Style Investing, Positive Feedback Loops, and Asset Pricing Factors

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

VERY PRELIMINARY, PLEASE DO NOT CITE

February 8, 2020

Abstract

We show that style investing has first-order causal effect on asset pricing factors (anomalies). Investment managers invest according to styles (e.g., growth or value) and investors direct flows to the best-performing funds. In turn, fund managers increase their investment in the same style, ultimately resulting in return factors and momentum in the underlying securities. We demonstrate this causal mechanism using the 2002 change in Morningstar's rating methodology, which weakened the link between capital flows and the performance of particular styles. This change largely accounted for the disappearance of return factors in the early 2000s, where the most affected factors were momentum and those most aligned with the treated styles.

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

*We thank Sylvester Flood (Morningstar), Paul Kaplan (Morningstar), and Andrei Shleifer for helpful comments. 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@email.arizona.edu, and songy18@uw.edu.

1 Introduction

Over the last four decades, the asset pricing literature discovered hundreds of factors[1] which span the cross-section of stock returns (Hou, Xue, and Zhang, 2020), and whose premia exhibit momentum (Ehsani and Linnainmaa, 2019). Financial economists debate about the origins of these factors. Some propose that factors reflect risk premia arising from fundamental economic constructs, such as state-dependent dividends[2] or firms' investment opportunities.[3] Others argue that factors are systematic mispricings,[4] driven by correlated demand from investors ("style investing:" De Long, Shleifer, Summers, and Waldmann, 1990; Hong and Stein, 1999; Barberis and Shleifer, 2003). As these explanations are non-mutually exclusive and often have similar predictions, it is hard to disentangle them. The literature has mostly been relying on sorts by characteristics and structural models to measure the effects of specific explanations, which are open to the critique that other, unobservable, factors could determine equilibrium prices. Furthermore, researchers have been puzzled in recent years by the fact that return factors have largely disappeared from the U.S. equity market in the early 2000s (e.g., Daniel and Moskowitz, 2016; Green, Hand, and Zhang, 2017; Arnott, Harvey, Kalesnik, and Linnainmaa, 2019b).

In this article, we present two important and related findings. We present *causal* evidence that stock market anomalies largely reflect mispricings driven by style investing and show that a simple institutional change dramatically disrupted the effects of style investing and therefore caused anomalies to disappear in the early 2000s. We identify the causal relationship by studying an apparently-innocuous change in rating methodology of Morningstar in 2002. This event effectively muted the "style" component from the chasing behavior of mutual fund investors. As a result, the aggregate investment by mutual funds became less correlated with investment styles, which ultimately led to the abrupt disappearance of return factors from the U.S. stock market. Figure 1 shows our main results. Following the rating methodology change in mid-2002, return factors lost most of their premium (Panel (a)). In the cross-section, when we sort factors by their associated with the impacted styles, and the most impacted ones experienced greater declines in premium (Panel (b)).

Style investing is a mechanism through which securities become mispriced through correlated demand. The idea is that investors tend to classify investments by broad categories or styles (e.g., value stocks, technology stocks), and that they chase past performance at the

[1] In this paper, we use factors and anomalies interchangeably.

[2] E.g., Fama and French (1993), Schwert (2003), and Fama and French (2008).

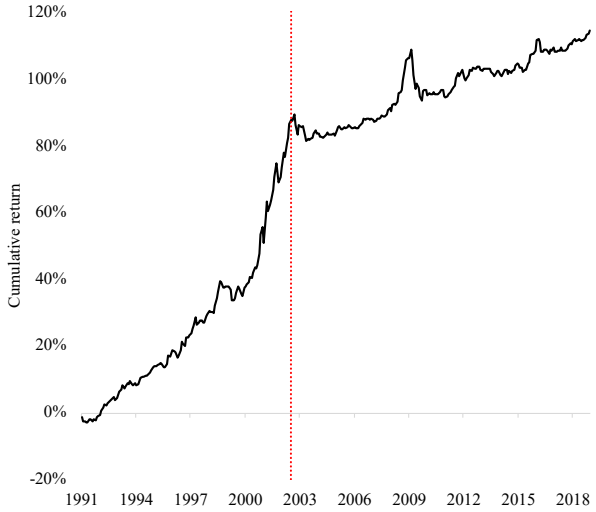
[3] E.g., Liu, Whited, and Zhang (2009), Li and Zhang (2010), Hou, Xue, and Zhang (2015), Betermier, Calvet, and Jo (2019), and Zhang (2019).

[4] E.g., Hirshleifer (2001), Lamont and Thaler (2003), and Stambaugh and Yuan (2016).

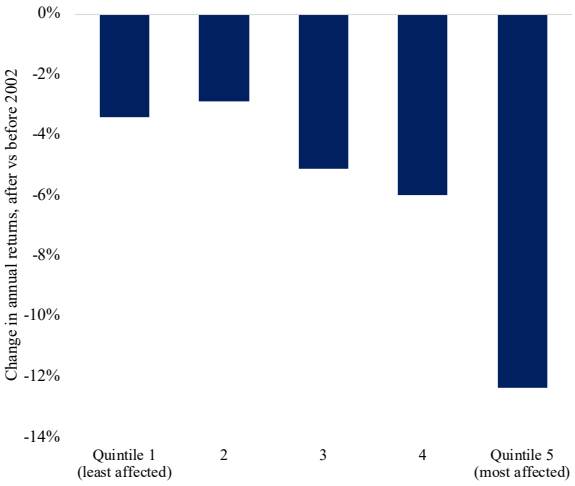
Figure 1. Return Factors Around Morningstar's Rating Methodology Change

Panel (a) shows the cumulative monthly returns of 49 stock market anomalies. Panel (b) shows the relation between the change in anomalies' monthly expected returns (12 months after June 2002 relative to 12 months before June 2012) and the predicted change in Morningstar rating as of June 2001.

(a) Cumulative averge factor returns.



(b) Change in factor returns after 2002.



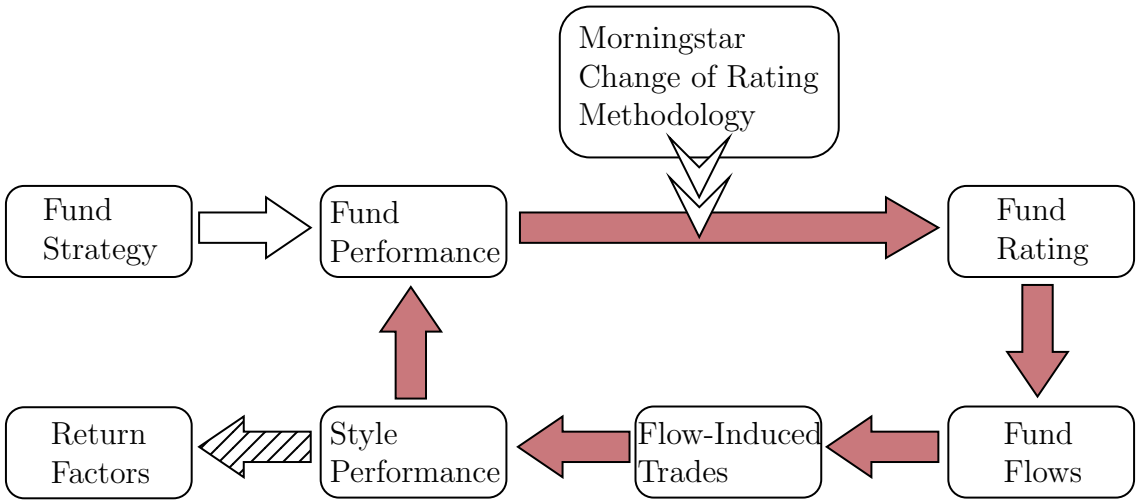
style level. As a result, investors chase styles that recently performed well which translates to further price increases, generating a positive feedback loop. Barberis and Shleifer (2003) propose that such behavior generates asset pricing factors, sometimes also called anomalies in the literature.

Figure 2 illustrates how the hypothesized causal links operate through mutual funds. Mutual funds specialize in investment styles such as growth, value, or can specialize in specific industries. Investors chase past performance, and therefore allocate capital to the best performing funds. Because much of mutual fund performance are simply due to style-level performance of stocks, funds in the best performing style are likely to receive a disproportionally large share of the flows. Mutual fund managers use the new capital to expand their portfolios, and by doing so, they push up stock prices in the winning style even more, creating a positive feedback loop. According to the hypothesis, return factors (anomalies) are a byproduct of this positive feedback loop. Because the positive feedback loop results in price pressure in a subset of stocks that are associated with a particular style, the returns of stocks in this style would display predictability in that is related to past returns (e.g., momentum) or to characteristics (e.g., value and growth).

We study a change in Morningstar's rating methodology of mutual funds that took place in June 2002. The change in Morningstar's rating methodology effectively muted the style

Figure 2. Style Investing Feedback Loop

First, funds with high (low) ratings experience in (out) flows. Second, fund managers buy (sell) their fund holdings in response to in (out) flows, leading to stock price pressures.



component in ratings, and therefore disrupted the positive feedback loop in Figure 2. Both before and after 2002, mutual fund investors strongly chase star ratings (Del Guercio and Tkac, 2008; Reuter and Zitzewitz, 2015; Ben-David, Li, Rossi, and Song, 2019). However, the methodology change created a regime shift in which funds are top-rated and as a consequence in the way capital flows allocated to mutual funds. Before 2002, top rated funds were those with the best past performance, which was often correlated with their investment style. During this period, we show, capital flows to mutual funds appeared as if they were chasing style performance.

After 2002, Morningstar began rating funds *within styles* (along value/growth, size, and industry dimensions), hence adjusting ratings to the performance of these styles. We show that as a result, capital flows did not chase style performance to a materially lesser degree, hence allowing us to test the relation between style performance chasing and return factors. We hypothesize that the change in Morningstar's rating methodology potentially disrupted the positive feedback loop that existed between style past performance and investors' flows.

We begin our empirical analysis by confirming that indeed the change to the Morningstar rating methodology led to a anticipated changes. Before 2002, top-rated funds were concentrated in a small subset of styles. For instance, among funds that invest in large stocks in 1999, about half of funds that were categorized as growth funds were ranked by Morningstar as 4 or 5 stars, while there almost were no value fund that was ranked as 4 or 5 stars. After 2002, all styles have roughly the same fraction of top-rated funds. The methodology change went unnoticed by investors investors, hence they continued pursuing ratings after

the change as they did before. flows, however, change their sensitivity to style performance.

In an intermediary step in the analysis, we present evidence for the channel we have in mind. Specifically, since we argue that changes in the allocation of flows across funds led to changes in the underlying correlation structure, we ought to show that changes in Morningstar ratings impact flows and those create price pressure on the underlying stocks. We estimate the impulse response functions of a change in Morningstar rating on fund flows, and of fund flows onto returns (due to flow-induced trading). A one-star increase is associated with a cumulative increase in flows of 6% over the following 36 months. A 1% increase in flows translate to an immediate price pressure of 0.8% and a cumulative effect of 0.2% in the long run. When estimated together, a one-star increase causes a 1% to 3% (depending on the method of estimation) increase in the underlying stock prices. Our results corroborate earlier studies that documented parts of the chain of effects. Del Guercio and Reuter (2014) finds that 4- and 5-star funds received most of investors flows. Reuter and Zitzewitz (2015) shows that increases in Morningstar ratings are associated with a sharp increase in flows over the following months. Lou (2012) documents that flow-induced trading by mutual funds create price pressure on stocks, resulting in momentum returns.

Then, we turn to the main part of the analysis, where we present evidence about the relation between the change in the rating methodology and return patterns in the stock market. We focus on three types of return patterns: *i*) momentum returns, *ii*) return factors (anomalies), and *iii*) factor momentum. Given our results that changes in Morningstar ratings translate to price impact at the underlying stocks, we calculate for each stock-month the aggregate rating changes of the mutual funds that hold it (weighted by fund size).

First, for momentum, we confirm the findings of earlier literature that the momentum factor stopped being profitable exactly around the date of the transition between the two rating methodologies of Morningstar (see Israel and Moskowitz, 2013; Daniel and Moskowitz, 2016; Ehsani and Linnainmaa, 2019; Gupta and Kelly, 2019). We explore the changes in Morningstar rating for the stocks along the momentum portfolios. We show that before 2002, stocks in the top deciles of momentum have the highest changes in Morningstar ratings, and those in the bottom deciles have the lowest changes in Morningstar ratings. In contrast, after 2002, there is little such association.

Second, we examine the performance of stock market anomalies around the 2002 event. We construct 49 characteristics-based anomalies, following the method in Hou et al. (2020). We document that the anomaly performance have been largely flat since the early 2000s. Our results echo results in previous studies. Green et al. (2017) document that the performance

of nearly a hundred anomalies in large stocks was flat since early 2000s.[5]

Third, we explore whether the 2002 Morningstar methodology change impacted the performance factor momentum. Several studies show that the return stock factors are persistent, i.e., have a momentum (Arnott, Clements, Kalesnik, and Linnainmaa, 2019a; Gupta and Kelly, 2019). As such, return factors that have recently performed well are likely to perform well in the short horizon. Our analysis shows that the performance of factor momentum was eliminated almost completely after 2002.

Overall, our results show that stock return factors are created due to performance chasing behavior. The beauty of 2002 methodology change of Morningstar is that retail investors did not change their behavior. They were blind to the change—just continued to pursue top-rated funds. What changed was the allocation of their capital to funds. Prior to the 2002 change, the flows of retail investors appear to behave *as if* they chase the best performing styles. Due to the rating methodology change, retail investors behave *as if* they chase idiosyncratic performance.

Our results have far-reaching implications for our understanding of investor behavior, asset prices, and the asset management industry. Over the period that we investigate, investors were oblivious to the change in Morningstar. Despite the fundamental change, investors kept chasing ratings without an attempt to adjust prices to reflect any risk preferences.[6] In the domain of asset pricing, factor premia appear to reflect little compensation for risk, if any. Our results cast doubt about models that attempt explaining return factors as economic fundamentals (e.g., compensation for economic risk, q-theory, information, frictions). Finally, our findings imply that the early evidence about persistence in mutual fund performance (Carhart, 1997; Cohen, Coval, and Pástor, 2005) and the later disappearance of this persistence (Choi and Zhao, 2020) are related to the breaking of the performance-flow positive feedback loop as a result of the change in the Morningstar rating methodology. We explore these issues further in Ben-David, Li, Rossi, and Song (2020).

2 Data and Variable Construction

In this section, we describe the dataset and how we construct the asset pricing factors.

[5]See similar results in Marquering, Nisser, and Valla (2006), Green, Hand, and Soliman (2011), Chordia, Subrahmanyam, and Tong (2014), Jones and Pomorski (2016), McLean and Pontiff (2016), and Arnott et al. (2019b).

[6]See response of Ben-David et al. (2019) to studies claiming that mutual fund investors use the CAPM as their asset pricing model (Barber, Huang, and Odean, 2016; Berk and van Binsbergen, 2016).

2.1 Mutual Fund Data

We obtain monthly fund return and total net assets (TNA) from the CRSP survivorship bias-free mutual fund dataset for the years 1991 to 2018. We use all equity mutual funds. While funds are often marketed to different clients through different share classes, they invest in the same portfolio, and typically the only difference is the fee structure. Therefore, all share classes are aggregated at the fund level using Russ Wermer's MFLINKS (Wermers, 2000). We also obtain latest quarterly fund holdings from Thomson Reuters' S12 data which is based on 13F filings.

Following the fund flow literature, the investment flow for fund j in month t is defined as the net flow into the fund divided by the lagged TNA:

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

We obtain Morningstar ratings and style categories from Morningstar Direct and merge it with the CRSP mutual fund data. Morningstar assign ratings at share class level, and we follow Barber et al. (2016) to aggregate them together at the fund level by AUM-weighting different share classes. We merge CRSP and Morningstar Direct using the matching table from Pástor, Stambaugh, and Taylor (2015).[7] The analysis is restricted to mutual funds with at least $1 million TNA, and flows are winsorized at the 0.5% and 99.5% levels. Because we investigate the impact of lagged ratings on stock returns, we require existence of 12 lagged monthly changes of fund flows, return, and ratings. The resulting sample comprises a total of 2,926 funds with 371,130 fund-month observations.

2.2 Stock-Level Rating and Flow-Induced Trading (FIT)

We are interested in how fund flows and ratings lead to stock-level pressures. Therefore, we define the average Morningstar rating of a stock i in month t as the holding-weighted rating of all funds that hold the stock:

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

To measure the amount of mutual fund trading caused by fund flows, we follow Lou

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

(2012) to calculate flow-induced trading (FIT) for each stock i in each month t:[8]

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

Put simply, FIT is the amount of trading in stock i by all mutual funds that is caused by (instrumented by) fund flows.

2.3 Asset Pricing Factors

Following prior literature, we compute 49 popular stock-level return-predicting characteristics. We then follow the exact procedure in Hou et al. (2020) to construct 49 characteristic-based asset pricing factors. To minimize the impact of microcaps, Hou et al. (2020) use NYSE break points to sort stocks into deciles, and then form factors as long the top decile and short the bottom decile. The decile portfolios are value-weighted to further reduce the impact of microcaps. We obtain stock data from CRSP and characteristic construction also uses accounting data from Compustat. A list of the 49 factors and their definition appear in Appendix Table A.I.

3 Rating Methodology Change and Mutual Fund Flows

In this section, we describe the June 2002 Morningstar methodology change and examine its impact on style-level fund flow pressures.

3.1 Morningstar Ratings Methodology Pre- and Post-2002

Morningstar introduced its mutual fund rating system in 1985 as a service to help investors choose mutual funds, and it quickly became the industry leader. Morningstar assigns 1 to 5 stars to mutual funds based on their ranking on past performance, with minor adjustments for total return volatility and expenses. Depending on the availability of data, the look back horizon for past performance can be 3, 5, or 10 years. For funds with over 10 years of history, the weights of the three horizons are set to be 20%, 30%, and 50%, respectively. The 1 to 5 star ratings are then assigned with fixed proportions (10%, 22.5%, 35%, 22.5%, and 10%). We provide further detail in Appendix A.1.

Morningstar's methodology changed abruptly in June 2002. Following the dotcom crash, many fund managers specializing in technology stocks complained that their fund ratings

[8]Lou (2012) also applies different scaling factors to in flows and out flows. We omit this for simplicity, but our results are robust to using his scaling factors.

dropped sharply from 5 stars to 3 stars or lower just because the technology sector crashed. Therefore, ratings barely reflect their own contributions but only reflect sector level returns which are outside of their control. As a result, Morningstar revised its methodology to allow ranking within style category. The research team, spearheaded by Dr. Paul Kaplan, a Northwestern economics graduate, redesigned the rating system. The main difference from the previous rating system is that the post-2002 fund rankings are done *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 (financial, utilities, etc.). 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.

3.2 The Impact on Style-Level Fund Flows

This seemingly innocent methodology change had far reaching consequences for the mutual fund industry. On an average month, 14% of funds change their rating. In a stark contrast, in June 2002, 54% of funds experienced a rating change. Since Morningstar's new method effectively "double-sorts," i.e., ranks mutual funds on performance within style, the resulting ratings become uncorrelated with the style components of performance. This is shown in Panel A of Figure 3. For instance, in 2000, at the height of the dotcom boom, large growth funds have an average rating of 4 stars while small value funds only have 1.9 stars. The style-level difference became negligible after 2002. Panel B of Figure 3 shows that there is more style-level fund flow dispersion before June 2002. Because investors chase ratings slowly, the flow dispersion does not collapse immediately on June 2002, but does become much more muted afterwards.

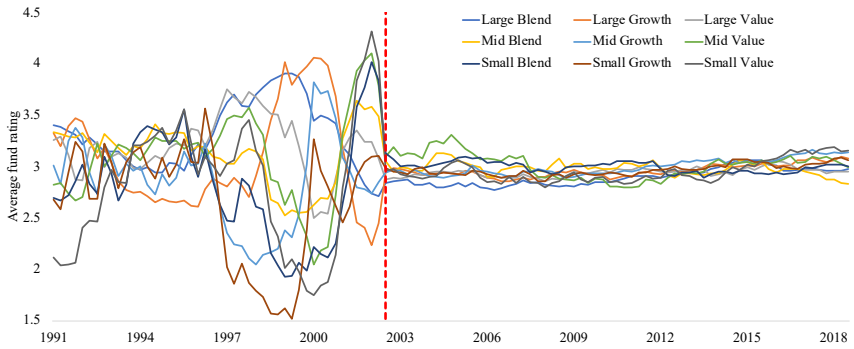
Importantly, for the identification of the style investing mechanism as generating factor anomalies (as opposed to a risk preference), we explore the behavior of the ultimate investors over time. Specifically, we estimate the flow response to lagged ratings at the fund level using five-year rolling window Fama-Macbeth regressions. The result, plotted in Panel C of Figure 3, shows that there is no visible change around 2002. This result shows that mutual fund investors (mostly retail investors) did not change their attitude towards Morningstar ratings over the years. Hence, their risk preference has not likely to have changed.

Although investors do not change the way they chase ratings, due to the methodology change, the allocation of flows across funds is altered. Before 2002, funds in styles that performed well recently receive significantly more flows. To examine this, in each month, we sort the 3×3 size-value styles into top to bottom performers by their recent 12-month

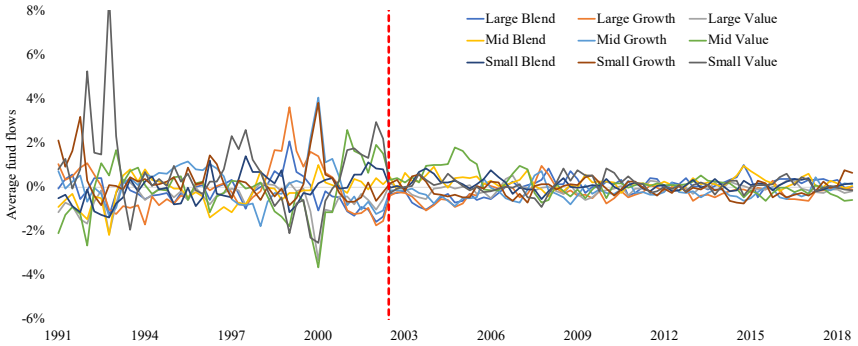
Figure 3. The June 2002 Morningstar Methodology Change

Panels A and B plot the quarterly average fund rating and fund flows by Morningstar 3 × 3 styles. There is large style-level dispersion ratings and flows across styles until June 2002. Panel C: fund flow response to lagged fund rating, estimated using rolling five-year Fama-Macbeth regressions. Dashed lines are two standard deviations. The graph starts in 1996 because we need five years to estimate the coefficient.

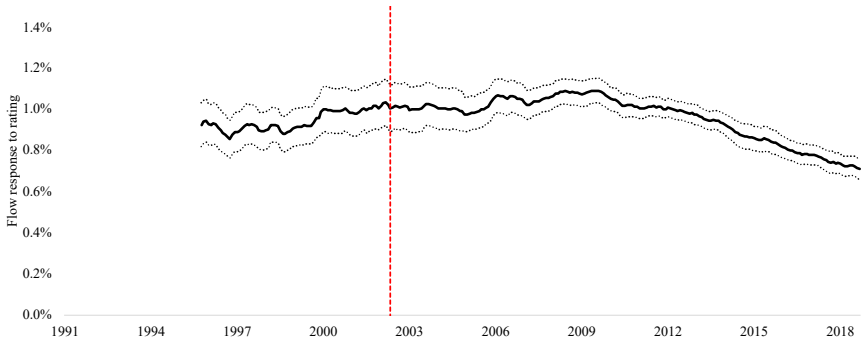
Panel A: Quarterly fund rating by style



Panel B: Quarterly fund flow by style



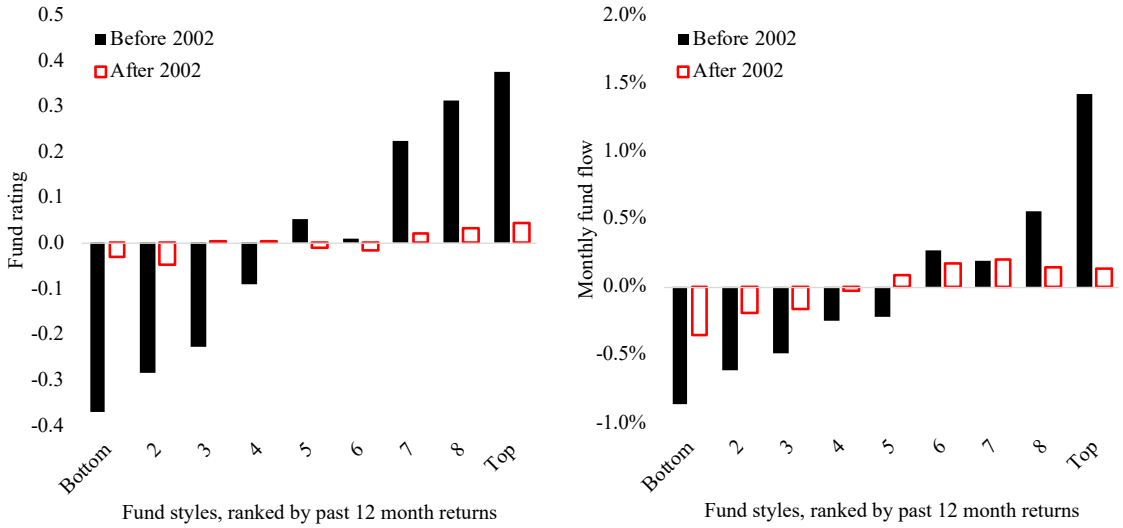
Panel C: Flow response to lagged rating



fund returns, and then we plot their subsequent ratings and flows in Figure 4. All style-level aggregation is AUM-weighted. The graphs are demeaned to focus on cross-sectional patterns

Figure 4. Rating and Flows by Fund Styles, Before versus After 2002

We sort the 3×3 fund styles by the past 12 month fund performance, so style 9 is the top performing style and style 1 is the bottom style. We plot the weighted average fund ratings (left) and fund flows (right) for 3×3 fund styles sorted by past 12 month performance.



across styles. The left panel shows that the average rating difference between funds in the best and worst performing styles is about 0.8 stars before 2002, and almost zero after 2002. Similarly, the right panel shows that funds in the top style receives 2.3% higher flows per month than the bottom style, and that difference reduced to 0.5% after 2002.[9] This result confirms the assertion that post-2002, mutual fund investors no longer appear as chasing style performance.

Overall, our findings point to the conclusion that mutual fund investors mechanically chase ratings. Pre-2002, ratings were highly driven by recent style-level performance, so flows end up chasing style returns. Post 2002, ratings stopped reflecting style performance, causing style-level flows to cease.

4 Effects of the 2002 Event on Momentum

In this section, we examine the effect of the 2002 event on the profitability of the momentum strategy. Momentum is arguably one of the most puzzling return factor to be studied because of its large magnitude (Jegadeesh and Titman, 1993) and because it cannot be easily

[9]Ben-David et al. (2019) show that, in addition to chasing ratings, investors also chase past fund returns, albeit with lower explanatory power.

explained by standard risk-based models.[10]

Momentum is also tied closely to the positive feedback mechanism that is proposed by Barberis and Shleifer (2003). According to the theory, assets that performed well in the past experience a further price pressure due to performance-chasing investors, hence generating momentum. Pre-2002, when ratings reflected mainly the style component of historical performance, the price pressure due to ratings-chasing was focused on a small subset of funds that invested in a similar subset of stocks. Post-2002, fund flows by investors who chase Morningstar ratings was spread across many funds in many different styles. As a result, the downstream price pressure was spread across a wider base of securities, muting the effect.

To see the relationship of momentum with positive feedback trading that is induced by style-level ratings, we examine the ten momentum decile portfolios in Figure 5. In each period, we rank the 3×3 fund styles by their recent 12-month rating changes. Then, for each momentum decile portfolio, we plot the fraction of its market capitalization held by funds in the top versus bottom rating styles in Figure 5. Before 2002, relative to the portfolio with lowest momentum (decile 1), the portfolio with highest momentum is 3% more likely to be held by funds in the top rating style and 7% less likely to be held by the bottom rating style.[11] After Morningstar changed its rating methodology, this difference vanished.[12].

Immediately, this implies that top (bottom) momentum decile portfolios should experience more (less) rating-induced price pressures before 2002, but not so afterwards. We examine the before versus after difference using both a short window of 24 months around the event, as well as a long sample using all available data. Because ratings revert to the mean after 2002, we anticipate some short-term reversal of the momentum factor in the short window. The full sample is intended to capture the steady state effects. The results are plotted in Figure 6.

Panel A plots the ΔRating by momentum deciles. As expected, the top (bottom) momentum portfolios had high ΔRating which reversed in the short window. In the long-run (full sample), the reversion becomes muted and the difference across decile portfolios become small. There is a similar effect on flow-induced trading as shown in Panel B. Over the full

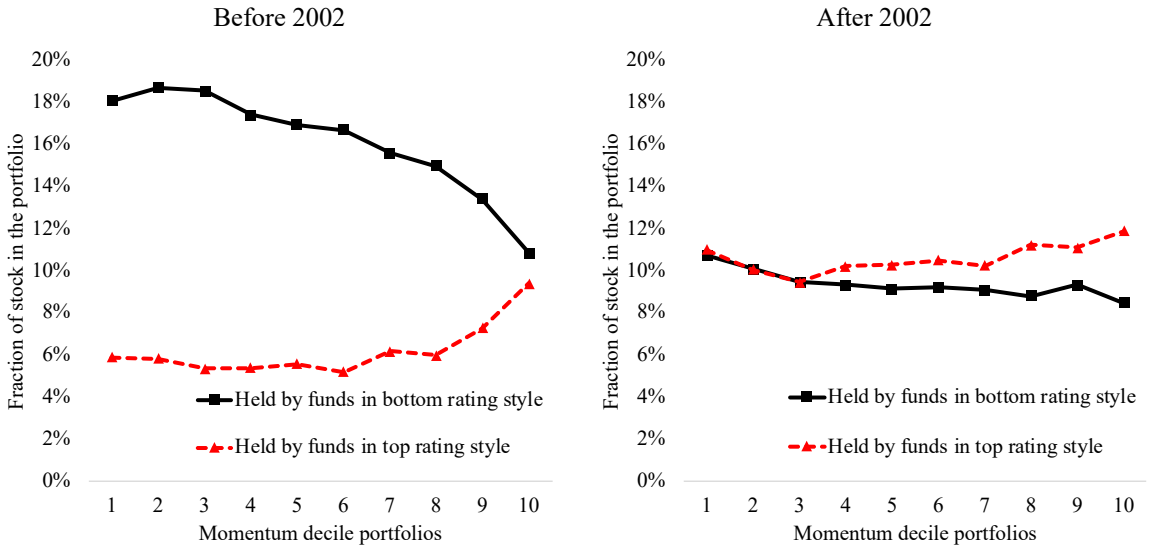
[10]Momentum has been observed for almost a century in the U.S. stock market (Daniel and Moskowitz, 2016) (until the early 2000s) as well as in many other asset classes (Asness, Moskowitz, and Pedersen, 2013). In an interview, Professor Eugene Fama stated that he views momentum as "the biggest embarrassment for efficient markets."

[11]We normalize the aggregate holding fractions by funds in the nine styles to sum to 1.

[12]While Morningstar makes the equal-weighted ratings the same across styles, there can be small differences when value-weighting portfolios, so the styles can still be ranked using rating changes. Also due to the differences in weighting, the resulting fractions in the right panel of Figure 5 are close to, but not exactly equal to, $1/9 \approx 11.1\%$.

Figure 5. The Alignment of Momentum with Top versus Bottom Fund Styles

In each period, we sort the 3×3 fund styles by their recent 12-month rating changes from bottom to top. Then, for each momentum decile portfolio, we plot the value-weighted fraction of stocks held by funds in the top (red) or bottom (black) styles.



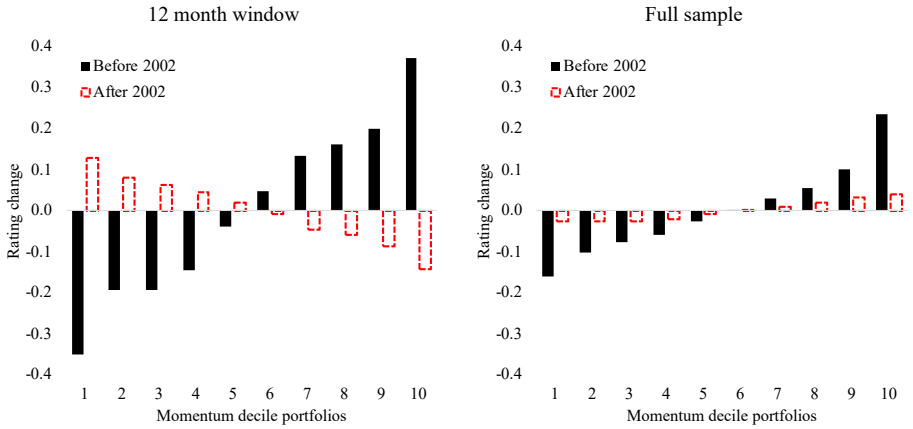
sample, before 2002, the top momentum decile experiences 0.72% higher monthly flows than the bottom decile, and that difference declined to 0.18% after 2002. We also see a similar effect, albeit noisier, when examining returns in Panel C.

We can see the structural break in 2002 in a time-series plot. Panel A of Figure 7 plots the cumulative return to the momentum factor as defined by top-minus-bottom decile, following Daniel and Moskowitz (2016) and Hou et al. (2020). Before 2002, the factor produced an average monthly return of 1.9%, which became a negligible 0.1% after 2002 with a clear kink around June 2002. In Panel B, we also see a clear kink for $\Delta Rating$ and a reduction of flow-induced trading post 2002.

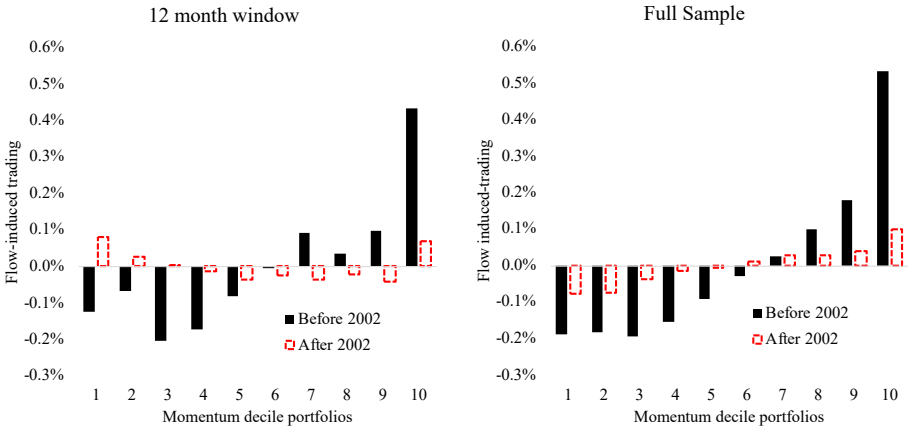
Figure 6. Momentum Factor Before vs After the 2002 Methodology Change

Left figures are based on a shorter window of 12 months before to 12 months after the event, and right figures are based on the full sample. The deciles are formed using NYSE break points and portfolios are value-weighted.

Panel A: 12 month lagged rating changes



Panel B: Flow-induced trading



Panel C: Monthly returns

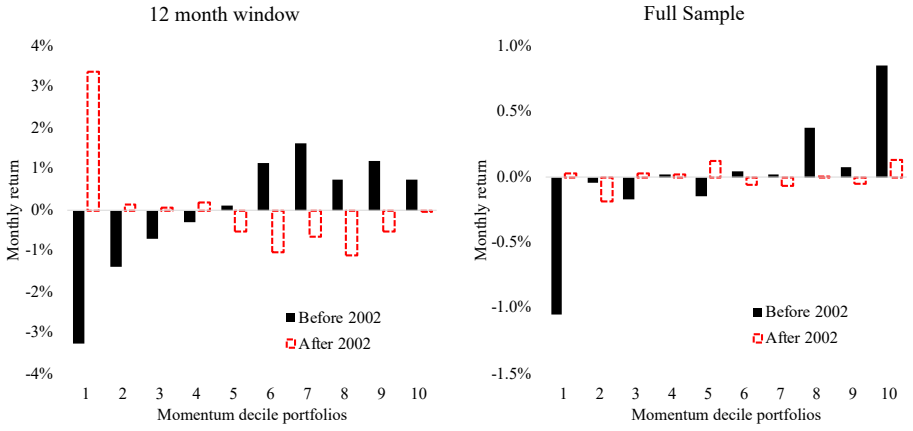
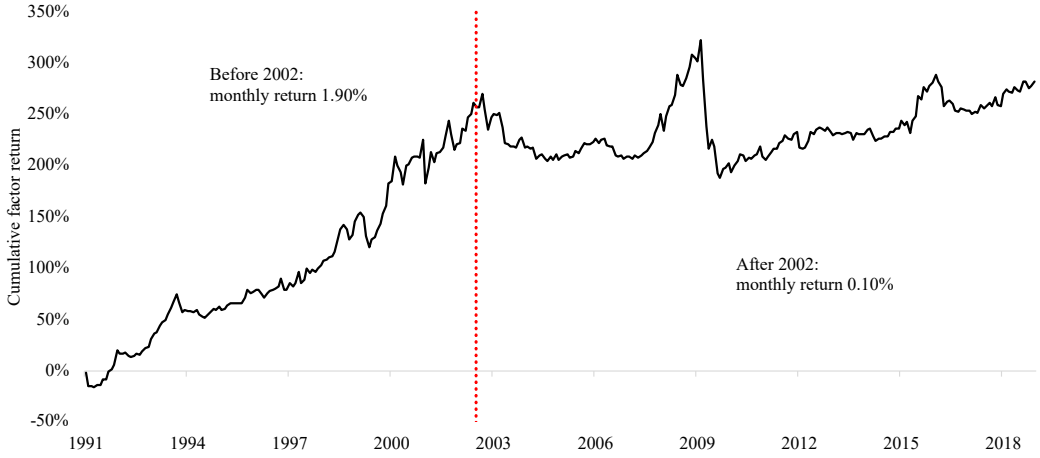


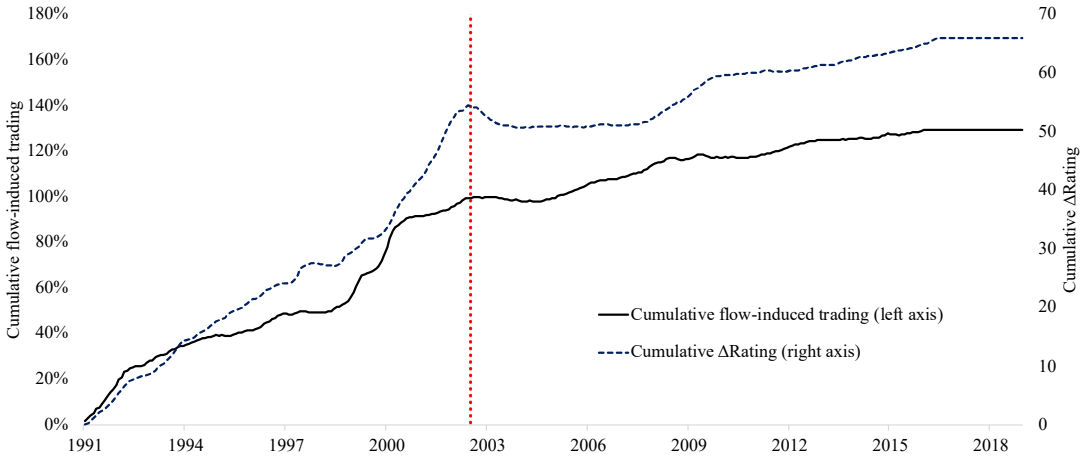
Figure 7. Momentum Factor Performance Over Time

The momentum factor is defined as long top decile and short bottom decile stocks. Following Hou et al. (2020), we use NYSE decile break points and value-weight the decile portfolios.

Panel A: Cumulative momentum factor return



Panel B: Cumulative lagged 12-month rating change (ΔRating) (LHS) and flow-induced trading for the momentum factor (RHS)



5 Effect of the 2002 Event on Asset Pricing Factors

Having shown suggestive evidence that the 2002 rating methodology change had a large impact on momentum, we now investigate its impact on all other asset pricing factors. Our main specification will examine the impact of ΔRating (lagged 12-month rating change) at the factor level on factor returns.[13] Our hypothesis is that many factors had high returns before 2002 because they load onto ΔRating caused by the style-level feedback trading from

[13]Section 5.4 gives the reasoning behind this specification.

the old methodology.

5.1 Heterogeneity in the Exposure to the Morningstar Event

We estimate the heterogeneity of each factor's exposure to the Morningstar methodology change. In principle, return factors would be affected more by the change in the Morningstar rating methodology if they sort stocks based on a variable that is correlated with the common style component of ratings.

To measure this common style component of ratings, we pursue the following procedure. First, for each fund j in month t, we decompose the pre-2002 fund ratings into two components—a common component to all funds within the same style and an idiosyncratic component:[14]

$$\text{Rating}_{j,t} = \underbrace{\text{StyleRating}_{j,t}}_{\substack{\text{due to pre-2002} \\ \text{methodology}}} + \text{IdiosyncraticRating}_{j,t}$$

where the style component isolates how the pre-2002 methodology changed the fund's rating. Luckily, documentation about Morningstar's pre- and post-2002 methodologies are publicly-available, so we use fund return data and can replicate what each funds' rating would have been under the two methodologies (Appendix A.2). We define $\text{StyleRating}_{j,t}$ as the difference of our replications:

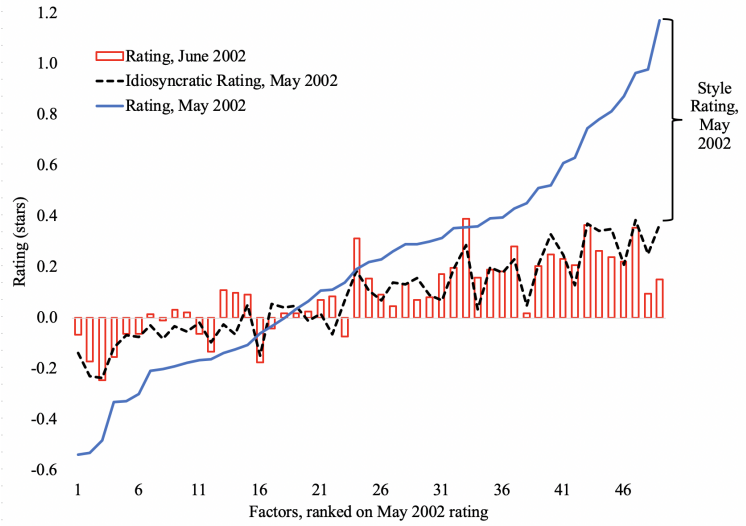
$$\text{StyleRating}_{j,t} = \widehat{\text{Rating}}^{\text{Pre-2002}}_{j,t} - \widehat{\text{Rating}}^{\text{Post-2002}}_{j,t}.$$

Because there are additional adjustments done by Morningstar that we do not fully capture, as well as potential differences in the historical database used by Morningstar and us, taking a difference between our two replications reduces the replication bias. The idiosyncratic component of ratings, $\text{IdiosyncraticRating}_{j,t}$, is the residual after subtracting $\text{StyleRating}_{j,t}$. Then, we follow the same steps as before to aggregate the two components of ratings at the factor level.

Figure 8 uses May and June 2002 data to show how our decomposition works at the factor level. We sort the 49 factors on their May 2002 rating, and the lines separate out the decomposed idiosyncratic and style components. As is shown in the graph, style rating

[14] For brevity, we use "style" to refer to the *Morningstar category* of funds which is assigned by Morningstar based on funds' portfolio holdings and fund objectives. As discussed in Appendix A.1, this span 3×3 styleboxes (combinations of value-blend-growth and small-midcap-large) for diversified U.S. equity funds, as well as a number of industry sectors for sector funds. This is different from the "style" used in Sensoy (2009) which only refers to the 3×3 style boxes.

Figure 8. Factor Rating Decomposition Around June 2002



accounts for the majority of the variation. For instance, in May 2002, the factor with highest pre-event rating (size) has a style rating of 0.80 and an idiosyncratic rating of 0.37; the factor with lowest pre-event rating (Ohlson O-Score) has a style rating of -0.40 and an idiosyncratic rating of -0.14.

If our estimation of style rating is effective at isolating the pre-2002 methodology-induced component, we expect this component of rating to vanish in June 2002, and only the idiosyncratic component—the residual—will remain. This is indeed the case. The red bars plot the actual factor ratings in May 2002 and they match up well with the idiosyncratic ratings, with an average estimation error of only 0.06.

5.2 Short-Term Effect on Factor Returns

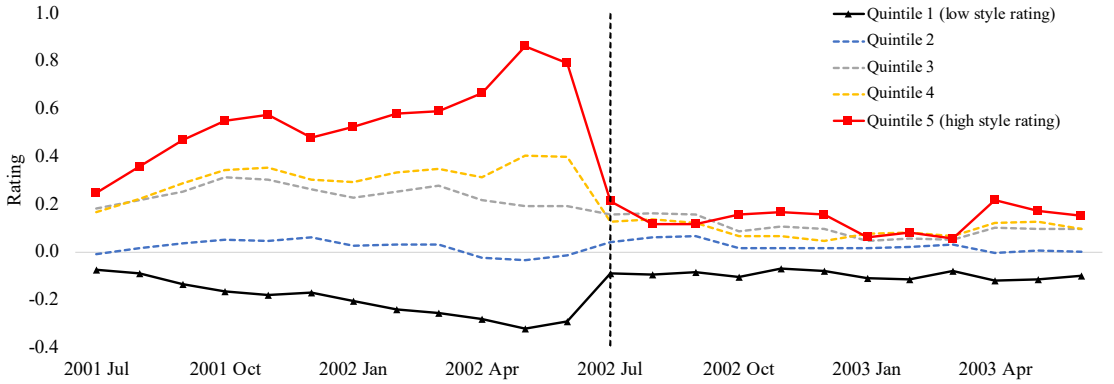
We now zoom in to a short window of 6 to 24 months around the June 2002 event to examine its impact on factor returns. We predict that, for factors positively exposed to the pre-2002 methodology, their returns should decline significantly after June 2002. The main benefit of using a short window is for identification because we can be reasonably certain that the Morningstar change is the main shock in the sample.

We first visualize the effect of the event in Figure 9. We sort factors into quintiles by their average style rating 12 months before the event. Thus, factors in quintile 5 have high ratings due to the pre-2002 methodology, and quintile 1 has low ratings due to the pre-2002 methodology. There is less dispersion within the middle three quintiles. Panel A plots the average rating of those factors. As expected, the top quintile factors experienced a sharp

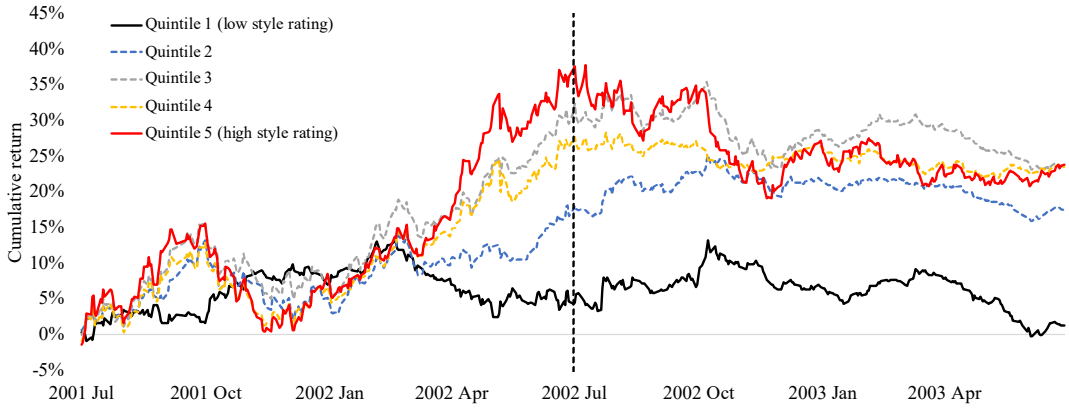
Figure 9. Stock Factors Around 2002

Factors are sorted into quintiles by their average StyleRating in the 12 months before June 2002. Panel A plots the average rating of those factors from 12 month before to 12 month after the event, and Panel B plots the cumulative returns.

Panel A: Factor ratings around June 2002



Panel B: Cumulative factor returns around June 2002



drop exactly at the event while the bottom quintile experienced a small increase.

Can this be related to factor-level stock returns? In Panel B, we plot the cumulative returns of factors in those five quintiles. Quintile 5 achieved an average of 36% returns in the year before the event, and that trend stopped and even slightly reversed after June 2002. In contrast, the bottom quintile had almost flat returns throughout.

Why did some factors have a much higher style rating exposure than others? We find that this arises from how factors load differently onto styles, and also the fact that some styles performed much better than others in this period. Panel A of Figure 10 shows the return, rating, and flow-induced trading of funds by the 3×3 styles. In the 12 month before the methodology change, small-value styles performed well while large-growth styles performed poorly, with a difference of 35% in returns and 1.04 difference in ratings. They also had a large 42% difference in flow-induced trading. To examine how this impacts

Figure 10. Factor Style Exposure Before the 2002 Event Panel A: the return, rating, and flow-induced trading of funds by 3 × 3 fund styles during the 12 months before June 2002. Larger values are associated with red while smaller values are associated with green. Panel B: the HML and SMB loadings of factors, sorted into quintiles by their style rating, during the 12 months before June 2002.

Panel A: Mutual fund returns, ratings, and flow-induced trading by style

Cumulative return

	Growth	Blend	Value
Small	-14%	-1%	9%
Mid	-23%	-6%	-1%
Large	-26%	-16%	-10%

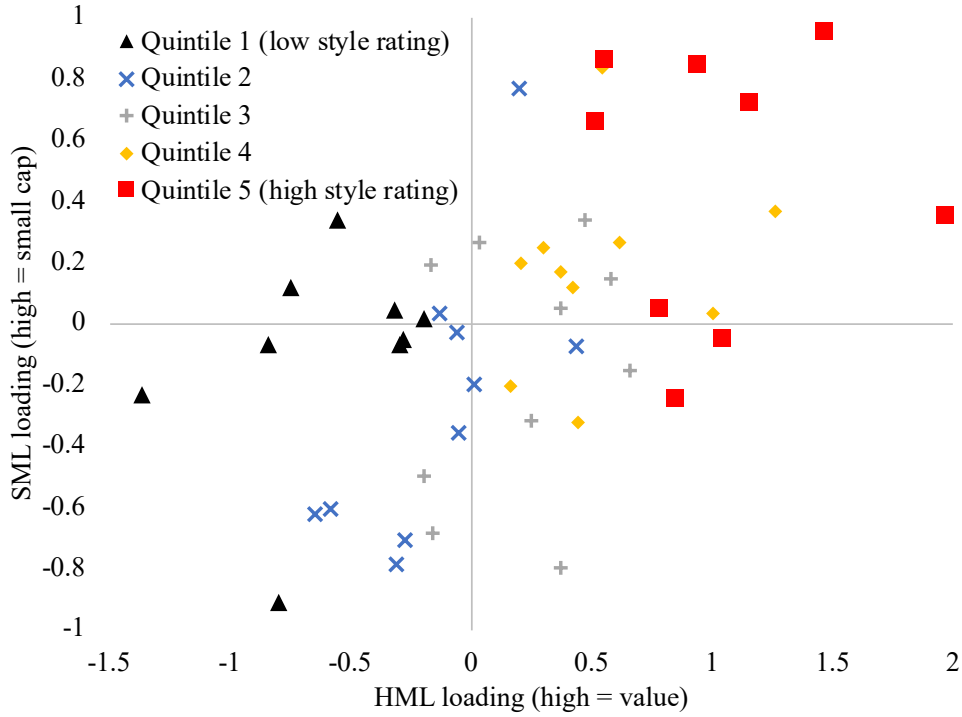
Rating

	Growth	Blend	Value
Small	3.36	3.82	4.14
Mid	3.32	3.65	4.09
Large	3.10	3.47	3.75

Cumulative flow-induced trading

	Growth	Blend	Value
Small	4%	21%	32%
Mid	-3%	11%	29%
Large	-10%	-8%	1%

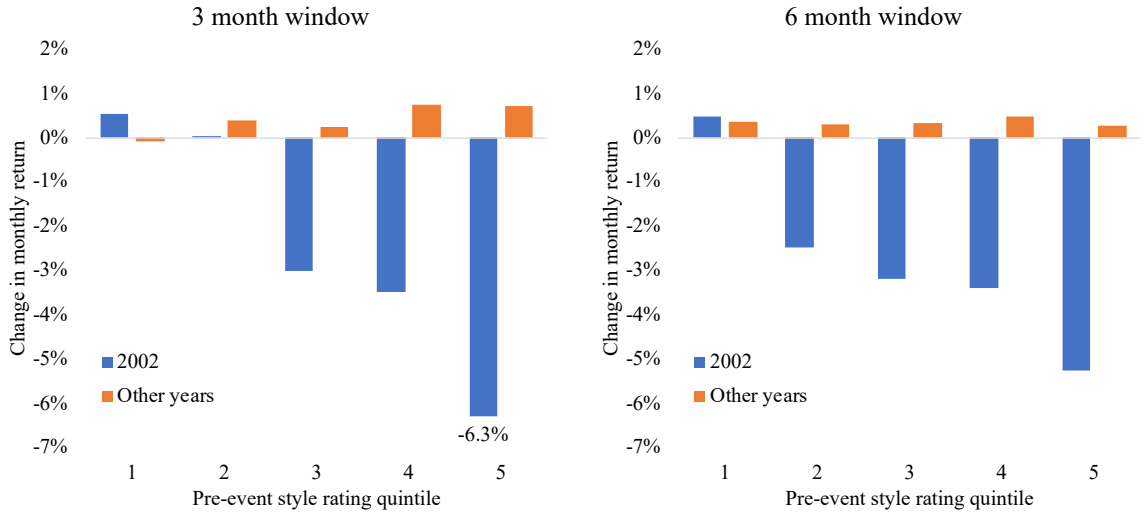
Panel B: Factor HML and SMB loadings



factor returns, we measure the style exposure of each factor using its SMB and HML factor exposures, estimated using daily returns in the 12 months before the event. We plot their factor exposures in Panel B. The results show that factors in the top style rating quintile are mostly loading onto the small and value styles, while those in the bottom quintile mostly load onto large and value styles.

Figure 11. Return Difference Around June 2002 by Pre-Event Style Rating

We examine 3 or 6 months before and after June of 2002. We sort factors into quintiles by the pre-event Morningstar Style Rating and then plot the average change in monthly factor returns.



We formally test the effect of style rating on cross-section of factor premia change around 2002 using a regression framework. For each factor f, we use h months before versus h months after to compute the change in monthly returns:

$$\Delta \text{Ret}_f^h = \underbrace{\frac{1}{h}\sum_{t=0}^{h-1} \text{Ret}_{f,t}}_{\text{after event}} - \underbrace{\frac{1}{h}\cdot\sum_{t=-h}^{-1} \text{Ret}_{f,t}}_{\text{before event}}.$$

where $t = 0$ is July 2002, the first month affected by the post-2002 Morningstar methodology.[15] To zoom in on the event, we use $h = 3$ or 6 months. In Figure 11, we plot the average return change for factors sorted by pre-event style rating. Top quintile experienced a decline of 5-6% in monthly returns, while the bottom quintile experienced a slight increase. To make sure this is unique to 2002, we also examine changes around June of other years as placebo tests. There is no similar effect in other years.

We use a cross-sectional regression to estimate the size of the effect. For each factor, we measure the change in ΔRating:

$$\Delta\text{Rating}_{f,\text{after}}^h - \Delta\text{Rating}_{f,\text{before}}^h = \underbrace{\frac{1}{h}\sum_{t=0}^{h-1} \Delta\text{Rating}_{f,t}}_{\text{after event}} - \underbrace{\frac{1}{h}\cdot\sum_{t=-h}^{-1} \Delta\text{Rating}_{f,t}}_{\text{before event}},$$

[15] The post-2002 Morningstar ratings are first released at the end of June 2002.

Table 1. Explaining Factor Return Changes Around 2002

We use 3- or 6-month periods before versus after June 2002 to estimate changes in factor premia. We regress factor premia change on changes in ΔRating using OLS or 2SLS where the instrument is the pre-event StyleRating. We also estimate the same regression using other years as a placebo test. In the 2SLS regressions, for the 2002 sample, the F-statistics are 1263 and 1685 for the 3- and 6-month specifications, respectively. For the non-2002 placebo sample, the F-statistics are 122.8 and 82.54, respectively.

Dependent variable:	Δ Monthly factor return (%)							
Window:	3 months				6 months			
Specification:	OLS		2SLS		OLS		2SLS	
Period:	2002	non-2002	2002	non-2002	2002	non-2002	2002	non-2002
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta\text{Rating}_{\text{after}} - \Delta\text{Rating}_{\text{before}}$	3.97***	−0.25	3.29***	−3.14	2.77***	0.18	2.79***	−1.65
	(1.29)	(0.42)	(1.27)	(2.01)	(0.84)	(0.41)	(0.78)	(2.06)
Difference	4.22***		6.44***		2.59***		4.43**	
	(3.11)		(2.71)		(2.78)		(2.02)	
Observations	49	1,323	49	1,323	49	1,323	49	1,323
R^2	20.0%	0.02%	15.54%	2.74%	13.45%	0.02%	13.38%	1.35%

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

and then estimate its effect on change in factor returns.

$$\Delta\text{Ret}_f^h = a^h + b^h \cdot \left(\Delta\text{Rating}_{f,\text{after}}^h - \Delta\text{Rating}_{f,\text{before}}^h\right) + u_f^h \tag{4}$$

In addition to the OLS regression, to narrow down to the effect of the pre-2002 methodology, we also estimate a two-stage least-squares regression (2SLS) using pre-event Style Rating. To account for the cross-sectional correlation of factor returns, we adjust the standard errors using the covariance matrix estimated from factor returns.[16]

The results are reported in Table 1. In response to one star change in ΔRating, monthly return changes by 3-4% in the 3 month window and 2-3% in the 6 month window. The coefficient estimates from OLS and 2SLS are similar to each other. As will be discussed in Section 5.4, this magnitude is similar to the estimates using stock-level data. The explanatory powers are large, with R^2 in the range of 15% to 20% for the 3-month window and around 13% for the 6-month window. We see no similar effects in the other Placebo years.

[16]Let $r = (r_1, ..., r_N)$ denote the vector of factor returns, and let $C = Var(r)$ be the covariance matrix. Under the null hypothesis that $r \sim N(\mu, C)$, and that returns are independent across periods, then the LHS of interest's distribution is $\bar{r}_{\text{post}} - \bar{r}_{\text{pre}} \sim N\left(0, \frac{2}{h} \cdot C\right)$.

Figure 12. Price impact of flow-induced trading over time.

The coefficients are estimated using 5 year rolling value-weighted Fama-MacBeth regressions of stock returns on flow-induced trading.



5.3 Long-Term Effect on Factor Returns

While using the narrow window around June 2002 achieves better identification, we now ask the more economically significant question: did the Morningstar change have an impact on long-run factor returns? We argue it is reasonable to expect the answer to be yes for two reasons. First, the mutual fund sector has been a significant holder of stocks since 1990s, and as shown in Panel C of Figure 3, mutual fund flows have been consistently chasing ratings throughout our sample. Therefore, Morningstar's pre-2002 methodology should have been generating style-level positive feedback flow pressures in the same way.

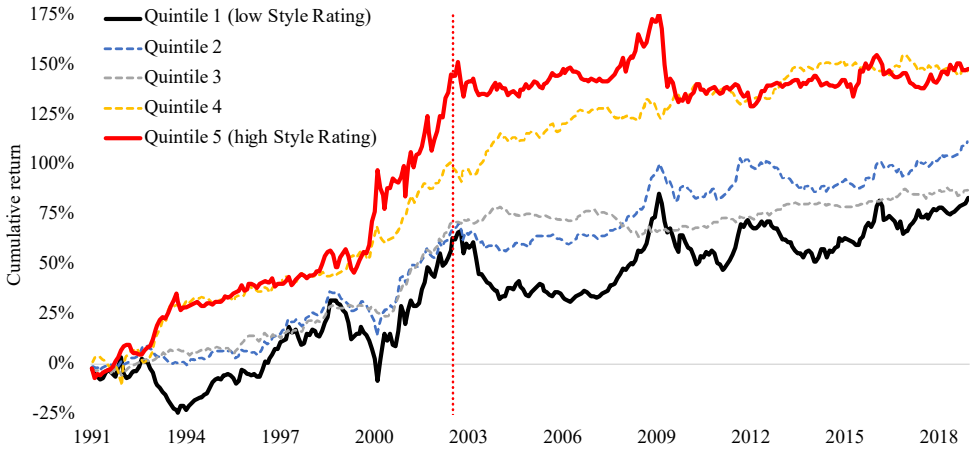
Second, one may worry that arbitrage forces or stock liquidity may have varied over time, so flow-induced trading may not have much impact in certain periods. To investigate how price impact varied over time, we estimate 5 year rolling value-weighted Fama-MacBeth regressions of stock returns on flow-induced trading. We report the rolling coefficient in Figure 12. The price impact is significant throughout the sample, even though it is highest around the dotcom period when the flows are the largest. Putting these two reasons together, we believe it is reasonable to expect and effect in long-term factor returns.

To estimate how much each factor is affected by the pre-2002 methodology, we compute the average StyleRating for each factor using the whole sample before June 2002. We then sort factors by this measure into quintiles. As shown in Appendix A.3, most factors load positive on StyleRating, and the ones most positively affected tends to be momentum types. The factors that are negatively affected tend to be related to value and profitability, consis-

Figure 13. Long-Sample Factor Returns by Pre-2002 Style Rating Exposure

The figure shows the cumulative return to factors sorted into quintiles using their *StyleRating* before 2002. Panel A plots the cumulative returns for each quintile. Panel B reports the average returns.

Panel A: Cumulative factor returns



Panel B: Summary statistics

	Factor quintile					
	1	**2**	**3**	**4**	**5**	**5 minus 1**
Style Rating Before 2002	-0.034	-0.002	0.019	0.053	0.135	0.170
Monthly return						
Before 2002	0.42%	0.47%	0.51%	0.73%	1.05%	0.63%
After 2002	0.13%	0.23%	0.08%	0.23%	0.02%	-0.11%
Change	-0.29%	-0.24%	-0.43%	-0.50%	-1.03%	-0.75%

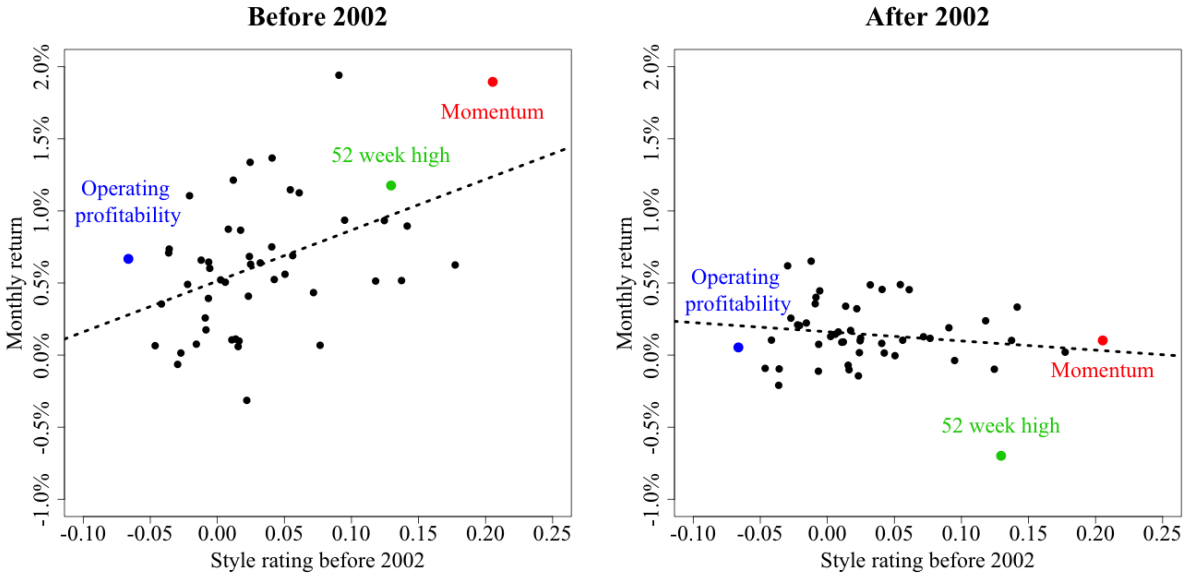
tent with the finding in Asness et al. (2013) that value and momentum tend to be negatively correlated to each other.

In Figure 13, we plot the cumulative average returns of factors in each quintile. The top quintile has an average of 0.135 style rating before 2002, and achieved an average of 1.05% monthly return before 2002 which dropped effectively to zero after 2002. To further investigate which exact factors are affected the most, we plot the average return for each factor before and after 2002 in Figure 14. The most heavily affected factor is momentum which had the highest pre-2002 style rating. Many other heavily affected factors are related to momentum, such as industry momentum and the 52-week high strategy. In contrast, strategies such as operating profitability are barely, if not slightly positively, affected by the methodology change.

We now do a regression analysis like that done in the short windows around 2002 done

Figure 14. Factor Returns Before versus After Morningstar Methodology Change

Average monthly factor returns before (left) and after (right) June 2002 (left) are plotted against the average pre-event Morningstar StyleRatings. Factors with high (low) StyleRatings had high (low) ΔRating before 2002 due to Morningstar's pre-2002 methodology. The dashed lines are best linear fits.



in Section 5.2. We regress changes in monthly factor returns before versus after 2002 on changes of ΔRating. We estimate both an OLS regression and a 2SLS where the instrument is the pre-event average Morningstar StyleRating. The results are reported in Table 2. The coefficient estimate is similar to those obtained in the short-horizon regressions (Table 1). The R^2 is around 30%, implying that Morningstar's methodology change can account for a large fraction of factor profitability change over time.

5.4 Comparing Factor-Level Estimates with Stock-Level Estimates

Our analysis thus far relies on reduced-form regressions of return changes on $\Delta Rating$ changes at the factor level. In this section, we use fund and stock-level data to estimate the chain of effects from bottom up: how flows chase ratings and how flow-induced trading cause price pressures at the stock level. We then compare the bottom-up results with the earlier factor level results. We have two goals. First, we check whether the 12-month rating change specification in the factor-level analysis is justified. Second, we compare the impact coefficients to examine whether the factor-level results are quantitatively plausible.

Translating these goals to empirical specifications, we estimate two impulse response functions: *i*) the response of fund flows to rating changes, and *ii*) the response of stock

Table 2. Explaining Long-Term Factor Premia Changes Around 2002

We regress the pre-vs-post 2002 change in average monthly factor return on changes in ΔRating. The 2SLS estimation uses the average pre-2002 StyleRating as the instrument. The first-stage F-statistic is 388.2. The standard errors are adjusted for the cross-sectional correlation between factor returns.

Dependent variable:	Δ Monthly factor return (%)	
Specification:	OLS	2SLS
	(1)	(2)
$\Delta\text{Rating}_{\text{after}} - \Delta\text{Rating}_{\text{before}}$	3.20^{***}	3.11^{***}
	(0.80)	(0.88)
Constant	-0.05	-0.04
	(0.05)	(0.05)
Observations	49	49
R^2	30.9%	28.6%

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

returns to flow-induced trading. Because we are eventually studying factor-level outcomes, our analysis is value-weighted: fund-level regressions are weighted by fund TNA and stock-level regressions are weighted by stock market cap. All regressions are estimated with the Fama and MacBeth (1973) procedure using monthly data. First, we estimate fund flow response to lagged fund rating changes using the full sample of fund data from 1991 to 2018:

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

where controls $X_{i,t}$ include 36 monthly lags of fund flows and ratings. The cumulative response coefficients $(b_1, b_1 + b_2, ...)$ are plotted in the left panel of Figure 15. In response to a one-star increase of fund rating, funds experience an average of 6% additional inflows, most of which take place within 12-18 months.

Second, we estimate the response of stock returns to stock-level flow-induced trading (FIT) defined in Equation (3):

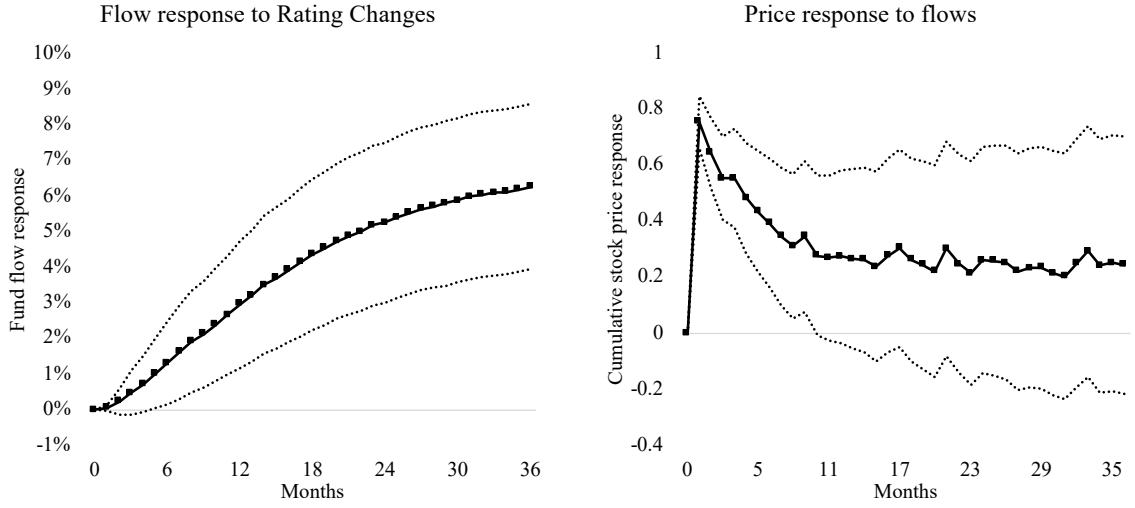
$$\text{Ret}_{i,t} = a + c_0 \cdot \text{FIT}_{i,t} + c_1 \cdot \text{FIT}_{i,t-1} + ... + c_{36} \cdot \text{FIT}_{i,t-36} + X_{i,t} + u_{i,t},$$

where the controls $X_{i,t}$ include 36 lags of past stock returns. The right panel of Figure 15 plots the cumulative response. Each increase in mutual fund ownership of 1% due to flow-induced trading causes price pressure resulting in immediate return of 0.8% in the first month, which gradually reverts in the subsequent one to two years. This finding is consistent with the earlier findings about the flow-induced-trades in Lou (2012).

The effect we are interested in—how changes in rating impact stock prices—is the com-

Figure 15. Impulse Response Functions: Rating→Flows, and Flows→Returns

Left: the cumulative response of fund flows to changes in fund ratings. Right: the cumulative response of stock returns to flow-induced trading (FIT).



bined effect of the two impulse response functions estimated above. We now estimate the combined effect in two fashions. First, we directly estimate at the stock level:

$$\text{Ret}_{i,t} = a + d_1 \cdot \Delta\text{Rating}^{\text{stock}}_{i,t-1} + ... + d_{36} \cdot \Delta\text{Rating}^{\text{stock}}_{i,t-36} + u_{i,t}, \tag{5}$$

where $\Delta\text{Rating}^{\text{stock}}_{i,t}$ are changes in stock-level Morningstar ratings defined in Equation (2). The results are plotted in blue in Panel A of Figure 16. In response to one-star of rating change, stock return response is in the order of 1-3%, and most of the effect happens within 12 months. Second, we also indirectly calculate the implied effect by cumulating the responses of the two separate effects, as illustrated in Panel B of Figure 16.[17]

These two methods each has its own merit. The first approach is more direct, but may capture spurious correlations between returns and ratings. The second is cleaner because it only captures the rating-induced effect, but suffers from stronger attenuation bias due to multiple steps of estimation, so it is likely to understate the effect.

Regardless of the method used, the conclusion is similar: the effect of ratings on returns is very strong initially and becomes weaker over time. The overall effect takes roughly 12 months to fully manifest. Guided by this exercise, we chose the 12 month rating change

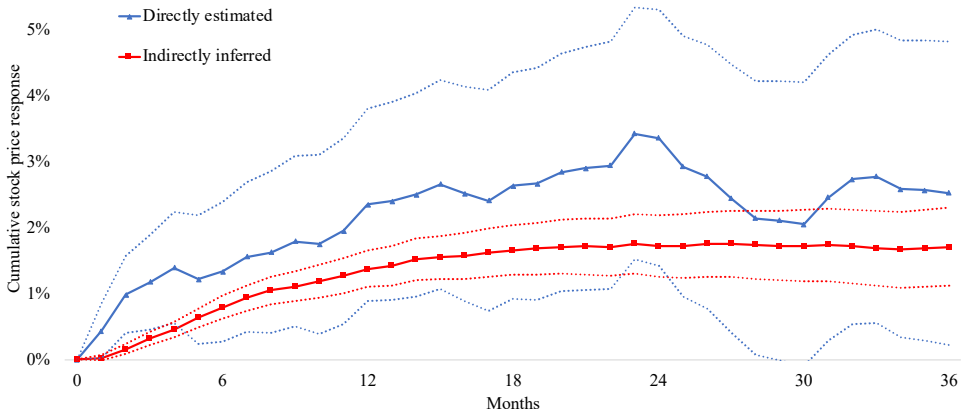
[17]Mathematically, the implied response coefficient for lag h is given by:

$$d^{\text{implied}}_h = \sum_{k=1}^{h} b_k \cdot c_{h-k}.$$

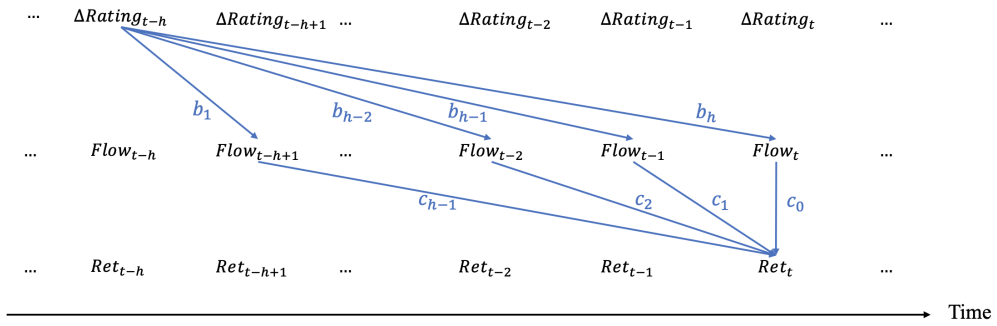
Figure 16. The Impact of Ratings on Stock Returns

Panel A: cumulative stock price response to rating changes. The blue line is based on directly regressing stock returns on lagged rating changes (Equation (5)). Dashed lines are two standard deviations. The red line coefficients are inferred by cumulating the coefficients from Figure 15. Panel B: Illustration of the calculation method indirectly estimating the effect of ratings on stock returns.

Panel A: Cumulative stock price response to rating changes



Panel B: Deriving inferred rating impact coefficients on stock prices



specification in factor-level analysis. Also, the size of the effect is approximately 1-3% return per one-star rating change, which is similar to the estimates obtained in the factor-level analysis.

It is worth explaining why earlier rating changes appear to have smaller price impact than recent rating changes. This is because, although the earlier rating changes continue to lead to flow pressures, its earlier price impact already started to revert (right plot of Figure 15), partially cancelling out the effect of continued price pressures.

5.5 Economic Significance of the Morningstar Methodology Change

Given our results thus far, we are interested to explore the extent to which the Morningstar change can explain the decline in the post-2002 factor profitability. Several papers found that the profitability of many stock return factors declined dramatically after June 2002 (Green et al., 2017; Khandani and Lo, 2011; Daniel and Moskowitz, 2016; Arnott et al., 2019b). Our analysis thus far strongly suggests that the Morningstar change explains part of the "kink." However, can the methodology change quantitatively explain the decline of profitability? We now do a simple back-of-envelope quantification exercise. It is worth noting that this is only a crude exercise and requires strong assumptions. We interpret this as an estimate whether our proposed mechanism can plausibly explain the decline.

We evaluate two counterfactual benchmarks. First, a "no StyleRating" counterfactual. This is a world in which Morningstar did not use its old methodology before 2002. The counterfactual returns are defined as:

$$\widehat{\text{Ret}}_{f,t}^{\text{no StyleRating}} = \text{Ret}_{f,t} - \lambda \cdot \text{StyleRating}_{f,t-1} \tag{6}$$

where λ is the price impact of ΔRating on factor returns. This coefficient has been estimated at both the factor level and at the stock level in the previous sections.

This "no StyleRating" counterfactual is rather realistic. As explained, Morningstar's methodology is decided by a few people in their Chicago office and is thus somewhat arbitrary. Investors blindly chase ratings regardless of the methodology, so this alternative world easily could have happened (Ben-David et al., 2019; Evans and Sun, 2020).

The second counterfactual is defined as a world in which investors do not chase Morningstar ratings. This would be true if investors realize that ratings are not necessarily good fund return predictors, as argued by a high profile 2017 Wall Street Journal article,[18] or if Morningstar did not exist and investors did not find other similar signals to chase. This counterfactual is, admittedly, less realistic than the first one. We define it as:

$$\widehat{\text{Ret}}_{f,t}^{\text{no rating-chasing}} = \text{Ret}_{f,t} - \lambda \cdot \Delta\text{Rating}_{f,t-1}$$

For the price impact coefficient λ, we have obtained three sets of estimates, one from stock-level regressions, and two from short- and long-window factor-level regressions. The range is around 2% to 4% per a change of one star rating.

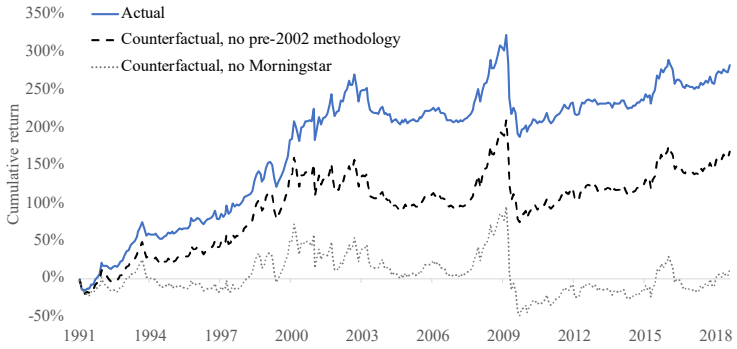
In Figure 17, we also examine the effect just on the momentum factor which is the

[18] See Kirsten Grind, Tom McGinty and Sarah Krouse, "The Morningstar Mirage," The Wall Street Journal, October 25, 2017, `https://www.wsj.com/articles/the-morningstar-mirage-1508946687`.

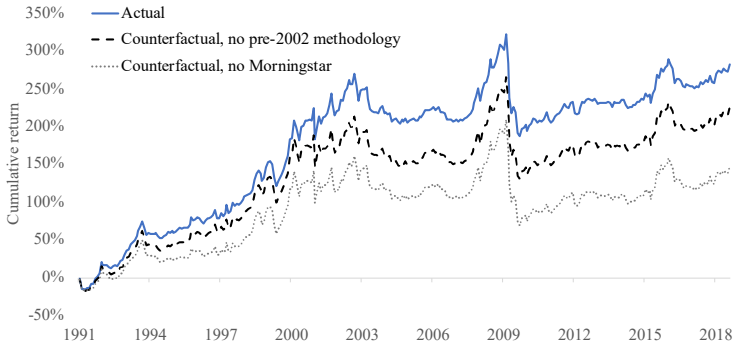
Figure 17. Counterfactual Cumulative Returns for the Momentum Factor

We evaluate two counterfactuals, one if Morningstar did not use the pre-2002 methodology (thick dashed black line) and one if investor flows does not chase ratings.

Panel A: Assuming high price impact of ratings: $\lambda = 0.04$



Panel B: Assuming low price impact of ratings: $\lambda = 0.02$



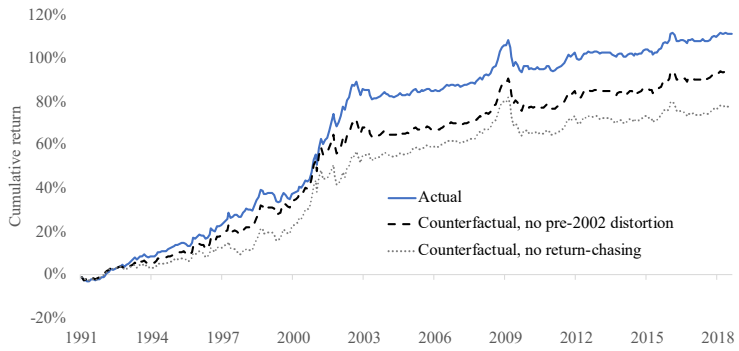
most profitable factor in our sample before 2002. If Morningstar did not use the pre-2002 methodology, around one-fifth to one half of its return would be gone; if investors do not chase ratings, than one half to the entire cumulative returns of momentum would have been gone.

We also a apply both 2% and 4% estimates to the average returns of all factors and present the cumulative returns in Figure 18. Our estimations show that Morningstar's methodology directly accounted for a quarter to half of anomaly returns. A similar magnitude of factors returns can be attributed to the rating chasing behavior of investors.

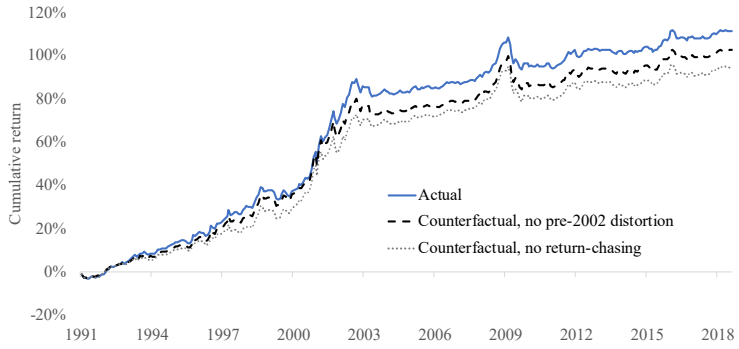
Figure 18. Counterfactual Average Factor Returns

We evaluate two counterfactuals, one if Morningstar did not use the pre-2002 methodology (thick dashed black line) and one if investor flows does not chase ratings.

Panel A: Assuming high price impact of ratings: $\lambda = 0.04$



Panel B: Assuming high price impact of ratings: $\lambda = 0.02$



6 Effects of 2002 Event on Factor Momentum

A recent literature documents that stock return factors exhibit momentum (Arnott et al., 2019a; Gupta and Kelly, 2019). That is, factors that recently performed well tends to continue doing so.

In this section, we show that the 2002 change of Morningstar methodology led to substantial decline in the profitability of factor momentum. We also show that this effect accounts for most of the Green et al. (2017) finding of stock characteristics-based return predictability decline.

6.1 Defining Time-Series Factor Momentum (TSFM) Factors

Inspired by Gupta and Kelly (2019), we define time-series factor momentum strategies by implementing the following trading rule. For each factor, trade in the direction of the

Table 3. Time-Series Factor Momentum Factors and the Original Factors

We vary the look-back window in constructing the TSFM factors. Column 2 shows the average fraction of time that they are long the original factors. Column 3 shows that 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	54%	0.12
3	54%	0.16
6	55%	0.20
12	58%	0.24
24	61%	0.26
36	62%	0.30
60	68%	0.38

sign of the recent w months of factor returns. Such a strategy will yield returns:

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

We can also define ΔRating and pre-2002 StyleRating for these TSFM factors in the same way, by flipping signs when needed:

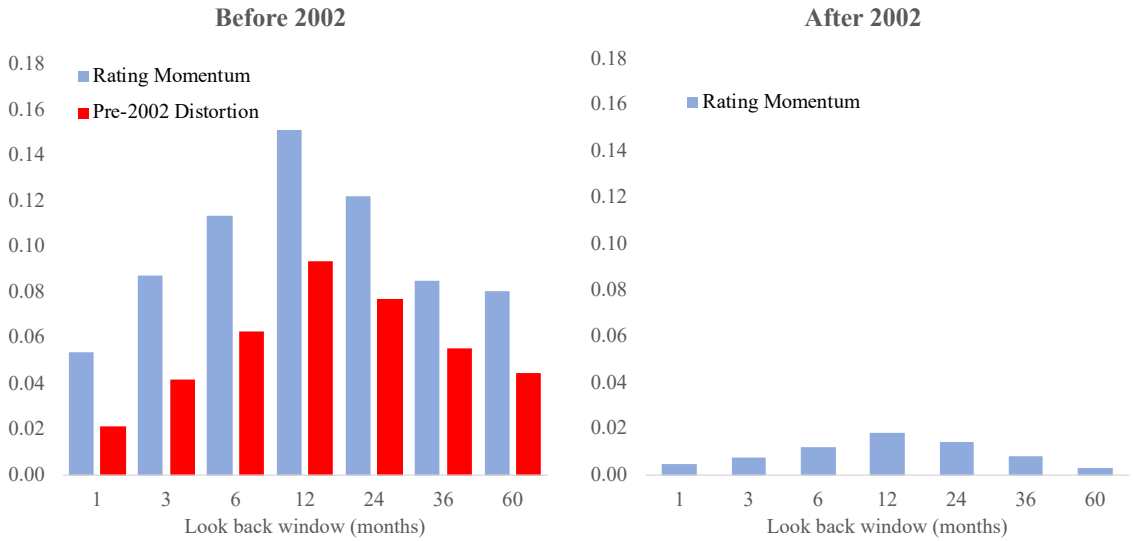
$$\Delta\text{Rating}_{f,t}^{\text{TSFM},w} = \Delta\text{Rating}_{ft} \cdot \text{sign}(\text{Ret}_{f,t-w\to t-1})$$
$$\text{StyleRating}_{f,t}^{\text{TSFM},w} = \text{StyleRating}_{ft} \cdot \text{sign}(\text{Ret}_{f,t-w\to t-1})$$

The literature has not converged on the look-back window. We thus try $w \in \{1, 3, 6, 12, 36, 60\}$ months.

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 54% to 68% 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 0.12 to 0.38. Therefore, judging from both criterion, they are quite different from the original factors.

Figure 19. ΔRating of Factor Momentum Strategies, by Look-Back Window

The figure shows results from analyses exploring the change in factor momentum returns with different look-back windows, before 2002 and after 2002.



6.2 Exploring Look-Back Window for Largest Profitability Drop

Since we focus on rating-induced feedback mechanism, we can pin down the look-back window for which we should see the largest profitability drop before versus after 2002. This is a further test of our mechanism: we should see the largest effect on the look back window we predict to have te effect, and not observe a large effect for the other windows.

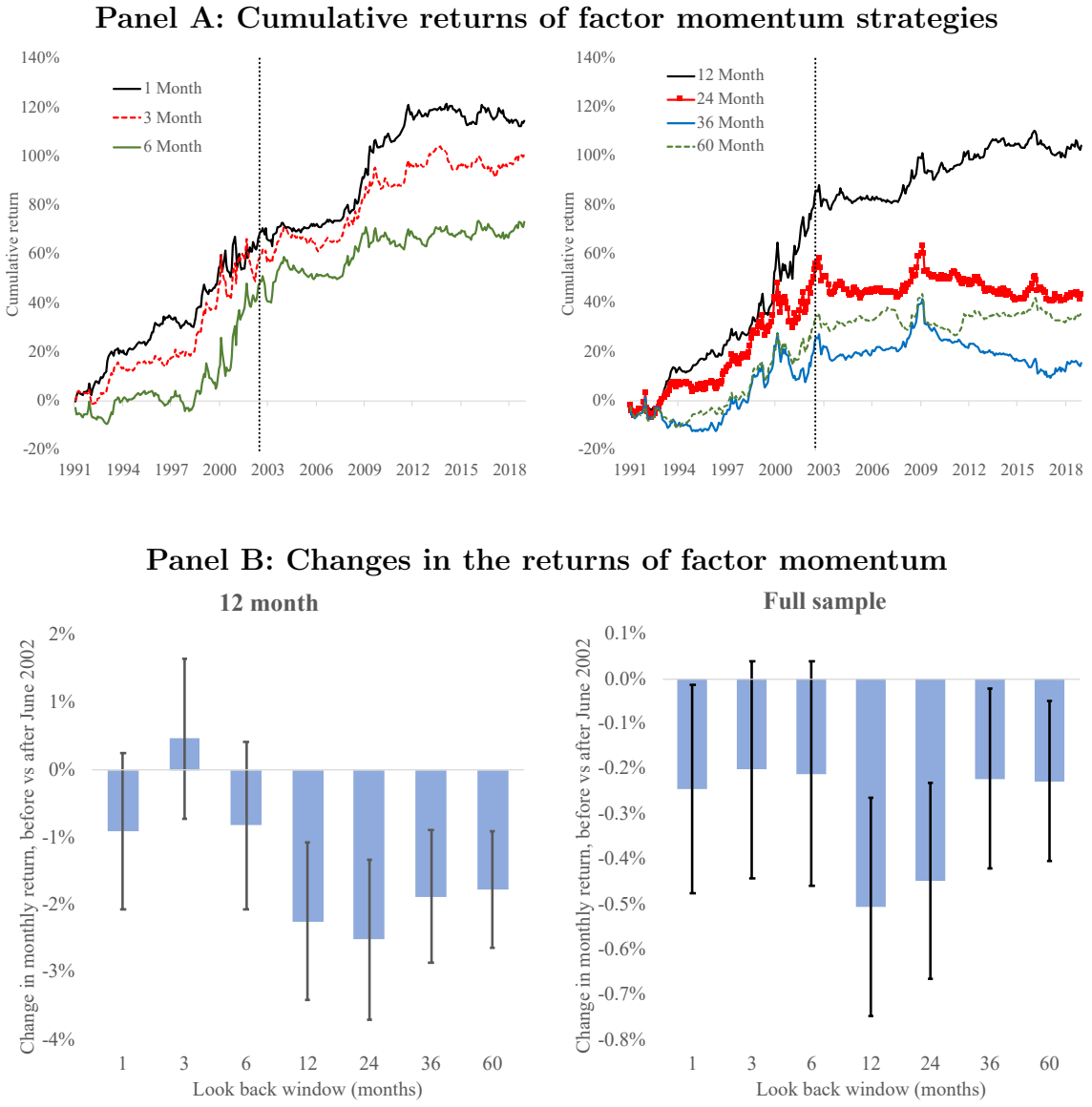
The effect of Morningstar ratings on factor momentum is akin to the one illustrated in Figure 2. High past returns for a factor leads to high ratings for the funds that pursue strategies that are aligned with this factor, which then leads to high inflows and consequent price pressures on the underlying securities in the factor portfolios. The speed of the last two links have already been quantified in Section 5.4 to be around 12 months. In Appendix B, we also estimate the speed of the first link and combine all. We conclude that the largest effect on factor momentum should happen for look-back windows of 12 to 24 months.

We first verify whether time series momentum experienced the sharpest drop of ΔRating. Figure 19 plots the average ΔRating and pre-2002 StyleRating measure. Indeed, the pre-2002 ΔRating is highest for the look-back windows of 12 and 24 months, and it is primarily explained by the pre-2002 methodology. After 2002, ΔRating of all these strategies became close to zero.

We now visually examine the factor momentum returns. Figure 20 plots the cumulative average returns for different window w. Visually, indeed, the largest "kink" happened in 12

Figure 20. Changes in the Returns of Factor Momentum

Panel A: cumulative returns of factor momentum strategies defined using different look-back windows. Panel B: the change of factor momentum strategies returns before versus after June 2002 with two standard error bars.



to 24 months look-back period. This is also shown visually using bar charts in Panel B. We also tabulate in Table A.III. Before 2002, 12- and 24-month TSFM has highest return which dropped afterwards. Factor momentum with other horizons also experienced return drops, but not as large, and many of them have statistically smaller drops.[19]

Our conclusion, therefore, is that factor momentum, as the momentum factor and host of

[19]Standard errors for comparisons with 12 months are computed as the unpaired t-stat of $\text{Ret}_{\text{w}} - \text{Ret}_{12\text{m}}$.

other return factors that we explores diminished almost entirely due to the 2002 Morningstar methodology change.

7 Conclusion

Stock market factors (anomalies) are perhaps the most researched topic in asset pricing. Scholars have suggested different explanations to why stock returns display cross-sectional and time-series predictable patterns. The existing explanations propose that return factors reflect risk premia, mispricing, or investment opportunities. Another, seemingly unrelated phenomenon, is the disappearance of factor premia since the early 2000s. Researchers proposed that changes in market structure, entry of hedge funds, improvement in liquidity, or simply the discovery of factors led to their elimination by arbitrageurs. To date, researchers were not able to pinpoint the exact drivers of factor premia and their disappearance due to lack of identification. As such, the debate has been ongoing for over two decades now.

In this article, we present causal evidence in support of the mispricing hypothesis: factors premia are driven by correlated demand ("style investing," see Barberis and Shleifer, 2003). This explanation suggests that factor premia are created when investors chase returns and invest in the best-performing styles (e.g., value or growth), and that these flows create price pressure which perpetuates the performance of the style.

We show that a change in Morningstar's rating methodology eliminated the relation between flows and the style-related-component of mutual funds' past performance. As a consequence, the feedback loop broken, and factor premia weakened dramatically, and that the effect is stronger for those factors that are more correlated with the dimensions affected by the Morningstar change, as well as the momentum factor which reflects the perpetuating time dimension in the feedback loop.

References

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, *The Journal of Portfolio Management* 45, 18–36.

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

Barber, Brad M, Xing Huang, and Terrance Odean, 2016, Which factors matter to investors? Evidence from mutual fund flows, *The 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, Where has mutual fund managers' skill disappeared?, Working paper, The Ohio State University.

Berk, Jonathan B., and Jules H. van Binsbergen, 2016, Assessing asset pricing models using revealed preference, *Journal of Financial Economics* 119, 1–23.

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.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *The Journal of Finance* 52, 57–82.

Choi, James J, and Kevin Zhao, 2020, Did mutual fund return persistence persist?, Technical report, National Bureau of Economic Research.

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.

Cohen, Randolph B, Joshua D Coval, and L'uboš Pástor, 2005, Judging fund managers by the company they keep, *The Journal of Finance* 60, 1057–1096.

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

De Long, J Bradford, Andrei Shleifer, Lawrence H Summers, and Robert J Waldmann, 1990, Positive feedback investment strategies and destabilizing rational speculation, *The Journal of Finance* 45, 379–395.

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

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

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, 2008, Dissecting anomalies, *The Journal of Finance* 63, 1653–1678.

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

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

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

Green, Jeremiah, John RM 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.

Hirshleifer, David, 2001, Investor psychology and asset pricing, *The 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, *The Journal of Finance* 54, 2143–2184.

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

Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating anomalies, *The Review of Financial Studies* forthcoming.

Israel, Ronen, and Tobias J Moskowitz, 2013, The role of shorting, firm size, and time on market anomalies, *Journal of Financial Economics* 108, 275–301.

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

Jones, Christopher S, and Lukasz Pomorski, 2016, Investing in disappearing anomalies, *Review of Finance* 21, 237–267.

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.

Lamont, Owen A, and Richard H Thaler, 2003, Anomalies: The law of one price in financial markets, *Journal of Economic Perspectives* 17, 191–202.

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, *The Review of Financial Studies* 25, 3457–3489.

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?, *The Journal of Finance* 71, 5–32.

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

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

Schwert, G William, 2003, Anomalies and market efficiency, *Handbook of the Economics of Finance* 1, 939–974.

Sensoy, Berk A, 2009, Performance evaluation and self-designated benchmark indexes in the mutual fund industry, *Journal of Financial Economics* 92, 25–39.

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

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

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

Appendix A Data and Measures

A.1 Morningstar Methodology

We explain Morningstar rating construction and the June 2002 change in detail here. Morningstar ratings are updated every month. There are two steps in Morningstar's rating calculation:

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

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

In June 2002, Morningstar changed both steps of the methodology. The steps are consecutive, however, independent. Our analysis shows that the change to the second step (described in Section A.1.2) made the biggest difference to the issues at interest.

A.1.1 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:

$$R_i^T = \prod_{t=1}^{T}(1 + r_{it}) - 1, \qquad T \in \{36, 60, 120\}. \tag{7}$$

where the monthly fund returns r_{it} are net of management fees but not yet adjusted for loads. Similarly, R_f^T is defined as the cumulative risk-free rate return for horizon T using three-month T-bills. Then, Morningstar adjusts the cumulative returns for loads to get a load-adjusted return over the risk-free return:

$$\text{LoadRet}_i^T = R_i^T L_i - R_f^T \tag{8}$$

where the load adjustment L_i is equal to 1 minus the sum of the front- and back-end load. Morningstar then standardizes the measure to get:

$$\text{MnLoadRet}_i^T = \frac{\text{LoadRet}_i^T}{\max(R_f, \text{AvgLoadRet}^T)}, \tag{9}$$

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:

$$\text{Performance}_{it} = \text{MnLoadRet}_{it}^{T} - \text{MnRisk}_{it}^{T} \tag{10}$$

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

$$\text{Risk}_{i}^{T} = \frac{\sum_{t=1}^{T} -\min(r_{it} - r_{t}^{f}, 0)}{T},$$

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

$$\text{MnRisk}_{t}^{T} = \frac{\text{Risk}_{i}^{T}}{\text{AvgRisk}^{T}}. \tag{11}$$

After June 2002, Morningstar conducts risk adjustment in a slightly different way.[20] Morningstar summarizes a fund's past performance using the so-called Morningstar risk-adjusted return (MRAR):

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

where $r_{it} - r_{t}^{f}$ is the geometric return in excess of the risk-free rate after adjusting for loads,[21] and $\gamma = 2$ is the risk aversion coefficient.

The formula penalizes funds with higher return volatility. To see this, notice that when γ converges to 0, $\text{MRAR}^{\text{T}}(0)$ is equal to the annualized geometric mean of excess returns.[22] 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 $\text{MRAR}^{\text{T}}(0) - \text{MRAR}^{\text{T}}(2)$.

[20] Morningstar explains its post-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).

[21] 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 past T months. For details, see "The Morningstar Rating Methodology," June 2006.

[22] 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 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.

A.1.2 Step two: Rank funds and assign ratings

Given rankings of funds, Morningstar calculates 3-year, 5-year, and 10-year ratings for funds with the necessary 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 3 years but less than 5 years of data, the overall rating is just the 3-year rating. For funds with more than 5 years but less than 10 years of data, the overall rating assigns 60% and 40% weights on the 5-year and 3-year ratings. For those with 10 years of data, 50%, 30%, and 20% weights are assigned on the 10-year, 5-year, and 3-year ratings, respectively.

The ratings are based on rankings of funds. Before June 2002, Morningstar 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 3).

A.2 Replicating Morningstar Ratings

To construct counterfactual ratings using the post-2002 methodology in the pre-2002 period, we start from fund returns and calculate Morningstar ratings from scratch. This allows us to isolate the effect that is due to the methodology change.

We can use observed ratings to gauge the accuracy of our replication. Due to differences in data sources, we cannot replicate ratings with 100% accuracy at the fund level.[23] However, when aggregating at the factor level, errors in replication cancels out and we are very close to being able to match actually observed ratings.

We first verify that the accuracy of rating replication changed sharply at June 2002. Panel A of Figure A.I plots the fraction of funds for which the replication is exact. Replicated ratings are correct around 80% of the time before June 2002, and that declined to around 40%, while accuracy of the new method went up. Panel B plots the average rating estimation error (in stars) and show a similar pattern. When using the right methodology, average errors

[23] While we use Morningstar Direct data, it is clear that our data source is not exactly the same as the historical data set used by Morningstar. For instance, there are fund-months where we cannot find ten years of past returns, but Morningstar does assign a ten year rating, so Morningstar must have more data internally.

are below 0.4 stars most of the time. Panel C shows the average rating estimation error when aggregating at the factor level. When using the correct methodology, the error is usually below 0.1 stars.

As a consequence, we can estimate the factor-level rating change on June 2002 quite well. Figure A.II plots the actual rating change from May to June 2002 against the predicted changes using May 2002 data.

A.3 Long-Term Impact of Morningstar Event on Factors

We sort factors by their average pre-2002 StyleRating into quintiles and tabulate the sorting in Table A.I. The most affected factors are, as hypothesized, mostly momentum type factors or related to momentum such as the 52-week high strategy in George and Hwang (2004). There are fewer factors being negatively distorted, but those tend to be value/profitability types.

Panel A in Figure A.III plots the evolution of ΔRating for the five quintiles. Two facts emerge. First, there is large dispersion of ΔRating before 2002, and it is almost entirely driven by Morningstar-exposure. Second, there is little dispersion after 2002. Panel B shows that, for factors in the top quintile, the effect is mainly coming from the Morningstar exposure. If Morningstar adopted the post-2002 methodology, those factors would have had very little ΔRating.

A.4 Results: Short Window Around 2002

Figure A.I. Accuracy of Rating Estimation

We estimate fund ratings using fund returns using both the old (pre-2002) and the new (post-2002) methodologies. Panel A plots the fraction of funds for which our estimated ratings exactly match the acual ratings. Panel B plots the average estimation error for funds and Panel C plots the average error when ratings are aggregated at the factor level.

Panel A: Accuracy rate at fund level



Panel B: Average estimation error at fund level



Panel C: Average estimation error at factor level

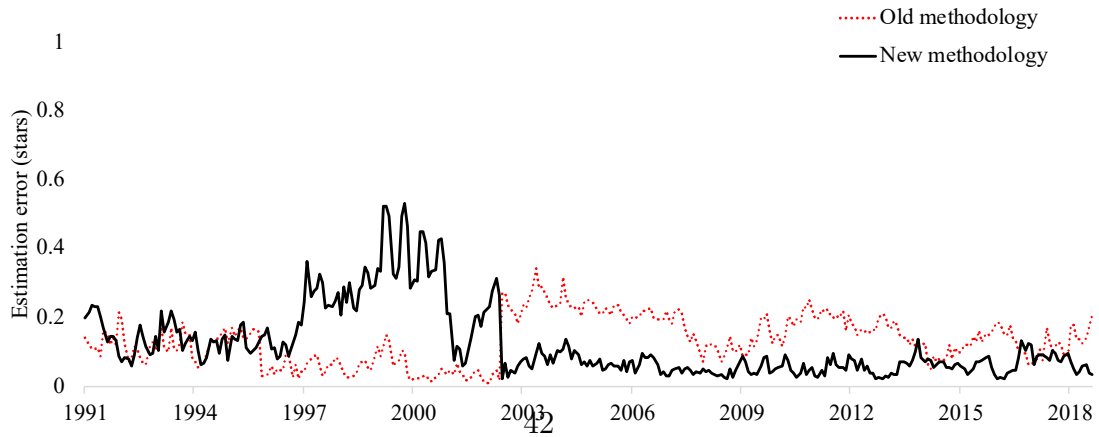


Figure A.II. Predicting Factor-Level Rating Changes in June 2002

The horizon axis plots the predicted rating changes by factor using May 2002 data. The vertical axis plots the actual change in June 2002. The green dashed line is the 45% line.

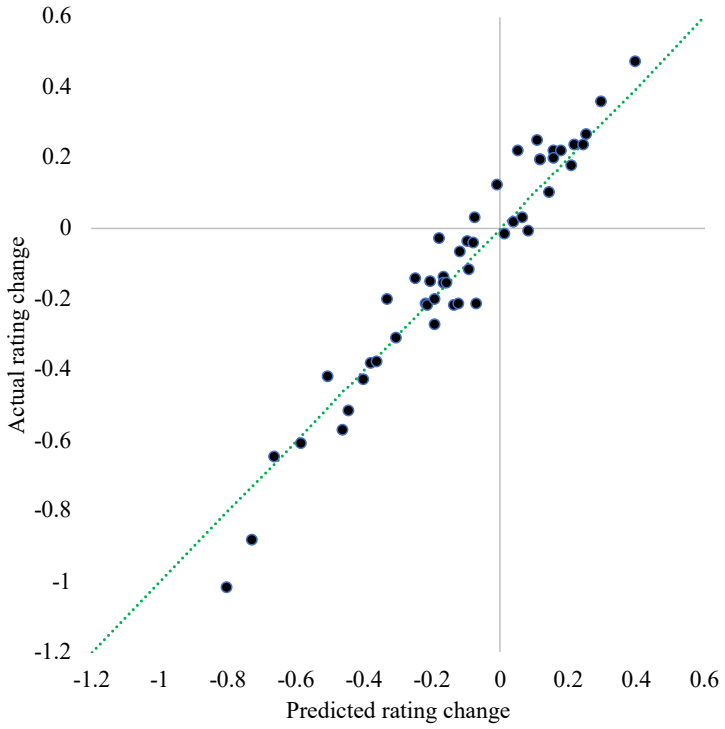
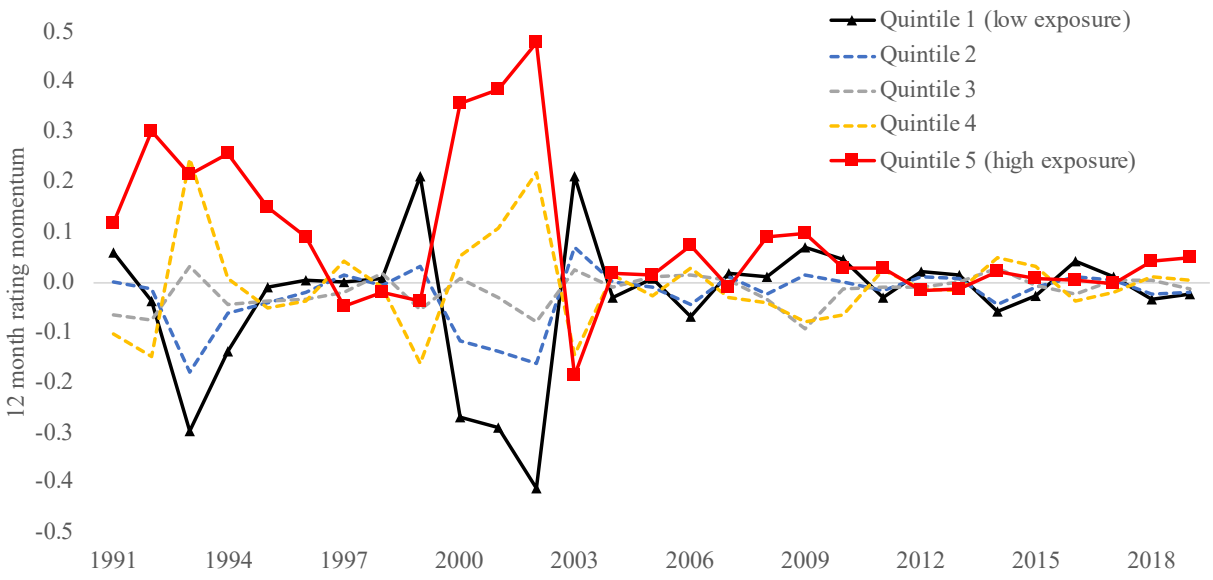


Figure A.III. ΔRating and Returns of Factors

Panel A: ΔRating of factors by exposure to old methodology



Panel B: Factors in top quintile of pre-2002 Morningstar StyleRating

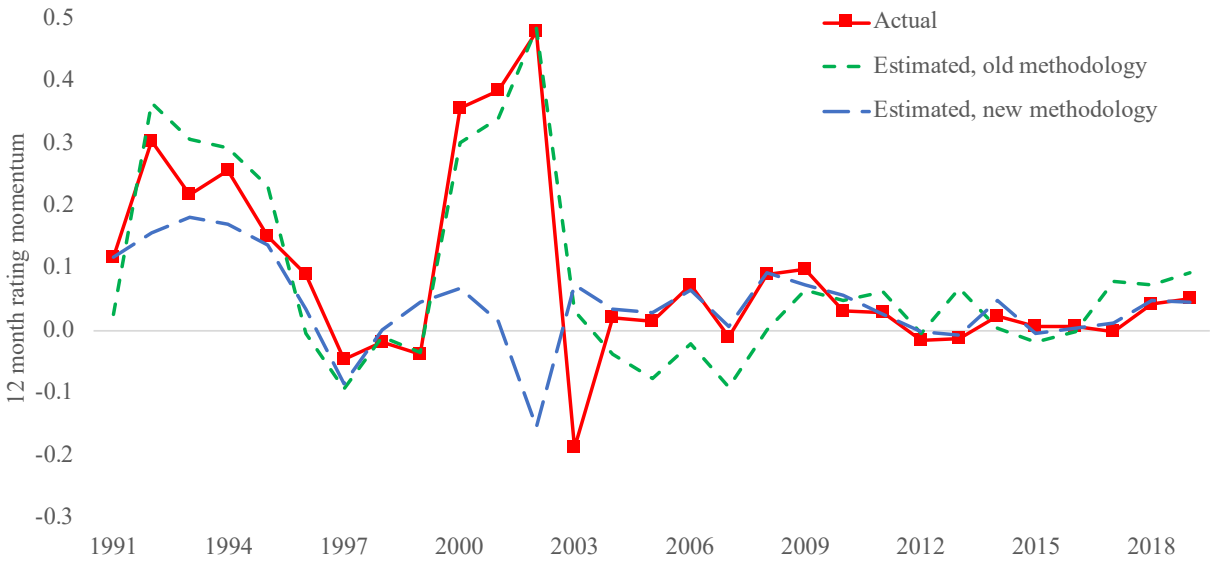


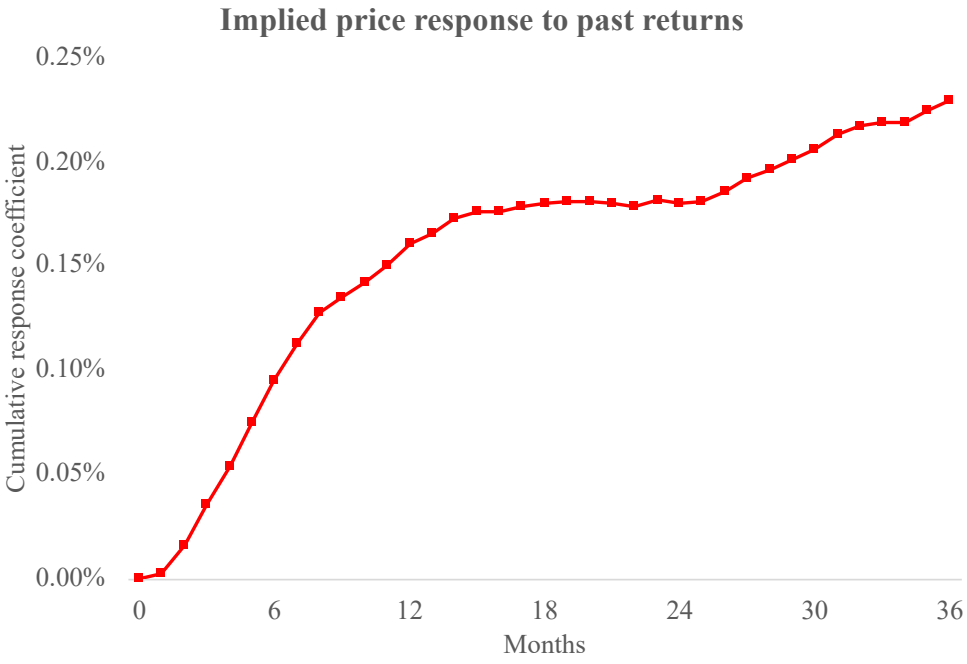
Table A.I. Factors, Ranked on Pre-2002 Exposure

Ranking	Factor	Pre-event StyleRating	Average Monthly Return Before	After 2002	Difference
1	Operating profitability	-0.066	0.667%	0.053%	-0.614%
2	Altman's Z-score	-0.046	0.065%	-0.092%	-0.157%
3	Maximum daily return	-0.042	0.354%	0.104%	-0.250%
4	Profit margin	-0.036	0.709%	-0.210%	-0.918%
5	Ohlson's O-score	-0.036	0.736%	-0.096%	-0.831%
6	Cash-based profitability	-0.030	-0.065%	0.619%	0.684%
7	Operating Leverage	-0.027	0.014%	0.256%	0.242%
8	Return on equity	-0.022	0.490%	0.210%	-0.281%
9	QMJ Profitability	-0.021	1.105%	0.205%	-0.900%
10	Piotroski's F-score	-0.016	0.076%	0.223%	0.147%
11	Earnings Persistence	-0.012	0.659%	0.651%	-0.007%
12	Five-year share issuance	-0.009	0.257%	0.356%	0.098%
13	Gross profitability	-0.008	0.175%	0.401%	0.226%
14	Industry adjusted CAPX growth	-0.007	0.393%	-0.112%	-0.505%
15	Return on assets	-0.006	0.645%	0.074%	-0.571%
16	Earnings-to-price	-0.006	0.602%	0.445%	-0.156%
17	Total external financing	0.002	0.522%	0.129%	-0.392%
18	Cashflow-to-price	0.006	0.505%	0.144%	-0.361%
19	Accruals	0.008	0.873%	0.161%	-0.713%
20	Sales growth	0.011	0.105%	0.089%	-0.017%
21	Net Payout Yield	0.012	1.213%	0.091%	-1.122%
22	Industry Concentration	0.014	0.111%	0.339%	0.228%
23	Investment-to-assets	0.016	0.059%	-0.070%	-0.129%
24	Investment-to-capital	0.016	0.098%	-0.102%	-0.200%
25	Net working capital changes	0.017	0.866%	0.171%	-0.695%
26	Firm age	0.022	-0.314%	0.321%	0.636%
27	Sales-minus-inventory growth	0.023	0.408%	-0.144%	-0.553%
28	Investment growth	0.024	0.684%	0.017%	-0.667%
29	One-year share issuance	0.024	1.337%	0.099%	-1.238%
30	Sustainable growth	0.025	0.630%	0.117%	-0.513%
31	Growth in Inventory	0.032	0.639%	0.487%	-0.152%
32	Asset Growth	0.041	0.750%	0.081%	-0.669%
33	Net operating assets	0.041	1.367%	0.455%	-0.912%
34	Book-to-market	0.042	0.524%	0.014%	-0.510%
35	Change in asset turnover	0.050	0.561%	-0.004%	-0.565%
36	R&D Expense	0.054	1.147%	0.489%	-0.658%
37	Sales-to-price	0.056	0.690%	0.103%	-0.587%
38	Abnormal capital investment	0.061	1.125%	0.454%	-0.672%
39	Advertising Expense	0.072	0.433%	0.127%	-0.306%
40	Enterprise multiple	0.077	0.068%	0.116%	0.048%
41	Distress risk	0.091	1.941%	0.189%	-1.752%
42	Long-term reversals	0.095	0.936%	-0.038%	-0.973%
43	Amihud Illiquidity	0.118	0.513%	0.238%	-0.275%
44	Momentum (t-2, t-6)	0.125	0.933%	-0.098%	-1.031%
45	52-week high	0.130	1.176%	-0.699%	-1.875%
46	Size	0.137	0.517%	0.101%	-0.415%
47	Intermediate momentum (t-7,t-12)	0.142	0.895%	0.333%	-0.562%
48	Industry Momentum	0.177	0.625%	0.020%	-0.605%
49	Momentum (t-1, t-12)	0.205	1.895%	0.102%	-1.794%

Table A.II. 12 Months Before and After Event

Ranking	Quintile	Factor	Pre-2002 Morningstar StyleRating	ΔRating Before 2002	ΔRating After 2002	Change
1	1	Altman's Z-score	-0.426	-0.552	0.263	0.815
2	1	Sales growth	-0.406	-0.454	0.184	0.637
3	1	Ohlson's O-score	-0.342	-0.347	0.282	0.629
4	1	Cash-based profitability	-0.324	-0.290	0.127	0.417
5	1	Industry Concentration	-0.316	-0.302	0.089	0.392
6	1	Operating Leverage	-0.270	-0.258	0.090	0.348
7	1	Piotroski's F-score	-0.191	-0.160	-0.010	0.150
8	1	Earnings Persistence	-0.188	-0.198	0.105	0.303
9	1	Gross profitability	-0.171	-0.149	0.142	0.291
10	1	Profit margin	-0.094	-0.071	0.077	0.149
11	2	QMJ Profitability	-0.068	-0.078	0.166	0.245
12	2	Return on equity	-0.038	-0.052	0.135	0.188
13	2	Return on assets	-0.032	-0.023	0.127	0.151
14	2	Firm age	-0.014	-0.060	-0.083	-0.023
15	2	Net working capital changes	0.025	-0.027	-0.097	-0.070
16	2	Accruals	0.030	-0.040	-0.017	0.022
17	2	Sales-minus-inventory growth	0.032	-0.032	0.085	0.117
18	2	Net operating assets	0.076	0.101	-0.113	-0.214
19	2	Investment-to-assets	0.132	0.173	-0.167	-0.340
20	2	Operating profitability	0.165	0.231	0.096	-0.135
21	3	Momentum (t-2, t-6)	0.171	0.210	-0.127	-0.338
22	3	Distress risk	0.173	0.267	-0.029	-0.296
23	3	Industry adjusted CAPX growth	0.182	0.272	-0.289	-0.561
24	3	Sustainable growth	0.186	0.215	-0.085	-0.299
25	3	Maximum daily return	0.230	0.337	-0.066	-0.403
26	3	Five-year share issuance	0.241	0.268	-0.035	-0.303
27	3	Earnings-to-price	0.248	0.429	-0.163	-0.592
28	3	Growth in Inventory	0.251	0.199	-0.114	-0.313
29	3	Net Payout Yield	0.254	0.366	-0.049	-0.415
30	3	R&D Expense	0.264	0.298	-0.228	-0.526
31	4	Abnormal capital investment	0.267	0.290	-0.360	-0.650
32	4	Total external financing	0.277	0.354	-0.085	-0.439
33	4	One-year share issuance	0.285	0.425	0.077	-0.349
34	4	Asset Growth	0.331	0.383	-0.117	-0.500
35	4	Change in asset turnover	0.350	0.437	-0.035	-0.472
36	4	Industry Momentum	0.386	0.418	-0.121	-0.540
37	4	Intermediate momentum (t-7,t-12)	0.414	0.466	-0.310	-0.777
38	4	Investment growth	0.453	0.583	-0.231	-0.814
39	4	Enterprise multiple	0.457	0.444	-0.276	-0.721
40	4	Cashflow-to-price	0.575	0.783	-0.221	-1.004
41	5	Momentum (t-1, t-12)	0.600	0.723	-0.272	-0.994
42	5	Investment-to-capital	0.601	0.740	-0.251	-0.991
43	5	52-week high	0.603	0.753	-0.168	-0.921
44	5	Book-to-market	0.670	0.831	-0.408	-1.239
45	5	Advertising Expense	0.794	0.929	-0.320	-1.249
46	5	Long-term reversals	0.916	1.185	-0.356	-1.541
47	5	Sales-to-price	0.941	1.129	-0.456	-1.585
48	5	Amihud Illiquidity	1.208	1.374	-0.553	-1.927
49	5	Size	1.265	1.489	-0.620	-2.109

Figure A.IV. Stock Return Feedback Loop, Impulse Response



Appendix B Additional Results on Factor Momentum

We provide additional results about factor momentum (Section 6). We use stock-level data to help pin down the horizon at which Morningstar should have the largest effect on factor momentum. When we put together the first link and the second two (coefficients in footnote), we get that the effect should be strongest around 12- to 24-month window. This magnitude cannot be taken literally, due to the severe attenuation bias, but can be used to figure out the appropriate look back window.

We tabulate the average return for factor momentum before versus after 2002 in Table A.III.

Table A.III. Change in Factor Momentum Strategy Returns

Panel A: 12 months before to 12 months after event					
Look-back window (w)	**Monthly return**			**Difference of change**	
	Before	**After**	**Change**	**from $w = 12$**	**from $w = 24$**
1	1.06%	0.16%	−0.90%	1.34%	1.61%
3	-0.19%	0.27%	0.47%	2.70%**	2.98%**
6	0.96%	0.15%	−0.81%	1.42%	1.69%
12	1.76%	-0.48%	−2.24%*		0.27%
24	1.67%	-0.84%	−2.51%**	−0.27%	
36	1.26%	-0.61%	−1.87%*	0.37%	0.64%
60	1.35%	-0.42%	−1.77%**	0.47%	0.74%*

Panel B: Full sample					
Look-back window (w)	**Monthly return**			**Difference of change**	
	Before	**After**	**Change**	**from $w = 12$**	**from $w = 24$**
1	0.48%	0.24%	−0.24%	0.26%	0.20%
3	0.42%	0.22%	−0.20%	0.30%	0.25%
6	0.34%	0.13%	−0.21%	0.30%*	0.24%
12	0.61%	0.10%	−0.51%**		−0.06%
24	0.39%	-0.05%	−0.45%**	0.06%	
36	0.18%	-0.04%	−0.22%	0.28%*	0.23%**
60	0.24%	0.02%	−0.23%	0.28%*	0.22%*

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