Non-Fundamental Demand and Style Returns

Itzhak Ben-David, Jiacui Li, Andrea Rossi, Yang Song*

VERY PRELIMINARY, PLEASE DO NOT CITE

Abstract

We present *causal* evidence that non-fundamental correlated demand exerts first-order impact on style returns. Mutual fund investors chase fund performance via Morningstar ratings. Until June 2002, funds pursuing the same investment style had highly correlated ratings. Therefore, rating-chasing investors directed capital into winning styles, generating style-level price pressures that reverted over time. In June 2002, Morningstar reformed its rating methodology so that ratings are equalized across styles. As a result, 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 also find that this source of correlated demand explains substantial variation in the size and value factors.

Keywords: Correlated demand, style investing, mutual funds, momentum

JEL Classification: G11, G24, G41

*We thank Sylvester Flood (Morningstar), Paul Kaplan (Morningstar), Alex Chinco, Bing Han, Juhani Linnainmaa, Andrei Shleifer, and Xin Wang for helpful comments. We thank Martin Lettau for spotting a data mistake in an earlier version. Ben-David is with The Ohio State University and the National Bureau of Economic Research, Li is with University of Utah, Rossi is with the University of Arizona, and Song is with the University of Washington. Emails: ben-david.1@osu.edu, jiacui.li@eccles.utah.edu, rossi2@arizona.edu, and songy18@uw.edu.

1 Introduction

Economists agree that prices of assets reflect investors' demand for these assets. What is subject to debate is the extent to which prices of securities are elastic and the nature of investors' demand. On one end of the spectrum, elastic prices combined with rational expectations about future cash flows yields an efficient market. On the other end of the spectrum, inelastic prices and non-fundamental demand results in a market with mispricings. If the non-fundamental demand is entirely idiosyncratic, then the mispricings may be inconsequential. However, if non-fundamental demand is systematically correlated, e.g., at the style level (such as value or growth), then prices can contain systematic inefficient components (Barberis and Shleifer, 2003; Koijen and Yogo, 2019). In the words of Paul Samuelson, markets can become *"macro-inefficient"* (Samuelson, 1998).

Prior studies found evidence that demand curves are downward sloping and that non-fundamental trading can cause prices to deviate from fundamentals. For instance, stock inclusions and exclusions cause non-fundamental price movements,[1] and so does the trading by mutual funds (Coval and Stafford, 2007).[2] When price-pressure is correlated with characteristics or price history, it may also create systematic comovement (Barberis, Shleifer, and Wurgler, 2005; Lou, 2012; Ben-David, Franzoni, Moussawi, and Sedunov, 2020). Koijen and Yogo (2019) propose that institutions have characteristics-based demand for securities and that this demand shapes common return factors. Despite the evidence about demand setting prices, it is a challenge to identify a *source* of non-fundamental demand and therefore test whether that demand causes systematic return patterns. Such a link is indispensable to precluding alternative explanations related to unobservable fundamental factors, e.g., time-varying risk aversion.

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-related price pressures in the stocks owned by the mutual funds. We identify the link between capital flows and style-level stock returns by exploiting a reform in Morningstar's

[1] See Harris and Gurel (1986), Shleifer (1986), Wurgler and Zhuravskaya (2002), and Chang, Hong, and Liskovich (2015).

[2] Other examples include Ben-David, Franzoni, and Moussawi (2018), Brown, Davies, and Ringgenberg (2018), and Huang, Song, and Xiang (2020).

rating methodology that took place in June 2002, in which mutual funds were rated against their peers that have the same investment styles, instead of against all mutual funds. The reform caused massive reallocation of ratings across funds, and as a consequence, a reallocation of capital flows across styles. We show that this non-fundamental shock to demand reshaped several systematic properties of the stock market, including style-level momentum and reversals and return variation across styles. Furthermore, Morningstar's reform impacted the properties of the size and value factors, which are considered by many economists as reflecting risk attitudes of investors. Our results demonstrate that correlated non-fundamental demand has a first-order causal impact on forming systematic return patterns in the equity market.

Mutual funds are the primary conduit for equity investment by U.S. households, and therefore channel most of their capital flows to the equity market. When our sample begins in 1991, equity mutual funds held about 10% of the U.S. stock market capitalization. Their ownership has tripled to about 30% by 2005 and remained steady since then. Hence, capital inflows and outflows are large enough to generate systematic price patterns. Indeed, Coval and Stafford (2007) document that fire sales by individual mutual funds lead to price pressure in the underlying stocks, and Lou (2012) documents that mutual flows predict momentum returns. Furthermore, capital flows from and to mutual funds are plausibly driven by non-fundamental factors. Prior literature has also found that mutual fund investors, around 90% of whom are retail, exhibit unsophisticated behavior.[3]

The reform of Morningstar ratings is an ideal experiment to evaluate the causal effect of demand on systematic return patterns. Throughout the sample period, mutual investors heavily rely on Morningstar ratings to guide the allocation of capital. In point of fact, Ben-David, Li, Rossi, and Song (2019) find that Morningstar rating is the most important driver of mutual fund flows among all factors hitherto studied. Morningstar ratings, however, have different interpretation before June 2002, and after that date. In other words, while the demand of ratings by investors remained unchanged, the allocation of flows to mutual funds,

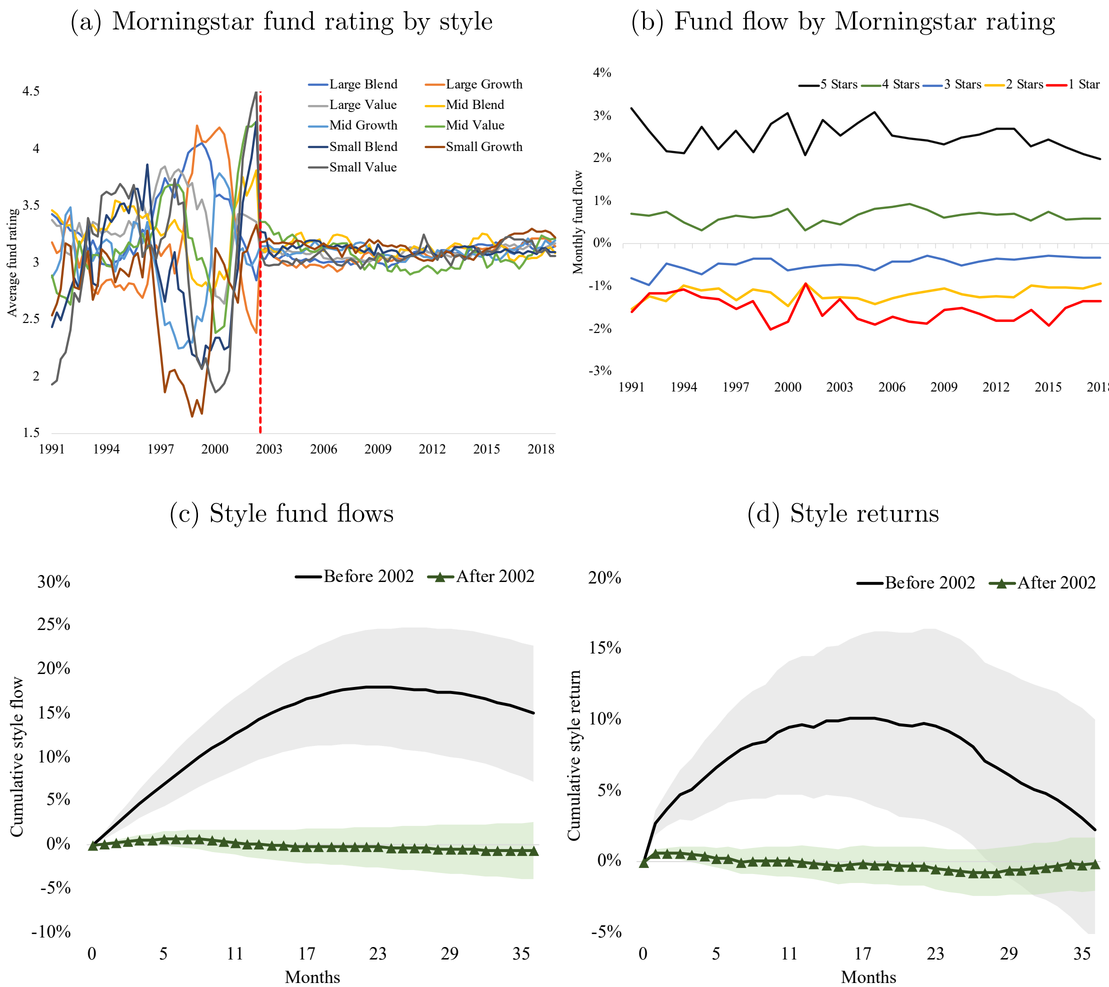
[3]For example, they allocate money to funds that subsequently underperform (Frazzini and Lamont, 2008; Song, 2020), invest in high-fee funds (Barber, Odean, and Zheng, 2005; Choi and Robertson, 2018), time the market poorly (Akbas, Armstrong, Sorescu, and Subrahmanyam, 2015; Friesen and Nguyen, 2018),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

and therefore to the underlying securities, was altered dramatically around June 2002.

Prior to June 2002, Morningstar ratings were broadly aligned with mutual funds' past performance over horizons of three to ten years. 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 fees and volatility. Since mutual funds are classified by nine investment styles (value/blend/growth $\times$ small/mid/large), funds within the same style had similar past returns, and as a consequence, similar ratings.

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 lines represent the June 2002 methodology change event. Panel (b) plots the average monthly fund flow by one to five star Morningstar ratings. In Panels (c) and (d), we sort the 3×3 style portfolios by their lagged rating changes (ExpSum(ΔRating), defined in Section 3.3), and then plot the cumulative fund flow and returns of those portfolios for the subsequent three years. The shaded areas are 95% bootstrapped confidence intervals.



In June 2002, Morningstar implemented an innocuous, yet impactful, change in its rating methodology. Instead of simply ranking all equity funds against each other, Morningstar began assigning ratings based on how funds rank *within* their styles. By design, fund ratings immediately became balanced across styles, with over half of mutual funds having their ratings being modified in June 2002. Panel (a) of Figure 1 shows how the dispersion of fund ratings across styles suddenly collapsed in June 2002 as a consequence of the methodology change.

Importantly for our identification, we show that investors continued to chase ratings regardless of the rating methodology.[4] Panel (b) of Figure 1 plots a time series of the monthly flows per star-rating: Morningstar ratings have a stable and sizeable impact on mutual fund flows over time. For an intuitive explanation of the flow magnitude, over a two-year period, five-star funds approximately double their sizes due to inflows, while one-star funds shrink nearly by half due to outflows. In particular, *changes* in fund ratings cause a permanent changes in flows (similar evidence is in Del Guercio and Tkac, 2008; Reuter and Zitzewitz, 2015). Therefore, our main variable of interest is the recent changes (upgrades or downgrades) in the ratings of mutual funds, aggregated at the style-month level.

Morningstar's reform, combined with investors' fixation on ratings, led to a redistribution of capital flows, and hence demand pressures, across styles. Pre-June 2002, ratings-driven demand was concentrated in the styles with the best recent performance. Post-June 2002, ratings-driven demand became balanced across styles. Because concentrated flows can generate large price pressure in stocks, style-level pressures should be present prior to the methodology reform and would be alleviated after it.

To establish the causal effect of non-fundamental correlated demand on style returns,[5] we first perform an event study using the one-year time window surrounding the methodology change in June 2002. In the months leading to June 2002, funds in top-rated styles gathered inflows and the underlying stocks performed well in the following months. Accordingly, funds

[4]Ben-David et al. (2019) and Evans and Sun (2020) have similar observations. These findings are most consistent with the explanation that investors view Morningstar's ratings as a recommendation about the best funds from an independent expert. As in Gennaioli, Shleifer, and Vishny (2015), Morningstar's is considered by investors a trusted advisor.

[5]Prior work found that style-level capital flows are associated with future style returns (Teo and Woo, 2004; Wahal and Yavuz, 2013; Li, 2020).

in bottom-rated styles experienced outflows and the underlying stocks performed poorly in the following months. The methodology reform caused rating dispersion across styles to sharply collapse, and so did flow dispersion across styles. As predicted, style return dispersion also disappeared immediately after June 2002, confirming the causal influence of demand on style returns.

Next, we explore the impact of the methodology reform on the stock market over the full sample of 1991–2018. Panels (c) and (d) of Figure 1 trace the differences in flows and returns of the most upgraded minus most downgraded styles, as of $t = 0$. Panel (c) shows that before 2002, the most upgraded style attracted approximately 20% more flows than the most downgraded style over the proceeding 12 months. In contrast, the spread in flows to the best versus worst styles disappeared completely after June 2002, when ratings are evenly distributed across styles. Panel (d) shows that the spread in style return performance mirrors the rating-induced flows. Pre-2002, there is a strong style return momentum and a subsequent reversal due to the positive feedback between style flows and returns. Conversely, post-2002, despite investors continuing to chase Morningstar ratings (Figure 1, Panel (b)), there is no positive feedback and thus neither style momentum nor reversal in flows or returns.

As ratings across styles became more homogeneous across styles as a consequence of Morningstar's reform, so did flows and returns. The average monthly style-level flow spread dropped from 3.3% before June 2002 to 1.4% after June 2002. Comparing the year after June 2002 to the year before that date, the monthly style return spreads dropped from 6.2% to 3.5%. Over the entire sample period, the spread in monthly style returns dropped from 5.5% pre-June 2002 to 3.0% post-June 2002.

We close our study with an empirical analysis of the explanatory power of rating-induced demand on the fluctuations in the Fama-French size and value stock factors (Fama and French, 1993). These factors are essentially long-short portfolios built on styles and are considered by many financial economists as mimicking risks that investors care about (Zhang, 2005). We 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 have a clean variation in rating that is caused by the methodology change and, therefore, mitigate endogeneity concerns. Then, we use the estimated parameter to assess the effect of style ratings on style returns for the entire sample period of 1991–2018. Admittedly, this would be a crude estimate, yet it is informative about the potential impact of rating-induced demand on style returns. Our analysis shows that rating-induced price pressure can explain 12% to 29% of the variation in monthly factor returns before 2002.[6] As expected, the explanatory power dropped precipitously after 2002.

To summarize, our analysis shows that a sizeable fraction of the common variation in stock returns can be attributed to non-fundamental correlated demand. We document that a seemingly-irrelevant reform in one rating firm created a long-lasting impact on the allocation of investors' capital across styles. As a consequence, the reform altered the cross-sectional variation of style returns, style-level momentum, and widely-used return factors. Our work identifies a non-fundamental permanent shift to investors' demand that transmuted the landscape of the equity market.

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 and shows that investors chase Morningstar rating to a similar extent after 2002. Section 4 establishes the casual effects of correlated demand on style returns through an event study approach and demonstrates that style return dynamics change dramatically since 2002. Section 5 quantifies the influence of correlated demand on the size and value factors. Section 6 concludes. Additional results and robustness checks are provided in the appendices.

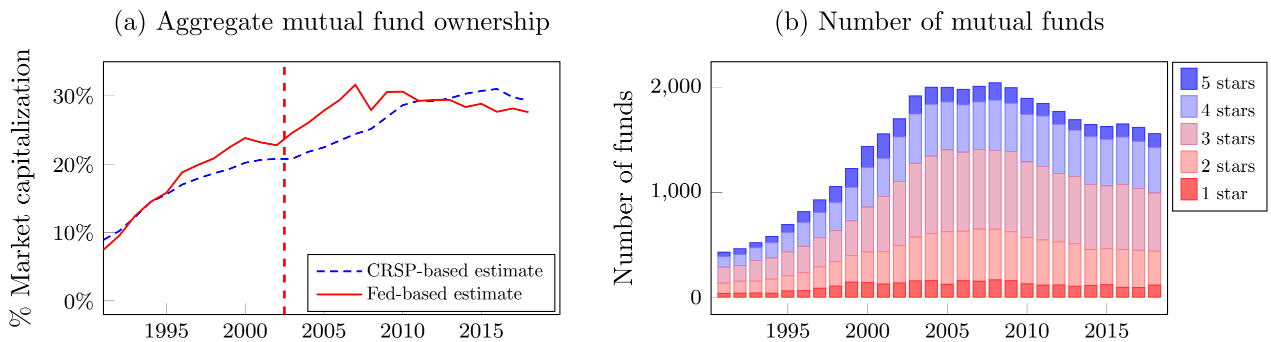
2 Data and Variable Construction

In this section, we describe the data set and explain how we measure the key variables: stock- and style-level ratings and flows.

[6] Because ratings are persistent, the explanatory power of rating-induced price pressure rises to an average of 40% at the quarterly frequency.

Figure 2. Summary Statistics of Mutual Funds

Panel (a) shows the aggregate U.S. stock holdings by domestic mutual funds as a fraction of the total U.S. stock 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.



2.1 Mutual Fund Data

Mutual funds are one of the largest classes of equity investors. Panel (a) of Figure 2 demonstrates their prominence in the U.S. equity market. Mutual funds held about 10% of the U.S. equity market in the early 1990s, and that rose steadily to around 30% after 2005.

Our analysis of the impact of Morningstar ratings on stock returns relies on detailed mutual fund-level data. We use monthly data from 1991 to 2018. The start year of 1991 is based on data availability: monthly fund flows from the Center for Research in Security Prices (CRSP) start in 1990, and some measures require one year of lagged data to construct.

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, which is based on 13F filings.

Following the fund flow 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 the lagged TNA:

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

We download Morningstar ratings and style categories from Morningstar Direct and merge them with the CRSP mutual fund data using the matching table from Pástor, Stambaugh, and Taylor (2015).[7] Morningstar assigns ratings at the share class level. We follow Barber, Huang, and Odean (2016) to 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. 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.

Panel (b) of Figure 2 summarizes the time series of the number of funds and their average size in our sample. The number of funds quadrupled from 1991 to 2005, and then plateaued and slightly declined from 2009 onward. Additional summary statistics are provided in Appendix A.1.

2.2 Stock- and Style-Level Rating and Flow-Induced Trading

Because the main focus on this study is on price pressure in stocks and styles, we summarize ratings, changes in ratings, and flows at the stock and style levels. We define the level and change of Morningstar rating of stock i in month t as a holding-weighted average of all funds that hold the stock:

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

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

We later drop the superscript "stock" when unambiguous.

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:[8]

$$\text{FIT}_{i,t} = \frac{\sum_{\text{fund j}} \text{SharesHeld}_{j,t-1} \cdot \text{Flow}_{j,t}}{\sum_{\text{fund j}} \text{SharesHeld}_{j,t-1}}. \tag{4}$$

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

[8] 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.

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 is likely related to fundamentals, FIT isolates the non-discretionary trading that is only attributable to fund flows and thus likely does not contain information. Consistent with this interpretation, Lou finds that FIT leads to price pressures that revert over time.[9]

Having computed stock-level ratings and flows, we then aggregate them up to the style-level. For a given style π, we calculate

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

$$\text{FIT}^{\text{style}}_{\pi,t} = \sum_{i \in \text{style } \pi} w^{\pi}_{i,t-1} \cdot \text{FIT}^{\text{stock}}_{i,t}, \tag{6}$$

where $w^{\pi}_{i,t-1}$ is the market cap weight of stock i in the corresponding style. We omit the superscript "style" when unambiguous.

3 Empirical Setting

In this section, we describe the simple, yet radical, methodological change in the popular Morningstar star rating system that took place in June 2002. We then show that investors continued to rely on the ratings with similar intensity both before and after the change. Because ratings are assigned within-styles after 2002, rating-induced correlated demand at the style level largely disappeared after 2002. We then present evidence that rating-induced flows to funds exert a large price impact on the underlying stocks. In later sections, we demonstrate that this exogenous rating change has a far-reaching impact on style return dynamics.

[9] Wardlaw (2019) 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.

3.1 Morningstar Ratings Methodology Pre- and Post-2002

After launching its mutual fund rating system in 1985, Morningstar quickly became the industry leader in guiding investor 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 return of funds and conducts minor adjustments for return volatility and expenses. Depending on a fund's age, the lookback horizon for past performance can be 3, 5, or 10 years, and more weight is applied to the most recent 3 years of returns. For funds with ten or more years of history, the weights of the three horizons are set at 20%, 30%, and 50%, respectively.[10] Then, Morningstar ranks funds by their performance and assigns 1 to 5 star ratings with fixed proportions (10%, 22.5%, 35%, 22.5%, and 10%).

Morningstar's 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. Initially 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 dotcom crash, many fund managers specializing in large/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.[11]

The main novelty in the post-June 2002 methodology is that ratings are based on fund rankings *within* style categories. Morningstar classifies diversified U.S. equity funds into the well-known 3×3 style matrix based on funds' holdings: combinations of value/blend/growth and small/midcap/large. For simplicity, our analysis focuses on ratings and flows of diversified U.S. equity funds, which constituted 87% of all mutual funds in 2002. Morningstar

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

[11] We learned this during phone conversations with Morningstar management. Making ratings more balanced across styles was also one of the stated objectives for this methodology change. 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.

also categorizes sector funds—the remaining 13% of funds—into 12 industrial sectors (e.g., financial, utilities). Since June 2002, the distribution of star ratings is the same for funds in each comparison group: either one of the 3×3 styles for diversified funds or one of the 12 industrial sectors for sector funds. The modified methodology was announced in early April 2002[12] and was implemented at the end of June 2002. Appendix B provides a detailed description of the pre- and post-2002 methodologies as well as a mapping of the holdings of funds in each of the styles to size and value spectra.

Before the change, fund ratings differed dramatically across styles based on recent style performance, as shown in Panel (a) of Figure 1 in the Introduction. For instance, in 2000, at the peak of the dotcom boom, large-cap growth funds had an average rating of 4 stars, while small-cap value funds only had 1.9 stars. After the change, ratings became uncorrelated with past style performance by construction, and the rating imbalance across styles became negligible. As a result, 54% of mutual funds experienced a change in rating at the end of June 2002, compared to an average of 14% of funds experiencing a rating change in other months.

3.2 Rating-Chasing Behavior Persists Before and After 2002

As depicted in Panel (a) of Figure 1, ratings changed dramatically in June 2002. Important for our identification purposes, the abrupt change only happened in the way in which ratings were distributed across styles, but not in the way in which fund flows respond to ratings. In this section, we formally show that investors chase Morningstar ratings to a similar extent both before and after June 2002.[13]

We start by presenting observational evidence. Panel (b) of Figure 1 simply 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 important as it implies that the AUM of a five-star fund increases by about 40% over one year and doubles in two years. In contrast, one-star funds

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

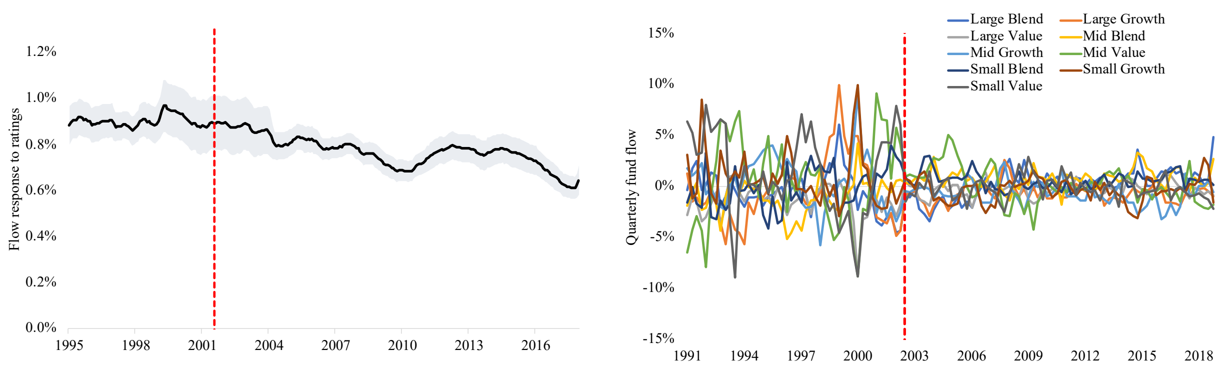
[13] Evans and Sun (2020) also provide evidence that fund flows respond strongly to Morningstar ratings, irrespective of the methodology used to assign the ratings.

Figure 3. The June 2002 Morningstar Methodology Change

This figure plots the time variation of relevant quantities over the full sample. The vertical dashed red lines mark the June 2002 methodology change event. Panel (a) explores the stability of the relationship between ratings and fund flows. Specifically, it plots the regression coefficient of fund flows on lagged ratings estimated using five year rolling windows. The shaded area is the two standard error band. Panel (b) plots the TNA-weighted average quarterly fund flow by Morningstar 3×3 styles. The flows are demeaned cross-sectionally to focus on the dispersion.

(a) Regression coefficient of fund flows on ratings

(b) Fund flow by styles



experience outflows of -1.5% to -2.0% of their AUM per month on average. Importantly, flows in and out of differently-rated funds do not appear to change before and after June 2002.

More formally, we estimate the response of fund flows to lagged fund ratings using a five-year rolling-window TNA-weighted Fama-MacBeth (Fama and MacBeth, 1973) regression. The results are plotted in Panel (a) of Figure 3. The coefficient estimate only varies slightly over the sample and there is no abrupt change before and after 2002. For example, the average flow-to-rating response was 0.90% before June 2002 and 0.77% after June 2002.

Ben-David et al. (2019) conduct additional tests that demonstrate the same point. For example, in a regression framework, they show that the marginal R^2 of ratings in explaining fund flows is nearly identical before and after the 2002 rating reform. Their results control for lagged fund returns and style fixed effects. On the other hand, once ratings are controlled for, style fixed effects lose most of their marginal explanatory power after 2002. This is consistent with the view that much of style-level flows are driven by the style-level rating variation, which largely disappeared after 2002.

In summary, these results indicate that mutual fund investors keep chasing Morningstar ratings regardless of the rating methodology. Because ratings are constructed within-styles after June 2002, style-level fund flow dispersion dropped after the methodology reform, as is easily visible in Panel (b) of Figure 3. In other words, the style-level correlated demand due to rating-chasing behaviors mostly disappeared since 2002.

3.3 Rating-Induced Price Pressures

Next, we confirm that Morningstar ratings can exert a large price impact on stocks through flow-induced trading. This is a preliminary stage before exploring the influence of rating-induced style demand on style returns in the following analysis.

We first note that aggregate mutual fund flows are large enough to generate a non-negligible price impact. Panel (a) of Figure 2 shows that mutual funds, as a prime investment vehicle for retail investors, hold a substantial and increasing share of the U.S. equity market. 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. These numbers grew steadily over time, and by 2018, the end of our sample period, equity mutual funds owned \$10,849 billion, which represented 29.3% of the entire market capitalization.

We use Fama-MacBeth regressions to estimate the chain of dynamic effects: *i*) the response of fund flows to Morningstar rating changes, and *ii*) the response of stock returns to flow-induced trading. All regressions are value-weighted: Fund-level regressions are weighted by lagged fund TNA, and stock-level regressions are weighted by lagged stock market capitalization.[14]

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

$$\text{Flow}_{j,t} = a + b_1 \cdot \Delta\text{Rating}_{j,t-1} + ... + b_{36} \cdot \Delta\text{Rating}_{j,t-36} + X_{j,t} + u_{j,t}, \tag{7}$$

where $\Delta\text{Rating}_{j,t}$ is the month t rating change of fund j, and controls $X_{i,t}$ include 36 monthly lags of fund flows and returns. The cumulative response coefficients ($b_1, b_1+b_2, ...$) are plotted in Panel (a) of Figure 4. In response to a one-star change in fund rating, funds experience

[14] To account for the growth of total market size over time, we re-normalize the weights by period. For instance, the weight of a stock-month equals the fraction of the total market cap it represents in that month.

an average of 6% additional flows, most of which take place within 24 months. This result is consistent with Del Guercio and Tkac (2008) and Del Guercio and Reuter (2014).

Second, we estimate the response of stock returns to the stock-level flow-induced trading (FIT), as defined in Equation (4):

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

We plot the cumulative response in Panel (b) of Figure 4. Each 1% increase in mutual fund ownership due to FIT leads to immediate price pressures of approximately 0.6% in the contemporaneous month, which gradually reverts 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 can expect that stock returns will also be affected by rating changes, particularly by more recent rating changes.[15] To facilitate our later analysis of rating-induced price impact, it is convenient to summarize recent rating changes into a weighted average sum where the weights correspond to how much each lagged rating change 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 4. As expected, more recent rating changes are more impactful and the coefficients converge towards zero with the horizon.

Since the impact primarily happens within 12 months, we summarize past rating changes using the following weighted sum:

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

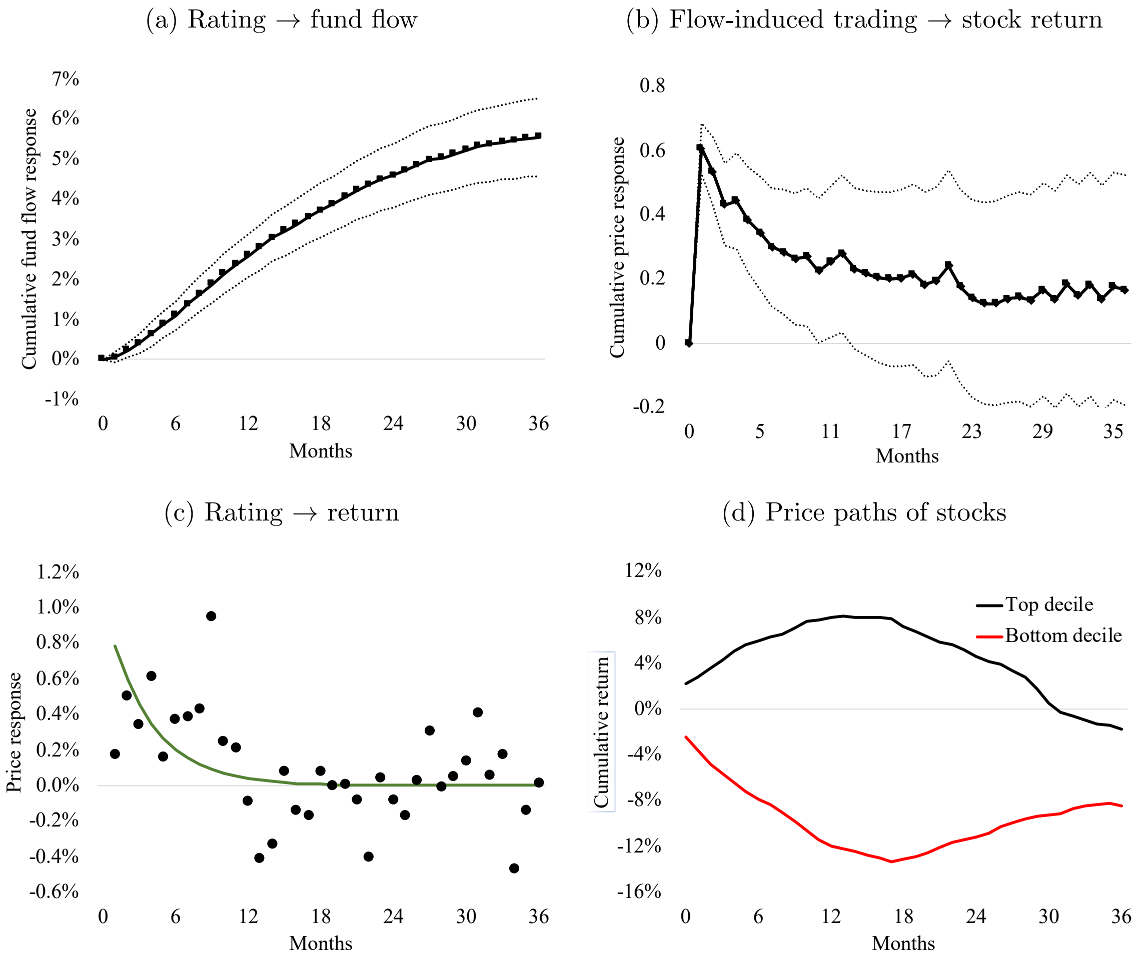
where $\sum_{k=1}^{12} \tau_k = 12$ and the weights decay with factor $\delta = 0.764$, which is estimated from a least-squares fit to the cumulative response (Panel (c) of Figure 4).[16] Because the weights sum to 12 (months), ExpSum(ΔRating) should be interpreted as the rating change over one

[15]The impact of more distant rating changes, such as those 24 months ago, should be weaker. While those rating changes may continue to generate flows (Figure 4, Panel (a)), the price pressures generated by their earlier impact are already reverting, so the two effects will partially cancel each other out.

[16]Therefore, $\tau_k = \frac{12\cdot(1-\delta)}{1-\delta^{12}} \cdot \delta^{k-1}$.

Figure 4. Price Impact of Rating and Flows

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



year. The estimated decay factor $\delta = 0.764$ implies a half-life of $-\ln(2)/\ln(\delta) \approx 2.58$ months. Our results are insensitive 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 4, we sort all stocks into deciles portfolios each month based on $\text{ExpSum}(\Delta\text{Rating})_{i,t-1}$, and track the performance over the next three years. Stocks in the top decile of past rating changes outperform stocks in the bottom decile by about 20% over the next 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 large price pressure on stocks.

4 Ratings, Flows, and Style Returns: Pre- vs Post-2002

In this section, we demonstrate that the dynamics of style returns changed dramatically after 2002 due to the exogenous rating methodology change. We first use an event-study approach with a short-window around June 2002 to establish the causal effects of rating-induced correlated demand on style returns. Then we show that style return momentum, reversal, and volatility experienced substantial changes after June 2002.

Because we are interested in examining the impact of Morningstar ratings, we use Morningstar's fund categories to define style portfolios.[17] For instance, the large-cap growth style portfolio is defined by the AUM-weighted aggregate holdings of all funds in that category.

In Appendix A.2, we compare the holdings of fund-based style portfolios against the academic definition of style portfolios. One can see that this fund-based style definition can be thought of as a "smoothed" version of the academic style definitions by, for instance, Fama and French (1993). In Section 5 we also show that the results based on Morningstar-defined styles also extend to the academic-defined styles.

4.1 Identification Using the 2002 Shock

To establish the causal effects of rating-induced correlated demand on style return dynamics, we use a one-year window around June 2002. The reform in Morningstar ratings can be traced down to the day (June 30$^{\text{th}}$, 2002), and therefore its impact should be traced

[17] Lettau, Ludvigson, and Manoel (2019) document that the holdings of mutual funds (and other investment vehicles) do not fully overlap with the stocks that make up risk factor portfolios as defined by financial economists.

around that date, or right around it. Although we do not have a control group, different subsets of the data have different sensitivities to Morningstar's methodology change. Hence, we measure the differential effect across portfolios, e.g., top rated style versus bottom rated. Furthermore, we repeat our tests in other years, as a falsification test.

To visualize style ratings, flows, and return variation in 2002, we sort the nine styles by their average lagged ratings during the six months before the event. We plot the evolution of style ratings, style flows, and style returns in Figure 5.

Panel (a), which plots average style ratings (demeaned cross-sectionally), shows a sharp methodology-induced drop exactly at the event. The top-rated style suffered a drop of about 0.4 stars, while the bottom-rated style experienced an increase of about 0.4 stars. This rating change is purely exogenous and is caused by the methodology change of Morningstar. Similar patterns can be observed when comparing the flows to the second- and penultimate-rated styles.

Consistent with flows chasing ratings, Panel (c) shows that the top-rated style experienced approximately 25% additional flows relative to the bottom-rated style during the six months before the event. However, right after the rating change in June 2002, the flow differences became muted.

The change in style returns is consistent with rating and flow changes exerting a causal effect. In Panel (e), we plot the cumulative style returns. The pre-event factor returns line up with pre-event rating but suffer slight reversals after the event. This phenomenon is clearest in the top-rated style: it experienced a staggering 15% return in the pre-event period, but that reversed subsequently. The bottom-rated style experienced a -12% return and also reversed after the rating methodology change.

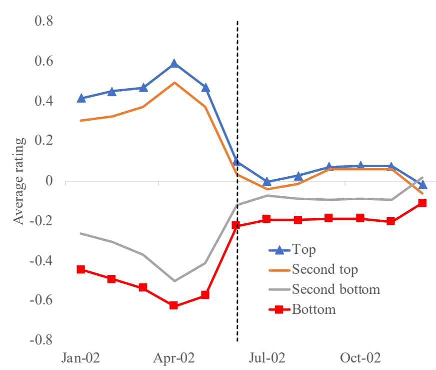
To alleviate the concern that the style return and flow changes could result from mean reversion (e.g., statistical mean-reversion as in Stambaugh, 1999), we also conduct a placebo test. Panels (b), (d), and (f) of Figure 5 show that the patterns observed in 2002 did not take place in other years. The orange bars show the same exercise in other years, together with the two standard error bands. Clearly, the sharp changes in style ratings, flows, and returns are unique to 2002.

In summary, this event study strongly indicates that rating-induced style demand has

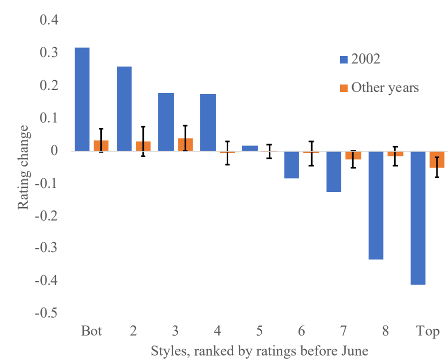
Figure 5. Event Study Around June 2002

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 average lagged ratings during the six months before the event and 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 in 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 2 standard error bands. To focus on cross-sectional dispersion, all variables—ratings, returns, and flows—are demeaned cross-sectionally.

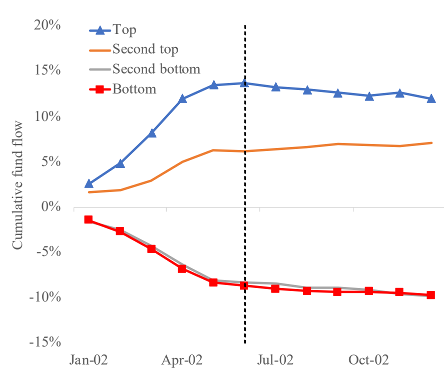
(a) Ratings, 2002



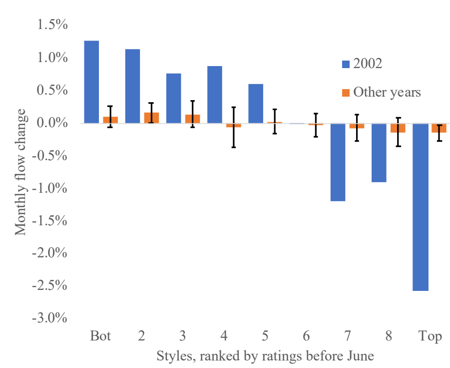
(b) Ratings, 2002 vs. placebo



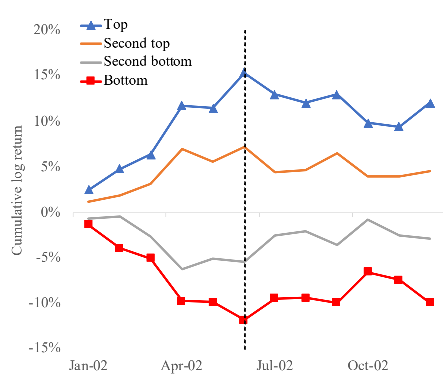
(c) Flow, 2002



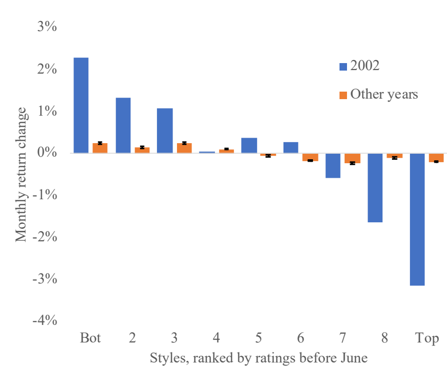
(d) Flow, 2002 vs. placebo



(e) Return, 2002



(f) Return, 2002 vs. placebo



a casual influence on style returns. As the effect of Morningstar ratings on fund flows is pervasive throughout the sample, in the next section, we will analyze the long-term effects of the dramatic shift in style-level demand on style return dynamics.

4.2 Style Dynamics Changed After 2002

In this section, we analyze the effect of the reform in Morningstar rating methodology on style-level flows and returns.

To illustrate the effects of Morningstar rating on style-level flows and returns, we first compare the performance of top and bottom styles sorted by past rating changes. Specifically, we summarize the stock-level rating changes for each style portfolio π:

$$\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1} = \sum_{\text{stock } i \in \text{ style } \pi} w_{i,t-1} \cdot \text{ExpSum}(\Delta\text{Rating})_{i,t-1}. \tag{10}$$

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

In Panel A of Table 1, we tabulate cumulative fund flows to funds in the top versus bottom styles over the next 36 months, and their difference. Before 2002, the top style experienced approximately 1% higher flows per month relative to the bottom style over the next 12 months. However, the differences in flows disappear after 2002. We observe similar patterns when comparing the three top and three bottom styles in Panel B.

Next, we show that these flow patterns translate into return patterns as well. In Table 1, we also examine the differences in returns between top and bottom style portfolios. Before June 2002, rating-induced style demand exerted large price pressure on style returns, which mostly dissipated after the rating system was changed. As a result, 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 after 36 months. Strikingly, the same long-short strategy generates zero return spread after June 2002. Again, we find similar patterns when comparing the top-three and bottom-three styles (Panel B).

The results in Table 1 and in Panels (c) and (d) of Figure 1 suggest that there exists a positive feedback loop between performance and ratings before June 2002. That is, high past

Table 1. Rating-Induced Price Pressures in Style Portfolios

We sort style portfolios using lagged exponential sum of rating changes ($\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$) and tabulate their monthly fund flow and return over the subsequent 36 months. Panel A shows the difference between the top and bottom styles, while Panel B shows the difference between the average of top three and bottom three styles. Bootstrapped standard errors are in the parentheses.

Panel A: Top 1 Minus Bottom 1 (%)

		Month 1 to 6	Month 7 to 12	Month 13 to 24	Month 25 to 36
Monthly Fund Flow (%)	Before 2002.06	1.14***	0.92***	0.38*	−0.25
		(0.33)	(0.28)	(0.23)	(0.19)
	After 2002.06	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 2002.06	0.76**	0.39	−0.04	−0.58***
		(0.31)	(0.35)	(0.22)	(0.22)
	After 2002.06	−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 (%)

		Month 1 to 6	Month 7 to 12	Month 13 to 24	Month 25 to 36
Monthly Fund Flow (%)	Before 2002.06	0.81***	0.66***	0.14	−0.14
		(0.22)	(0.19)	(0.16)	(0.09)
	After 2002.06	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 2002.06	0.47**	0.28	−0.10	−0.39***
		(0.21)	(0.22)	(0.17)	(0.13)
	After 2002.06	−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)

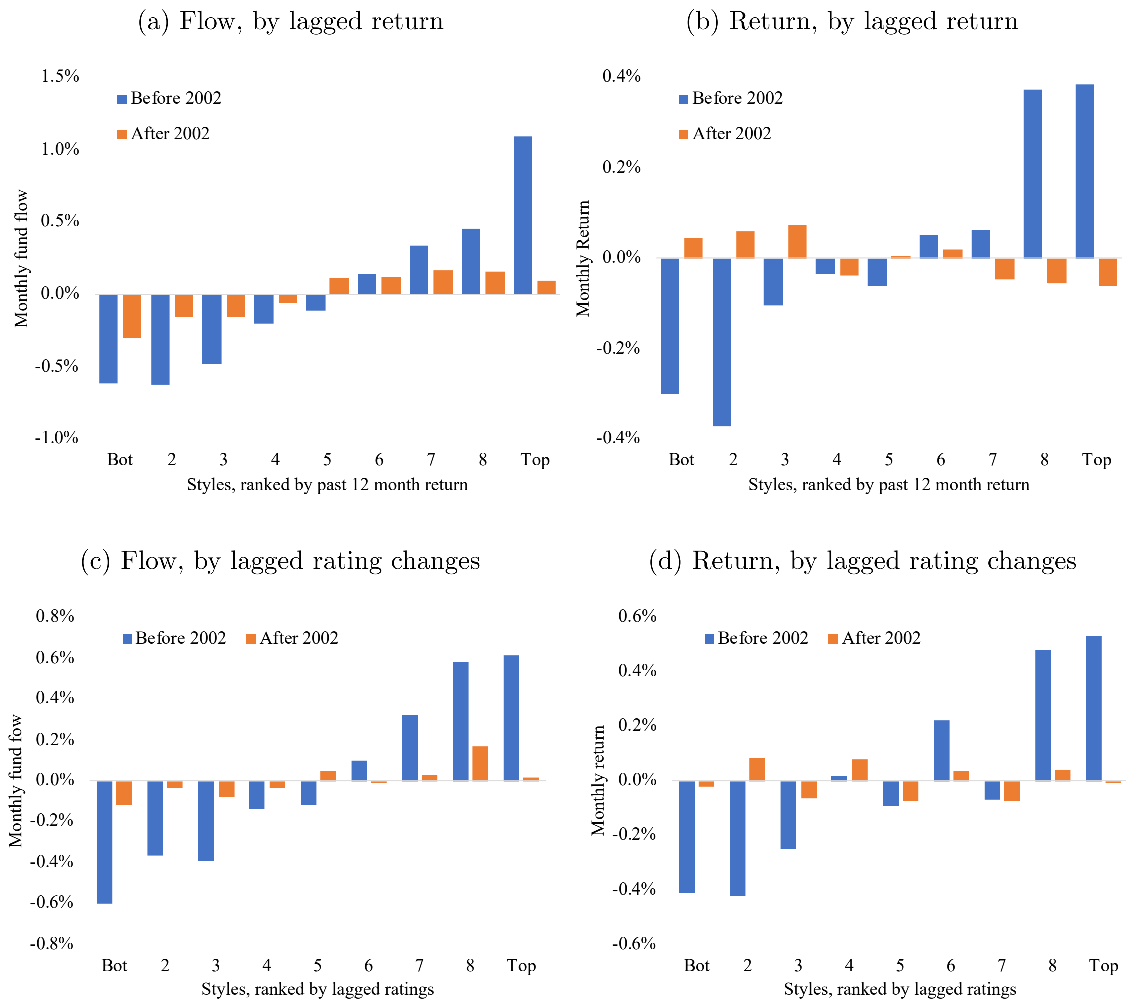
$^{***}p < 1\%$, $^{**}p < 5\%$, $^{*}p < 10\%$

performance of a particular style leads to rating upgrades, which in turn attracts inflows to the mutual funds in that style. The resulting price pressure further pushes up the price of the underlying stocks, which leads to even higher ratings, and so on. Due to the Morningstar methodology change, the positive feedback loop in style returns via the mutual fund channel was severed in June 2002. Thus, we expect that style momentum would become weaker or even muted after that date.

We find that this is indeed the case. Specifically, we analyze the long-short style momen-

Figure 6. Style Momentum Before and After 2002

This figure shows momentum strategy based on styles. We sort the 3×3 styles each month using lagged 12 month returns or lagged exponential sum of rating changes ($\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$). The bars plot the average monthly flows and returns of those style portfolios before and after June 2002.



tum strategy based on either past style-level rating changes or past 12-month style returns. Panels (a) and (c) of Figure 6 confirm that after June 2002, the flow spread between styles with high and low past return (rating) shrank significantly. Panels (b) and (d) show that the style momentum strategy was profitable before June 2002 with a monthly return of about 70 to 90 bps per month (top minus bottom styles). However, after June 2002, the style momentum strategy became entirely unprofitable.

4.3 Cross-Sectional Dispersion in Style Returns Shrank after 2002

Due to the sharp decline in the cross-sectional dispersion of style ratings after 2002, we also predict that the dispersion in style flows and returns should decrease.

We formally test this prediction in this section. We use two definitions of dispersion: the spread between the top and the bottom style, and the standard deviation across all nine 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, we also use two-year, four-years, and ten-year windows centered around the methodology change event.

In Panel A of Table 2, we first confirm that rating and flow dispersion across styles dropped sharply after 2002. This result is robust to the choice of time period around the event and is a formal confirmation of the patterns illustrated in earlier sections of the paper.

Crucially, column (5) of Table 2 shows that style return dispersion also dropped sharply after 2002. For example, within one year after the rating methodology change, the monthly return spread between the highest-return style and lowest-return style dropped by 2.77% per month (from 6.27% to 3.50%). Over the entire sample, the monthly return spread between top and bottom styles dropped by 2.54% after 2002 (from 5.48% to 2.94%). Column (6) further shows a significant drop of 0.89% in the monthly cross-sectional standard deviation of style returns within one year after 2002 (from 2.09% to 1.20%) and a drop of 0.90% over the entire sample period (from 1.90% to 1.01%).

To ensure that the result is robust to using alternative style definitions, we also conduct the same exercise for the nine "academic" styles defined by market capitalization and book-to-market ratio (Fama and French, 1993). In Panel B of Table 2, we observe similar drops in cross-sectional variations of style ratings, flow-induced trading,[18] and returns after the rating methodology change in 2002.

To summarize, the results in this section indicate that the rating-chasing behavior of mutual fund investors, together with the rating methodology change, have significantly changed how fund flows are distributed across investment styles. As a result, style return dynamics

[18] Unlike fund-based styles where we can directly measure fund flows, we aggregate up stock-level flow-induced trading in these academic style portfolios (Equation (6)).

Table 2. Dispersion of Styles Before and After June 2002

We regress dispersion measures of monthly rating, flow, and return 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. Panel A uses the styles defined by Morningstar and Panel B uses the academically-defined 3 × 3 styles. Across the different rows, we vary the sample size used in the regressions. For instance, the first specification only uses one year before to one year after the methodology change in June 2002.

	Regression coefficient on the post-June 2002 dummy					
Dependent variables:	Rating		Flow (%)		Return (%)	
	Spread	Std	Spread	Std	Spread	Std
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Fund-based Styles						
2001Q3–2003Q2	−0.37***	−0.15***	−2.42***	−0.90***	−2.77***	−0.89***
	(0.08)	(0.02)	(0.41)	(0.15)	(0.77)	(0.26)
2000Q3–2004Q2	−0.53***	−0.20***	−1.74***	−0.63***	−4.45***	−1.53***
	(0.06)	(0.02)	(0.29)	(0.10)	(0.81)	(0.28)
1997Q3–2007Q2	−0.84***	−0.32***	−1.81***	−0.62***	−4.83***	−1.69***
	(0.04)	(0.02)	(0.22)	(0.08)	(0.68)	(0.24)
Full sample	−0.61***	−0.22***	−1.88***	−0.60***	−2.54***	−0.90***
	(0.03)	(0.01)	(0.15)	(0.05)	(0.32)	(0.11)
Panel B: Academic Styles						
	Rating		Flow-Induced Trading (%)		Return (%)	
	Spread	Std	Spread	Std	Spread	Std
2001Q3–2003Q2	−0.23***	−0.06***	−1.41***	−0.51***	−2.62**	−0.95**
	(0.08)	(0.02)	(0.39)	(0.14)	(1.24)	(0.37)
2000Q3–2004Q2	−0.53***	−0.17***	−0.77***	−0.28***	−4.52***	−1.69***
	(0.08)	(0.03)	(0.24)	(0.08)	(1.00)	(0.33)
1997Q3–2007Q2	−0.95***	−0.34***	−0.75***	−0.26***	−5.47***	−1.94***
	(0.06)	(0.03)	(0.18)	(0.06)	(0.90)	(0.31)
Full sample	−0.57***	−0.21***	−0.74***	−0.24***	−2.50***	−0.87***
	(0.04)	(0.01)	(0.09)	(0.03)	(0.43)	(0.15)

***$p < 1\%$, **$p < 5\%$, *$p < 10\%$

changed dramatically after 2002.

5 Demand Effects on the Size and Value Factors

So far, we have focused primarily on styles based on Morningstar's definition. We have seen that rating-induced style-level demand creates large price pressures in the pre-June 2002 period. In this section, we turn the spotlight to styles as defined by financial economists: size, based on market capitalization, and value/growth, based on book-to-market ratio (Fama and French, 1993). We aim to answer another economically important question: how much variation in the Fama-French size and value factor returns can be attributed to rating-induced correlated demand?

To answer this question, we first use a short window around June 2002 to causally estimate the price impact of Morningstar ratings on style returns—the building blocks of the size and value factors. There are two reasons for choosing a short window. First, we want to ensure that rating changes primarily come from the methodology change.[19] Second, we want to avoid the impact of other market-level changes, such as the dotcom bubble burst in early 2000 or the "momentum crash" period in 2008 (Daniel and Moskowitz, 2016). Those events are not contained in our window of estimation.

We now estimate the price impact of rating-induced demand on style returns using the 2002 shock. Figure 7 illustrates the ratings and returns of the academic styles within this one year window when sorted on average pre-event ratings. The patterns are similar to those depicted in Figure 5, where style portfolios are instead based on Morningstar-defined styles. Formally, we estimate a panel regression using six months before to six months after the event:

$$\mathrm{Ret}_{\pi,t} = \lambda \cdot \mathrm{ExpSum}(\Delta\mathrm{Rating})_{\pi,t-1} + X_{\pi,t-1} + \epsilon_{\pi,t}, \tag{11}$$

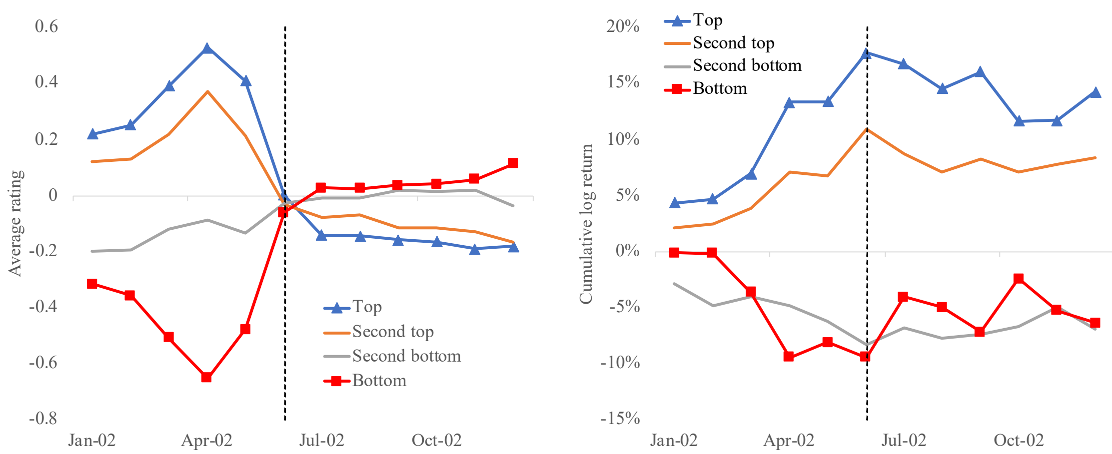
where the controls $X_{\pi,t-1}$ include style returns over $t-1$, $t-2$ to $t-6$, and $t-7$ to $t-12$ months as well as style and time fixed effects. By controlling for past returns, we can better capture

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

Figure 7. Behavior of Academic Styles around the June 2002 Event

We perform event studies on the 3×3 size-value stock styles during the 6 months before and after the June 2002 methodology change. The styles are sorted by their average rating during the 6 months before the event. Note that these style portfolios use the standard academic definition by sorting on size and value stock characteristics. To focus on cross-sectional dispersion, all variables are demeaned cross-sectionally.

(a) Ratings, 2002 (b) Returns, 2002



the marginal effect of ratings in addition to possible style momentum effects (Gupta and Kelly, 2019; Ehsani and Linnainmaa, 2019). To account for the cross-sectional style return correlation, we adjust the standard errors using a feasible generalized least squares (FGLS) approach. Specifically, we use the full sample of style returns to estimate the covariance matrix of style returns and incorporate it into the estimation.[20]

The estimation results are shown in Table 3. For each star rating change, the style-level price impact is 2.88% per month with a t-statistic of 3.91 after controlling for style and time

[20] Let y be the vector of style returns stacked together so that the first 9 entries are the first month, the next 9 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 covariance matrix of the 9 styles. Let X denote the matrix of independent variables. Then, we estimate the regression coefficients and covariance using

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

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

We estimate the rating price impact coefficient λ through a forecasting panel regression of monthly returns of the 3×3 (academic) stock styles on lagged rating changes ($\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$). The sample spans the six months before to six months after the methodology change. The four specifications differ only by the inclusion of style and/or time fixed effects.

Dependent variable:	Monthly style return $\text{Ret}_{\pi,t}$(%)			
	(1)	(2)	(3)	(4)
$\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$	2.881***	2.875***	2.594***	2.613***
	(0.736)	(0.732)	(0.703)	(0.699)
Past Return Controls	Yes	Yes	Yes	Yes
Style FE	Yes	Yes	No	No
Time FE	Yes	No	Yes	No
Observations	108	108	108	108
Adj R^2	39.9%	30.9%	21.1%	11.7%

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

fixed effects. The result is 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 Table A.2 of Appendix A.3. In the analyses that follow, we use the estimated $\lambda = 2.88\%$ in column (1) to quantify the influences of rating-induced flows on style returns.

To quantify the influence of rating-induced correlated demand on factor returns, we use the following price-impact specification:

$$\text{Ret}_{\text{SMB},t} = \underbrace{\lambda \cdot \text{ExpSum}(\Delta\text{Rating})_{\text{SMB},t-1}}_{\text{Rating-induced price pressure}} + \text{Ret}^{\text{counterfactual}}_{\text{SMB},t} \tag{12}$$

and

$$\text{Ret}_{\text{HML},t} = \underbrace{\lambda \cdot \text{ExpSum}(\Delta\text{Rating})_{\text{HML},t-1}}_{\text{Rating-induced price pressure}} + \text{Ret}^{\text{counterfactual}}_{\text{HML},t}, \tag{13}$$

where

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

and

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

In the equations above, we use $\mathcal{S}$, $\mathcal{L}$, $\mathcal{V}$, 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 example, $\mathcal{S}$ combines the value-small, blends-small, and growth-small.

To visualize the influence of rating-induced price pressure on the factors, in Panels (a) and (b) of Figure 8, 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 (SVMLG) that is long the small-value style and short the large-growth style. It is clear that rating-induced price pressures can explain a large portion of factor return variation before 2002. After 2002, rating-induced demand largely lost explanatory power as expected. Noticeably, most of the size and value premia were also realized before 2002.

To quantify the explanatory power on factor return variation, we compute the modified "R-squared" of rating-induced price pressures where we use the cleanly-identified λ estimate in Equation (11):

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

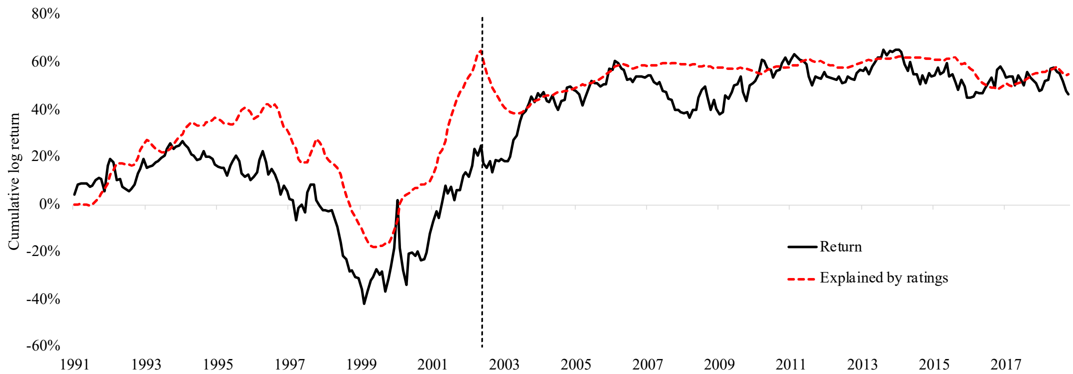
We compute this "R-squared" measure for both before and after 2002. Before 2002, we find that R-squared$^{\text{SMB}}$ = 23%, R-squared$^{\text{HML}}$ = 12%, and R-squared$^{\text{SVMLG}}$ = 29%. Because the explanatory variable $\text{ExpSum}(\Delta\text{Rating})_{t-1}$ is persistent, "R-squared" further rises to an average of 40% among these factors at the quarterly frequency. After 2002, these figures drop to 2%, 9%, and 9%, respectively.

While this exercise admittedly delivers a crude estimate, we find that correlated demand of styles can explain a significant fraction of factor return variation. The drop of explanatory power after 2002 further adds validity to our interpretation.

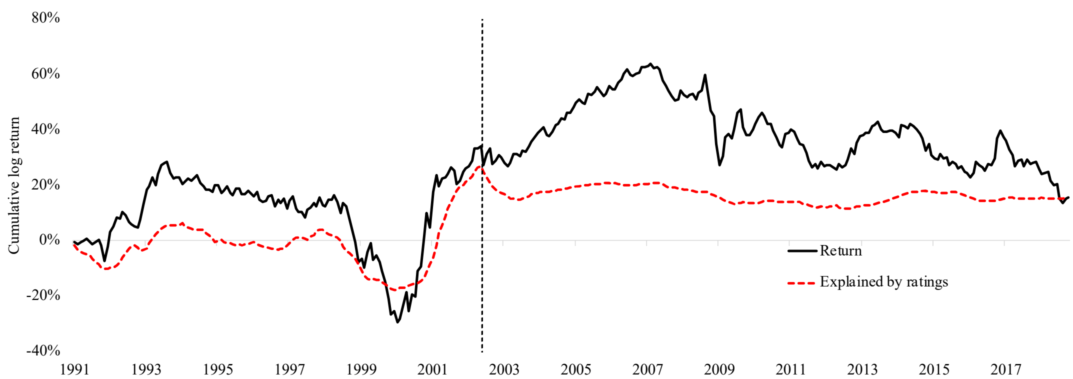
Figure 8. Explanatory Power of Ratings on Size and Value Factors

We quantify the explanatory power of rating pressures 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”), while Panel (c) plots the small cap-value style minus the large cap-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.

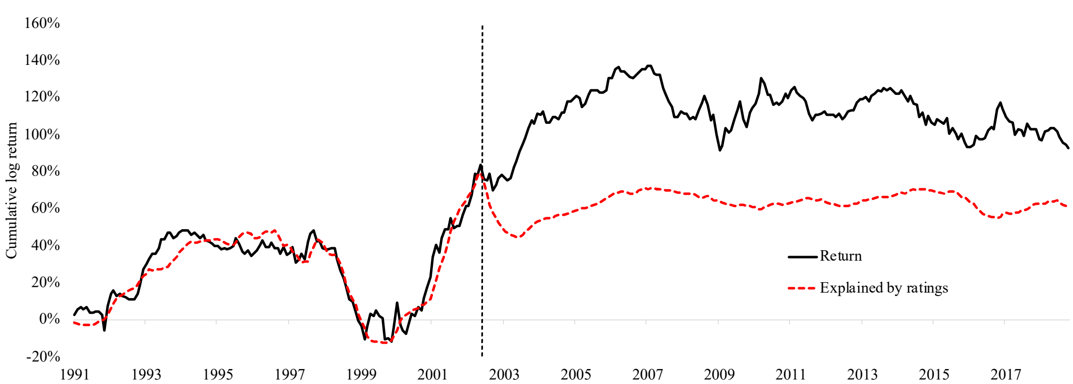
(a) Small Cap − Large Cap



(b) Value − Growth



(c) Small Value − Large Growth



6 Conclusion

In recent years, there is mounting evidence that price fluctuations may be impacted by price pressure due to investor demand. It is difficult, however, to identify the source of the demand and determine the extent to which it is independent of risk attitudes and whether it is the cause of observed systematic fluctuations in prices.

Our study shows that an innocuous modification to a rating methodology of mutual funds drove non-fundamental demand and had far-reaching implications to the underlying equity market. The reform in Morningstar ratings equalized ratings across styles. Prior to the reform, style portfolios displayed momentum and reversal in a pattern that matched the pattern of flows to a great extent. After the reform, these persistent patterns disappeared. Furthermore, the dispersion in style-level ratings, flows, and stock returns declined sharply and materially following the reform, as expected. Importantly, we can pinpoint the effects of ratings to style-level flows and returns the specific date—June 2002.

We use our estimates from June 2002 for the effects of rating-driven flows on the performance of style portfolios and explore the other implications. We show that pre-2002, the time-series variation in the Fama-French size and book-to-market factors can be explained well by the Morningstar ratings of the funds that hold the factor portfolio stocks. In contrast, the correlation dissipated in later years, as expected.

Overall, our results show that one specific source of non-fundamental demand—rating-chasing regardless of content—can drive significant fluctuations in style-level performance. These findings can alter the way in which the performance of factor portfolios are viewed by economists: instead of reflecting some unobservable risks, they may be driven by non-fundamental demand.

It is possible that the role of correlated demand in determining asset pricing is even greater than what is documented here. Our empirical setting allows us to identify one specific source of non-fundamental demand—mutual fund rating-chasing irrespective of the determinants of the ratings. This single source of non-fundamental demand drove meaningful fluctuations in style-level performance in the 1990s and until June 2002. Correlated demand, however, can arise from sources other than Morningstar ratings, such as institutional demand for certain styles driven by inertia (Froot and Teo, 2008; Koijen and Yogo, 2019) or performance-chasing in index-linked products (Broman, 2016). In totality, these findings should alter the way in which the performance of factor portfolios are viewed by economists: instead of reflecting some unobservable risks, they may be driven 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.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler, 2005, Comovement, *Journal of Financial Economics* 75, 283–317.

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, 2018, ETF arbitrage and return predictability, Working paper, University of Utah.

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, 2018, What matters to individual investors? Evidence from the horse's mouth, Working paper, Yale University.

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.

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

Del Guercio, Diane, and Jonathan Reuter, 2014, Mutual fund performance and the incentive to generate alpha, *Journal of Finance* 69, 1673–1704.

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, Technical report, 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.

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* forthcoming.

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.

Li, Jiacui, 2020, What drives the size and value factors?, Working paper, University of Utah.

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

Pástor, Ľuboš, Robert F. Stambaugh, and Lucian A. Taylor, 2015, Scale and skill in active management, *Journal of Financial Economics* 116, 23–45.

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* Forthcoming.

Stambaugh, Robert F, 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.

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

Wahal, Sunil, and M. Deniz Yavuz, 2013, Style investing, comovement and return predictability, *Journal of Financial Economics* 107, 136–154.

Wardlaw, Malcolm, 2019, Measuring mutual fund flow pressure as shock to stock returns, Working paper, University of Georgia.

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.

Zhang, Lu, 2005, The value premium, *Journal of Finance* 60, 67–103.

Appendix A Additional Results

A.1 Detailed statistics of our sample

Table A.1 shows the detailed statistics of our mutual fund sample from 1991 to 2018. There were 433 mutual funds included at the beginning of our sample, and the number peaked in 2008. Since then, the number of funds decreases a bit, yet the total assets managed keep on increasing to about 4 trillion in 2018. Columns (4) to (8) report the distribution of funds in each rating category, and columns (9) to (13) report the fraction of funds in different styles.

A.2 Morningstar Style Classification versus Characteristics

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 little value stocks as 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-to-market ratios into 10×10 portfolios using NYSE breakpoints. The heatmaps in Panel (a) show the academic style definitions, which are strictly based on stock characteristics. The stocks in those style portfolios are concentrated in a "rectangular region" by construction. 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 less, stocks with other characteristics.

A.3 Additional Empirical Results

Table A.2 estimates the price-pressure coefficient in Equation (11) with two alternative time windows. The results do not change materially if we decrease or increase the length of the estimation window.

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 each Morningstar rating. Note that this can differ from (10%, 22.5%, 35%, 22.5%, 10%) because Morningstar assigns those fixed fraction of ratings at the share-class level, while we follow Barber et al. (2016) to aggregate 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. For the other U.S. domestic equity funds which are classified as diversified, they are classified into the 3×3 style box categories, and columns (10) to (13) indicate the fraction of funds in different styles.

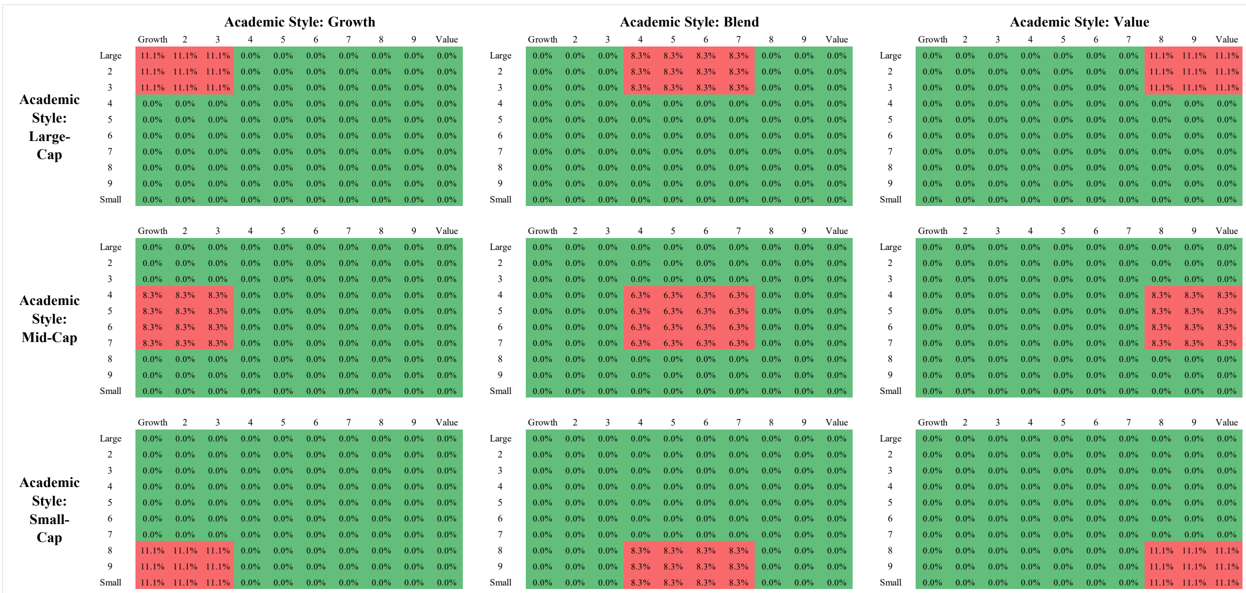
Year	Number funds	AUM ($ million)	Fraction by rating					Sector funds	Diversified fund style			
			1 star	2	3	4	5 star		Large	Small	Growth	Value
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1991	433	458.2	9%	23%	36%	23%	10%	19%	51%	16%	30%	28%
1992	466	600.8	9%	25%	32%	23%	11%	18%	52%	17%	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%

Figure A.2 reproduces panels (c) and (d) in Figure 1 with academic-defined styles. While the patterns are similar, results based on Morningstar-defined style are slightly sharper than academic-defined styles, consistent with that the change happened directly on the Morningstar styles.

Figure A.1. Comparison of Fund-based and Academic Stock Style Definitions

We sort stocks into 10×10 portfolios based on NYSE size and book-to-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.

(a) Academic Style Definition



(b) Morningstar Fund-based Style Definition

Table A.2. Robustness: Estimating Price Impact of Ratings (λ) around the June 2002 Event

This is a robustness check of Table 3 by varying the sample window length. We estimate the rating price impact coefficient λ through a forecasting panel regression of monthly returns of the 3×3 (academic) stock styles on lagged rating changes ($\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$). Specification (2) is the same as the the main specification (column (1)) in Table 3.

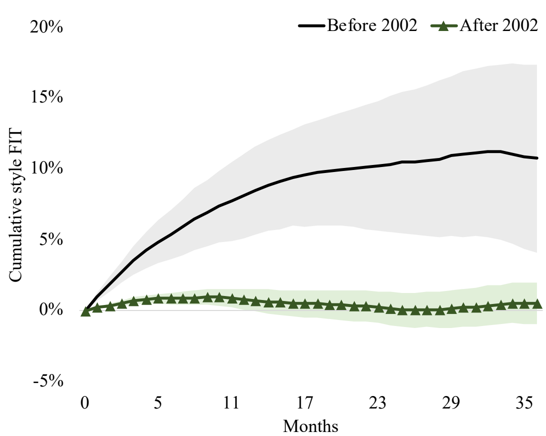
Dependent variable:	Monthly style return $\text{Ret}_{\pi,t}(\%)$		
	6 month	12 month	18 month
	(1)	(2)	(3)
$\text{ExpSum}(\Delta\text{Rating})_{\pi,t-1}$	2.543***	2.881***	2.181***
	(0.865)	(0.736)	(0.625)
Style FE	Yes	Yes	Yes
Style FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Observations	108	108	108
Adj R^2	97.8%	65.9%	52.4%

***$p < 1\%$, **$p < 5\%$, *$p < 10\%$

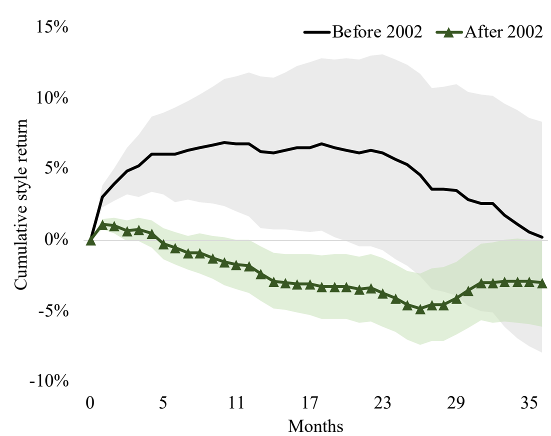
Figure A.2. Price Pressure in Academic Style Portfolios

Stocks are sorted into 3×3 size-value styles using NYSE breakpoints. 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 separately estimate 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



Appendix B Morningstar Methodology

We explain Morningstar rating construction and the June 2002 change in detail here. 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, Morningstar calculates the cumulative return over the three horizons:

$$R_i^T = \prod_{t=1}^{T}(1 + r_{i,t}) - 1, \qquad T \in \{36, 60, 120\}, \tag{14}$$

where the monthly fund returns $r_{i,t}$ are net of management fees but not yet adjusted for loads. Then, Morningstar adjusts the cumulative returns for loads to get a load-adjusted return over the risk-free return:

$$\text{LoadRet}_i^T = R_i^T L_i - R_f^T, \tag{15}$$

where the load adjustment L_i is equal to 1 minus the sum of the front- and back-end load, and R_f^T is defined as the cumulative risk-free rate return for horizon T using three-month T-bills. Morningstar then standardizes the measure to get:

$$\text{MnLoadRet}_i^T = \frac{\text{LoadRet}_i^T}{\max(R_f, \text{AvgLoadRet}^T)}, \tag{16}$$

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

Second, Morningstar subtracts a risk-adjustment term to arrive at the final performance measure:

$$\text{Performance}_{i,t} = \text{MnLoadRet}_{i,t}^T - \text{MnRisk}_{i,t}^T. \tag{17}$$

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

$$\text{Risk}_i^T = \frac{\sum_{t=1}^T - \min(r_{i,t} - r_t^f, 0)}{T}, \tag{18}$$

and then normalizes it by the average risk for the investment class:

$$\text{MnRisk}_t^T = \frac{\text{Risk}_i^T}{\text{AvgRisk}^T}. \tag{19}$$

After June 2002, Morningstar began to conduct risk adjustment in a slightly different way.[21] Morningstar summarizes a fund's past performance using the so-called Morningstar risk-adjusted return (MRAR):

$$\text{MRAR}_i^T(\gamma) = \left[\frac{1}{T}\sum_{t=1}^{T}(1 + r_{i,t} - r_t^f)^{-\gamma}\right]^{-\frac{12}{\gamma}} - 1, \tag{20}$$

where $r_{i,t} - r_t^f$ is the geometric return in excess of the risk-free rate after adjusting for loads,[22] and $\gamma = 2$ is the risk aversion coefficient.

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

[21] 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).

[22] 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, $\mathrm{MRAR}^{\mathrm{T}}(0)$ is equal to the annualized geometric mean of excess returns.[23] 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 $\mathrm{MRAR}^{\mathrm{T}}(0) - \mathrm{MRAR}^{\mathrm{T}}(2)$.

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 amount of historical returns at those horizons, and then take 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. 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 ranks the past performance of all equity funds together and assign them ratings with fixed proportions: 10%, 22.5%, 35%, 22.5%, and 10%. After June 2002, Morningstar ranks funds within each style ("Morningstar category") and assigns 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). Because much of fund performance is due to style-level stock return variation, before the change, there is significant variation of ratings across styles. That variation became much smaller after June 2002 (Panel (b) in Figure 3).

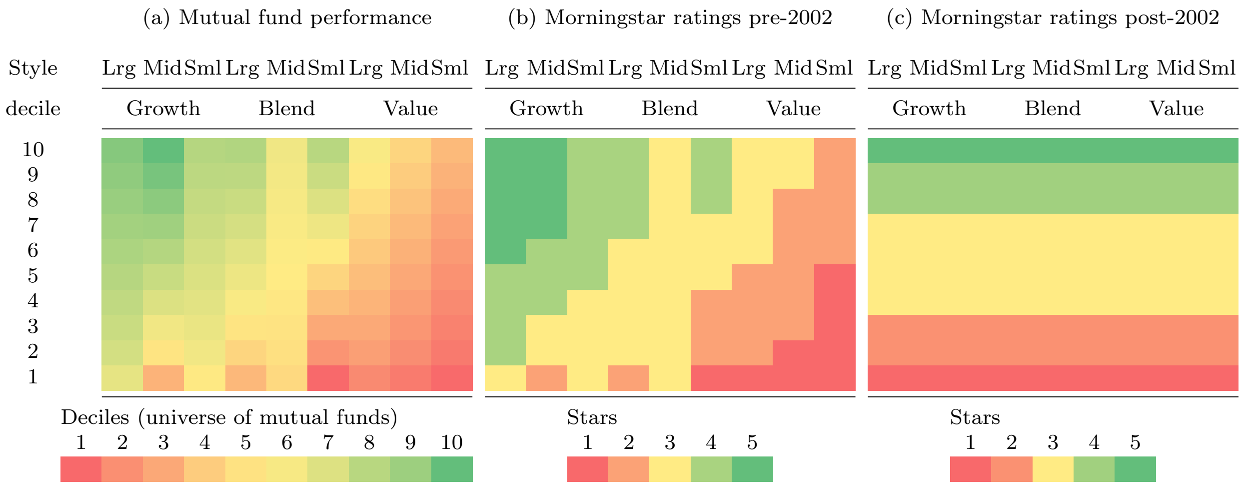
The style-level demand in our study originates from the mutual fund industry. Because mutual funds follow investment strategies ("styles"),[24] their past performance contains a large style-level component. Before June 2002, Morningstar's mutual fund ratings

[23] 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.

[24] E.g., value and growth as in Graham and Dodd (1934) and Fisher (1958), respectively.

Figure B.1. 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-2002 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 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.



closely map absolute past fund performance into star ratings, as illustrated in Figure B.1. Panel (a) shows a snapshot of mutual funds' past performance (colors) for funds within styles. Panel (b) shows how Morningstar would translate funds' performance into star ratings. Because mutual fund investors chase unadjusted past performance and Morningstar ratings (e.g., Del Guercio and Reuter, 2014; Reuter and Zitzewitz, 2015; Ben-David et al., 2019; Evans and Sun, 2020), their flows would appear *as if* they chase style returns, and aggregate mutual fund investment puts price pressures on stocks associated with the common style in demand.