Correlated Demand, Positive Feedback Trading, and Asset Pricing Factors

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

VERY PRELIMINARY, PLEASE DO NOT CITE

Abstract

We present *causal* evidence that correlated demand driven by performance chasing accounts for a substantial fraction of the profits of asset pricing factors and factor momentum. Until June 2002, funds pursuing the same investment style had highly correlated ratings by Morningstar. Hence, rating-chasing investors generated large style-level correlated demand which led to positive-feedback price pressures. Factors, especially those related to momentum strategy, and factor momentum benefit from the tailwind of Morningstar ratings. A 2002 reform in Morningstar's methodology equalized ratings across styles and caused investor demand to spread evenly across styles. Once the link between style performance and investors' demand was disrupted, there was a precipitous drop in factor profitability. We estimate that the decline in correlated demand induced by rating methodology change explains approximately half of the probability decline in momentum-related factors and most of the decline in factor momentum.

Keywords: Morningstar, style investing, mutual funds, anomalies, momentum, factor momentum
JEL Classification: G11, G24, G41

*We thank Sylvester Flood (Morningstar), Paul Kaplan (Morningstar), Andrei Shleifer, Juhani Linnainmaa, and Xin Wang for helpful comments. We thank seminar participants at The Ohio State University, the University of Utah, University of Washington, and Arrowstreet Capital, as well as the National Bureau of Economic Research Behavior Finance Workshop for comments and George Aragon for sharing data. 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

Over the last four decades, students of asset pricing have identified hundreds of factors that span the cross-section of stock returns.[1] Financial economists have incessantly debated the origins of factor profitability. Some propose that the profits reflect compensation for bearing fundamental economic risk.[2] Others claim that factors stem from systematic mispricings,[3] and one specific strand of that literature—the "style investing hypothesis" introduced by Barberis and Shleifer (2003)—argues that factors could be driven by correlated demand from investors. Scholars have also recently proposed that some factors were "discovered" through data mining and therefore do not really exist (Harvey, Liu, and Zhu, 2016; Harvey, 2017; Hou, Xue, and Zhang, 2019). So far, the literature has mostly relied on reduced-form tests or structural models to examine the validity of specific explanations. These techniques, however, are indirect and thus open to the critique that other unobservable forces may also determine equilibrium prices.

Explanations of the underlying mechanism of factor profitability also ought to explain the perplexing sharp drop in factor return profitability since mid-2002. The "profitability kink" has been documented in a few earlier studies[4] and is clearly visible in Panel (a) of Figure 1. The average monthly return of 49 popular factors went from 0.43% during the earlier period of 1991 to June 2002 to merely 0.09% during the later period of July 2002 to 2018. This profitability drop is particularly sharp for the momentum factor and momentum-related factors (e.g., industry momentum) which accounted for substantial factor profits

[1] "Factors" in this paper also include what many would call anomalies. For a comprehensive list of factors identified in the academic literature, see Harvey and Liu (2019).

[2] The literature proposing risk-based explanations of factor profitability is vast. For examples, see Fama and French (1993), Jagannathan and Wang (1996), Berk, Green, and Naik (1999), Gomes, Kogan, and Zhang (2003), Pástor and Stambaugh (2003), Campbell and Vuolteenaho (2004), Zhang (2005), Hansen, Heaton, and Li (2008), Liu, Whited, and Zhang (2009), Li and Zhang (2010), Cochrane (2011), Kogan and Papanikolaou (2014), Hou, Xue, and Zhang (2015), Campbell, Giglio, Polk, and Turley (2018), Betermier, Calvet, and Jo (2019), and Zhang (2019).

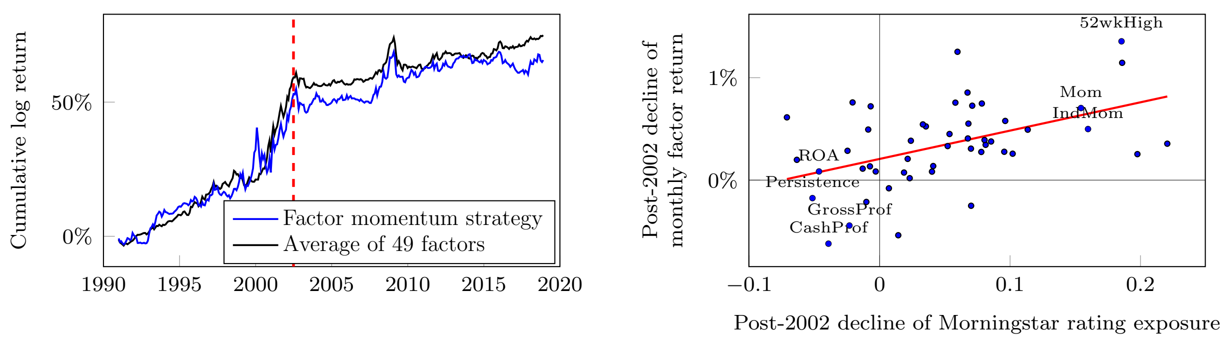
[3] For examples of mispricing-based explanations of factor profitability, see Lakonishok, Shleifer, and Vishny (1994), Daniel, Hirshleifer, and Subrahmanyam (1998), Hong and Stein (1999), Daniel, Hirshleifer, and Subrahmanyam (2001), Hirshleifer (2001), Baker and Wurgler (2006), Stambaugh, Yu, and Yuan (2012), and Stambaugh and Yuan (2016).

[4] The finding of this kink is not specific to our factor construction. It is clearly visible in several previous papers on factors. See, for example, Daniel and Moskowitz (2016), Green, Hand, and Zhang (2017), and Arnott, Harvey, Kalesnik, and Linnainmaa (2019b). For the readers' convenience, we present screenshots from Green et al. (2017) and Daniel and Moskowitz (2016) in Appendix A.

Figure 1. Morningstar Rating Methodology Change and Factor Returns

The figure shows the main results in this study. Panel (a) shows the cumulative returns of 49 popular asset pricing factors (black line) and the factor momentum strategy (blue line) which is defined as the strategy that is long (short) factors that experienced positive (negative) returns in the previous 12 months. Panel (b) plots the decline of average factor return after June 2002 against the decline in their exposure to Morningstar ratings. From left to right, the labeled factors include the 52 week high strategy, industry momentum, price momentum, cash profitability, earnings persistence, and return on assets. Rating exposure measures how much each factor benefits from the positive feedback induced by Morningstar ratings and is defined formally in in Section 5 using variable ExpSum(ΔRating).

(a) Asset pricing factor returns (b) Factor returns and exposure to Morningstar



before 2002. Also plotted in the figure, factor momentum strategies experienced a sharp profitability decline since mid-2002 (Gupta and Kelly, 2019; Ehsani and Linnainmaa, 2019). For example, the monthly return of the factor momentum strategy with a one-year lookback window dropped from 0.39% to 0.08% since mid-2002.

In this study, we present two novel and related findings regarding the origins of factor returns and the 2002 profitability kink. First, we provide *causal* evidence that correlated, style-level positive-feedback demand has a first-order impact on the returns of asset pricing factors. Second, we show that a seemingly innocuous institutional change—Morningstar's mutual fund rating reform in June 2002—caused the sharp factor profitability decline mentioned above. The rating reform effectively removed the style-level correlated demand of mutual funds and disrupted positive feedback trading. As a result, factors that benefited more from the tailwind of the pre-June 2002 Morningstar ratings, such as the momentum-related factors, experienced a larger profitability decline (Panel (b) of Figure 1). Overall, the effects we identify support the view that asset pricing factors are driven by uninformed demand and thus reflect systematic mispricings.

The correlated demand in our study originates from the mutual fund industry. Because mutual funds follow investment strategies ("styles"),[5] their past performance contains a large style-level component. Before June 2002, Morningstar's mutual fund ratings closely mapped total past fund performance into ratings (1 to 5 stars). As a consequence, ratings of funds in the same style are highly correlated, and funds in the style with good recent stock performance ends up having high rating. Because mutual fund investors chase Morningstar ratings (e.g., Reuter and Zitzewitz, 2015; Ben-David, Li, Rossi, and Song, 2019; Evans and Sun, 2020), their flows lead to positive-feedback price pressures on the stocks making up the style.

A specific event allows us to identify the link between Morningstar's ratings and factor returns. In June 2002, Morningstar abruptly changed its rating methodology, causing a halt in the style-level positive feedback trading. Instead of ranking all funds together as before, after June 2002 Morningstar began ranking funds *within* 3×3 size-value style categories[6] in order to compare fund managers to their peer group. Because investors continued to chase Morningstar ratings like before, investor flows became evenly distributed across styles. Thus, aggregate mutual fund trading was no longer concentrated in the best-performing stock styles but rather became balanced across all styles. As a consequence, momentum-related factors and factor momentum strategies, which particularly benefited from the pre-2002 style-level positive feedback trading, abruptly became less profitable.

Our empirical analysis proceeds in three steps. First, we demonstrate the effect of the Morningstar reform on fund flows and show that style-level positive-feedback trading largely halted after the reform. Before June 2002, past style performance was a major driver of Morningstar ratings and aggregate fund flows. During this early period, mutual funds in the top performing styles, on average, received flows that were higher *per month* by 1.7% of assets under management (AUM) than funds in the bottom style performance. As mutual funds trade stocks in response to these correlated flows, creating style-level positive-feedback price pressure, the past-winning style continues to outperform the past-losing style by 70 to 80 bps per month. However, after June 2002, there was virtually no relation between past style performance and Morningstar ratings, leading fund flows and thus price pressure to

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

[6] Morningstar's style categories are combinations of value/growth investment philosophy (value, blend, and growth) and stock size focus (small, mid, and large cap).

become evenly distributed across styles. Consequently, style-level momentum also stops being profitable.

Second, we demonstrate that the rating-induced style-level price pressures have a substantial causal impact on asset pricing factors. To establish the causality, we first examine the behavior of 49 popular factors in an event study with a short window of 12 months around the rating reform. The benefit of an event study is that most rating changes in this period are due to the reform, and we can also be reasonably certain that there are no other major shocks to expected factor profitability. The rating reform created a heterogeneous impact on factors. As predicted, factors that were positively (negatively) affected by the rating reform experienced an increase (reduction) in flows and an increase (reduction) in returns. Using all other years as placebo tests, we confirm that the changes on factor-level flows and returns around June 2002 are unique to that period, both economically and statistically. We also estimate that each star rating revision leads to a price impact of around 2.2% per month at the factor level.

Finally, we study the long-term impact of rating-induced correlated demand on factor profitability through the full sample of 1991 to 2018. We observe that factors whose return benefited more from the pre-2002 rating system experienced a larger profitability drop after 2002. As we hypothesized, the momentum factor and factors related to momentum—such as industry momentum, the 52-week high strategy, etc.—were the most negatively impacted by the disappearance of positive-feedback trading. Morningstar ratings served as an important tailwind for the profitability of those factors before the reform. For example, before June 2002, the top quintile of winner stocks experienced higher ratings changes and also an additional 0.5% of flow-induced trading per month relative to the loser stock quintile. After the methodology change, the difference became much more muted (around 0.1%). We estimate that the Morningstar methodology change explains approximately half of the decline of profitability of those momentum-related factors after June 2002.

We close our analysis with an illustration of the impact on factor momentum strategy. Recently, researchers found that factors also exhibit momentum. That is, factors that recently experienced high returns tend to continue to outperform other factors (Gupta and Kelly, 2019; Ehsani and Linnainmaa, 2019). We show that factor momentum strategies also benefited heavily from the rating-induced positive feedback trading before June 2002. Consequently, they also suffered a sudden profitability decline after the rating reform took place. We find that Morningstar rating changes can explain a majority share of the factor momentum profitability decline.

Our paper has important implications for understanding asset pricing factor profitability. First, our findings imply that a substantial fraction of factor profitability arose from correlated performance-chasing demand and thus does not reflect compensation for risk. We contribute to the existing papers by providing identification. Second, our finding about Morningstar rating reform also contributes to understanding why factor profitability declines over time. Our explanation is non-mutually-exclusive of the existing theories and is the only one that attributed the mid-2002 kink in factor profitability to a specific dated event. Existing papers on the gradual decline of factor profitability emphasize the improvement in liquidity and decrease in arbitrage costs (Khandani and Lo, 2011; Chordia, Subrahmanyam, and Tong, 2014; Lee and Ogden, 2015) or the entry of arbitrageurs (e.g., hedge funds, see Green, Hand, and Soliman, 2011; Hanson and Sunderam, 2013), and the elimination of factors by market participants after they have been publicised by academics (Marquering, Nisser, and Valla, 2006; McLean and Pontiff, 2016; Calluzzo, Moneta, and Topaloglu, 2019). It is also possible that some of the factors reported in the literature were data-mined and therefore stopped working after their "discovery" (Harvey et al., 2016; Harvey, 2017).

The rest of the paper is organized as follows. Section 2 details the data, factor universe, and variable construction. Section 3 shows the mechanism—ratings lead to fund flows, and flow-induced trading leads to price pressures—and also explains how the 2002 methodology change caused a sudden stop of style-level positive feedback trading of mutual funds. Section 4 uses a short window around the 2002 methodology change to establish the causal impact of Morningstar ratings on factor profits. Section 5 quantifies Morningstar's impact on the post-June 2002 factor profitability decline. Section 6 analyzes the influence on factor momentum. Section 7 concludes. Robustness checks and additional tests are provided in the Appendix.

2 Data and Variable Construction

This section describes the data set and how we construct the asset pricing factors.

2.1 Mutual Fund Data

We obtain monthly fund return and total net assets (TNA) from the CRSP survivorship bias-free mutual fund data set. Our data are at a monthly frequency and spans 1991 to 2018. We start in 1991 because monthly fund flow in CRSP starts in 1990, and some measures require one year of lagged data to construct. 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 which is based on 13F filings.

We obtain Morningstar ratings and fund style categories from Morningstar Direct and merge them with the CRSP data using the matching table from Pastor, Stambaugh, and Taylor (2020).[7] Because Morningstar assigns ratings at share class level, and 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 we 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.

2.2 Asset Pricing Factors

Mostly following Arnott, Clements, Kalesnik, and Linnainmaa (2019a), we compute 49 popular stock-level characteristics that have been shown to predict returns. We restrict our attention to those that can be constructed using CRSP and Compustat data. Using the classification categories proposed in Hou et al. (2019), these characteristics include 14 in the profitability category (e.g., return on assets), 13 in the investments category (e.g., share issuance), eight in the value/growth category (e.g., book-to-market), six in the intangibles

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

category (e.g., industry concentration), five in the momentum category (e.g., momentum of Jegadeesh and Titman, 1993), and three in the trading frictions category (e.g., Amihud illiquidity).

We follow the procedure in Hou et al. (2019) to construct long-short factor portfolios using these characteristics. To minimize the impact of microcaps, we use NYSE breakpoints to sort stocks into quintile, and then form factors as long the top quintile and short the bottom quintile. The quintile portfolios are value weighted to further reduce the impact of microcaps. Appendix Table B.1 lists all asset pricing factors.

2.3 Morningstar Rating and Flow-Induced Trading

We are interested in how Morningstar ratings and rating-induced fund flows lead to trading pressures on asset pricing factors. To this end, we first measure ratings and flows at the stock level, and then aggregate them up to the factor level.

As ratings are assigned to mutual funds, we define the average Morningstar rating of stock i in month t as the holding-weighted rating of all funds that hold the stock:

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

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

To measure the amount of mutual fund trading caused by fund flows, we follow Lou (2012) to calculate flow-induced trading (FIT) for each stock i in each month t:[8]

$$\text{FIT}_{i,t} = \frac{\sum_{\text{fund} j \in J} \text{SharesHeld}_{i,j,t-1} \cdot \text{Flow}_{j,t}}{\sum_{\text{fund} j \in J} \text{SharesHeld}_{i,j,t-1}}. \tag{3}$$

Here, flow of fund j in month t is defined as the net flow into the fund divided by the lagged TNA, following the literature (e.g., Coval and Stafford, 2007):

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

[8] Lou (2012) also applies slightly 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 total amount of non-discretionary mutual fund trading in stock i caused by fund flows. As explained in Lou (2012), whereas discretionary trading is likely to be related to fundamentals, FIT isolates the nondiscretionary trading that is only attributable to fund flows and thus likely does not contain value-relevant information. Consistent with this interpretation, Lou finds that FIT leads to price pressures that revert over time.[9]

In addition to computing the standard long-short portfolio returns, we also aggregate up stock-level variables to compute factor-level ratings and FIT. For each factor f, we calculate

$$\text{Rating}_{f,t} = \sum_{i \in \text{top quintile}} w^f_{i,t-1} \cdot \text{Rating}^{\text{stock}}_{i,t} - \sum_{i \in \text{bottom quintile}} w^f_{i,t-1} \cdot \text{Rating}^{\text{stock}}_{i,t} \tag{4}$$

$$\text{FIT}_{f,t} = \sum_{i \in \text{top quintile}} w^f_{i,t-1} \cdot \text{FIT}^{\text{stock}}_{i,t} - \sum_{i \in \text{bottom quintile}} w^f_{i,t-1} \cdot \text{FIT}^{\text{stock}}_{i,t}, \tag{5}$$

where $w_{i,t-1}$ is the lagged market cap weight of stock i in the corresponding quintile portfolio.

3 Morningstar Reform and Positive-Feedback Trading

Next, we first briefly introduce the Morningstar rating system and its methodology reform in June 2002. We then show that that mutual fund flows chase ratings regardless of the methodology. Because of the persistent rating-chasing behavior and the rating reform, style-level positive-feedback trading through mutual funds mostly disappeared since June 2002.

3.1 Rating Methodology Reform

After introducing its mutual fund rating system in 1985, Morningstar quickly became the industry leader in helping investors choose mutual funds. To assign ratings, Morningstar first summarizes the past return performance of funds and conducts minor adjustments for total return volatility and expenses. Depending on the availability of data, the lookback horizon for past performance can be three, five, or ten years, but more weight is applied to more recent returns. For funds with over 10 years of history, the weights of the three

[9]Wardlaw (2019) recently shows that some flow measures, such as that in Edmans, Goldstein, and Jiang (2012), inadvertently include the contemporaneous stock return. This does not apply to our flow measure which follows Lou (2012) and does not use any price data.

horizons are set at 20%, 30%, and 50%, respectively.[10] Then, Morningstar ranks funds by their performance summary and assigns 1 to 5 star ratings with fixed proportions (10%, 22.5%, 35%, 22.5%, and 10%). The Morningstar methodology is fully transparent, and we provide further details in Appendix B.3.

Morningstar's methodology changed abruptly in June 2002. Many funds follow certain specific investment styles (e.g., large-cap growth) by mandate. Because style performance is a significant part of fund performance, fund ratings became highly dependent on style performance. Following the dotcom crash, many fund managers specializing in technology stocks complained that their fund ratings dropped sharply from 5 stars to 3 stars or even lower just because the technology sector had crashed. Consequently, ratings were barely reflecting their own contributions and instead were only echoing sector-level returns that were outside of their control. As a result, the research team at Morningstar, which is spearheaded by the economist Dr. Paul Kaplan, redesigned the rating system.[11]

The main change from the previous rating system is that the post-June 2002 fund ratings are based on fund rankings *within* style categories. For U.S. diverse equity funds (87% of all mutual funds in 2002), Morningstar classified them into the well-known 3×3 matrix: value-blend-growth and small-midcap-large. Sector funds—the remaining 13% of funds—were classified into 12 sectors (e.g., financial, utilities). The change in methodology was announced in February 2002 and was first implemented in Morningstar's monthly ranking of funds at the end of June 2002.

This seemingly innocent methodology change had far-reaching consequences for the mutual fund industry. Before the change, fund ratings differed dramatically across styles based on recent style performance, as shown in Panel (a) of Figure 2. For instance, at the height of the dotcom boom in 2000, 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, and the rating imbalance across styles became negligible. Consistent

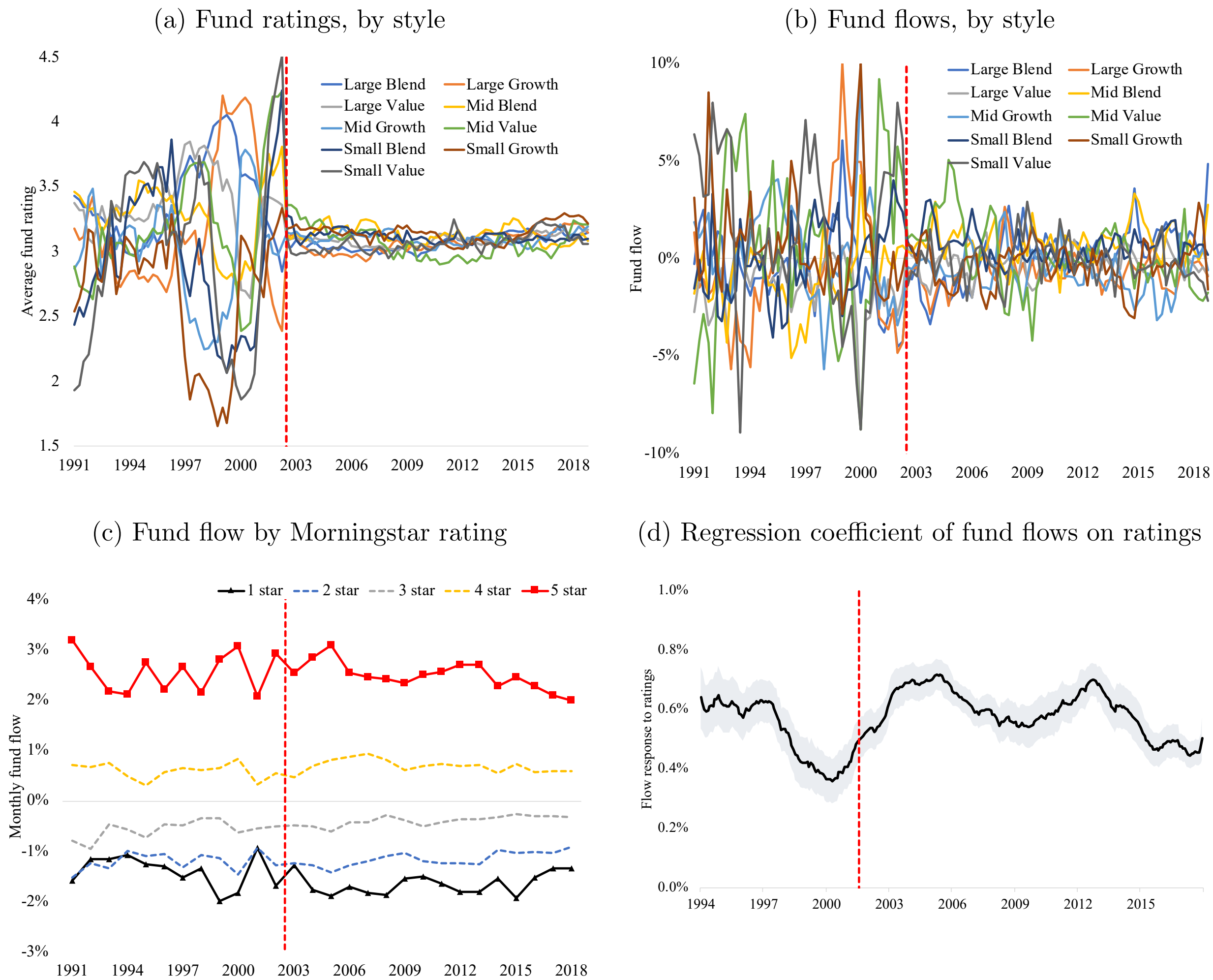
[10]Because the five-year history contains the three-year history, the three most recent years are effectively given more weight than more distant history.

[11]We learned this from a phone conversation 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." (Floyd Norris, *Morningstar to Grade on a Curve*, New York Times, April 23, 2002.)

with flows chasing ratings, Panel (b) shows that style-level fund flow dispersion became much more muted after the change.

Figure 2. The June 2002 Morningstar Methodology Change

Panel (a) and (b) plot the quarterly average mutual fund ratings and TNA-weighted average fund flows by the 3×3 size-value Morningstar styles. Panel (c) plots the average monthly fund flow by Morningstar rating in each year. In both Panels (b) and (c), fund flows are demeaned by period to focus on the cross-sectional variation. Panel (d) plots the regression coefficient of fund flows on lagged ratings estimated using rolling three year windows, with the black dashed lines representing two standard error bands. In all panels, the red dashed vertical line marks the June 2002 Methodology change event.



Despite the dramatic change in rating distribution, investors chase ratings regardless of the rating methodology. Panel (c) of Figure 2 plots the monthly flows to mutual funds by Morningstar ratings. Throughout our sample period, 5-star funds receive flows that amount to +2.5% of their AUM per month, while 1-star funds experience outflows amounting to

-1.5% of their AUM per month. This is a large difference: relative to one-star funds, a five-star fund would *double* its size in approximately one and a half years. As is clearly seen in the figure, the flow response to ratings did not change after the reform. In a more formal test in Panel (d), we estimate the response of fund flows to lagged fund ratings using a three-year rolling-window TNA-weighted Fama-MacBeth regression (Fama and MacBeth, 1973) that controls for lagged 12 months of fund returns. The coefficient estimate only varies slightly over the sample, and there is no material drop around or following the introduction of the reform.[12]

3.2 Disappearance of Positive-Feedback Trading at the Style Level

In this section, we show that the rating reform caused the style-level positive-feedback trading and thus price pressure to halt after June 2002.

To start, it is worth noting that aggregate mutual fund flows are large enough to generate systematic price impact. The mutual fund sector, as a prime investment vehicle for retail investors, holds a substantial and increasing share of the U.S. equity market. When our sample begins in 1991, U.S. equity mutual funds had 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 (Figure B.2 in the Appendix).

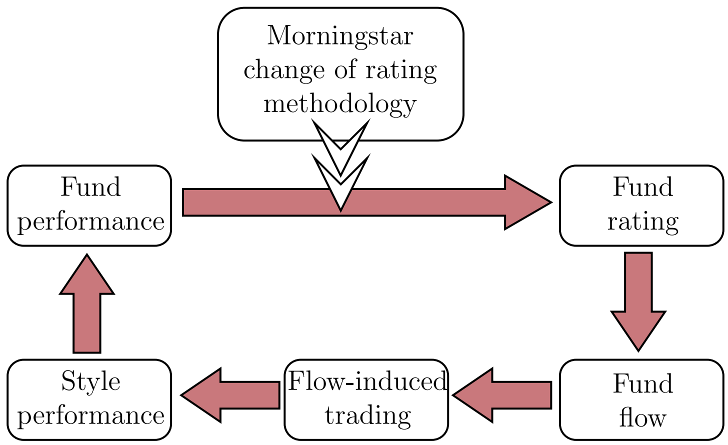
As illustrated in Figure 3, investors' rating chasing behavior creates a style-level positive feedback loop before 2002: funds in styles that performed well in the recent past all get high ratings and attract correlated flows. They use the new flows to increase their investments in the same style of stocks, pushing the prices of securities associated with the styles even further. The mechanism also works in the other direction: Funds in underperforming styles experience correlated outflows, resulting in downward price pressure on stocks associated with these styles. The post June 2002 rating methodology, however, should lead to a sudden reduction in such positive feedback trading.

We confirm the disappearance of positive-feedback trading in Figure 4. Specifically, we

[12] Ben-David et al. (2019) and Evans and Sun (2020) provide further evidence that Morningstar ratings are the mostly important determinant of fund flows.

Figure 3. Style Investing Feedback Loop

This flow chart illustrates how Morningstar ratings generate positive style-level positive feedback trading. First, funds holding stocks in the styles that recently performed well (poorly) also exhibit good (poor) performance, causing Morningstar to assign high (low) ratings to them. Second, investors chase ratings, so high- (low-) rated funds experience in (out) flows. Third, fund managers buy (sell) fund holdings in response to flows, leading to stock price pressures.



sort the 3×3 styles based on past 12-month returns. Before June 2002, funds in styles that performed well recently received higher average ratings and higher fund flows. Panel (a) shows that the average rating spread between funds in the top and bottom styles was about 0.8 stars before 2002 and shrank to almost zero after June 2002.[13] Because rating attracts flows, Panel (b) shows that funds in the top style, on average, received about 1.7% higher flows per month than the bottom style before June 2002, and that difference dropped to around 0.4% after June 2002.

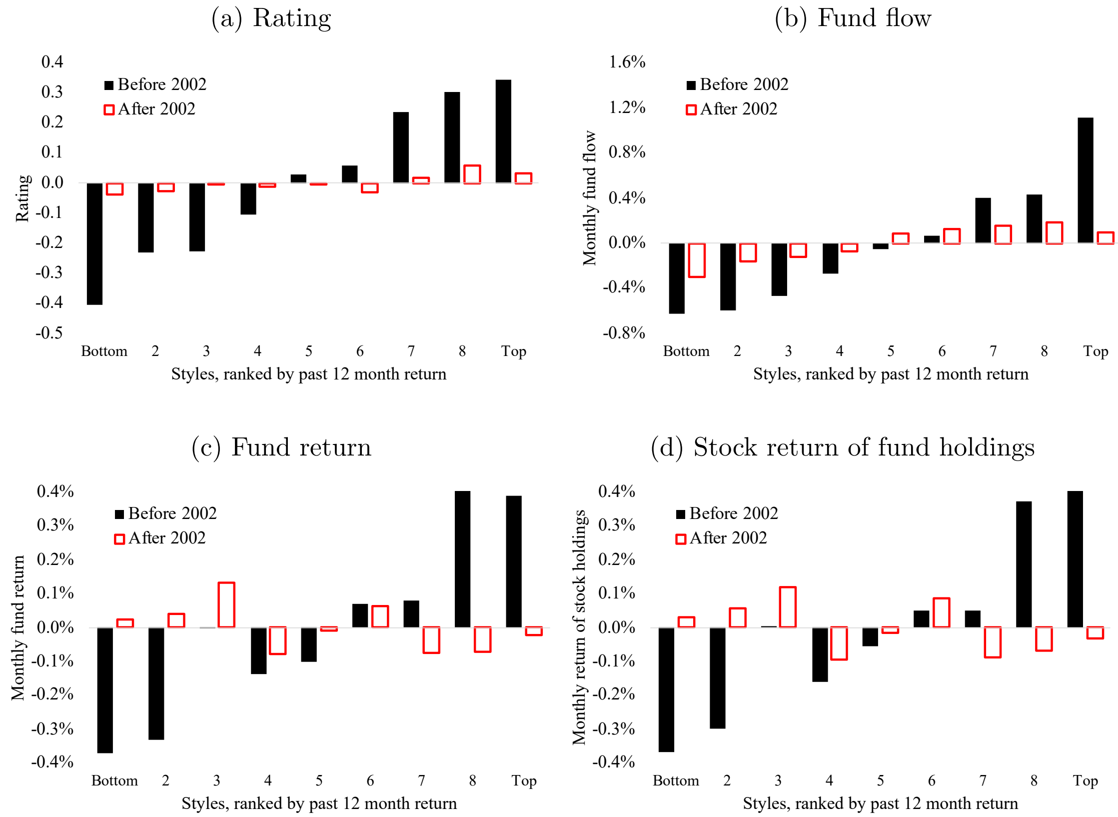
Before the reform, the style-level correlated demand further generates positive price pressure and creates style-level return momentum. Based on either fund returns and stock returns of fund holdings, Panels (c) and (d) show that the style momentum strategy was profitable before 2002 with a monthly return of about 70 to 80 bps per month (top minus bottom style), but became entirely unprofitable after June 2002.[14]

If asset pricing factors are neutral to styles, then they would not be affected by this rating-induced positive feedback trading. However, even a casual look at them suggests that many of them should be affected. In particular, the momentum factor is long stocks with

[13]The graphs are demeaned to focus on cross-sectional patterns across styles.
[14]Ben-David, Li, Rossi, and Song (2020) provide further evidence that rating-induced demand caused systematic patterns in style returns.

Figure 4. Style-level Positive Feedback Trading, Before and After 2002

This figure shows that the style-level positive feedback trading largely halted after the Morningstar methodology change in June 2002. In each month, we sort the 3×3 Morningstar styles by their lagged 12 month returns. Panels (a) and (b) plots the TNA-weighted average rating and fund flows of the sorted styles. Panels (c) and (d) plot the return of those styles, with (c) showing fund returns and (d) showing the returns of the stock holdings of those funds. All variables are demeaned to focus on the difference across styles.



high past returns and short stocks with low past returns. Because a significant amount of stock returns arise from style-level variation Fama and French (1996), we would expect the momentum factor to benefit heavily from the pre-2002 Morningstar ratings. There are also a number of factors that are of the momentum-type, such as industry momentum, that also should be affected. In Section 5, we will show that factors that are more influenced by this rating-induced structural change experienced larger return decline after 2002. Section 6 shows that the factor momentum strategy also lost a large fraction of its profitability after the Morningstar reform.

4 Event Study Using the 2002 Shock

Before we analyze the influence of the Morningstar reform on long-term factor profitability, in this section, we use an event study to demonstrate that Morningstar ratings indeed exert a first-order causal effect on factor returns.

4.1 Causal Effects of Rating Reform on Factor Returns

We conduct the event study using a short window of 12 months around the rating reform. There are two benefits to using a short window. First, the rating changes over this period are primarily caused by the methodology reform. Second, by using a short window, we reduce the concern that factor returns were impacted by other events. For instance, the decimalization of U.S. stock trading—all U.S. exchanges moved to quoting prices in cents (rather than in 1/8s of a dollar) in April 2001—may also have reduced factor profitability through increasing liquidity (Chordia et al., 2014), but it is not included in our sample.

While we later study asset pricing factors, it is important to note that the impact of rating reform happens at the style level, and the effect on factors is derived through the dependence of factors on styles. We thus first examine what happened to style ratings, flows, and returns around 2002 in Figure 5. In the six months before the methodology change, small-value stocks performed well and large-growth stocks performed poorly, and the ratings mirrored the style-level performance: small value funds have ratings that are on average 2.07 stars higher than large growth funds. The fund flows and returns reflected similar cross-style differences. However, in the six months after the event, ratings become balanced due to the reform, and so did fund flows and returns.

We now examine how asset pricing factors are affected in 2002. First, we sort factors by how much their rating is predicted to change—due to the reform—based on data in December 2001 which is *before* the event study window. Specifically, we estimate fund ratings from ground up under the pre-2002 and post-2002 Morningstar rating methodologies, and then aggregate these ratings at the factor level. We then predict that each factor f will experience

Figure 5. Style Level Changes Around the 2002 Event

The tables show the average rating, monthly flow, and monthly return of funds by 3 × 3 fund styles during the six months before and after the rating methodology change event. All variables are demeaned by period to focus on the differences across styles. Larger values are colored in red and smaller values are colored in green.

		Rating				**Flow**				**Return**		
		Growth	Blend	Value		Growth	Blend	Value		Growth	Blend	Value
Before	Small	-0.36	0.25	0.96	Small	-0.4%	1.0%	2.3%	Small	-0.5%	1.1%	2.4%
	Mid	-0.47	0.33	0.79	Mid	-0.9%	0.0%	1.5%	Mid	-1.2%	-0.2%	0.7%
	Large	-1.11	-0.45	0.05	Large	-1.4%	-1.4%	-0.7%	Large	-1.8%	-0.7%	0.2%
		Growth	Blend	Value		Growth	Blend	Value		Growth	Blend	Value
After	Small	-0.01	-0.15	0.39	Small	0.2%	0.1%	-0.3%	Small	-0.5%	-0.4%	-0.3%
	Mid	0.08	-0.33	0.21	Mid	-0.1%	0.0%	0.3%	Mid	-0.3%	0.3%	0.3%
	Large	-0.28	-0.02	0.10	Large	-0.2%	-0.2%	0.1%	Large	0.2%	0.4%	0.2%

a rating change of

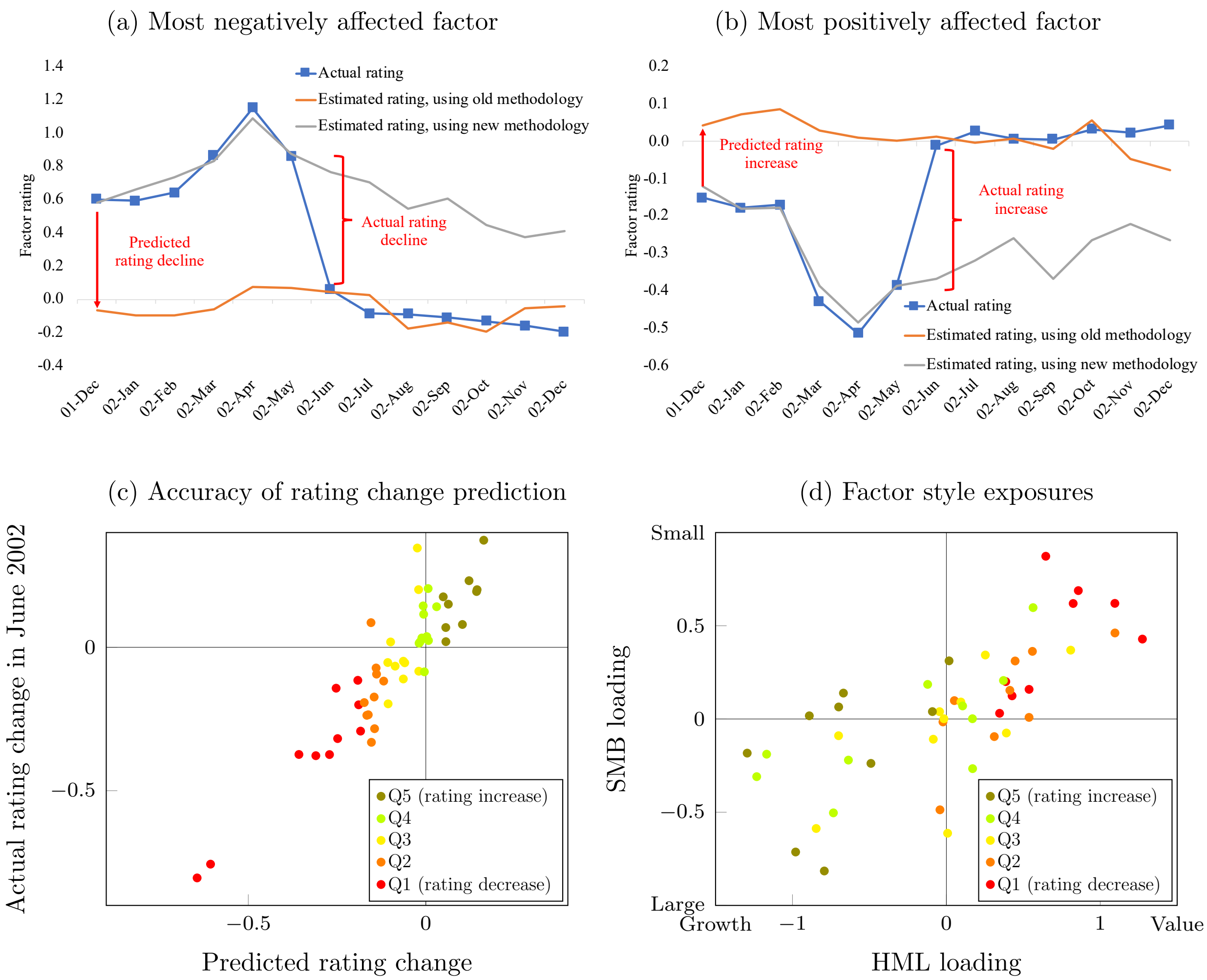
$$\text{PredictedChange}_f = \widehat{\text{Rating}}_{f,\text{Dec 2001}}^{\text{new methodology}} - \widehat{\text{Rating}}_{f,\text{Dec 2001}}^{\text{old methodology}}.$$

This is illustrated in the Panels (a) and (b) in Figure 6 using the two factors whose ratings are predicted to experience the largest decline (size factor) and the largest increase (O-score factor). Our estimated ratings match actual ratings quite well. Before June 2002, the actual ratings closely match the estimated ratings under the old methodology (grey lines), and after June 2002, the actual ratings closely match the estimated ratings under the new methodology (orange lines). Further, because the changes of ratings of factors are small, the predicted rating change using December 2001 data ends up being a reasonable predictor of the actual rating change in June 2002. This is further shown in Panel (c) where we plot the actual June 2002 rating changes of factors against the predicted changes. The latter explains the former with an R^2 of 84%.

To exploit the heterogeneity across factors, we sort factors into quintiles based on the predicted rating changes. Why are some factors predicted to experience rating declines and others predicted to experience rating increases? This depends on how factors load onto styles around the event. In Panel (d) of Figure 6, we measure the size and value exposure of

Figure 6. Sorting Factors by Predicted Rating Change at the Event

Panels (a) and (b) illustrated how we predict rating changes of factors at the June 2002 event using data in December 2001. Following Morningstar's rating construction process, we estimate ratings from ground up using fund returns. The grey lines plot the estimated rating under the old (pre-change) methodology and the orange lines plot the estimated rating under the new (post-change) methodology. We use the difference between the two estimates in December 2001 (marked using red arrows) as the predicted rating change. The blue lines are the actual ratings. Panel (a) and (b) plots the factor with the largest predicted rating decline and increase, respectively. Panel (c) compares the actual rating change in June 2002 against the predicted change using data in December 2001. Panel (d) plots the SMB and HML factor exposures of those factors, estimated using time series regressions with daily returns during the six months before the event. In both (c) and (d), factors are sorted into quintiles based on the predicted rating change and colored differently.



each factor using their SMB and HML loadings, estimated using time-series regressions with daily returns during the six months before the methodology change event. The loadings are plotted in Panel (d) and the factors are colored based by quintile of predicted rating changes. As expected, the factors predicted to experience the largest rating decline loads onto small and value styles, and factors predicted to face largest increase load onto large and growth styles.

We are now ready to examine what happened to factors during 2002, in Figure Figure 7. Panel (a) plots average ratings: it shows a sharp methodology-induced drop exactly at the event. Factors in quintile 1 suffer a drop of 0.43 rating stars, while factors in quintile 5 experiences a small increase of 0.19. Panels (c) and (e) and plot monthly factor FIT and returns around the event, respectively. Quintile 1—the factors that benefited from ratings pre-event but suffered post-event—experienced a decline of 1% in monthly flow-induced trading and a sharp decline of −3.7% in monthly returns. At the same time, quintile 5 experienced an increase of 0.14% in monthly FIT and a slight increase of 0.75% in monthly returns.

To alleviate the concern that the return and FIT changes could result from mean reversion due to other reasons, we also show that similar effects do not happen in other years. Panels (b), (d), and (f) show the same exercise in other years and plot the two standard error bands. Clearly, the shock is unique to 2002.

4.2 Alternative Hypotheses

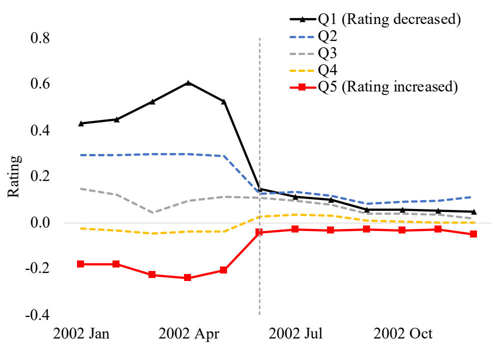
Can the factor return changes around June 2002 be caused by other forces? A number of papers have pointed out that factor returns are related to arbitrageur activity. For instance, Hanson and Sunderam (2013) argue that when more capital is devoted to value and momentum strategies, the returns to those strategies decreased. Related to this, Lou and Polk (2018) use a return correlation-based measure of arbitrageur trading in momentum negatively predict momentum profits. However, in this section, we estimate two measures of arbitrage activity in factors and show they do not experience any meaningful change in 2002. Therefore, our findings are unlikely to be explained by time-varying arbitrage activity.

We consider two arbitrage measures in the literature. First, we construct the net arbitrage activity (NAT) measure in Chen, Da, and Huang (2019). For each stock, the authors measure the long position of arbitrageurs using aggregate 13F holdings of hedge funds and the short

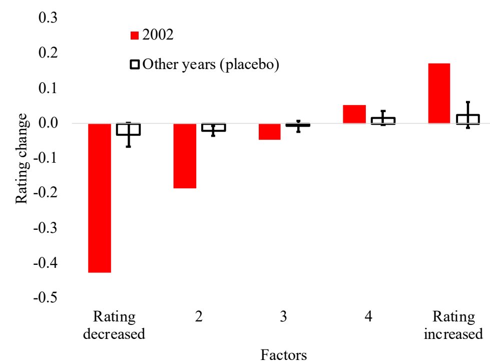
Figure 7. Stock Factors around the June 2002 Event

We perform event studies on the 49 factors during the six months before and after the June 2002 methodology change. In the left panels, we sort factors by their *predicted* event rating change using December 2001 data in quintiles, and then plot the evolution of their ratings in Panel (a), cumulative fund flow-induced trading (FIT) 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 red bars plot the average rating, FIT, and return changes after June in 2002 (average of July to December 2002 minus average of January to June 2002), while the white 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 by period.

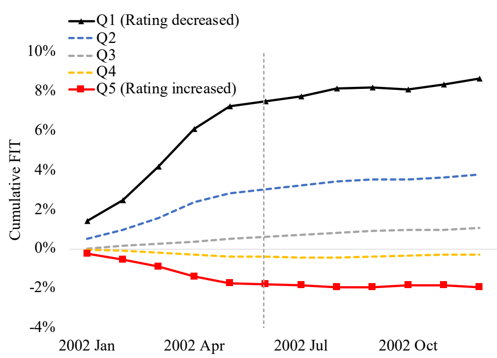
(a) Rating in 2002



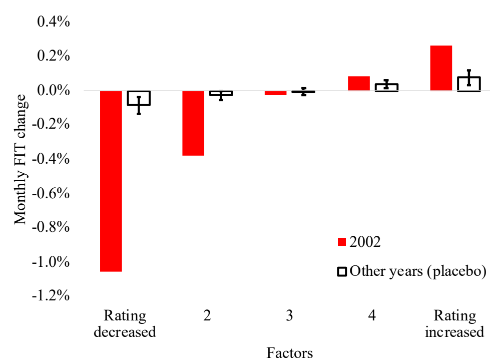
(b) Rating change, 2002 vs placebo



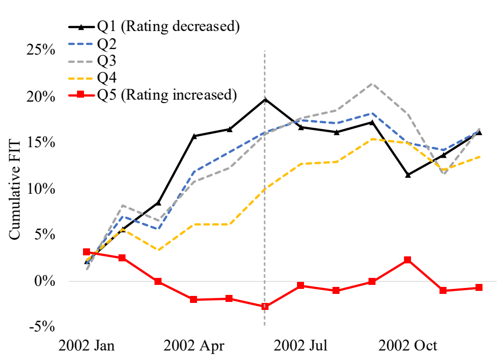
(c) Cumulative FIT in 2002



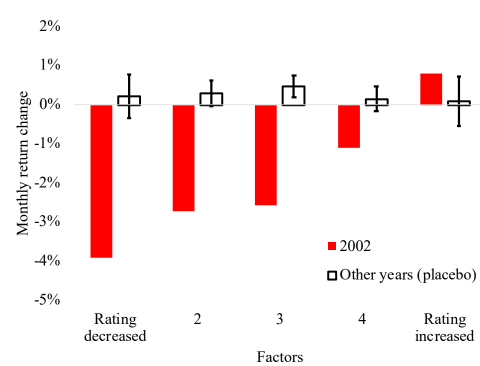
(d) FIT change, 2002 vs placebo



(e) Cumulative return in 2002



(f) Return change, 2002 vs placebo



position using aggregate short interest from Compustat.[15] The authors combine the long and short positions into a net position, and subtract the past four-quarter average to arrive at a measure of arbitrageur position changes which they call NAT. We follow them to compute stock-level NAT and aggregate it at the factor level.

Second, we also follow Lou and Polk (2018) and construct "CoFactor" measures of arbitrage activity in factors. In that paper, the authors measure arbitrage activity in the momentum strategy by estimating excess return correlation within the long and short portfolios. Specifically, in any given month, they use the previous 52 weeks of data to compute a "comomentum" measure:

$$\begin{aligned}\text{CoMomentum}_t = \frac{1}{2} \cdot \Bigg[& \frac{1}{N^L(N^L-1)} \sum_i \sum_{j \neq i} \text{PartialCorr}(\text{Ret}_i, \text{Ret}_j) \\ & + \frac{1}{N^S(N^S-1)} \sum_i \sum_{j \neq i} \text{PartialCorr}(\text{Ret}_i, \text{Ret}_j) \Bigg],\end{aligned}$$

where N^L and N^S are the number of stocks in the long and short portfolios, respectively. To compute the partial return correlations, they first subtract Fama-French 30 industry returns from weekly stock returns, and then regress the residuals on the Fama-French three factors to obtain alphas. Finally, they compute equal-weighted averages of the pairwise correlations of the alphas within the portfolios and take an average. Following their path, we compute this measure for all factors.[16]

We plot the evolution of these measures in the 12 month event window in Figure 8. As in Section 4.1, we sort factors into quintiles by their predicted rating change using data in December 2001. Panel (a) plots the NAT measure and Panel (b) plots the CoFactor measure. There is no noticeable change of arbitrage activity in factors using either measure.

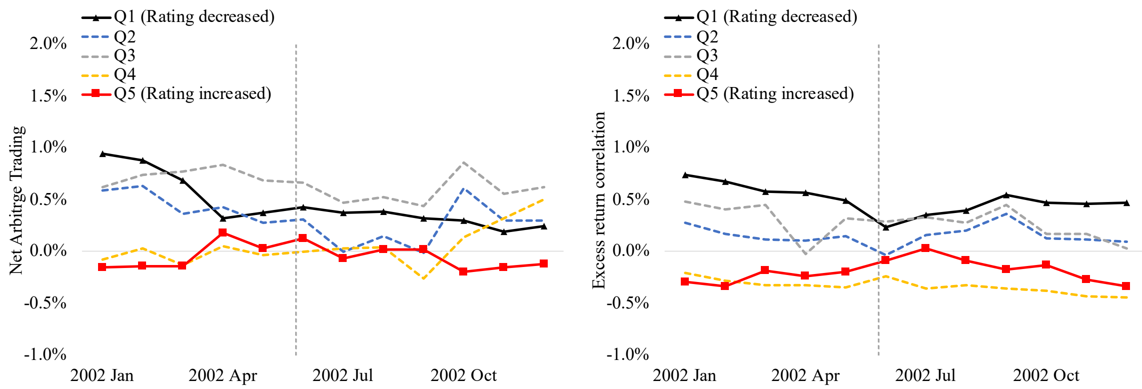
[15] We use the list of 13F institutions identified as hedge funds in Aragon, Li, and Lindsey (2018). We thank the authors for kindly sharing the data. While the short side of NAT is updated monthly, the long side relies on 13F holdings and is only updated quarterly.

[16] Consistent with Lou and Polk (2018), we find that this measure negatively predicts returns of factors in the momentum category.

Figure 8. Arbitrage Activity in Stock Factors around 2002

As in Figure 7, factors are sorted into quintiles by the predicted rating change using data in December 2001. Panel (a) plots the net arbitrage trading measure in Chen et al. (2019). Panel (b) plots excess return correlation in extreme factor deciles, a measure of arbitrage activity developed in Lou and Polk (2018). The vertical red dashed lines mark the methodology change event.

(a) Net Arbitrage Trading (b) Excess Correlation in Extreme Deciles



5 Long-Term Impact on Factor Profitability

The event study in the previous section demonstrates the existence of the effect on ratings on factor returns. We now ask another economically important question: can the Morningstar methodology change explain a sizeable fraction of the *long-run* factor profitability decline after June 2002 (Panel (a) of Figure 1)? If so, which factors were more affected than others?

5.1 Which Factors Suffered Ratings Drop Post-June 2002?

To capture the effect of rating changes on future factor returns, we use a simple specification of rating-induced price impact

$$\mathrm{Ret}_{f,t} = \mathrm{Ret}_{f,t}^{\mathrm{counterfactual}} + \underbrace{\lambda \cdot \mathrm{ExpSum}(\Delta\mathrm{Rating})_{f,t-1}}_{\text{Rating-induced price pressures}}.$$

Specifically, we summarize past rating changes using a weighted sum:

$$\text{ExpSum}(\Delta\text{Rating})_{f,t-1} = \sum_{k=1}^{12} \tau_k \cdot \Delta\text{Rating}_{f,t-k}, \tag{6}$$

where the weights τ_k decay at an exponential rate $\delta = 0.764$[17] which is estimated from a least-squares fit to the cumulative response of stock returns to past rating changes (see Appendix Figure C.3 for details). We use 12 lags because the impact primarily happens within 12 months, and the estimated decay rate implies a half-life of 2.58 months. We normalize the weights to sum to 12 (i.e., $\sum_{k=1}^{12} \tau_k = 12$) so ExpSum(ΔRating) should be interpreted as the rating change over one year. Thus, the price impact coefficient λ reflects the impact on monthly returns by one-year change in ratings.[18] The reasoning behind this specification is explained in detail in Appendix Section C.1.

Intuitively, because rating changes attract fund flows and fund flows create price pressures, ExpSum(ΔRating) reflects the extent to which factor profitability depends on rating-induced trading. In Panel (a) of Figure 9, we plot each factor's average post-June 2002 ExpSum(ΔRating) against the pre-2002 values over the full sample. We mark factors from different categories using different colors. Clearly, before June 2002, Morningstar serves as an important tailwind for many factors, especially those in the momentum categories (colored blue). After June 2002, the ExpSum(ΔRating) across factors collapsed to close to zero. Clearly, momentum-category factors were mostly positively affected by rating-induced positive-feedback trading before 2002 and also suffered the sharpest drop after June 2002. This is consistent with our hypothesis in Section 3.2. We thus expect the momentum-category factors to suffer the sharpest profitability decline after the rating reform.

In Panels (b) and (c), we plot the pre-2002 and post-June 2002 average factor returns against the pre-June 2002 ExpSum(ΔRating). As expected, factors that benefit from pre-2002 ratings experienced high returns before the reform but not afterwards. For instance, the momentum factor experienced close to a 1% monthly return before June 2002 but became negligible after June 2002. Other momentum-category factors, such as the 52-week high

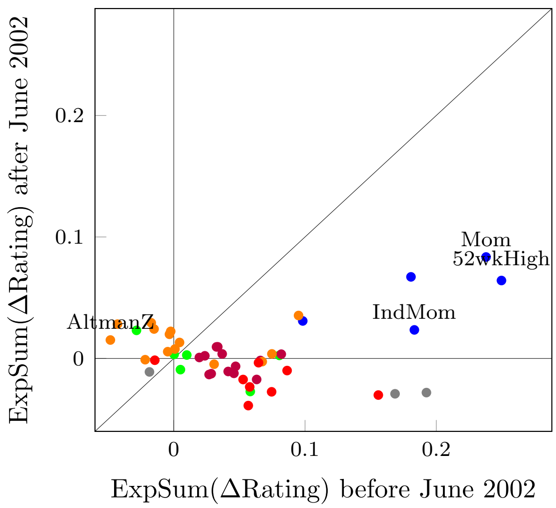
[17] Therefore, $\tau_k = \frac{12\cdot(1-\delta)}{1-\delta^{12}} \cdot \delta^{k-1}$. Our results are insensitive to reasonable variations in the parameter δ.

[18] For instance, if $\lambda = 1$, then for ever 1 star increase in the exponentially-weighted rating change in the previous year ExpSum(ΔRating), subsequent monthly return is higher by 1%.

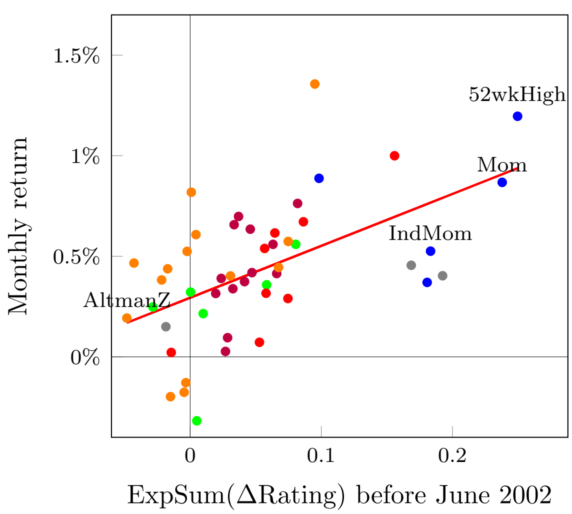
Figure 9. Factors before versus after June 2002

We compare factor statistics before versus after June 2002 over the full sample (1991 to 2018). Panel (a) plots the post-June 2002 ExpSum(ΔRating) (the exponentially-weighted sum of past-12-month rating changes) against the pre-2002 values. Panels (b) and (c) plot average monthly factor returns before and after June 2002 against pre-2002 ExpSum(ΔRating). The red lines in Panels (b) and (c) are best linear fits. The different colors represent the return factor classifications in Hou et al. (2019). The factors with data labels include momentum, 52-week high, industry momentum, and Altman's Z-score.

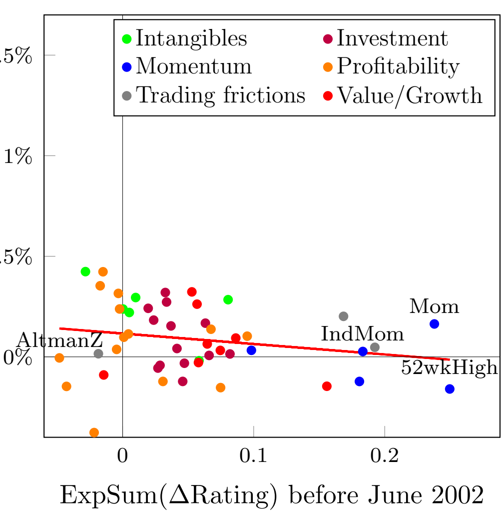
(a) ExpSum(ΔRating) pre- and post-June 2002



(b) Factor returns before June 2002



(c) Factor returns after June 2002



factor, suffered similar declines in profitability.

5.2 Effects of the Rating Reform on Momentum-Related Factors

In this section, we dig deeper into the effect of rating reform on the momentum factor and other momentum-category factors. Momentum is arguably one of the most puzzling factors because of its high profits and the difficulty to rationalize it using risk-based explanations.[19]

Figure 10 compares ExpSum(ΔRating), FIT, and returns of the five momentum quintile portfolios before and after June 2002.[20] Panel (a) shows that, before the rating reform, the winner portfolio experiences significant upward rating changes while the loser portfolio experiences significant downward changes. This pattern became much smaller after June 2002. There is a similar effect on FIT, as shown in Panel (c). Before June 2002, the winner portfolio experiences 0.51% higher monthly flows than the loser portfolio; that difference declined to a meager 0.14% after June 2002.

Finally, Panel (e) shows similar patterns in returns. Before June 2002, the winner quintile portfolio enjoyed 0.87% higher monthly return relative to the loser quintile, and that difference declined to 0.16% after June 2002. In Panels (b), (d), and (f), we confirm that similar post-2002 changes also happened for other factors that fall into the momentum category.

5.3 Quantifying Explanatory Power on the Post-2002 Factor Profitability Decline

In this section, we quantify the exploratory power of Morningstar rating reform on the factor profitability decline after 2002.

To this end, we first use the short 12-month window around June 2002 to causally estimate the price impact of Morningstar ratings on factor returns. Specifically, we estimate a panel

[19]Momentum has been observed for almost a century in the U.S. stock market (Daniel and Moskowitz, 2016) (until the early 2000s) as well as in many other asset classes (Asness, Moskowitz, and Pedersen, 2013). In an interview, Eugene Fama stated that he views momentum as "the biggest embarrassment for efficient markets." See "Fama on Momentum," AQR 2016, accessible at `https://www.aqr.com/Insights/Perspectives/Fama-on-Momentum`.

[20]We follow Jegadeesh and Titman (1993) to define momentum by sorting stocks using their lagged $(t-1, t-12)$ month returns. To avoid the impact of microcaps, we follow Hou et al. (2019) in using NYSE breakpoints and value-weight each portfolio.

regression using six months before to six months after the event:

$$\text{Ret}_{f,t} = \lambda \cdot \text{ExpSum}(\Delta\text{Rating})_{f,t-1} + X_{f,t-1} + \epsilon_{f,t}, \tag{7}$$

where the control $X_{f,t-1}$ includes factor returns over $t-1$, $t-2$ to $t-6$, and $t-7$ to $t-12$ months as well as factor and time fixed effects. There are two reasons for choosing a short window. First, we want to ensure that rating changes primarily come from the methodology change.[21] Second, we want to avoid the impact of other market-level changes such as the dotcom bubble burst in early 2000s or the "momentum crash" event in 2008 (Daniel and Moskowitz, 2016). Those events are not contained in our estimation window.

To account for the cross-sectional factor return correlation, we adjust the standard errors using a feasible generalized least squares (FGLS) approach. Specifically, we use the full sample of factor returns to estimate the covariance matrix of factor returns and incorporate it into the estimation.[22]

The estimation results are shown in Table 1. For each star rating change, the factor-level price impact is 2.27% with a t-statistic of 4.28. The result is both statistically and

[21]This would not be true if we use a longer sample because ratings are, after all, (complex) functions of past returns. Recent papers find that factors can exhibit momentum (Gupta and Kelly, 2019; Arnott et al., 2019a); earlier papers found that returns tend to exhibit long-term reversion (De Bondt and Thaler, 1985). Thus, regressing returns on lagged ratings over a longer sample may simply be picking up some combination of momentum and reversal effects.

[22]Specifically, let y be the vector of factor returns stacked together so that the first 49 entries are the first month, the next 49 entries are the second month, and so forth. Then, we estimate the covariance matrix of y to be

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

where $\hat{C}$ is the estimated contemporaneous return covariance matrix of the 49 factors. Let X denote the matrix of independent variables. Then, we estimate the regression coefficients and covariance using

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

In Appendix Figure C.4, we show that some categories of factors have positive correlations among themselves, such as the momentum-related factors. However, there are also negative correlations across factors categories. For example, the value factor is negatively correlated with the momentum-related factors, consistent with Asness et al. (2013).

economically significant. The estimates are robust to whether we include time and factor fixed effects, as shown in columns (2) to (4).

Table 1. Estimating Price Impact Coefficient (λ) Around the June 2002 Event

We use a panel regression to estimate the predictive relationship on monthly returns of 49 factors using lagged ratings ($\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$) during the six months before versus six months after the methodology change. We control for lagged factor returns in months $t-1$, $t-6$ to $t-2$, and $t-12$ to $t-7$. The four specifications differ in whether factor and month fixed effects are included. The standard errors in parentheses are adjusted for the cross-sectional correlation between factor returns using a feasible generalized least squares approach.

Dependent variable:	Monthly factor return $\text{Ret}_{f,t}(\%)$			
	(1)	(2)	(3)	(4)
$\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$	2.270***	2.057***	2.388***	2.033***
	(0.534)	(0.493)	(0.516)	(0.478)
Lagged Returns	Yes	Yes	Yes	Yes
Factor FE	Yes	No	Yes	No
Month FE	Yes	Yes	No	No
Observations	588	588	588	588
Adjusted R^2	14.36%	7.21%	12.73%	5.86%

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

We now quantify, over the full sample of 1991–2018, how much of the post-June 2002 factor profitability decline can be explained by Morningstar.[23] In our framework, the change of factor f returns explained by Morningstar equals

$$\lambda \times \left(\overline{\text{ExpSum}(\Delta\text{Rating})}_{f,\text{after 2002}} - \overline{\text{ExpSum}(\Delta\text{Rating})}_{f,\text{before 2002}}\right).$$

Admittedly, this is a crude exercise and requires strong functional form assumptions. For our main estimate, we use the $\lambda = 2.27\%$ estimated from the short window around the 2002 shock. We believe this to be the best estimate because it exploits exogenous methodology-induced rating variation.

How much can the factor profitability changes be explained under this framework? Ta-

[23]The fact that factors became less profitable after June 2002 has been documented by a number of papers and so far defies explanation (Khandani and Lo, 2011; Daniel and Moskowitz, 2016; Green et al., 2017; Arnott et al., 2019b).

ble 2 shows that the momentum factor and other momentum-category factors experienced a drop of monthly return of 0.70% to 0.80% after 2002. The other factors experienced much smaller return decline, averaging at 0.29% per month. For the momentum factor, rating-induced price pressures can explain close to half of the profitability decline. Relative to non-momentum category factors in column (3), momentum experienced an additional decline of 0.41% of which ratings can explain slightly more than 60%.

Table 2. Explanatory Power on Factor Profitability Decline After 2002

We tabulate the change of average monthly factor returns from the pre-change period (1991–June 2002) to the post-change period (July 2002–2018). Column (1) examines the momentum factor; column (2) examines the other factors in the momentum category; column (3) examines all the other factors. The "Explained by Morningstar" values are computed by multiplying the price impact parameter λ, estimated in specification (1) in Table 1, with the change of lagged rating changes ($\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$) after June 2002. The standard errors in the parenthesis accounts for the standard error in estimating λ.

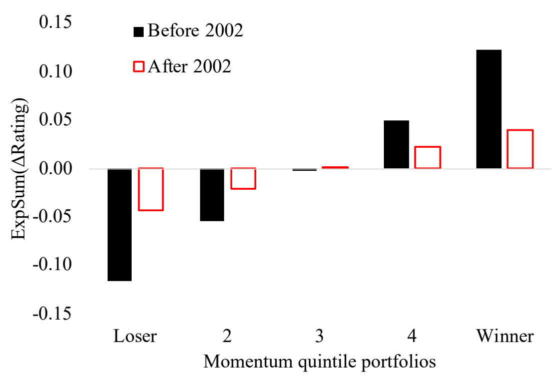
	Factors			Difference	
	Momentum Factor	Other Momentum-Category Factors	All other Factors	(1) − (3)	(2) − (3)
	(1)	(2)	(3)		
Actual return drop	−0.70%	−0.80%	−0.29%	−0.41%	−0.51%
Explained by Morningstar	−0.35% (0.08%)	−0.30% (0.07%)	−0.09% (0.02%)	−0.26%	−0.21%
Fraction explained	49.8%	37.3%	30.6%	63.6%	41.2%

While the 2002 shock is the best identified variation in our exercise, one may still be worried that the price impact of ratings λ varies over time. In Appendix C.2, we repeat this quantification exercise using λ estimated from rolling-window factor return–predicting regressions. Under that specification, we find that ratings can explain approximately one-third of the decline in momentum profits.

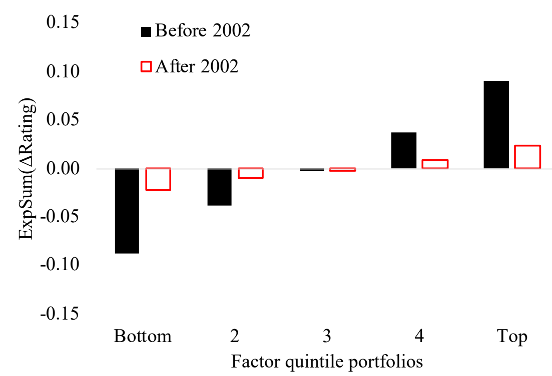
Figure 10. Momentum Category Factors, Before versus After 2002

The left panels plot the ExpSum$(\Delta\text{Rating})_{t-1}$ (exponentially-weighted sum of past-12-month rating changes), fund flow-induced trading (FIT), and returns of the five momentum quintile portfolios before (from 1991) versus after June 2002 (until 2018). The right panels plot the same for the other factors in the momentum category: industry momentum, 52 week high, 7–12 months momentum, and intermediate (2–6 months) momentum. The quintiles break points are formed only using NYSE stocks and the portfolios are value-weighted. All variables are demeaned to emphasize cross-sectional differences.

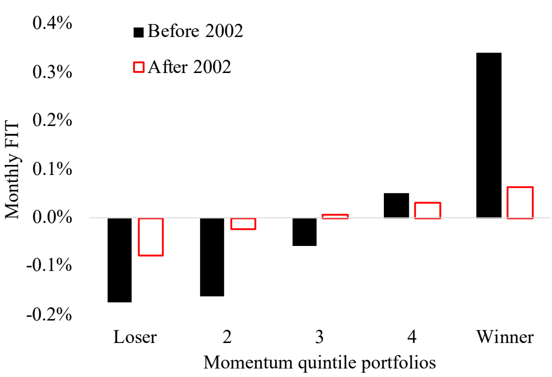
(a) ExpSum(ΔRating), Momentum



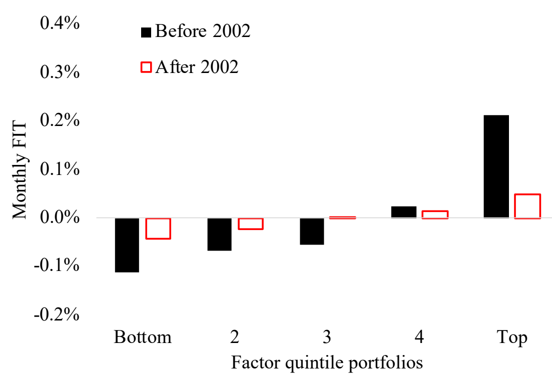
(b) ExpSum(ΔRating), other momentum-category factors



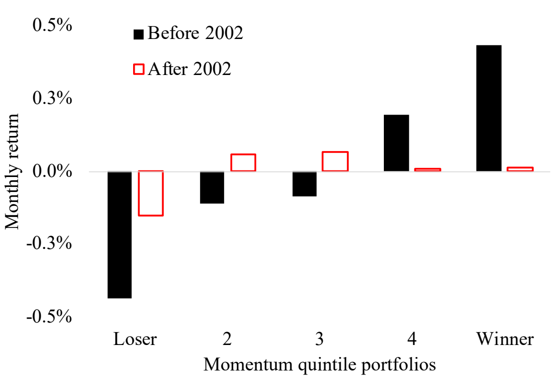
(c) FIT, Momentum



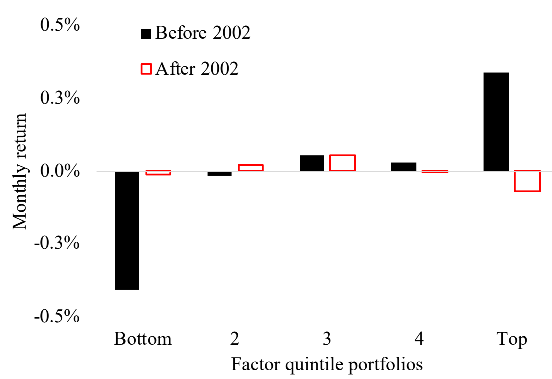
(d) FIT, other momentum-category factors



(e) Return, Momentum



(f) Return, other momentum-category factors



6 Factor Momentum

A recent literature documents that asset pricing factors exhibit momentum (Arnott et al., 2019a; Gupta and Kelly, 2019). That is, factors that recently performed well tends to continue doing so. Strikingly, factor momentum strategy also experienced a sharp profitability with a clear kink in mid-2002 (Panel (a) of Figure 1). In this section, we show that the 2002 change of Morningstar methodology can explain a substantial portion of decline in the profitability of factor momentum.

6.1 Defining Time-Series Factor Momentum (TSFM) Factors

Following Gupta and Kelly (2019), we define time-series factor momentum (TSFM) strategies by implementing the following trading rule: for each factor, trade in the direction of the sign of the recent w months of factor returns.[24] Such a strategy will yield returns:

$$\mathrm{Ret}_{f,t}^{\mathrm{TSFM},w} = \mathrm{Ret}_{f,t} \cdot \mathrm{sign}(\mathrm{Ret}_{f,t-w\to t-1}). \tag{8}$$

Because the literature has not converged on a single look back window, we try look-back windows of $w = 1, 12, 24$, and 60 months. We also define the factor momentum strategy with look back window w as the investment strategy that invests equally in all TSFM factors with window w.

Consistent with the prior literature, we find that these TSFM factors are quite different from the original factors. First, we examine the fraction of time that the TSFM factors are long, rather than short, the original factors. If they are long 100% of the time, then they are identical to the original factors; if they are unrelated to the original factors, they would be long 50% of the time. The second column of Table 3 shows that, for the look-back windows considered, the TSFM factors are only long during 52% to 65% of the time. Second, we find that the average monthly return correlation of TSFM factors and the original factors are only in the range of 7% to 33%. Therefore, judging from both criterion, they are quite different from the original factors.

[24]We focus on TSFM rather than cross-sectional factor momentum (CSFM) as TSFM is found to deliver higher returns (Gupta and Kelly, 2019; Ehsani and Linnainmaa, 2019).

Table 3. Time-Series Factor Momentum Factors versus Original Factors

For each asset pricing factor (e.g., size), we define time-series factor momentum (TSFM) factors as the strategy that is long (or short) the original factor if the previous w month return is positive (or negative), as in equation (8). In this table, we vary the look-back window w and examine how similar are the TSFM factors relative to the original factors. Column 2 shows the average fraction of time that they are long the original factors. Column 3 shows the average monthly return correlation of the TSFM factors with the original factors.

Look-back window	Fraction of time long	Return correlation with original factor
1	52.3%	6.7%
12	56.7%	22.1%
24	58.4%	22.3%
60	65.1%	33.5%

6.2 Rating Reform and Profitability Drop of Factor Momentum

We now examine how much the factor momentum is affected by the 2002 methodology change. We compute ExpSum(ΔRating) for the factor momentum strategies and plot the values before and after 2002 in Panels (a) and (b) of Figure 11. For all look back windows, there is a marked decline in ExpSum(ΔRating).

Further, we confirm that the decline is almost entirely due to the disappearance of style-level rating variation after 2002. For each mutual fund j, we decompose its rating into a style and an idiosyncratic component:

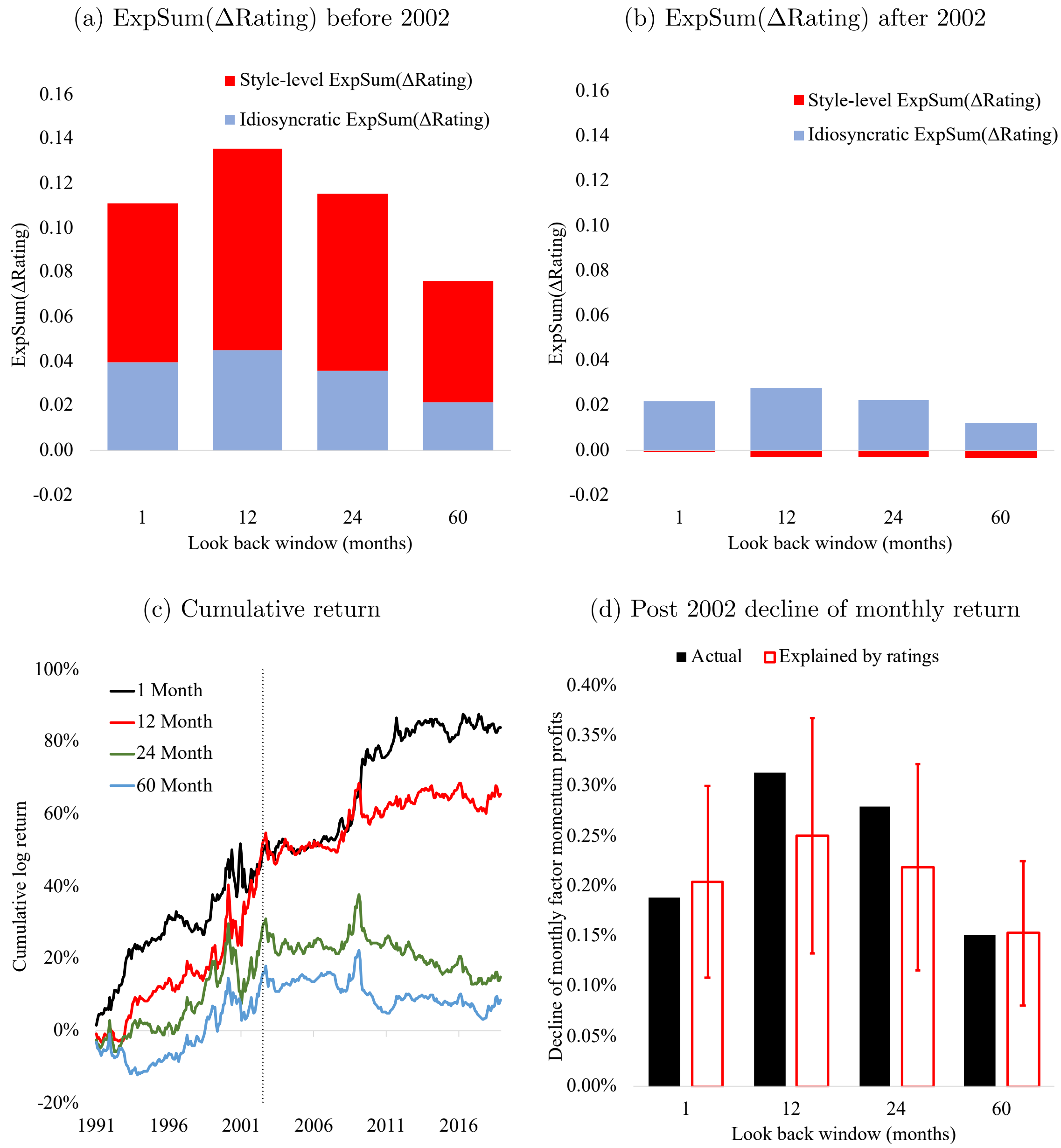
$$\text{Rating}_{j,t} = \text{Style Rating}_{j,t} + \text{Idiosyncratic Rating}_{j,t}$$

where Style Rating$_{j,t}$ is the value-weighted average rating of all funds in the same style and the idiosyncratic rating is defined as a residual. We then use these two components to separately decompose ExpSum(ΔRating) into style and idiosyncratic level variation. They are shown using different colors in Panels (a) and (b) in Figure 11. As expected, the reduction of ExpSum(ΔRating) is almost entirely due to the style-level component.

Panel (c) of Figure 11 plots the cumulative log return of the various factor momentum strategies. They all experienced a "kink" in 2002, as expected. Panel (d) shows the post 2002 decline of their monthly returns. All forms of factor momentum expe-

Figure 11. Factor Momentum, Before versus After June 2002

Panels (a) and (b) shows the lagged rating changes (ExpSum(ΔRating) for factor momentum strategies with different look back windows before and after June 2002. Rating variation is further decomposed into style-level variation (red bars) and the residual (idiosyncratic variation, blue bars). Panel (c) plots the cumulative log return of the factor momentum strategies, and Panel (d) shows the decline of monthly returns after June 2002 and our estimate of how much can be explained by ratings ($\lambda \times$ (ExpSum(ΔRating)$_{\text{before 2002}}$ − ExpSum(ΔRating)$_{\text{after 2002}}$)). The red whiskers are two standard error bands that take into account uncertainty in estimating λ.



rienced declines in profitability, with the largest declines happening for the 12- and 24-month look-back windows. For example, monthly returns of the factor momentum strategy with 12-month look-back window dropped from 39 bps per month to 8 bps per month, a drop of 31 bps. The red bars are our estimate of how much can be explained by ratings ($\lambda \times (\text{ExpSum}(\Delta\text{Rating})_{\text{before 2002}} - \text{ExpSum}(\Delta\text{Rating})_{\text{after 2002}})$), with the whiskers representing two standard error bands that take into account uncertainty in λ estimation. As in Section 5.3, we use the λ estimated using the 12-month window around June 2002. Based on this estimate, Morningstar can explain most of the declines in factor momentum profitability.

7 Conclusion

Stock market factors are perhaps the most researched topic in asset pricing and are central in modeling the cross-section of expected stock returns. Scholars continue to debate the source of cross-sectional return predictability. Some argue that factor profits reflect compensation for risk, while others argue for mispricing-based explanations. Recently, researchers have also questioned whether these factors came from data-mining in the research process.

In this study, we show causal evidence that a significant fraction of factor profitability during the 1991–2002 period can be attributed to mispricing driven by style-level performance-chasing demand. Before June 2002, Morningstar rated funds using absolute fund returns; therefore, funds pursuing investment strategies associated with recently-outperforming styles were rated higher than funds in recently-underperforming styles. As investors chased fund ratings, their behavior led to large style-level positive feedback trading. The price pressure on the best-performing styles caused by this mechanism led to higher returns for momentum and factors related to momentum—the most profitable factor strategies during the 1991–2002 period.

In June 2002, Morningstar changed its methodology to make ratings unrelated to past style performance. As a consequence, the positive feedback flow pressures halted, and factor premia weakened dramatically ever since. This methodology change can explain the mysterious factor profitability "kink" in mid-2002, and can also quantitatively account for around half of the momentum profits decline after 2002 relative to the full sample. This rating methodology change also provides clean causal identification which is rarely achieved in asset pricing research.

It is possible that the role of correlated demand in determining asset pricing is even bigger than what is documented here. We estimate that between a third to half of the momentum factor premium during the 1991–2002 period can be explained solely by the correlated demand driven by Morningstar ratings. Correlated demand, however, can arise from sources other than Morningstar ratings, such as institutional demand for certain styles (Froot and Teo, 2008; Koijen and Yogo, 2019) or the performance-chasing behavior of index-linked products (Broman, 2016). Hence, it is possible, and even likely, that correlated demand plays a central role in explaining the cross-section of asset returns.

References

Aragon, George O, Emma Li, and Laura Anne Lindsey, 2018, Exploration or exploitation? hedge funds in venture capital, *Hedge Funds in Venture Capital (September 18, 2018)* .

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

Arnott, Robert D., Campbell R. Harvey, Vitali Kalesnik, and Juhani T. Linnainmaa, 2019b, Alice's adventures in factorland: Three blunders that plague factor investing, *Journal of Portfolio Management* 45, 18–36.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen, 2013, Value and momentum everywhere, *Journal of Finance* 68, 929–985.

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *Journal of Finance* 61, 1645–1680.

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

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

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

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, 2020, Non-fundamental demand and style returns, Working paper, The Ohio State University.

Berk, Jonathan B., Richard C. Green, and Vasant Naik, 1999, Optimal investment, growth options, and security returns, *Journal of Finance* 54, 1553–1607.

Betermier, Sebastien, Laurent E. Calvet, and Evan Jo, 2019, A supply and demand approach to equity pricing, Working paper, McGill 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.

Calluzzo, Paul, Fabio Moneta, and Selim Topaloglu, 2019, When anomalies are publicized broadly, do institutions trade accordingly?, *Management Science* 65, 4555–4574.

Campbell, John Y., Stefano Giglio, Christopher Polk, and Robert Turley, 2018, An intertemporal CAPM with stochastic volatility, *Journal of Financial Economics* 128, 207–233.

Campbell, John Y., and Tuomo Vuolteenaho, 2004, Bad beta, good beta, *American Economic Review* 94, 1249–1275.

Chen, Yong, Zhi Da, and Dayong Huang, 2019, Arbitrage trading: The long and the short of it, *Review of Financial Studies* 32, 1608–1646.

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.

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

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

Daniel, Kent D., David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under-and overreactions, *Journal of Finance* 53, 1839–1885.

Daniel, Kent D., David Hirshleifer, and Avanidhar Subrahmanyam, 2001, Overconfidence, arbitrage, and equilibrium asset pricing, *Journal of Finance* 56, 921–965.

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 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 Kenneth R French, 1996, Multifactor explanations of asset pricing anomalies, *The journal of finance* 51, 55–84.

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).

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

Gomes, Joao, Leonid Kogan, and Lu Zhang, 2003, Equilibrium cross section of returns, *Journal of Political Economy* 111, 693–732.

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

Green, Jeremiah, John R.M. Hand, and Mark T. Soliman, 2011, Going, going, gone? The apparent demise of the accruals anomaly, *Management Science* 57, 797–816.

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

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

Hansen, Lars Peter, John C. Heaton, and Nan Li, 2008, Consumption strikes back? Measuring long-run risk, *Journal of Political Economy* 116, 260–302.

Hanson, Samuel G., and Adi Sunderam, 2013, The growth and limits of arbitrage: Evidence from short interest, *Review of Financial Studies* 27, 1238–1286.

Harvey, Campbell R., 2017, Presidential address: The scientific outlook in financial economics, *Journal of Finance* 72, 1399–1440.

Harvey, Campbell R., and Yan Liu, 2019, A census of the factor zoo, Working paper, Duke University.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, . . . and the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

Hirshleifer, David, 2001, Investor psychology and asset pricing, *Journal of Finance* 56, 1533–1597.

Hong, Harrison, and Jeremy C. Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *Journal of Finance* 54, 2143–2184.

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

Hou, Kewei, Chen Xue, and Lu Zhang, 2019, Replicating anomalies, *Review of Financial Studies* forthcoming.

Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3–53.

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

Khandani, Amir E., and Andrew W. Lo, 2011, Illiquidity premia in asset returns: An empirical analysis of hedge funds, mutual funds, and US equity portfolios, *Quarterly Journal of Finance* 1, 205–264.

Kogan, Leonid, and Dimitris Papanikolaou, 2014, Growth opportunities, technology shocks, and asset prices, *Journal of Finance* 69, 675–718.

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

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

Lee, Jieun, and Joseph P. Ogden, 2015, Did the profitability of momentum and reversal strategies decline with arbitrage costs after the turn of the millennium?, *Journal of Portfolio Management* 41, 70–83.

Li, Dongmei, and Lu Zhang, 2010, Does Q-theory with investment frictions explain anomalies in the cross section of returns?, *Journal of Financial Economics* 98, 297–314.

Liu, Laura Xiaolei, Toni M. Whited, and Lu Zhang, 2009, Investment-based expected stock returns, *Journal of Political Economy* 117, 1105–1139.

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

Lou, Dong, and Christopher Polk, 2018, Comomentum: Inferring arbitrage activity from return correlations, Working paper, London School of Economics.

Marquering, Wessel, Johan Nisser, and Toni Valla, 2006, Disappearing anomalies: A dynamic analysis of the persistence of anomalies, *Applied Financial Economics* 16, 291–302.

McLean, R. David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability?, *Journal of Finance* 71, 5–32.

Pástor, Ľuboš, and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

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.

Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan, 2012, The short of it: Investor sentiment and anomalies, *Journal of Financial Economics* 104, 288–302.

Stambaugh, Robert F., and Yu Yuan, 2016, Mispricing factors, *Review of Financial Studies* 30, 1270–1315.

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.

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

Zhang, Lu, 2019, Q-factors and Investment CAPM, Working paper, National Bureau of Economic Research.

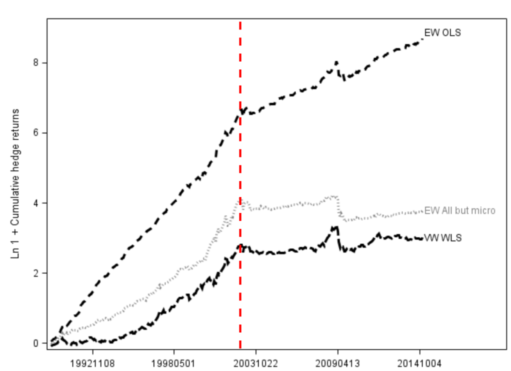
Appendix A Previous Evidence of Factor Returns and Kink around June 2002

Scholars have previous identified a kink around the performance of return factors and specifically the momentum factor. We present relevant charts from two recent publications in Appendix Figure A.1. Panel (a) shows a chart from Green et al. (2017), summarizing the average performance (equally-weighted as well as value-weighted) of 94 factors. Panel (b) shows a chart from Daniel and Moskowitz (2016), summarizing the performance to momentum strategy. In both charts, we added a dashed line for June 2002.

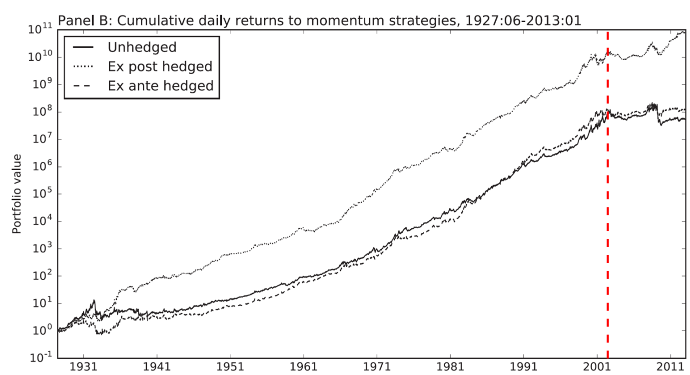
Figure A.1. Previous Evidence of Factor Returns and Kink around June 2002

The figure presents charts that appeared in Green et al. (2017) (Panel (a)) and in Daniel and Moskowitz (2016) (Panel (b)), showing a kink in the cumulative returns of 94 factors and of momentum strategy, respectively. In both panels, we added a red dashed line representing the approximate location June 2002 on the timeline.

(a) Green, Hand, and Zhang (2017, Fig 3)



(b) Daniel and Moskowitz (2016, Fig 4b)



Appendix B Data and Measures

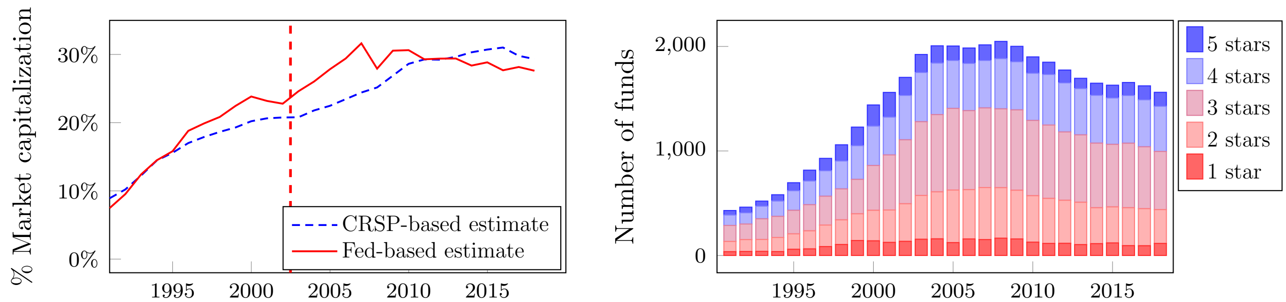
B.1 Mutual fund data

To gauge the importance of the mutual fund sector in the U.S. stock market, Panel (a) of Figure B.2 plots the value-weighted fraction of stocks held by mutual funds during 1991–2018. We present two estimates, one based on Federal Reserve Flow of Funds and the other based on CRSP mutual fund database. Both estimates show that the aggregate holding of

Figure B.2. Number of Funds and the Average TNA over Time

The figure shows the number of funds in each Morningstar star classification (bars; left-hand scale), as well as the average TNA (line; right-hand scale). Fund TNA (total net assets) data comes from CRSP, and Morningstar ratings come from Morningstar Direct.

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



mutual funds increased dramatically over our sample period, starting from holding approximately 10% of the U.S. stock market in 1991 to approximately 30% by 2018. Panel (b) of Figure B.2 shows the number of mutual funds used in our sample along with the distribution of Morningstar ratings.

B.2 Asset pricing factors

Table B.1 shows the list of 49 asset pricing factors used in this paper. Following Hou et al. (2019), we classify them into six categories: intangibles, investment, momentum, profitability, trading frictions, and value/growth.

Table B.1. Asset Pricing Factors

The table lists the factors that are used in this study. The categorization is based on Hou et al. (2019).

Category	Factor	Publication
Intangibles (6)	Industry concentration	Hou and Robinson (JF 2006)
	Operating leverage	Novy-Marx (RF 2010)
	Firm age	Barry and Brown (JFE 1984)
	Advertising expense	Chan, Lakonishok, and Sougiannis (JF 2001)
	R&D expense	Chan, Lakonishok, and Sougiannis (JF 2001)
	Earnings persistence	Francis, LaFond, Olsson, and Schipper (AR 2004)
Investment (13)	Abnormal capital investment	Titman, Wei, and Xie (JFQA 2004)
	Accruals	Sloan (AR 1996)
	Asset growth	Cooper, Guylen, and Schill (JF 2008)
	Five-year share issuance	Daniel and Titman (JF 2006)
	Growth in inventory	Thomas and Zhang (RAS 2002)
	Industry-adjusted CAPEX growth	Abarbanell and Bushee (AR 1998)
	Investment growth	Xing (RFS 2008)
	Investment-to-assets	Hou, Xue, and Zhang (RFS 2015)
	Investment-to-capital	Xing (RFS 2008)
	Net operating assets	Hirshleifer, Hou, Teoh, and Zhang (JAE 2004)
	Net working capital changes	Soliman (AR 2008)
	One-year share issuance	Pontiff and Woodgate (JF 2008)
	Total external financing	Bradshaw, Richardson, and Sloan (JAE 2006)
Momentum (5)	52-week high	George and Hwang (JF 2004)
	Intermediate momentum $(t-7, t-12)$	Novy-Marx (JFE 2012)
	Industry momentum	Grinblatt and Moskwotiz (1999)
	Momentum $(t-2, t-6)$	Jegadeesh and Titman (JF 1993)
	Momentum $(t-1, t-12)$	Jegadeesh and Titman (JF 1993)
Profitability (14)	Cash-based profitability	Ball, Gerakos, Linnainmaa, and Nikolaev (JFE 2016)
	Change in asset turnover	Soliman (AR 2008)
	Distress risk	Campbell, Hilscher, and Szilagyi (JF 2008)
	Gross profitability	Novy-Marx (JFE 2013)
	Ohlson's O-score	Griffin and Lemmon (JF 2002)
	Operating profitability	Ball, Gerakos, Linnainmaa, and Nikolaev (JFE 2016)
	Piotroski's F-score	Piotroski (AR 2000)
	Profit margin	Soliman (AR 2008)
	QMJ profitability	Asness, Frazzini, Israel, Moskowitz, and Pederson (JFE 2018)
	Return on assets	Haugen and Baker (JFE 1996)
	Return on equity	Haugen and Baker (JFE 1996)
	Sales-minus-inventory growth	Abarbanell and Bushee (AR 1998)
	Sustainable growth	Lockwood and Prombutr (JFR 2010)
	Altman's Z-score	Dichev (JFE 1998)
Trading frictions (3)	Size	Banz (JFE 1981)
	Amihud illiquidity	Amihud (JFM 2002)
	Maximum daily return	Bali, Cakici, and Whitelaw (JF 2010)
Value/Growth (8)	Book-to-market	Fama and French (JF 1992)
	Cash flow-to-price	Lakonishok, Shleifer, and Vishny (JF 1994)
	Earnings-to-price	Basu (JF 1977)
	Enterprise multiple	Loughran and Wellman (JFQA 2011)
	Sales growth	Lakonishok, Shleifer, and Vishny (JF 1994)
	Sales-to-price	Barbee, Mukherji, and Raines (FAJ 1996)
	Long-term reversals	Debondt and Thaler (JF 1985)
	Net payout yield	Boudoukh, Michaely, Richardson, and Roberts (JF 2007)

Journals: AR: Accounting Review, FAJ: Financial Analysts Journal, JAE: Journal of Accounting and Economics, JF: Journal of Finance, JFE: Journal of Financial Economics, JFQA: Journal of Financial and Quantitative Analysis, JFR: Journal of Financial Research, RAS: Review of Accounting Studies, RFS: Review of Financial Studies, RF: Review of Finance.

B.3 Morningstar Methodology

We explain Morningstar rating methodology 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.3.2) made the biggest difference to the issues of interest in the study.

B.3.1 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:

$$R_i^T = \prod_{t=1}^{T}(1 + r_{i,t}) - 1, \qquad T \in \{36, 60, 120\}, \tag{9}$$

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:

$$\text{LoadRet}_i^T = R_i^T L_i - R_f^T, \tag{10}$$

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:

$$\text{MnLoadRet}_i^T = \frac{\text{LoadRet}_i^T}{\max(R_f, \text{AvgLoadRet}^T)}, \tag{11}$$

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:

$$\text{Performance}_{i,t} = \text{MnLoadRet}_{i,t}^T - \text{MnRisk}_{i,t}^T. \tag{12}$$

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

$$\text{Risk}_i^T = \frac{\sum_{t=1}^{T} - \min(r_{i,t} - r_t^f, 0)}{T}, \tag{13}$$

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

$$\text{MnRisk}_t^T = \frac{\text{Risk}_i^T}{\text{AvgRisk}^T}. \tag{14}$$

After June 2002, Morningstar began to conduct risk adjustment in a slightly different way.[25] Morningstar summarizes a fund's past performance using the so-called Morningstar risk-adjusted return (MRAR):

$$\text{MRAR}_i^T(\gamma) = \left[\frac{1}{T}\sum_{t=1}^{T}(1 + r_{i,t} - r_t^f)^{-\gamma}\right]^{-\frac{12}{\gamma}} - 1, \tag{15}$$

where $r_{i,t} - r_t^f$ is the geometric return in excess of the risk-free rate after adjusting for loads,[26] and $\gamma = 2$ is the risk aversion coefficient.

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

[25] 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).

[26] 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.[27] 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.3.2 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 the 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 negligible after June 2002 (Panel (b) in Figure 2).

[27] 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.

Appendix C Additional Empirical Results

C.1 Estimating Stock-Level Effect of Ratings on Returns

In this section, we first show the stock-level effect of rating on flows and then flows on returns. We then estimate the decay parameter δ in specifying ExpSum(ΔRating).

We use Fama-MacBeth regressions to estimate the chain of dynamic effects: *i*) the response of fund flow-induced trading (FIT) to Morningstar rating changes, and *ii*) the response of stock returns to flow-induced trading. The regressions at the the stock level and value-weighted.[28]

First, we estimate the FIT response to lagged stock-level rating changes:

$$\mathrm{FIT}_{i,t} = a + b_1 \cdot \Delta\mathrm{Rating}_{i,t-1} + ... + b_{36} \cdot \Delta\mathrm{Rating}_{i,t-36} + X_{i,t} + u_{j,t}, \tag{16}$$

where $\Delta\mathrm{Rating}_{i,t}$ is the month t rating change of stock i, and controls $X_{i,t}$ include 36 monthly lags of FIT and stock returns. The cumulative response coefficients $(b_1, b_1+b_2, ...)$ are plotted in Panel (a) of Figure C.3. In response to a one-star change in rating, stocks experience an average of 8% additional FIT in the subsequent 3 yaers. This result is consistent with prior research showing that, controlling for past fund performance, discrete changes in ratings create continued fund flow response that last for months (Del Guercio and Tkac, 2008; Reuter and Zitzewitz, 2015).

Then, we estimate the response of stock returns to FIT:

$$\mathrm{Ret}_{i,t} = a + c_0 \cdot \mathrm{FIT}_{i,t} + c_1 \cdot \mathrm{FIT}_{i,t-1} + ... + c_{36} \cdot \mathrm{FIT}_{i,t-36} + u_{i,t}. \tag{17}$$

We plot the cumulative response in Panel (b) of Figure C.3. Each 1% increase in mutual fund ownership due to flows leads to immediate price pressures of approximately 0.8% 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).

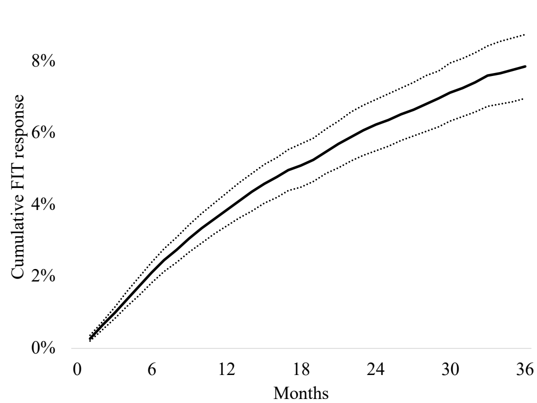
Combining these two effects, we expect that stock returns will also be affected by rating

[28] 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.

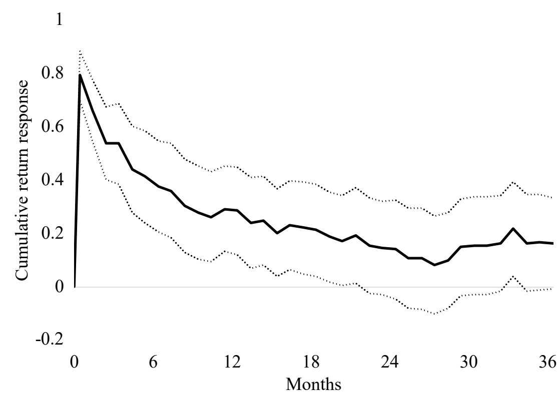
Figure C.3. Price Impact of Rating and Flows.

Panel (a) shows the cumulative response of flow-induced trading (FIT) to changes in stock-level ratings. Panel (b) shows the cumulative response of stock returns to FIT. Following Lou (2012), FIT is defined as the nondiscretionary trading induced by mutual fund managers proportionally adjusting existing portfolio holdings in response to fund flows.In both (a) and (b), the dashed lines show two standard error bands. Panel (c) shows the *non*-cumulative response coefficients 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 in the 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. The decile porfoilo returns are demeaned by period to remove the overall trend of stock value increasing over time.

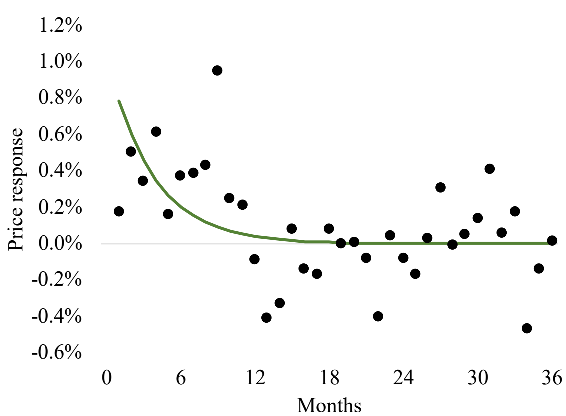
(a) FIT response to rating changes



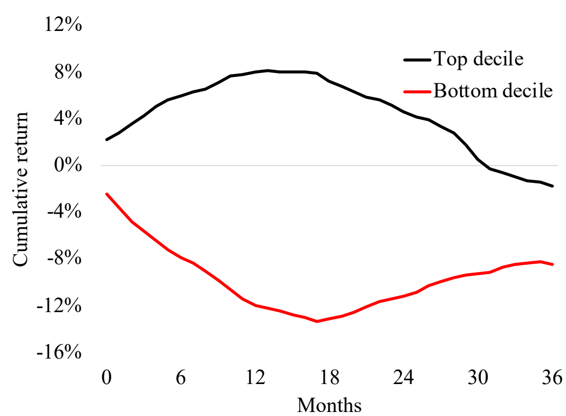
(b) Stock price response to FIT



(c) Rating → return



(d) Price paths of stocks



changes, and particularly by *the most recent* rating changes. 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, the price pressures generated by their initial price impact are already reverting, so the two effects will partially cancel each other out.

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 and plot the coefficients in Panel (c) of Figure C.3. As expected, more recent rating changes are more impactful, and the coefficients on more distant rating changes converge towards zero.

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

$$\text{ExpSum}(\Delta\text{Rating})_{i,t-1} = \sum_{k=1}^{12} \tau_k \cdot \Delta\text{Rating}_{i,t-k}, \tag{18}$$

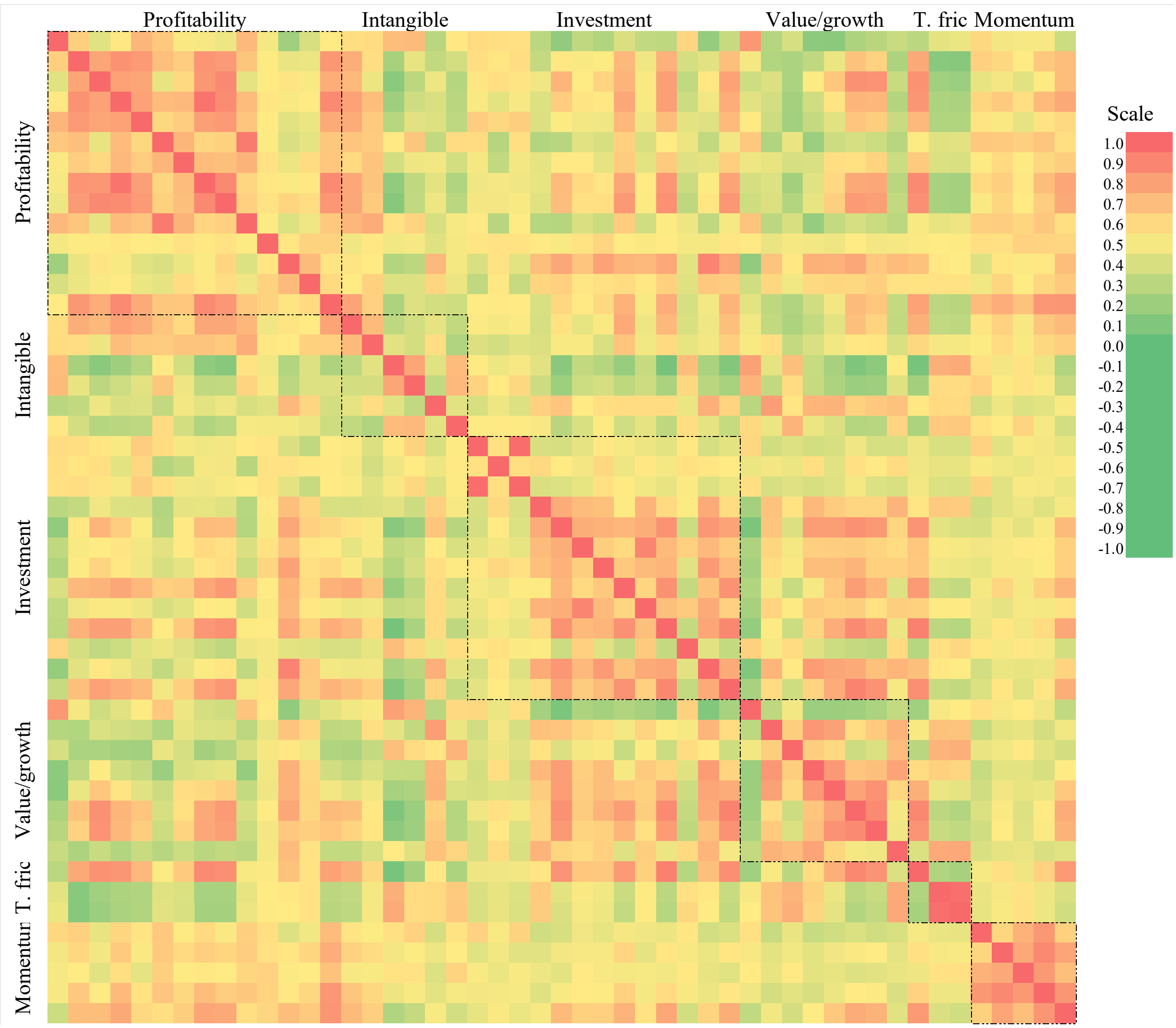
where $\sum_{k=1}^{12} \tau_k = 12$ and the weights decay with factor $\delta = 0.764$, which is estimated from a least-squares fit to the cumulative response (Panel (c) of Figure C.3).

We end this section with an illustration of price pressures. If the price movements predicted by ExpSum(ΔRating) are truly price pressures, we expect it to eventually revert. In Panel (d) of Figure C.3, we sort stocks in each month t by $\text{ExpSum}(\Delta\text{Rating})_{t-1}$ into NYSE deciles and plot the cumulative returns of the top and bottom deciles. To focus on the cross-sectional dispersion, the returns of the decile portfolios are demeaned by period. As shown in the Figure, the rating-induced price movements in the top and bottom decile portfolios do revert after 3 years.

In estimating factor-level price impact parameter λ in Section 5.3, we adjust the return correlation between factors using a feasible general least squares approach. Figure C.4 uses a heatmap to visualizes the return correlation of factors clustered by the Hou et al. (2019) categories. Higher values are shown in red and lower values are shown in green. There appears to be high correlations within certain categories of factors, such as the momentum category, the trading friction category, and the profitability categories. There are also negative correlations across certain categories.

Figure C.4. Correlation Matrix of Factor Returns

This correlation matrix is estimated using monthly returns of all factors over the full sample of 1991–2018. Higher correlations are colored red and lower correlations are colored green. Following Hou et al. (2019), we classify the 49 asset pricing factors into six categories and order them accordingly. The second-to-last category is the trading friction category.



C.2 Estimating Explanatory Power using Time-Varying λ

In Section 5.3, we estimate the explanatory power of Morningstar ratings on the post-June 2002 factor profitability decline using the rating price impact parameter λ estimated from the 2002 shock. To accommodate possible changes in λ, we estimate time-varying λ through factor return-predicting regression using rolling 10 year windows. The results are plotted in Panel (a) of Figure C.5. There is indeed variation over the sample period, with λ being higher in the earlier half of the sample and dropping after the pre-2002 period exits the estimation window. We then calculate the return explained by ratings in month t as:

$$\hat{\lambda}_t \cdot \text{ExpSum}(\Delta\text{Rating})_{f,t-1},$$

where $\hat{\lambda}_t$ is estimated using a 10-year window centered around month t.[29] The results, as a parallel to Table 2, are shown in Panel (b) of Figure C.5. Morningstar ratings can approximately explain a third of the overall drop of momentum profits after June 2002.

[29]For years before 1995, we use the λ estimate using the 10 year window centered at 1995; for years after 2013, we use the λ estimate centered at 2013.

Figure C.5. Explanatory Power of Morningstar Ratings on Post-June 2002 Profitability Decline using Time-Varying λ Estimate

Panel (a) plots estimations of the rating price impact parameter λ through predictive regressions of factor returns on $\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$ using rolling 10-year centered windows. The graph starts in 1995 and ends in 2013 due to needing 10 years for each estimation. The dashed blue lines represent two standard error bands. The vertical red dashed lines mark the methodology change event. Panel (b) tabulates how much the post-2002 factor profitability decline can be explained when applying the time-varying λ estimate. Column (1) examines the momentum factor; column (2) examines the other factors in the momentum category; column (3) examines all the other factors. The "Explained by Morningstar" values are computed by multiplying the time-varying λ, estimated in Panel (a) with the lagged rating changes ($\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$).

(a) Price impact parameter λ, estimated from 10 year rolling windows

(b) Explanatory power on factor return decline using time-varying λ

	Factors			Difference	
	Momentum Factor	Other Momentum-Category Factors	All other Factors	(1) − (3)	(2) − (3)
	(1)	(2)	(3)		
Actual return drop	−0.70%	−0.80%	−0.29%	−0.41%	−0.51%
Explained by Morningstar	−0.21%	−0.16%	−0.04%	−0.17%	−0.12%
Fraction explained	30.1%	19.9%	14.6%	41.3%	23.0%