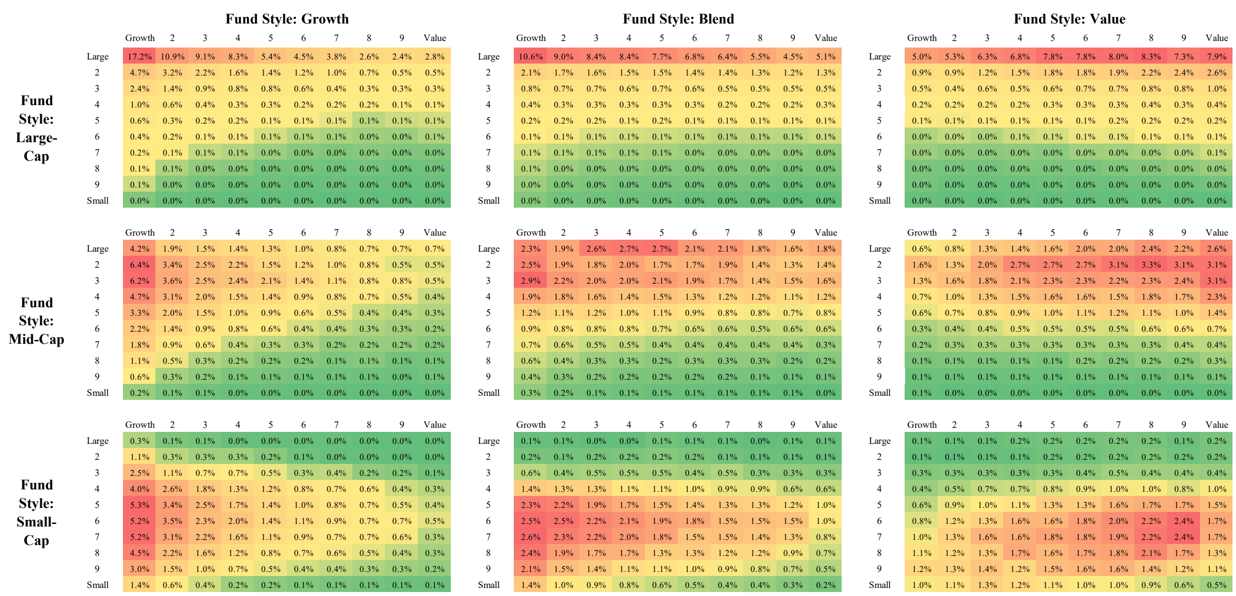
Figure A.1. Comparison of Morningstar and Academic Stock Style Definitions

We sort stocks into 10 × 10 portfolios based on NYSE size and book/market 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 Definition



(b) Morningstar Fund-based Style Definition



A.3 Empirical Results Using Academic Styles

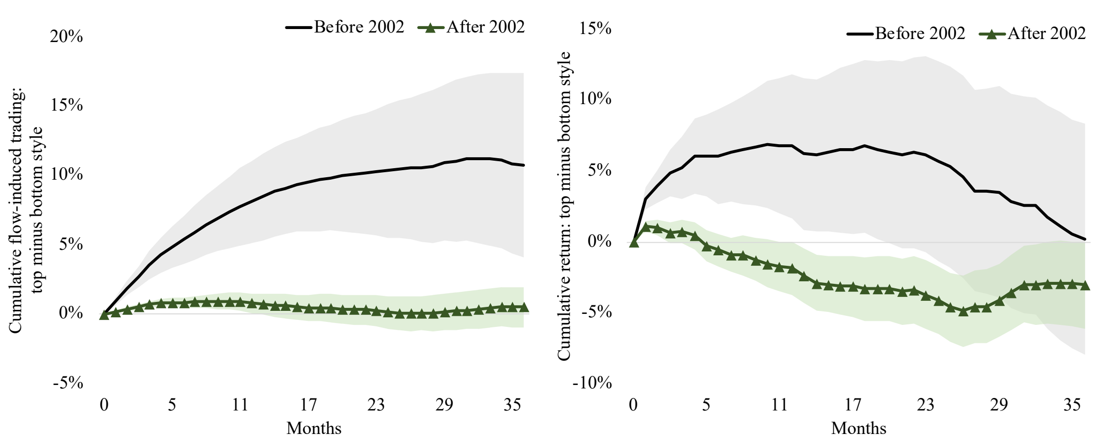
This section shows that the key results based on Morningstar-defined styles also extend to the academic-defined styles using size and book/market characteristics.

Figure A.2 reproduces Panels (c) and (d) in Figure 1 with academic-defined styles. The general patterns are similar.[34]

Figure A.2. Price Pressure in Academic Style Portfolios

This is similar to Panels (c) and (d) in Figure 1 but performed using style portfolios defined using stock characteristics. Stocks are sorted into 3×3 size-value styles using NYSE breakpoints of market cap 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 are 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 6. Figure A.3 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 8, where style portfolios are instead based on Morningstar-defined styles.

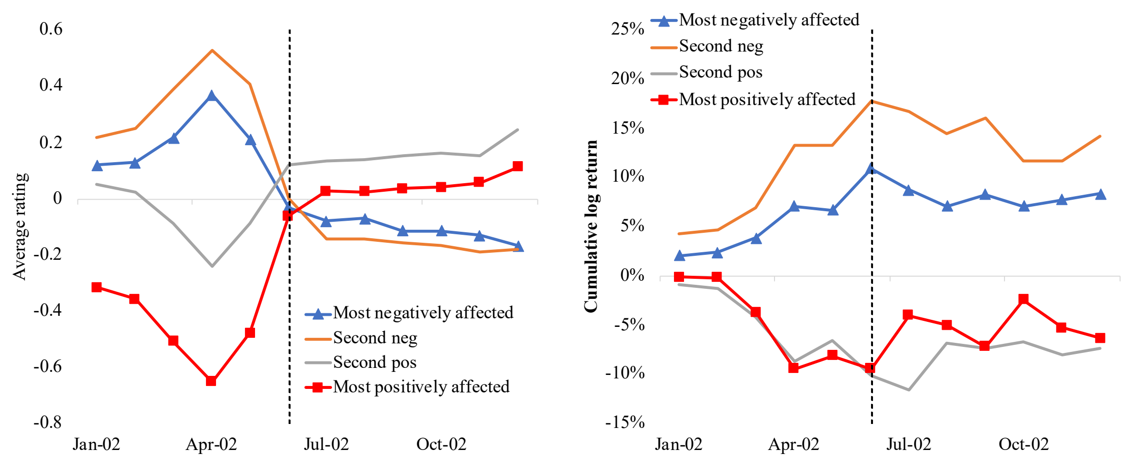
We also quantify the influence of rating-induced demand on the size and value factors that are constructed using the academic-defined styles. In Figure A.4, we plot the cumulative returns of the "academic" factors together with the price pressure driven by the rating-chasing demand as in Figure 11. We find that the rating-induced demand also explains a

[34] 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.3. Behavior of Academic Styles Around the June 2002 Event

We perform event studies on the 3×3 size-value academic style portfolios during the six 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). To focus on cross-sectional dispersion, all variables are demeaned cross-sectionally.

(a) Ratings (b) Returns



large part of the variation in academic factors before 2002. The explanatory power disappears after June 2002 as the demand is spread out across styles.

A.4 Estimating Price Impact λ

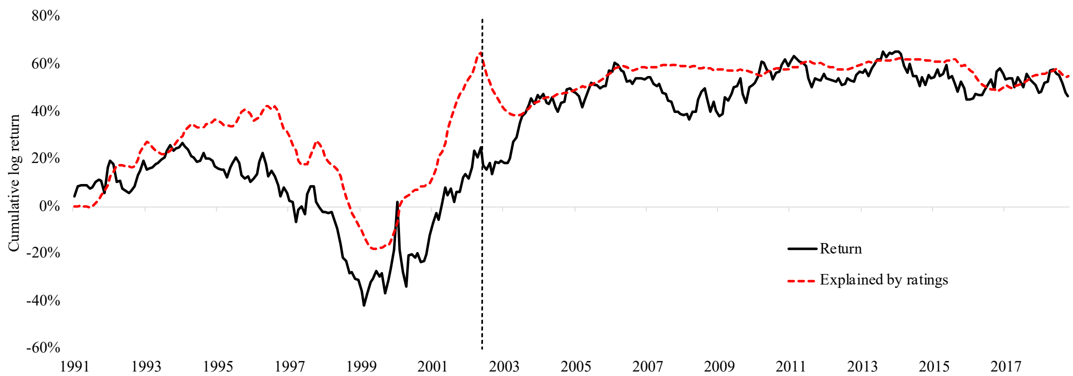
This section provides additional results for the estimation of price impact parameter λ in Section 7. Table A.3 estimates the price-pressure coefficient in Equation (15) using two alternative time windows. The results do not change materially if we decrease or increase the length of the estimation window.

Figure A.5 shows the estimate of the price impact λ using five-year rolling windows. Compared to the short-window estimate of λ (see Table A.3), the coefficient estimated using rolling-window regressions is slightly lower before 2002 and significantly lower after June 2002. These results are entirely expected for two reasons. First, in a typical month, only around 14% of funds experience a change in rating (Ben-David et al., 2019; Evans and Sun, 2020), whereas in June 2002 54% of funds experienced a long-lasting change in rating. Thus, the estimates from windows that do not contain June 2002 likely understate the true price

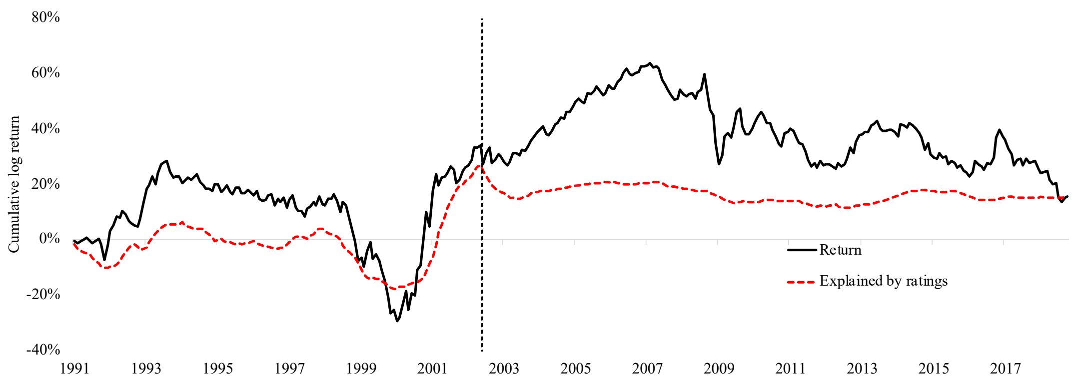
Figure A.4. Explanatory Power of Ratings on Academic Size and Value Factors

We quantify the explanatory power of rating pressure on long-short portfolios based on the 3 × 3 academic styles. Panel (a) plots the average returns of the three small capitalization styles minus the three large capitalization styles ("small-minus-big"). Panel (b) plots the average of the three value styles minus the average of the three growth styles ("high-minus-low"), and Panel (c) plots the small-value style minus the large-growth style. The solid black lines are the actual cumulative log returns, and the red dashed lines are the returns explained by ratings ($\lambda \cdot \text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$), where λ is estimated in Column (1) of Table 3.

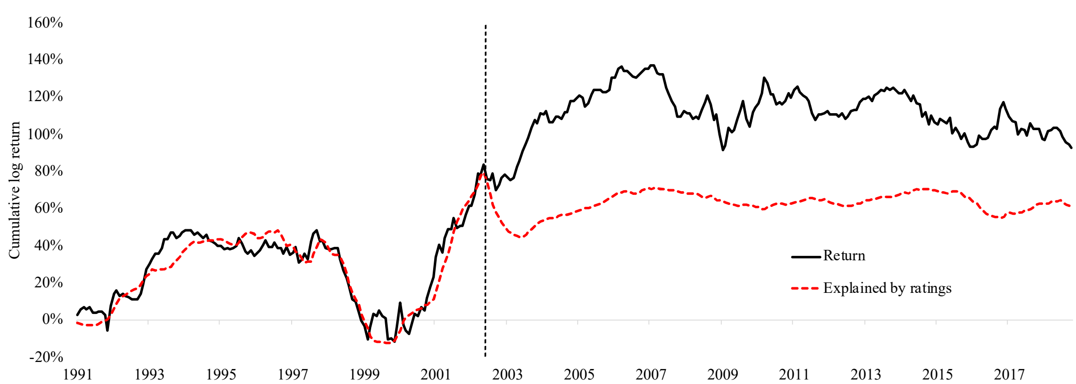
(a) Small cap − Large cap



(b) Value − Growth



(c) Small value − Big growth



impact due to noise in the measurement of the right-hand-side variable. Second, a corollary of the hypothesis we explore in this paper is that changes in average style-level ratings should have a smaller price impact after the rating methodology reform. This is exactly what we observe in Figure A.5.

Table A.3. Estimating the Price Impact of Ratings (λ) Around June 2002

This robustness check of Table 3 uses alternative sample window lengths. We estimate the rating price impact coefficient λ through a forecasting panel regression of monthly returns of the 3×3 styles on lagged rating changes ($\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$), controlling for past style returns. Columns (1) to (3) are estimated using panel regressions with standard errors clustered by month, while Columns (4) to (6) are estimated using feasible general least squares (FGLS).

Dependent variable:	Monthly style return $\text{Ret}_{\pi,t}$(%)					
	Panel regression			FGLS		
	6 months	12 months	18 months	6 months	12 months	18 months
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$	3.29***	2.89**	2.77***	2.32***	2.49***	2.20***
	(1.20)	(1.16)	(0.90)	(0.86)	(0.71)	(0.60)
Past Return Controls	Yes	Yes	Yes	Yes	Yes	Yes
Style, Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	54	108	162	54	108	162
Adjusted R^2	84.5%	93.0%	91.7%	63.6%	80.6%	75.2%

$^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$

A.5 Predicting Rating Changes

We follow Morningstar's methodology in Section 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 historically used 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.6 plots the actual style-level rating change in June 2002 against our predictions (computed following the method in Section 6.1). As shown in the figure, we correctly predict that small-value is the most negatively affected style and small-blend is the

Figure A.5. Price Impact Parameter (λ) Estimated from Rolling Five-Year Windows

We estimate panel regressions of style returns on the exponential sum of past rating changes ($\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$) using five-year rolling windows. The windows are centered so that the estimate on month t uses data from month $t-29$ to month $t+30$. The dashed lines indicate two standard error bands.

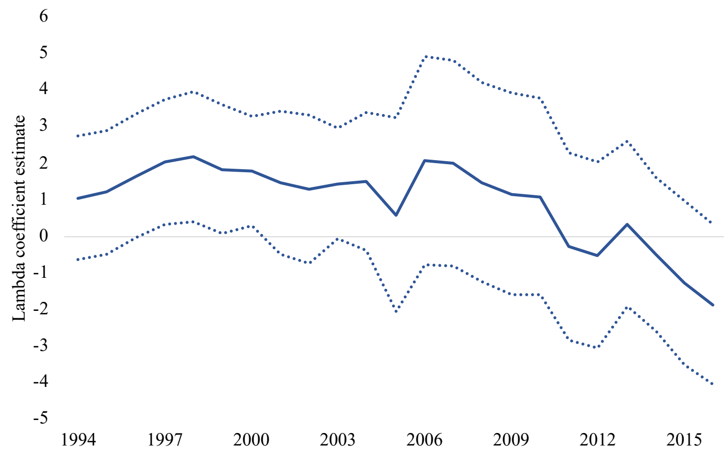
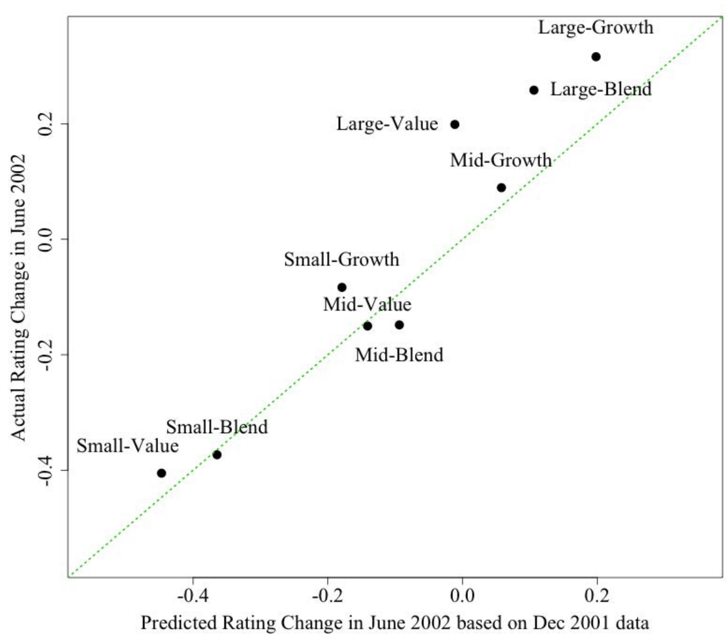


Figure A.6. Predicting June 2002 Style Rating Change using End of 2001 Data

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 up 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 perfect match.



second negatively affected, while large-growth is the most positively affected and large-blend is the second. The predictions are also reasonably accurate in magnitudes.

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. There are two steps in Morningstar's rating calculation:

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):

$$R_i^T = \prod_{t=1}^{T}(1 + r_{i,t}) - 1, \qquad T \in \{36, 60, 120\}, \tag{18}$$

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:

$$\text{LoadRet}_i^T = R_i^T L_i - R_f^T, \tag{19}$$

where the load adjustment L_i equals to 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 three-month T-bills. The measure is standardized to:

$$\text{MnLoadRet}_i^T = \frac{\text{LoadRet}_i^T}{\max(R_f, \text{AvgLoadRet}^T)}, \tag{20}$$

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:

$$\text{Performance}_{i,t} = \text{MnLoadRet}_{i,t}^T - \text{MnRisk}_{i,t}^T. \tag{21}$$

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

$$\text{Risk}_i^T = \frac{\sum_{t=1}^T - \min(r_{i,t} - r_t^f, 0)}{T} \tag{22}$$

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

$$\text{MnRisk}_t^T = \frac{\text{Risk}_i^T}{\text{AvgRisk}^T}. \tag{23}$$

After June 2002, Morningstar changed the way it adjusts for risk.[35] Morningstar summarizes a fund's past performance using the so-called Morningstar risk-adjusted return (MRAR):

$$\text{MRAR}_i^T(\gamma) = \left[\frac{1}{T}\sum_{t=1}^{T}(1 + r_{i,t} - r_t^f)^{-\gamma}\right]^{-\frac{12}{\gamma}} - 1, \tag{24}$$

where $r_{i,t} - r_t^f$ is the geometric return in excess of the risk-free rate after adjusting for loads,[36] and $\gamma = 2$ is the risk aversion coefficient.

The formula penalizes funds with higher return volatility. To see this, notice that when

[35] 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).

[36] 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} (and V_{unadj}) is the load-adjusted (unadjusted) cumulative fund return over the past T months. For details, see "The Morningstar Rating Methodology," June 2006.

γ converges to 0, $\text{MRAR}^{\text{T}}(0)$ is equal to the annualized geometric mean of excess returns.[37] 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 three-year, five-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 three years but less than five years of data, the overall rating is just the three-year rating. For funds with more than five years but less than 10 years of data, the overall rating assigns 60% and 40% weights on the five-year and three-year ratings.[38] For those with 10 years of data, 50%, 30%, and 20% weights are assigned on the 10-year, five-year, and three-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-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).

[37] 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.

[38] Because the five-year history contains the three-year history, the three most recent years are effectively given more weight than more distant history.