Style Investing, Positive Feedback Loops, and Asset Pricing Factors

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

VERY PRELIMINARY, PLEASE DO NOT CITE

May 1, 2020

Abstract

We show that correlated demand that is driven by performance chasing creates positive feedback in stock returns, and explains a substantial fraction of the premia of asset pricing factors. Between 1991 and 2018, mutual fund investors chased Morningstar's fund ratings regardless of methodology. Until mid-2002, funds in the best-performing styles received high ratings, and as a result, investors' flows introduced style-level price pressures and positive feedback loops in the underlying equity market. A 2002 revision to Morningstar's methodology equalized investors' demand across styles. We show that the decline in correlated demand explains half of factor profitability, especially for factors that were most exposed to the change in Morningstar's methodology and for momentum.

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

*We thank Sylvester Flood (Morningstar), Paul Kaplan (Morningstar), Andrei Shleifer, and Xin Wang for helpful comments. We thank seminar participants at the University of Utah for comments and George Aragon for sharing data. Ben-David is with The Ohio State University and the National Bureau of Economic Research, Li is with University of Utah, Rossi is with the University of Arizona, and Song is with the University of Washington. Emails: ben-david.1@osu.edu, jiacui.li@eccles.utah.edu, rossi2@email.arizona.edu, and songy18@uw.edu.

1 Introduction

Over the last four decades, the asset pricing literature has identified hundreds of factors that span the cross-section of stock returns.[1] Since the discovery of the size factor in 1981, financial economists have strongly debated the origins of factor profitability. Some propose that the profits reflect compensation for bearing fundamental economic risk.[2] Others claim that factors stem from systematic mispricings.[3] One specific strand of the literature—the "style investing hypothesis" introduced by Barberis and Shleifer (2003)—argues that factors can be driven by correlated demand from investors. Recent studies have also proposed that some factors that have been identified through data mining and therefore may not truly reflect return factors out-of-sample (Harvey, Liu, and Zhu, 2016; Harvey, 2017). So far, the literature has mostly relied on reduced-form tests and structural models to measure the validity of specific explanations. These techniques, however, are open to the critique that other, unobservable, forces could also determine equilibrium prices.

Explanations of the underlying economics of factor profitability ought also to the explain the perplexing sharp drop in factor return profitability since the early 2000s. The "profitability kink" in mid-2002 has been documented in earlier work[4] and is visible in Figure 1, Panel (a). The average monthly return of 49 popular factors went from 0.63% during the period of 1991 to June 2002 to 0.14% during the later period of July 2002 to 2018. The abrupt drop in performance is especially puzzling because it was specific to the U.S. and not observed in most other advanced markets (Asness, Moskowitz, and Pedersen, 2013).

In this study, we present two novel and related findings. First, we show *causal* evidence that style-investing price pressures have a first-order impact on asset pricing factor profitability. Second, we show that an seemingly innocuous institutional change—Morningstar's

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

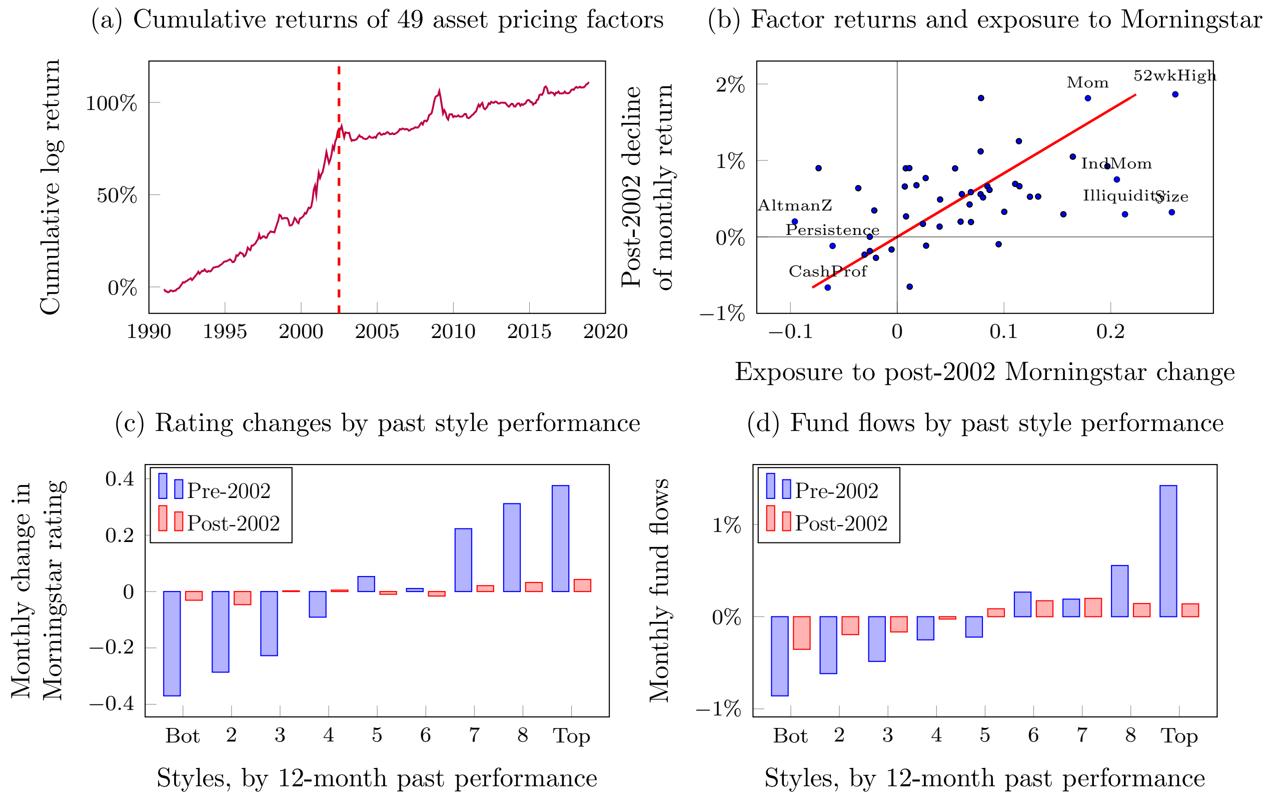
[2] The literature proposing risk-based explanations of factor profitability is vast. For examples, see Fama and French (1993), Jagannathan and Wang (1996), Berk, Green, and Naik (1999), Gomes, Kogan, and Zhang (2003), Pástor and Stambaugh (2003), Campbell and Vuolteenaho (2004), Zhang (2005), Hansen,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

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

[4] See Marquering, Nisser, and Valla (2006), Green, Hand, and Soliman (2011), Israel and Moskowitz (2013), Chordia, Subrahmanyam, and Tong (2014), Daniel and Moskowitz (2016), Jones and Pomorski (2016), McLean and Pontiff (2016), Green, Hand, and Zhang (2017), Arnott, Harvey, Kalesnik, and Linnainmaa (2019b), Gupta and Kelly (2019), and Ehsani and Linnainmaa (2019). For the readers' convenience, we present screenshots from Green et al. (2017) and Daniel and Moskowitz (2016) in Appendix Section A.

Figure 1. Morningstar Rating Methodology Change and Factor Returns

The figure shows the main results in this study. Panel (a) shows the cumulative average returns of 49 popular asset pricing factors. Panel (b) shows the change in factor premia (pre-2002 minus post-June 2002) as a function of pre-2002 exposure to the Morningstar methodology change. Panel (c) shows the total net assets (TNA)-weighted average 12-month rating change of funds ($\text{Rating}_t - \text{Rating}_{t-12}$) by the past 12-month performance of their style (e.g., large cap-growth), separately for pre- and post-June 2002. Panel (d) shows the TNA-weighted flows to funds by their style's 12-month past performance pre- and post-June 2002.



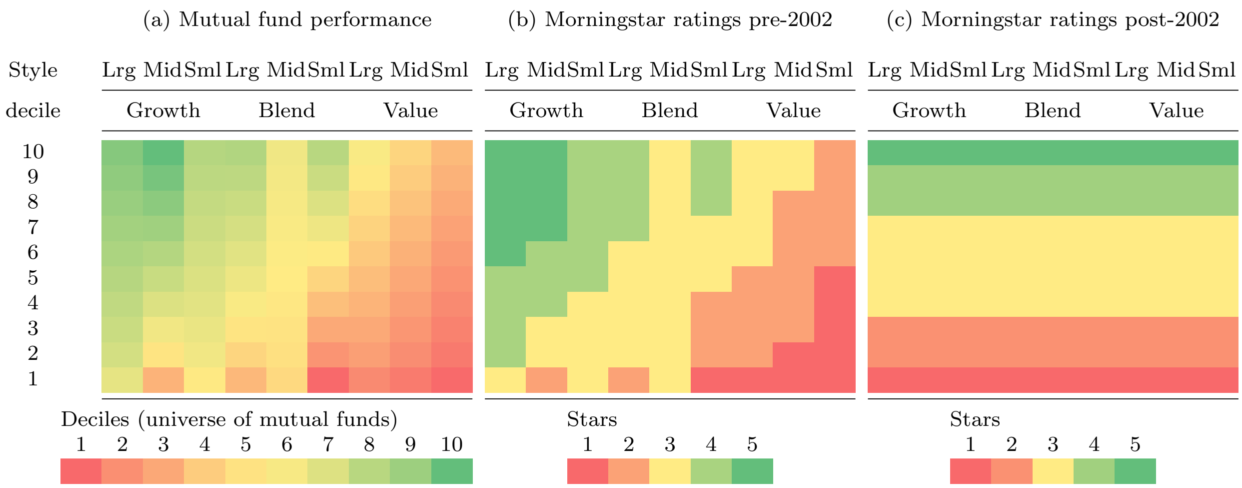
fund rating methodology change in June 2002—dramatically disrupted style-level feedback loops, causing the sharp decline in the profitability of a large group of factors. The change in Morningstar's methodology effectively removed a component of correlated demand (related to investment styles) from the aggregate investment by mutual funds. As a result, factor profitability experienced a reversal in the short run and lost steam in the long run. We estimate that this change contributed to around half of the factor profitability decline post-June 2002. Overall, our results are consistent with the hypothesis in Barberis and Shleifer (2003) that asset pricing factors can emerge out of style-level correlated demand.

The style-level demand in our study originates from the mutual fund industry. Because mutual funds follow investment strategies ("styles"),[5] their past performance contains a large style-level component. Before June 2002, Morningstar's mutual fund ratings closely map

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

Figure 2. Illustration of Morningstar Methodology Pre- and Post-June 2002

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



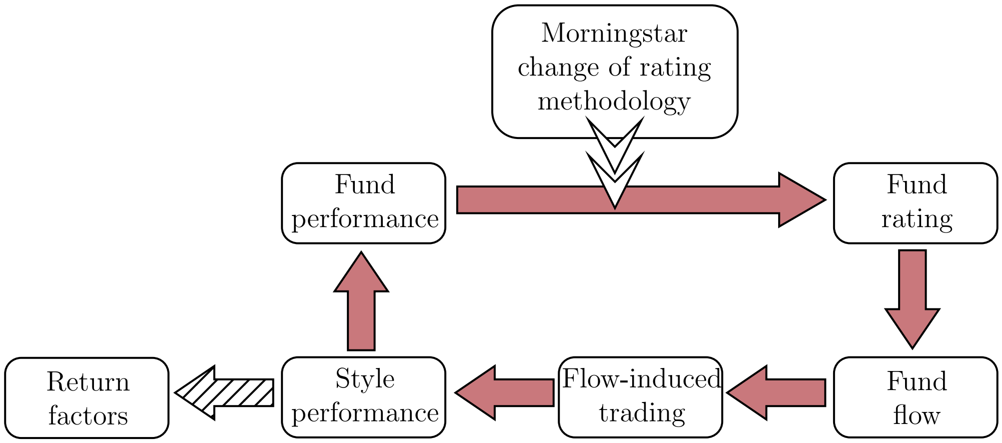
absolute past fund performance into star ratings, as illustrated in Figure 2. Panel (a) shows a snapshot of mutual funds' past performance (colors) for funds within styles. Panel (b) shows how Morningstar would translate funds' performance into star ratings. Because mutual fund investors chase unadjusted past performance and Morningstar ratings (e.g., Del Guercio and Reuter, 2014; Reuter and Zitzewitz, 2015; Ben-David, Li, Rossi, and Song, 2019; Evans and Sun, 2020), their flows would appear *as if* they chase style returns, and aggregate mutual fund investment puts price pressures on stocks associated with the common style in demand.

Performance chasing creates a positive feedback loop: Funds that pursue styles that performed well in the recent past attract new capital. They use the new capital to increase their investments in the same set of styles, pushing the prices of securities associated with the styles even further. Stocks associated with the styles in demand, therefore, would exhibit *i*) return momentum and *ii*) predictable return patterns that are associated with styles. The mechanism is illustrated in Figure 3. The mechanism also works in the other direction: Funds in underperforming styles experience outflows, resulting in downward price pressure on stocks associated with these styles.

Our study revolves around a shock: In June 2002, Morningstar abruptly changed its rating methodology, which caused a halt in ratings-induced style investing. Since then,

Figure 3. Style Investing Feedback Loop

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



Morningstar ranks funds *within* 3×3 style categories[6] in order to compare fund managers to their peer group. Based on the modified methodology, Morningstar would rate funds as exemplified in the Panel (c) of Figure 2. Because investors were inattentive to the methodology change and continued to chase Morningstar ratings as before, flows became evenly distributed across styles. After June 2002, mutual fund investors appear *as if* they no longer chased style-level returns. As a direct consequence, mutual fund aggregate investment was no longer concentrated in the best-performing styles but rather became more distributed across all styles. Factors benefiting from style-related price pressure abruptly stopped being profitable.

Our empirical analysis proceeds in three steps. First, we demonstrate the effect of the Morningstar methodology change on fund flows. Before June 2002, past style performance was a major driver of Morningstar ratings and aggregate fund flows. During this early period, mutual funds in the top four performing styles received flows that were higher by an average of 1.1% of assets under management (AUM) than funds in the bottom half of style performance, *per month*. Looking at the best- and worst-performing styles, funds in the top-performing style received flows that are 2.3% of AUM higher than those in the worst-performing style, per month. After June 2002, there was virtually no relation between past style performance and Morningstar ratings or fund flows. These dramatic effects are depicted in Figure 1, Panels (c) and (d).

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

We end the first part of our analysis with an illustration of the impacts on the momentum factor. For each momentum decile portfolio, we summarize the lagged ratings changes of the mutual funds that hold the portfolio stocks. Before June 2002, the decile portfolio of winner stocks experienced higher ratings changes and also an additional 0.7% of inflows per month relative to the loser decile. After the methodology change, the difference became much more muted. The profitability of the momentum factor also disappeared: Its monthly returns declined from 1.90% during the period of 1991 to June 2002 to a meager 0.10% during the later period of July 2002 to 2018.

Second, we use a short window of 12 months around the methodology change to causally estimate the impact of ratings on the returns of 49 popular factors that are based on characteristics sortings. The benefit of this exercise is that the rating changes mostly come from the methodology change, and we can also be reasonably certain that there are no other major shocks to factor returns. The methodology change created a heterogeneous impact on factors. As predicted, factors that were positively (negatively) affected experienced an increase (reduction) in flows and an increase (reduction) in returns. We estimate that each star rating revision leads to a price impact of around 2% per month at the factor level.

Finally, we study the long-term impact of ratings on factor profitability through our full sample of 1991 to 2018. Over the full sample, lagged ratings changes predict factor returns above and beyond existing factor return predictors, consistent with the price pressure interpretation. Further, the methodology change had long-lasting effects on many factors. As we hypothesized, factors related to momentum—such as industry momentum, the 52-week high strategy, etc.—were the most impacted. Some other factors, such as those based on trading illiquidity, were similarly affected, but to a lesser degree. Morningstar pre-2002 methodology served as an important tailwind for the profitability of those strategies, and the methodology change explains around half of the decline of profitability of those strategies after June 2002.

Our study has important implications for understanding the origins of asset pricing factor profitability. Our findings imply that over our sample period, a substantial fraction of factor profitability arose from style-level price pressures and thus does reflect compensation for risk. These results also contribute to our understanding of the decline in factor profitability over time. While our explanation is orthogonal to existing theories, it is the only interpretation that explains the abrupt and one-time change in factor returns (the "kink") and pinpoints the exact date of profitability decline. Existing papers on factor profitability decline emphasize reductions in arbitrage costs (Khandani and Lo, 2011; Chordia et al., 2014; Lee and Ogden, 2015) or the entry of arbitrageurs (e.g., hedge funds, see Green et al., 2011; Hanson and Sunderam, 2013), especially after academic studies documented the profitability of factors

(Marquering et al., 2006; McLean and Pontiff, 2016; Calluzzo, Moneta, and Topaloglu, 2019). It is also possible that some of the factors reported in the literature were data-mined and therefore stopped working after their "discovery" (Harvey et al., 2016; Harvey, 2017).

The rest of the paper is organized as follows. Section 2 details the data and variable construction. Section 3 shows the mechanism—ratings lead to fund flows, and flow-induced trading leads to price pressures—and also explains how the 2002 methodology change affected factor strategies. Section 4 uses a short window around the 2002 methodology change to estimate the factor-level price impact of Morningstar ratings. Section 5 examines Morningstar's impact throughout the sample period and quantifies its effect on post-June 2002 factor profitability decline. Section 6 concludes. Robustness checks and additional tests are provided in the Appendix.

2 Data and Variable Construction

This section describes the data set and how we construct the asset pricing factors. Our data are at a monthly frequency and span 1991 to 2018. We start in 1991 because monthly fund flow data from the Center for Research in Security Prices (CRSP) starts in 1990, and some measures require one year of lagged data to construct.

2.1 Mutual Fund Data

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

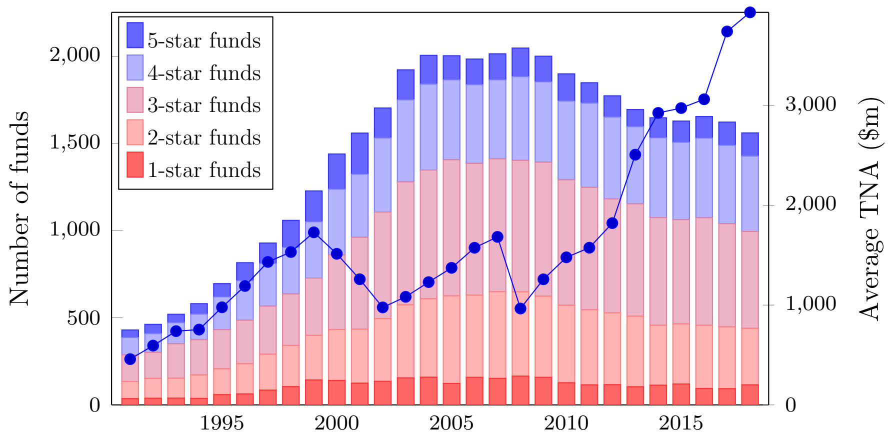
Following the fund flow literature (e.g., Coval and Stafford, 2007), the fund flow for fund j in month t is defined as the net flow into the fund divided by the lagged TNA:

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

We obtain Morningstar ratings and style categories from Morningstar Direct and merge them with the CRSP mutual fund data using the matching table from Pástor, Stambaugh,

Figure 4. Number of Funds and the Average TNA over Time

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



and Taylor (2015).[7] Morningstar assigns ratings at share class level, and we follow Barber, Huang, and Odean (2016) in aggregating them at the fund level by TNA-weighting different share classes. We restrict our analysis to mutual funds with at least \$1 million TNA, and we winsorize fund flows at the 0.5% and 99.5% levels. We require the existence of 12 lags of monthly flows, returns, and ratings. The resulting sample comprises a total of 3,305 funds with 454,787 fund-month observations. We present the time series of the number of funds and their average size in Figure 4. The figure shows that the number of funds quadrupled from 1991 to 2005, then declined a bit by the end of the sample. The average TNA of funds increased by a factor of eight from the beginning to the end of the sample.

2.2 Stock-Level Rating and Flow-Induced Trading

We are interested in how fund flows and ratings lead to stock-level trading pressures. Therefore, we also aggregate flows and ratings at the stock level. To measure the amount of mutual fund trading caused by fund flows, we follow Lou (2012) to calculate flow-induced trading (FIT) for each stock i in each month t:[8]

$$\mathrm{FIT}_{i,t} = \frac{\sum_{\text{fund j}} \mathrm{SharesHeld}_{j,t-1} \cdot \mathrm{Flow}_{j,t}}{\sum_{\text{fund j}} \mathrm{SharesHeld}_{j,t-1}}. \tag{2}$$

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

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

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

Because we are interested in the flow pressure induced by ratings, we also define the average Morningstar rating or the rating change of stock i in month t as the holding-weighted rating or the rating change of all funds that hold the stock:

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

$$\Delta\text{Rating}_{i,t}^{\text{stock}} = \frac{\sum_{\text{fund j}} \text{SharesHeld}_{j,t-1} \cdot (\text{Rating}_{j,t} - \text{Rating}_{j,t-1})}{\sum_{\text{fund j}} \text{SharesHeld}_{j,t-1}}. \tag{4}$$

We later drop the superscript "stock" when unambiguous.

2.3 Asset Pricing Factors

Following prior literature, we compute 49 popular stock-level characteristics that have been shown to predict returns. We restrict our attention to those that can be constructed using CRSP and Compustat data. Using the classification categories proposed in Hou, Xue, and Zhang (2019), these characteristics include 14 in the profitability category (e.g., return on assets), 14 in the investments category (e.g., share issuance), eight in the value/growth category (e.g., book-to-market), six in the intangibles category (e.g., industry concentration), five in the momentum category (e.g., momentum of Jegadeesh and Titman, 1993), and three in the trading frictions category (e.g., Amihud illiquidity). Table 1 lists all asset pricing factors.[10]

We follow the procedure in Hou et al. (2019) to construct long-short factor portfolios using these characteristics. To minimize the impact of microcaps, Hou et al. use NYSE breakpoints 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.

In addition to computing the standard long-short portfolio returns, we also aggregate up

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

[10] One factor is binary in nature, hence it is exlcuded from the analysis.

Table 1. Asset Pricing Factors

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

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

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

stock-level variables to compute factor-level ratings and FIT. For each factor f, we calculate

$$\text{Rating}_{f,t} = \sum_{i \in \text{top decile}} w_{i,t-1}^{f} \cdot \text{Rating}_{i,t}^{\text{stock}} - \sum_{i \in \text{bottom decile}} w_{i,t-1}^{f} \cdot \text{Rating}_{i,t}^{\text{stock}} \tag{5}$$

$$\text{FIT}_{f,t} = \sum_{i \in \text{top decile}} w_{i,t-1}^{f} \cdot \text{FIT}_{i,t}^{\text{stock}} - \sum_{i \in \text{bottom decile}} w_{i,t-1}^{f} \cdot \text{FIT}_{i,t}^{\text{stock}}, \tag{6}$$

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

3 Mechanism and Hypothesis Development

In this section, we investigate the dynamic impact of ratings on fund flows and price pressures. We then describe the June 2002 Morningstar methodology change and develop our hypothesis about its impact on asset pricing factor profitability. Finally, we use the momentum factor to illustrate the impact.

3.1 Mutual Fund Sector Size and Flows

We first note that aggregate mutual fund flows are large enough to generate a nonnegligible price impact at the factor level. Mutual funds, as a prime investment vehicle for retail investors, hold a substantial and increasing share of the U.S. equity market. When our sample begins in 1991, U.S. equity mutual funds had total AUM of \$326 billion, which was 8.9% of the entire market capitalization. These numbers grew steadily over time, and by 2018, the end of our sample period, equity mutual funds owned \$10,849 billion, which represented 29.3% of the entire market capitalization (Panel (a) of Figure 5). Panel (b) shows the flows to mutual funds by Morningstar ratings. Throughout our sample period, 5-star funds receive flows that amount to +1.5% of their AUM per month on average, and 1-star funds experience outflows amounting to −1.2% of their AUM per month on average. Since these flows translate to trades in the stock market, these two panels suggest that Morningstar ratings can generate substantial flow pressures on stocks.

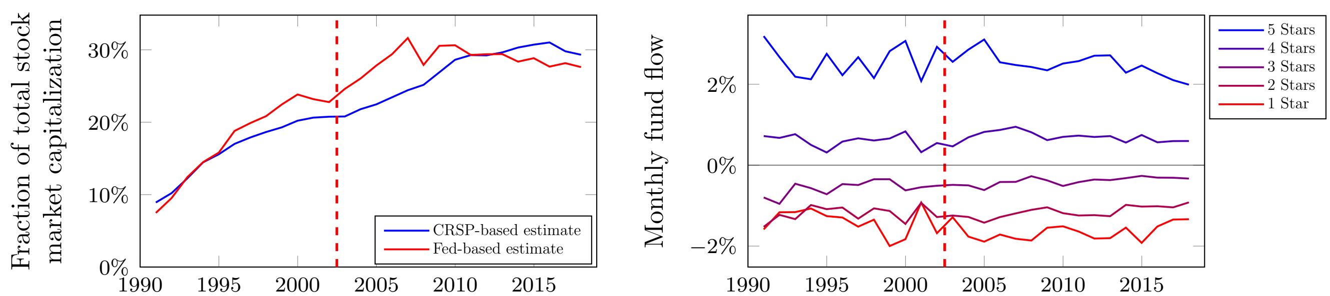
3.2 Mechanism: Rating-Induced Price Pressures

We use Fama-MacBeth regressions (Fama and MacBeth, 1973) to estimate the chain of dynamic effects: *i*) the response of fund flows to Morningstar rating changes, and *ii*) the response of stock returns to flow-induced trading. Because we are ultimately interested in studying factor-level outcomes, all regressions are value-weighted: Fund-level regressions are

Figure 5. Aggregate Mutual Fund Sector Size and Flows

The figure shows summary statistics about the size of the U.S. equity mutual fund space over time and flows to those funds by Morningstar rating. Panel (a) shows the aggregate domestic stock holdings by mutual funds over time as a fraction of the total U.S. stock market. The blue line is based on the CRSP mutual fund database, and the red line is based on Federal Reserve Board flow of fund reports (L.223). Panel (b) shows the TNA-weighted average monthly flows to funds by lagged Morningstar ratings.

(a) Aggregate mutual fund ownership — (b) Mutual fund flows by Morningstar rating



weighted by fund TNA, and stock-level regressions are weighted by stock market capitalization.[11]

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

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

where $\Delta\text{Rating}_{j,t}$ is the month t rating change of fund j, and controls $X_{i,t}$ include 36 monthly lags of fund flows and returns. The cumulative response coefficients $(b_1, b_1+b_2, ...)$ are plotted in Panel (a) of Figure 6. In response to a one-star change in fund rating, funds experience an average of 6% additional flows, most of which take place within 24 months.

Second, we estimate the response of stock returns to the stock-level FIT, as defined in Equation (2):

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

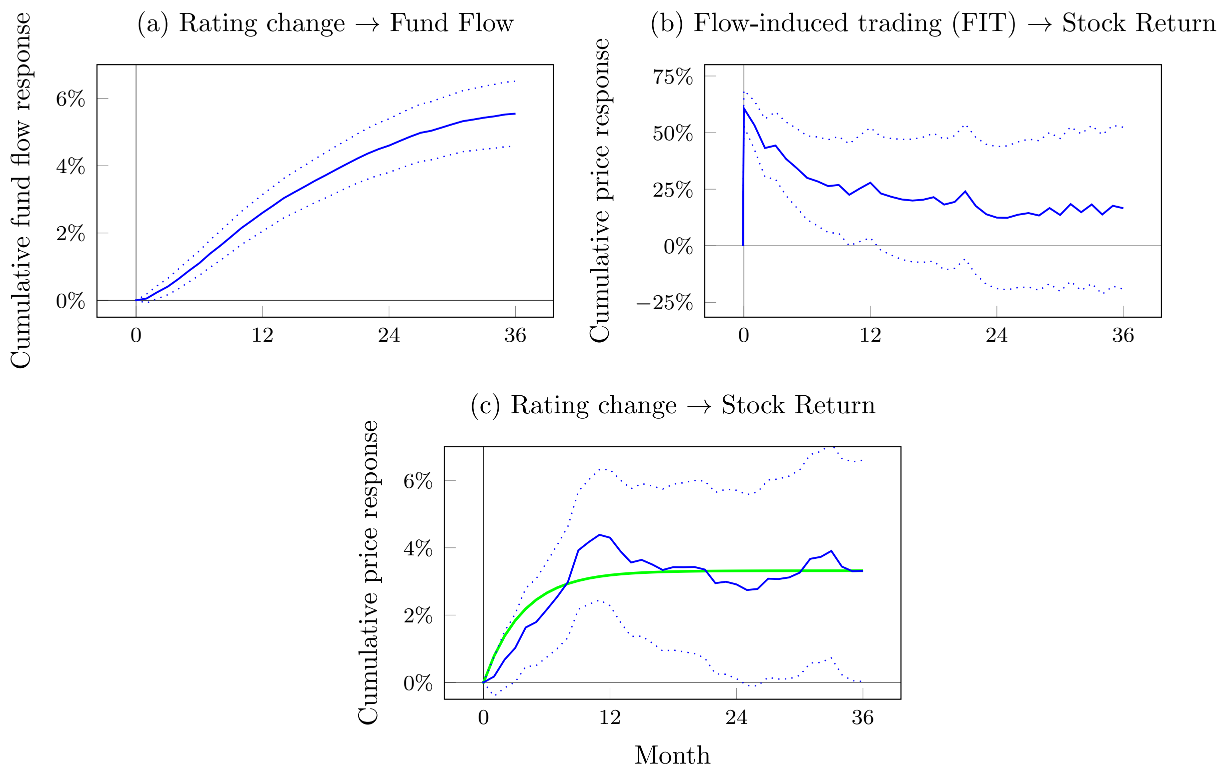
We plot the cumulative response in Panel (b) of Figure 6. Each 1% increase in mutual fund ownership due to FIT leads to immediate price pressures of approximately 0.6% in the contemporaneous month, which gradually reverts in the subsequent one to two years. This finding is consistent with the finding related to FIT in Lou (2012).

Combining these two effects, we can expect that stock returns also respond to rating

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

Figure 6. Cumulative Responses: Rating Change → Flows, and Flows → Returns

Panel (a) shows the cumulative response of fund flows to changes in fund ratings. Panel (b) shows the cumulative response of stock returns to flow-induced trading (FIT), defined as the nondiscretionary trading induced by mutual fund managers proportionally adjusting existing portfolio holdings in response to fund flows. Panel (c) shows the cumulative response of stock returns to changes in fund ratings, as well as the fitted exponential response (green line). In all panels, the dashed blue lines show two standard errors bands.



changes, particularly more recent rating changes.[12] Thus, we directly estimate the response of stock returns to stock-level rating changes:

$$\text{Ret}_{i,t} = a + d_1 \cdot \Delta\text{Rating}^{\text{stock}}_{i,t-1} + ... + d_{36} \cdot \Delta\text{Rating}^{\text{stock}}_{i,t-36} + u_{i,t}, \tag{9}$$

where $\Delta\text{Rating}^{\text{stock}}_{i,t}$, defined in Equation (4), is the holding-weighted average rating change experienced by all funds that hold stock i. The cumulative response is plotted in Panel (c) of Figure 6. In response to a one-star rating change, stock returns respond by around 4%, and the effect happens within 12 months.

For subsequent exercises, it is convenient to summarize the effect on the lagged 12 monthly

[12]The impact of more distant rating changes, such as those 24 months ago, should be weaker. While those rating changes may continue to generate flows (Figure 6, Panel (a)), the price pressures generated by their earlier impact are already reverting, so the two effects will partially cancel each other out.

rating changes using the following weighted average:[13]

$$\text{ExpSum}(\Delta\text{Rating})_{i,t-1} = \sum_{k=1}^{12} w_k \cdot \Delta\text{Rating}_{i,t-k}, \tag{10}$$

where $\sum_{k=1}^{12} w_k = 12$ and the weights decay with factor $\delta = 0.764$, which is estimated from a least-squares fit to the cumulative response (Panel (c) of Figure 6).[14] This implies a half-life of $-ln(2)/ln(\delta) \approx 2.58$ months. Our results are not sensitive to the choice of the parameter δ.

In this analyses that follow, we focus on the following price impact specification:

$$\text{Ret}_{i,t} = \text{Ret}_{i,t}^{\text{counterfactual}} + \underbrace{\lambda \cdot \text{ExpSum}(\Delta\text{Rating})_{i,t-1}}_{\text{Rating-induced price pressure}}, \tag{11}$$

where $\text{Ret}_{i,t}^{\text{counterfactual}}$ is what returns would have been in the absence of rating-induced price pressures. The explanatory power of this mechanism critically depends on the price impact coefficient λ, which we estimate in multiple ways in Section 4 and Appendix D.

3.3 Morningstar Ratings Methodology Pre- and Post-June 2002

After introducing its mutual fund rating system in 1985, Morningstar quickly became the industry leader in helping investors choose mutual funds. To assign ratings, Morningstar first summarizes the past return performance of funds and conducts minor adjustments for total return volatility and expenses. Depending on the availability of data, the lookback horizon for past performance can be three, five, or 10 years, but more weight is applied to recent history. For funds with over 10 years of history, the weights of the three horizons are set at 20%, 30%, and 50%, respectively.[15] Then, Morningstar ranks funds by their performance and assigns 1 to 5 star ratings with fixed proportions (10%, 22.5%, 35%, 22.5%, and 10%). The Morningstar methodology is fully transparent, and we provide further details in Appendix B.1.

Morningstar's methodology changed abruptly in June 2002. Many funds follow certain specific investment styles (e.g., large-cap growth) by mandate. Because style performance is a significant part of fund performance, fund ratings became highly dependent on style

[13]We only intend to capture the short-term ($\leq$ 12 months) momentum effects of rating-induced price pressures. Figure E.5 in the Appendix verifies that, if one examines buy-and-hold returns over longer horizons, these short-term price effects revert over 2-3 years.

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

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

performance. Following the dotcom crash, many fund managers specializing in technology stocks complained that their fund ratings dropped sharply from 5 stars to 3 stars or lower just because the technology sector had crashed. Consequently, ratings were barely reflecting their own contributions and instead were only echoing sector-level returns that were outside of their control. As a result, the research team at Morningstar, which is spearheaded by the economist Dr. Paul Kaplan, redesigned the rating system.[16]

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

3.4 Effects of the Methodology Change on Fund Flows

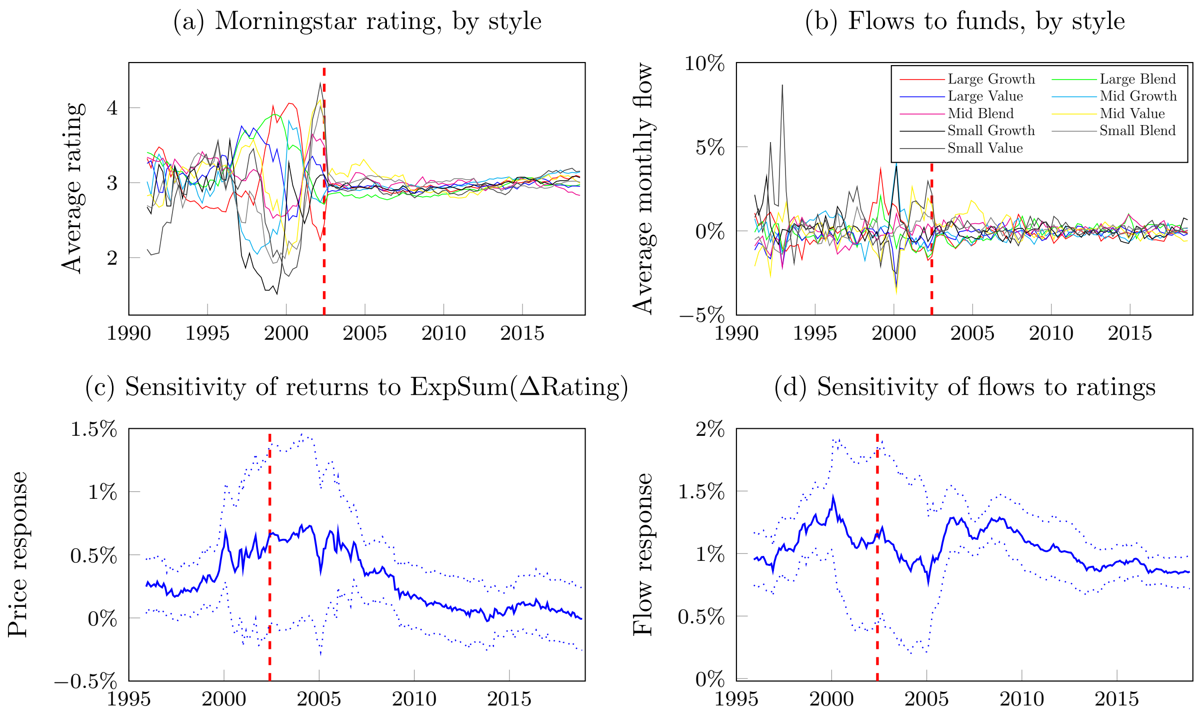
This seemingly innocent methodology change had far-reaching consequences for the mutual fund industry. Before the change, fund ratings differed dramatically across styles based on recent style performance, as shown in Panel (a) of Figure 7. For instance, in 2000, at the height of the dotcom boom, large-cap growth funds had an average rating of 4 stars, while small-cap value funds only had 1.9 stars. After the change, ratings became uncorrelated with past style performance, and the rating imbalance across styles became negligible. Consistent with flows chasing ratings, Panel (b) of Figure 7 shows that style-level fund flow dispersion also declined after the change. Because the flow response to ratings takes 12 months or more to manifest, the flow dispersion did not collapse immediately in June 2002 but became much more muted afterward. Therefore, this methodology change led to an abrupt change in the distribution of style-level price pressures in the stock market.

Important for our identification purposes, the abrupt change only happened in the ratings distribution, not in how ratings impact returns or fund flows. We estimate the time-varying λ coefficient at the stock level through five-year rolling value-weighted Fama-MacBeth regressions of stock returns $\text{Ret}_{i,t}$ on $\text{ExpSum}(\Delta\text{Rating})_{i,t-1}$. The results are plotted in Panel (c) of Figure 7. While λ does vary throughout the sample and reaches the highest level after the

[16] We learned this from a phone conversation with Morningstar management. Making ratings more balanced across styles was also one of the stated objectives for this methodology change. For instance, in a *New York Times* interview, Don Phillips, a managing director of Morningstar, said, "Two years ago, every growth fund looked wonderful... Now, none does." See Floyd Norris, Morningstar to Grade on a Curve, *New York Times*, April 23, 2002.

Figure 7. The June 2002 Morningstar Methodology Change

This figure plots the time variation of relevant quantities over the full sample. The red dashed lines mark the June 2002 methodology change event . Panels (a) and (b) plot the TNA-weighted average fund rating and monthly fund flows by Morningstar 3×3 styles by quarter. There is large style-level dispersion in ratings and flows across styles until June 2002. The flows in Panel (b) are demeaned by quarter to focus on the cross-sectional dispersion. Panel (c) plots the time variation in stock-level λ, estimated using five-year rolling Fama-MacBeth regressions of stock returns on $\text{ExpSum}(\Delta\text{Rating})_{i,t-1}$, the exponentially weighted sum of past-12-month ratings changes. Panel (d) plots the time variation of fund-level flow responses to lagged fund ratings. Panels (c) and (d) start in 1995 because five years of data are needed for estimation. The blue dashed lines show two standard deviations.



dotcom period, there is no abrupt change around June 2002. Similarly, Panel (d) estimates the five-year TNA-weighted Fama-MacBeth response of fund flows to lagged fund ratings and shows no abrupt change around June 2002.

The post-June 2002 methodology led to a sudden reduction in positive feedback fund flows at the style level. Before June 2002, funds in styles that performed well recently received significantly more flows. To examine the effects of this change in flows, in each month, we sort the 3×3 size-value styles into the top to bottom performers by their recent 12-month-average fund returns. We plot their subsequent TNA-weighted ratings and flows in Figure 1, in the Introduction. Panel (c) shows that the average rating spread between funds in the best- and worst-performing styles is about 0.8 stars before 2002, and almost zero after June 2002.[17] Similarly, Panel (d) shows that funds in the top style receive 2.3%

[17]The graphs are demeaned to focus on cross-sectional patterns across styles.

higher flows per month than the bottom style, and that difference dropped to 0.5% after June 2002. This result confirms the assertion that post-June 2002, mutual fund investors mostly stopped chasing style performance.

3.5 Effects of the Methodology Change on Momentum

Because the 2002 methodology change drastically reduced ratings-induced positive feedback trading, we expect it to have the largest negative effect on momentum profitability. We graphically investigate the effect on the momentum factor in this section. Apart from its relation with positive feedback trading, in asset pricing, understanding the origins of momentum profits is also important in its own right. Momentum is arguably one of the most puzzling factors because of its high profits (Jegadeesh and Titman, 1993) and because it has proven difficult to rationalize using risk-based explanations.[18]

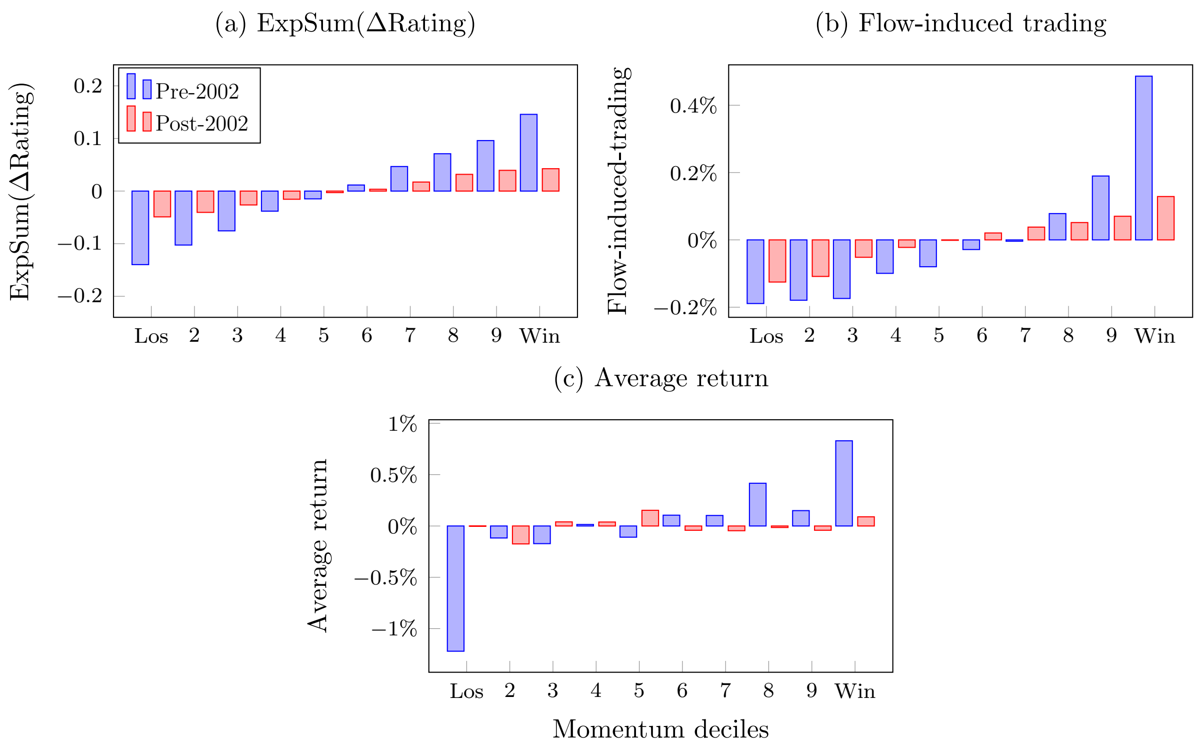
Figure 8 compares ExpSum(ΔRating), FIT, and returns of the 10 momentum decile portfolios before and after June 2002 over the entire sample. We follow Jegadeesh and Titman (1993) to define momentum by sorting stocks using their lagged $(t-1, t-12)$ month returns. To avoid the impact of microcaps, we follow Hou et al. (2019) in using NYSE decile breakpoints and value-weight each decile portfolio.

Panel (a) of Figure 8 plots ExpSum(ΔRating), the exponentially weighted sum of past-12-month rating changes, aggregated at the level of momentum portfolios. As expected, before the change, the winner portfolio experiences significant upward rating changes, and the loser portfolio experiences significant downward changes, but that pattern becomes muted after June 2002. There is a similar effect on FIT, as shown in Panel (b). Before June 2002, the winner portfolio experiences 0.72% higher monthly flows than the loser portfolio; that difference declined to 0.18% after June 2002. Finally, Panel (c) shows similar, albeit nosier, patterns in returns. This is related to the finding that the momentum factor, defined as long in the winner decile and short in the loser decile, experienced a dramatic decline in profitability after June 2002 (Panel (a) of Figure 1). The momentum factor, defined in this fashion, enjoyed a monthly return of 1.9% before June 2002 and only a negligible 0.1% after June 2002, with a clear kink around June 2002.

[18]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 et al., 2013). In an interview, Eugene Fama stated that he views momentum as “the biggest embarrassment for efficient markets.” See “Fama on Momentum,” AQR 2016, accessible at `https://www.aqr.com/Insights/Perspectives/Fama-on-Momentum`.

Figure 8. Momentum Factor before versus after the 2002 Methodology Change

We plot the ExpSum(ΔRating)$_{t-1}$ (exponentially-weighted sum of past-12-month rating changes), flow-induced trading, and returns of the 10 momentum decile portfolios before (from 1991) versus after June 2002 (until 2018). The deciles are formed using NYSE break points, and portfolios are value-weighted. All variables are demeaned to emphasize cross-sectional differences.



4 Event Study Using the 2002 Shock

As explained in Section 3.2, the explanatory power of ratings on factor profits crucially depends on the price impact parameter λ at the factor level. In this section, we use a short window of 12 months around June 2002 to estimate the impact of Morningstar ratings on factor profitability. There are two benefits to using a short window. First, the rating changes over this period are primarily caused by the methodology change. Second, by using a short window, we reduce the concern that factor returns were impacted by other events. For instance, the fact that all U.S. exchanges moved to quote prices in cents (rather than in 1/8s of a dollar) in April 2001 (decimalization event) may also have reduced factor profitability (Chordia et al., 2014), but it is not included in our sample.

4.1 A Visualization of Factors around June 2002

We first visualize rating, flow, and return variation in 2002. We sort factors into quintiles by their average lagged ExpSum(ΔRating) over the six months before the event. Thus, by construction, quintile 5 factors experience upward rating changes in the pre-event period, and quintile 1 factors experience downward rating changes.

We plot the evolution of factors in Figure 9. Panel (a), which plots average ratings, shows a sharp methodology-induced drop exactly at the event. Factors in quintile 5 suffer a drop of 0.43 rating stars, while factors in quintile 1 experiences a small increase of 0.19. Consistent with flows chasing ratings, Panel (b) shows that top-quintile factors experienced cumulative inflows of around 7.9% during the six months before the event, while quintile 1 experienced mild outflows. The flow differences became muted after June 2002. A casual look at the data suggests that this could be related to factor returns. In Panel (c), we plot the cumulative returns of factors in those five quintiles. The pre-event factor returns line up with pre-event ExpSum(ΔRating) but suffer reversals after the event. This phenomenon is clearest in the top quintile factors: They experienced a staggering 30.4% return in the pre-event period but that reversed subsequently. Moreover, this reversal happened in both the long and the short legs, as shown in Panel (d).

Figure 10 plots the change in monthly factor FIT and returns around the event. Quintile 5—the factors that benefited from ratings pre-event but suffered post-event—experienced a decline of -0.91% in monthly FIT and a sharp decline of -6.61% in monthly returns. At the same time, quintile 1 experienced an increase of 0.14% in monthly FIT and a slight increase of 0.75% in monthly returns. To alleviate the concern that the return and FIT changes could result from mean reversion due to other reasons, we also show that similar effects do not happen in other years. The red bars show the same exercise in other years and plot the two standard error bands. Clearly, the shock is unique to 2002.

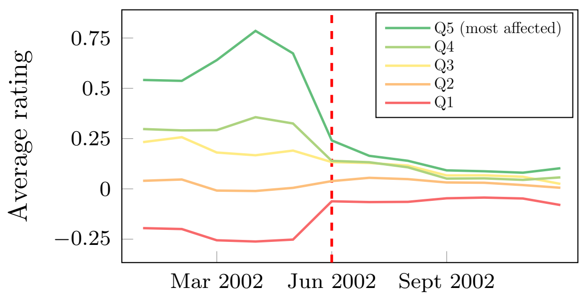
Why did some factors have a much higher style rating exposure than others? We find that this arises from how factors load differently onto stock styles in the period before June 2002. We first examine which styles performed well before the event. Panel (a) of Figure 11 shows the average rating, flow, and return of funds by the 3×3 styles in the six months before the event. In that period, small-value funds performed well, but large-growth funds performed poorly, with a difference of 24% in returns, a 1.9 star difference in ratings, and a 25% difference in fund flows. These differences almost entirely reflected the performance of the underlying stock styles.

How each factor is affected depends on its style exposure. We measure the size and value exposure of factors using their SMB and HML loadings, estimated using time-series

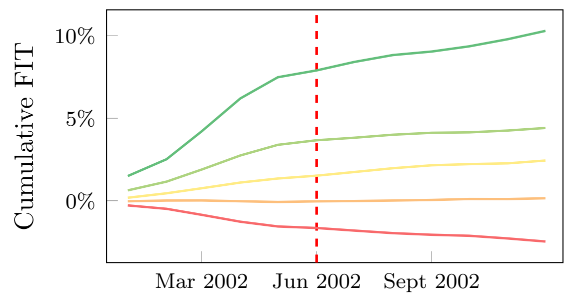
Figure 9. Stock Factors around the June 2002 Event

Factors are sorted into quintiles by their average ExpSum(ΔRating), defined in equation (10), in the six months before the methodology change. Thus, by construction, quintile 5 factors experience above-average rating changes, and quintile 1 factors experience low rating changes before the event. Panel (a) plots the average rating of those factors from six months before to six months after the event. Panel (b) plots the cumulative monthly flow-induced trading (FIT), and Panel (c) plots the cumulative daily returns. Panel (d) plots the average monthly return by decile portfolios for the top quintile factors. Returns before the event are shown by the blue bars and after the event are shown by the red bars. To focus on cross-sectional dispersion, both of the data series are demeaned.

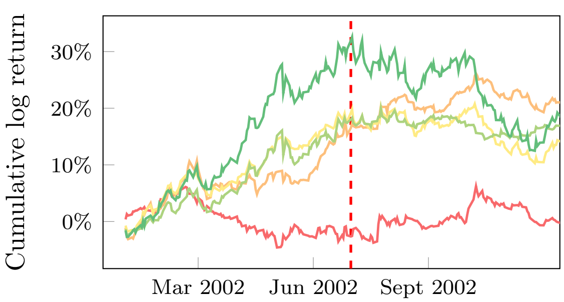
(a) Factor ratings



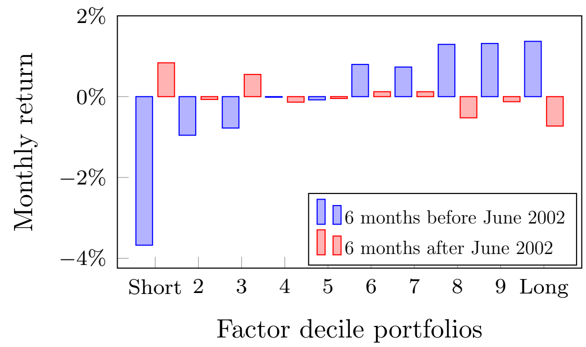
(b) Factor FIT



(c) Factor returns



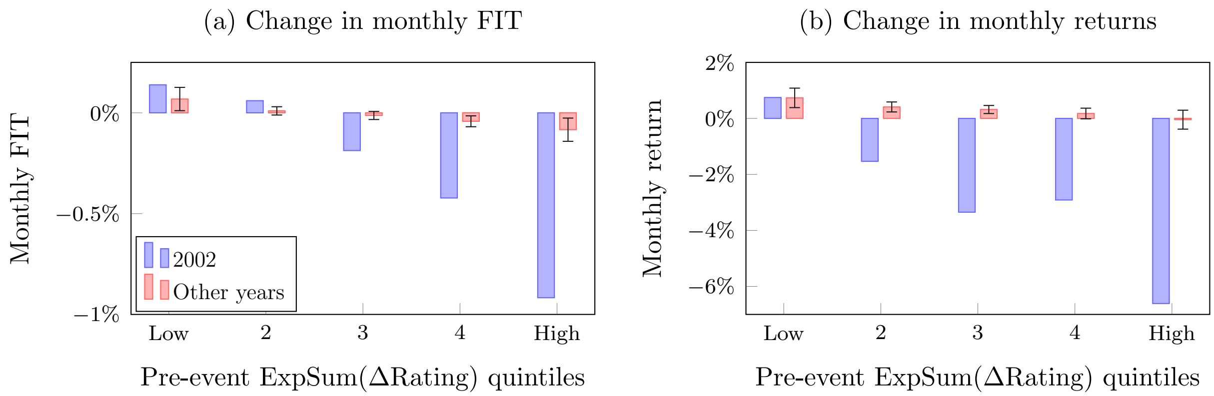
(d) Average quintile 5 factor returns by deciles



regressions with daily returns in the six months before the event. We plot the factors' style exposures in Panel (b). Clearly, factors with the highest pre-event $\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$ values (quintile 5) are mostly loading onto the outperforming small and value styles, while those in the bottom quintile mostly load onto the underperforming large and growth styles. We show additional details about factor loadings and ratings in Appendix Table B.1.

Figure 10. Factor FIT and Return Change around June 2002

Panel (a) plots the difference in monthly flow-induced trading (FIT). Panel (b) plots the difference between the average monthly factor returns six months after versus six months before the event. Factors are sorted into quintiles by their pre-event average ExpSum(ΔRating). The blue bars plot the results for 2002, and the red bars plot the average over the other years, serving as a placebo test. The whiskers are two standard error bars.



4.2 Estimating the Price Impact of Ratings Using the 2002 Shock

We now estimate the price impact coefficient λ using the 2002 shock. We estimate a panel regression using six months before to six months after the event:

$$\text{Ret}_{f,t} = \lambda \cdot \text{ExpSum}(\Delta\text{Rating})_{f,t-1} + \sum_f \mu_f \cdot \mathbf{I}_{\text{factor } f} + \epsilon_{f,t}, \tag{12}$$

where μ_f represents factor fixed effects. To account for the cross-sectional factor return correlation, we adjust the standard errors using a feasible generalized least squares (FGLS) approach. Specifically, we use the full sample of factor returns to estimate the covariance matrix of factor returns and incorporate it into the estimation.[19] In Appendix Figure C.3,

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

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

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

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

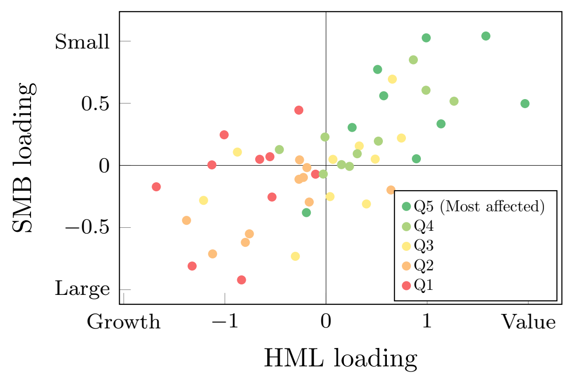
Figure 11. Factor Style Exposure before the 2002 Event

Panel (a) shows the rating, flow, and return of funds by 3 × 3 fund styles during the six months before the rating methodology change event. Larger values are shaded green, and smaller values are shaded red. Panel (b) shows the HML and SMB loadings of factors. The factors are sorted into quintiles by their average $\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$ value during the six months before the event such that quintile 5 has high ratings before the event and quintile 1 has low ratings.

(a) Mutual fund ratings, flows, and returns by style

	Rating				**Fund flow**				**Fund return**		
	Growth	Blend	Value		Growth	Blend	Value		Growth	Blend	Value
Small	3.39	4.06	4.69	Small	3%	13%	22%	Small	−13%	−3%	5%
Mid	3.22	4.09	4.41	Mid	1%	8%	17%	Mid	−17%	−10%	−4%
Large	2.79	3.30	3.79	Large	−3%	−1%	2%	Large	−19%	−13%	−8%

(b) Factor style exposures



we show that some factor groups have positive correlations between themselves, such as the momentum-related factors. However, there are also negative correlations; for example, the value factor is negatively correlated with the momentum-related factors, consistent with Asness et al. (2013).

The results are shown in Table 2. For each star rating change, the factor-level price impact is 1.92% with a t-statistic of 3.18. The result is both statistically and economically significant.

In addition to estimating λ via panel regression, we also estimate a two-stage least squares (2SLS) regression in which we instrument $\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$ using the predicted at-event rating change using seven-months-ago data. Specifically, in December 2001, we compute

$$\text{ExpectedChange}_{f,\text{Dec 2001}} = \widehat{\text{Rating}}_{f,\text{Dec 2001}}^{\text{post-June 2002 methodology}} - \widehat{\text{Rating}}_{f,\text{Dec 2001}}^{\text{pre-2002 methodology}}, \quad (13)$$

Table 2. Explaining Factor Returns around the June 2002 Event

We regress monthly returns on $\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$ during the six months before versus six months after the methodology change. Column (1) uses a panel regression, and Column (2) uses an instrumented version of the independent variable. The average first-stage F-statistic is 209.4. The standard errors in parentheses are adjusted for the cross-sectional correlation between factor returns using a feasible generalized least squares approach.

Dependent variable:	Monthly factor return $\text{Ret}_{f,t}(\%)$	
Regression:	Panel regression	2SLS
	(1)	(2)
$\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$	1.920^{***}	1.937^{***}
	(0.605)	(0.476)
Factor FE	Yes	Yes
Month FE	Yes	Yes
Observations	588	588
Adj R^2	7.82%	7.71%

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

where $\widehat{\text{Rating}}^{\text{post-June 2002 methodology}}$ and $\widehat{\text{Rating}}^{\text{pre-2002 methodology}}$ are our own estimates of factor-level Morningstar ratings under different methodologies using fund returns data. Thus, the difference is our prediction of how much each factor's rating should change at the event month. Because Morningstar is completely transparent about its rating methodology, we can replicate Morningstar ratings with high degrees of accuracy. In the 2SLS, we first regress $\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$ in each month on ExpectedChange_f, and then use the predicted values as the independent variable in Regression (12). More details about the first stage are given in Appendix Section C.

The 2SLS estimate of λ is similar to the panel regression estimate. For each star rating change in the previous 12 months, the monthly factor return changes by 1.94% with a t-statistic of 4.07. In Appendix C, we show a battery of robustness checks. Our results are not sensitive to including or excluding factor and/or time fixed effects or using alternative estimates of return covariance.

While using the 2002 shock achieves clear identification, we also show that ratings significantly predict future factor returns based on the entire sample of the data, even including various controls that have been indicated in the literature to have power in predicting factor returns. To see this, we estimate the following regression throughout our sample period:

$$\text{Ret}_{f,t} = \lambda \cdot \text{ExpSum}(\Delta\text{Rating})_{f,t-1} + X_{f,t-1} + \epsilon_{f,t}. \tag{14}$$

Here, $X_{f,t-1}$ is a vector of control variables. Specifically, to control for possible factor-level momentum or reversals, we include each factor's past $t-1$, $t-12$ to $t-2$, and $t-60$ to $t-13$ returns (Arnott, Clements, Kalesnik, and Linnainmaa, 2019a; Gupta and Kelly, 2019). We also control for the lagged book/market spread between the long and short legs of each factor. Motivated by the finding that factors become less profitable out-of-sample and post-publication (McLean and Pontiff, 2016), we include factor × period fixed effects where *period* is defined as in-sample, out-of-sample, and post-publication periods. In Appendix D, we show that our results are not sensitive to the set of controls we choose to include. Similar to regression (13), we employ an FGLS approach to estimate the regression using the estimated covariance matrix of factor returns.

We report the results in Table 3. In Column (1), the estimated full-sample λ indicates that if ExpSum(ΔRating) is higher by 1 star, the subsequent monthly factor return is higher by 0.57%, with t-statistics of approximately 3.9. We also separately estimate the regression using data before and after June 2002 and get similar results, albeit with slightly larger standard errors. In Appendix D, we also use rolling windows to estimate how factor-level λ varies over time . We find that the relation is positive and statistically significant throughout most of the sample.

4.3 Alternative Hypotheses

Given that the June 2002 methodology change is central to our analysis, we want to make sure it does not coincide with other events that may affect anomaly returns. In this section, we examine proxies of arbitrage activity in the factor portfolios and show that there is no abrupt change around the event.

We consider two arbitrage measures in the literature. First, we construct the net arbitrage activity (NAT) measure in Chen, Da, and Huang (2019). For each stock, the authors measure the long position of arbitrageurs by aggregating 13F holdings of hedge funds and the short position by using aggregate short interest from Compustat.[20] The authors then combine the long and short positions into a net position and subtract the past four-quarter average values to effectively arrive at a measure of the position change that they call NAT. We aggregate NAT at the factor level and verify that it appears to positively predict factor returns in our sample although without statistical significance.

We also follow Lou and Polk (2018) and construct "CoFactor" measures of arbitrage activity in all factors. The authors propose measuring arbitrage activity in the momentum strategy by measuring excess return correlation within the extreme decile portfolios.

[20]We use the list of 13F institutions identified as hedge funds in Aragon, Li, and Lindsey (2018). We thank the authors for kindly sharing the data.

Table 3. Explaining Factor Returns over the Full Sample

We regress monthly factor returns on ExpSum(ΔRating)$_{f,t-1}$, the exponentially-weighted sum of past-12-month rating changes. Column (1) uses the full sample of 1991 to 2018, while Columns (2) and (3) use parts of the sample. Control variables include lagged factor returns of various horizons and the lagged book/market spread between the long and short factor portfolios. Factor $\times$ period fixed effects, are indicator to whether the factor-month is in the factor's in-sample, out-of-sample, and post-publication periods. The standard errors in parentheses are adjusted for the cross-sectional return correlations using a feasible generalized least squares approach.

Dependent variable:	Monthly factor return $\text{Ret}_{f,t}$ (%)		
Sample period:	Full sample	Before June 2002	After June 2002
	(1)	(2)	(3)
ExpSum(ΔRating)$_{f,t-1}$	0.568***	0.465***	0.639*
	(0.144)	(0.164)	(0.339)
$\text{Ret}_{f,t-1}$	−0.002	−0.007	−0.004
	(0.008)	(0.011)	(0.011)
$\text{Ret}_{f,t-12\to t-2}$	−0.004*	−0.006*	−0.007**
	(0.002)	(0.003)	(0.003)
$\text{Ret}_{f,t-60\to t-13}$	−0.007***	−0.013***	−0.007***
	(0.001)	(0.002)	(0.002)
Book/Market Spread$_{f,t-1}$	0.112	0.628***	−0.073
	(0.121)	(0.217)	(0.149)
Factor $\times$ Period FE	Yes	Yes	Yes
Observations	16,415	6,762	9,653
Adj R^2	0.73%	1.41%	0.88%

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

Specifically, in any given month, they use the previous 52 weeks of data to compute a "co-momentum" measure as follows:

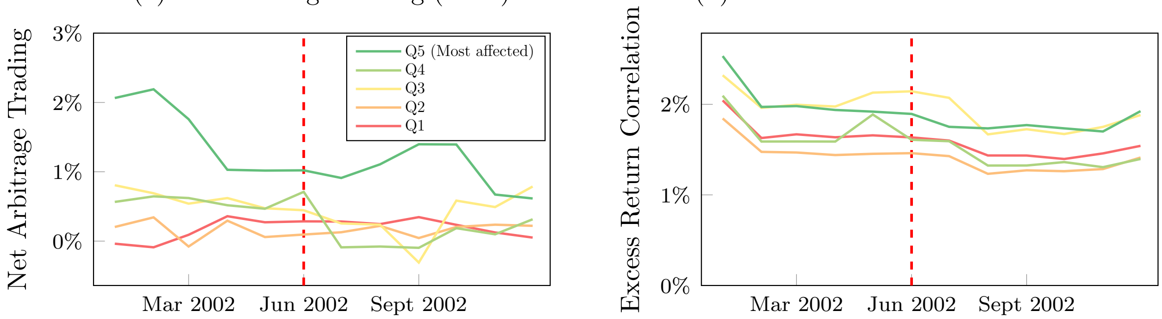
$$\text{CoMomentum}_t = \frac{1}{2} \cdot \left[\frac{1}{N^L(N^L-1)} \sum_i \sum_{j \neq i} \text{PartialCorr}(\text{Ret}_i, \text{Ret}_j) + \frac{1}{N^S(N^S-1)} \sum_i \sum_{j \neq i} \text{PartialCorr}(\text{Ret}_i, \text{Ret}_j) \right],$$

where N^L and N^S are the number of stocks in the long and short decile portfolios, respectively. To compute the partial return correlations, the authors first subtract Fama-French 30 industry returns from weekly stock returns, and then regress the residuals on the Fama-French three factors to obtain alphas. Finally, they compute equal-weighted averages of the pairwise correlations of the alphas within the portfolios and take an average. Whereas Lou and Polk only define this measure for momentum, we also calculate it for all other factors. Consistent with the authors, we find that this measure negatively predicts returns of fac-

Figure 12. Arbitrage Activity in Stock Factors around 2002

As in Figure 9, factors are sorted into quintiles by their average ExpSum(ΔRating) in the six months before the methodology change, and we plot the evolution of two factor-level arbitrage activity proxies. Panel (a) plots the net arbitrage trading measure in Chen et al. (2019). Panel (b) plots excess return correlation in extreme factor deciles, a measure of arbitrage activity developed in Lou and Polk (2018). The vertical red dashed lines mark the methodology change event.

(a) Net arbitrage trading (NAT) (b) Excess correlation in extreme deciles



tors in the momentum category—which the authors interpret as the effect of arbitrageur crowding—but not for other factors.

We plot the evolution of these measures in the 12 months around June 2002 in Figure 12. As in Figure 9, we sort factors into quintiles by their average ExpSum(ΔRating) before the event. Panel (a) plots the NAT measure and Panel (b) plots the CoFactor measure. There is no evidence that arbitrage activity in those factors changed around June 2002.

5 Long-Term Effect on Factor Profitability

While using the narrow window around June 2002 achieves better identification, we now ask the more economically important question: Did the Morningstar methodology change explain a sizeable fraction of the long-run factor profitability decline after June 2002?

Recalling our main specification of rating price impact,

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

we first show that there is a material change in factor-level ExpSum(ΔRating) after June 2002. We then estimate how much the Morningstar methodology change can explain the post-June 2002 factor profitability decline.

5.1 Which Factors Suffered Ratings Drop Post-June 2002?

In Panel (a) of Figure 13, we plot each factor's average post-June 2002 against their pre-2002 ExpSum(ΔRating) over the full sample. We mark factors from different Hou et al. (2019) categories using different colors. Clearly, before June 2002, Morningstar serves as the tailwind for many factors, especially those in the momentum and trading frictions categories. After June 2002, the ExpSum(ΔRating) across factors collapsed to close to zero.

In Panels (b) and (c), we plot pre-2002 and post-June 2002 average factor returns against the pre-June 2002 ExpSum(ΔRating). As expected, factors that benefit from pre-2002 ratings experienced high returns before June 2002 but not afterward. For instance, the momentum factor experienced close to a 2% monthly return before June 2002 but became a negligible 0.1% after June 2002. Other momentum-category factors, such as the 52-week high factor, suffered similar declines in profitability.

5.2 Role of Morningstar in Explaining the Post-June 2002 Factor Return Decline

We now quantify, over the full sample of 1991–2018, how much of the post-June 2002 factor profitability decline can be explained by Morningstar. The fact that factors became less profitable after June 2002 has been documented by a number of papers and so far defies explanation (Khandani and Lo, 2011; Daniel and Moskowitz, 2016; Green et al., 2017; Arnott et al., 2019b). We now do a simple back-of-envelope quantification exercise. Note that this is a crude exercise and requires strong functional form assumptions.

In our framework, the decline of factor f returns explained by Morningstar equals

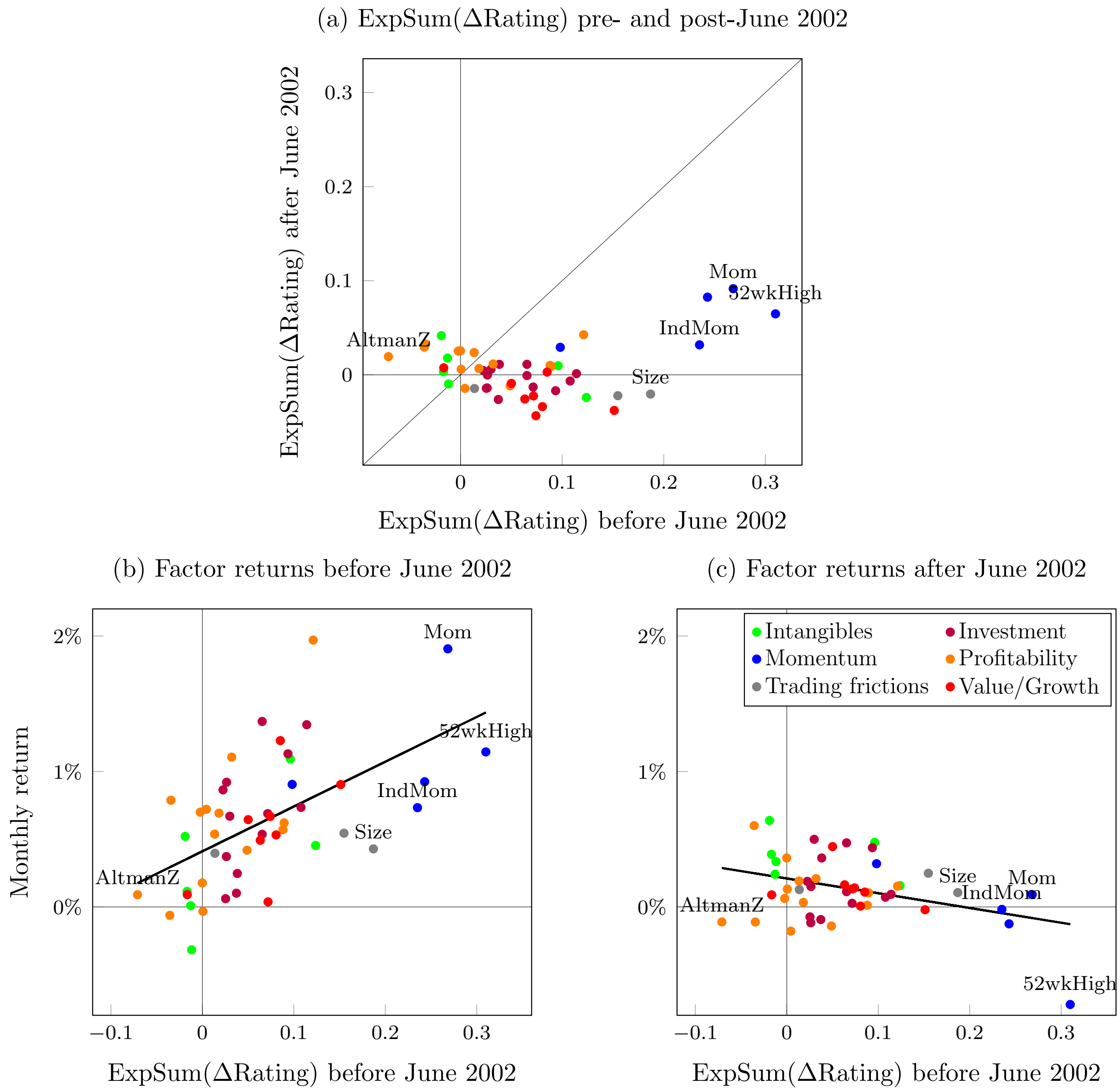
$$\lambda \times \left(\overline{\text{ExpSum}(\Delta\text{Rating})}_{f,\text{pre 2002}} - \overline{\text{ExpSum}(\Delta\text{Rating})}_{f,\text{post 2002}} \right).$$

For our main specification, we use the λ estimated from the 2SLS short window around the 2002 shock (Section 4.2). We believe this is the best estimate because this exercise exploits exogenous methodology-induced ratings variation.

In Figure 14, we plot the post-June 2002 change in monthly factor profitability against that explained by Morningstar. To visualize the heterogeneity, we sort factors by their post-June 2002 ExpSum(ΔRating) decline into quintiles such that the top quintile contains factors that suffered the sharpest decline in Morningstar tailwind after June 2002. Details of how each individual factor is affected are listed in Appendix Table D.4. We plot the actual return change in blue bars and the part explained by Morningstar in red, with whiskers showing two standard error bands based on $\hat{\lambda}$ standard errors. To focus on cross-sectional

Figure 13. Factors before versus after June 2002

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

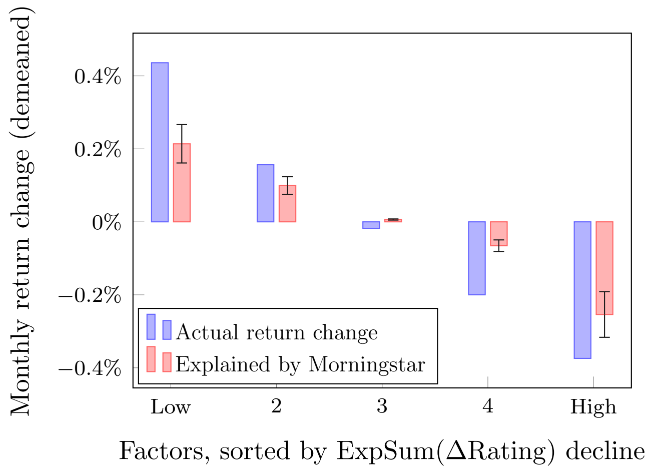


heterogeneity, we plot the demeaned (relative to the cross-sectional average factor returns) version of both data series.

The results show that ratings-induced price pressures can explain approximately half of the cross-sectional variation of the factor profitability decline. Relative to the bottom quintile, factors in the top quintile experienced a 0.81% additional monthly return decline.

Figure 14. Role of Morningstar in Explaining the Post-June 2002 Factor Return Decline

We examine the post-June 2002 factor return decline over the full sample of 1991 to 2018. We sort factors into quintiles by their average decline of $\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$ (the exponentially-weighted sum of lagged 12-month rating changes) after June 2002. Thus, the top quintile factors experience the largest ratings-induced profitability decline. The actual returns changes are plotted in blue. The part explained by the Morningstar methodology change, estimated using λ times the before-versus-after change in $\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$, is plotted in red. λ is based on the factor-level 2SLS estimate in Section 4.2, and the whiskers are two standard error bands. To focus on cross-sectional variation across factors, the data series are demeaned.



Ratings explain an additional 0.47% decline, which is $\approx$ 58% of the overall difference.

While the 2002 shock is the best identified variation in our exercise, one may still be worried that λ varies over time. In Appendix D, we repeat this exercise using λ estimated from rolling-window factor return–predicting regressions (equation (14)). Under that specification, we find that the explanatory power over the difference between the top and bottom quintile return is 30%.

6 Conclusion

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

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

In June 2002, Morningstar changed its methodology to make ratings unrelated to past style performance. As a consequence, the positive feedback flow pressures halted, and factor premia have weakened dramatically ever since. This methodology change also provided clean causal identification, which is rarely achieved in asset pricing research.

It is possible that the role of correlated demand in determining asset pricing is even bigger than what is documented here. We estimate that between 30% and 58% of the factor premium during the 1991–2002 period can be explained solely by the correlated demand driven by Morningstar ratings. Correlated demand, however, can arise from sources other than Morningstar ratings, such as institutional demand for certain styles (Froot and Teo, 2008; Koijen and Yogo, 2019) or the performance-chasing behavior of index-linked products (Broman, 2016). Hence, it is possible, and even likely, that correlated demand has a central role in explaining the cross-section of asset returns.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

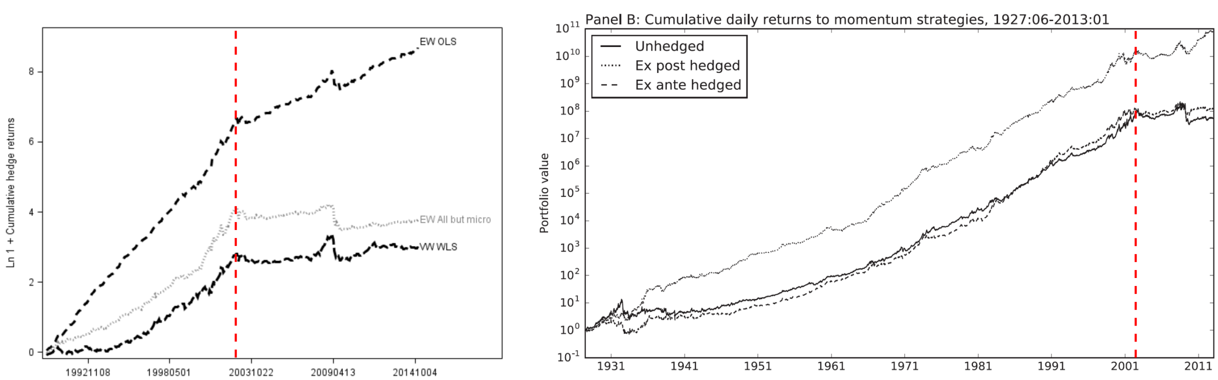
Appendix A Previous Evidence of Factor Returns and Kink around June 2002

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

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

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

(a) Green, Hand, and Zhang (2017, Fig 3) (b) Daniel and Moskowitz (2016, Fig 4b)



Appendix B Data and Measures

B.1 Morningstar Methodology

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

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

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

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

B.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:

$$R_i^T = \prod_{t=1}^{T}(1 + r_{i,t}) - 1, \qquad T \in \{36, 60, 120\}, \tag{15}$$

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

$$\text{LoadRet}_i^T = R_i^T L_i - R_f^T, \tag{16}$$

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

$$\text{MnLoadRet}_i^T = \frac{\text{LoadRet}_i^T}{\max(R_f, \text{AvgLoadRet}^T)}, \tag{17}$$

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

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

$$\text{Performance}_{i,t} = \text{MnLoadRet}_{i,t}^T - \text{MnRisk}_{i,t}^T. \tag{18}$$

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

$$\text{Risk}_i^T = \frac{\sum_{t=1}^{T} - \min(r_{i,t} - r_t^f, 0)}{T}, \tag{19}$$

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

$$\text{MnRisk}_t^T = \frac{\text{Risk}_i^T}{\text{AvgRisk}^T}. \tag{20}$$

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

$$\text{MRAR}_i^T(\gamma) = \left[\frac{1}{T}\sum_{t=1}^{T}(1 + r_{i,t} - r_t^f)^{-\gamma}\right]^{-\frac{12}{\gamma}} - 1, \tag{21}$$

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

The formula penalizes funds with higher return volatility. To see this, notice that when γ converges to 0, $\text{MRAR}^{\text{T}}(0)$ is equal to the annualized geometric mean of excess returns.[23] When γ is set to be greater than 0, holding the geometric mean return constant, the formula yields a lower MRAR value for funds whose monthly returns deviate more from their mean. Specifically, the risk adjustment can be expressed as $\text{MRAR}^{\text{T}}(0) - \text{MRAR}^{\text{T}}(2)$.

B.1.2 Step Two: Rank Funds and Assign Ratings

Given rankings of funds, Morningstar calculates three-year, five-year, and 10-year ratings for funds with the necessary amount of historical returns at those horizons, and the take a weighted average of them (rounded to the nearest integer) to form an overall rating—the rating most commonly reported and used. For funds with more than three years but less than five years of data, the overall rating is just the three-year rating. For funds with more than five years but less than 10 years of data, the overall rating assigns 60% and 40% weights on the five-year and three-year ratings. For those with 10 years of data, 50%, 30%, and 20%

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

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

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

weights are assigned on the 10-year, five-year, and three-year ratings, respectively.

The ratings are based on rankings of funds. Before June 2002, Morningstar ranks the past performance of all equity funds together and assign them ratings with fixed proportions: 10%, 22.5%, 35%, 22.5%, and 10%. After June 2002, Morningstar ranks funds within each style ("Morningstar category") and assigns ratings based on the within-style ranking. Styles include the standard 3×3 size-value categories in the Morningstar style box and also a number of specialized sector categories (e.g., financial, technology). Because much of fund performance is due to style-level stock return variation, before the change, there is significant variation of ratings across styles. That variation became negligible after June 2002 (Panel (b) in Figure 7).

Table B.1. Factors, Six Months before versus Six Months after the Event

The 49 asset pricing factors are ranked and sorted into quintiles using their average ExpSum(ΔRating) in the six months before the methodology change.

Quintile	Ranking	Factor	ExpSum(ΔRating)		Monthly return	
			Pre June 2002	Post June 2002	Pre June 2002	Post June 2002
1	1	Ohlson's O-score	-0.288	0.684	4.64%	-3.56%
	2	Altman's Z-score	-0.192	0.456	1.15%	-0.02%
	3	Sales growth	-0.18	0.216	0.38%	1.88%
	4	Cash-based profitability	-0.144	0.252	1.69%	-1.84%
	5	Operating leverage	-0.12	0.216	-3.17%	-0.66%
	6	Industry concentration	-0.12	0.12	4.10%	1.61%
	7	Earnings persistence	-0.108	0.3	-0.83%	-0.35%
	8	Profit margin	-0.096	0.168	0.88%	-1.24%
	9	Piotroski's F-score	-0.084	0.12	-1.73%	1.34%
	10	Gross profitability	-0.048	0.288	5.57%	-1.64%
2	11	Net working capital changes	-0.048	-0.096	3.50%	1.67%
	12	Accruals	-0.036	0.024	-0.48%	0.66%
	13	Return on equity	-0.024	0.456	0.45%	1.50%
	14	Sales-minus-inventory growth	0	0.084	1.50%	-0.05%
	15	QMJ profitability	0.012	0.372	6.63%	3.76%
	16	Net operating assets	0.012	-0.108	-1.24%	-0.77%
	17	Return on assets	0.024	0.54	7.12%	0.93%
	18	Growth in inventory	0.048	-0.156	2.15%	1.57%
	19	Distress risk	0.048	0.348	10.40%	0.99%
	20	Investment-to-assets	0.048	-0.108	-1.35%	-0.85%
3	21	Operating profitability	0.072	0.192	-0.36%	-0.50%
	22	Maximum daily return	0.084	0.168	2.76%	0.76%
	23	R&D expense	0.084	-0.432	-3.23%	0.51%
	24	Industry adjusted CAPX growth	0.096	-0.3	3.36%	0.17%
	25	Sustainable growth	0.132	-0.168	3.77%	-1.54%
	26	Firm age	0.132	-0.192	1.27%	0.94%
	27	Momentum (t-2, t-6)	0.144	-0.24	5.64%	0.82%
	28	Abnormal capital investment	0.144	-0.3	5.25%	0.90%
	29	Earnings-to-price	0.144	0.024	4.86%	1.28%
	30	Net payout yield	0.156	0	7.78%	0.90%
4	31	Intermediate momentum (t-7,t-12)	0.156	0.072	6.81%	-0.45%
	32	Industry momentum	0.156	-0.06	4.20%	-1.54%
	33	Five-year share issuance	0.168	0.012	4.17%	3.38%
	34	Total external financing	0.18	-0.048	3.86%	0.47%
	35	Change in asset turnover	0.18	0	2.50%	1.40%
	36	Asset growth	0.192	-0.192	3.22%	1.00%
	37	One-year share issuance	0.204	0.096	-0.77%	-0.59%
	38	Enterprise multiple	0.216	-0.588	3.36%	1.16%
	39	Investment growth	0.252	-0.24	6.15%	2.28%
	40	Advertising expense	0.3	-0.396	-0.25%	-0.05%
5	41	Investment-to-capital	0.336	-0.372	3.64%	1.68%
	42	Momentum (t-1, t-12)	0.36	-0.036	4.68%	2.65%
	43	Cash flow-to-price	0.384	-0.3	-1.30%	-1.91%
	44	Book-to-market	0.396	-0.648	4.25%	-0.44%
	45	52-week high	0.444	0.12	4.84%	2.22%
	46	Long-term reversals	0.576	-0.696	3.56%	1.91%
	47	Sales-to-price	0.612	-0.66	-4.03%	-0.24%
	48	Amihud illiquidity	0.864	-1.068	0.92%	0.00%
	49	Size	0.96	-1.176	2.68%	0.14%

Appendix C Explaining Factor Returns around June 2002

C.1 Regression Details

The instrument. Taking advantage of the fully transparent Morningstar rating methodology, we can construct an instrument using data seven months before the methodology change event. Because of differences in the fund data source between us and Morningstar, we cannot replicate their ratings exactly, but our estimates as sufficiently useful to form an instrument. Note that the instrument is calculated using data *outside* of the regression window of six months prior to six months after the event. We use the instrument to predict rating changes.

We start with fund returns and calculate our own estimates of Morningstar ratings for each fund under their two methodologies. Then, we aggregate those ratings at the factor level through fund holdings, as described in Section 2.3. The instrument is defined as

$$\text{ExpectedChange}_{f,\text{Dec 2001}} = \widehat{\text{Rating}}_{f,\text{Dec 2001}}^{\text{post-June 2002 methodology}} - \widehat{\text{Rating}}_{f,\text{Dec 2001}}^{\text{pre-2002 methodology}}, \quad (22)$$

where $\widehat{\text{Rating}}^{\text{post-June 2002 methodology}}$ and $\widehat{\text{Rating}}^{\text{pre-2002 methodology}}$ are our own factor-level rating estimates. Thus, the difference should be correlated with the post-event factor rating changes. Figure C.2, Panels (a) and (b), demonstrate this effect. Panel (a) shows that the expected change in ratings has a slight negative relation, close to zero, with the May 2002 actual ExpSum(ΔRating). In contrast, our instrument predicts actual June 2002 ExpSum(ΔRating) well.

We also present month-by-month predictions and plot the coefficients. For each month t used in the regression, we estimate one different first-stage regression,

$$\text{ExpSum}(\Delta\text{Rating})_{f,t} = a_t + b_t \cdot \text{ExpectedChange}_{f,\text{Dec 2001}} + \epsilon_{f,t}, \quad (23)$$

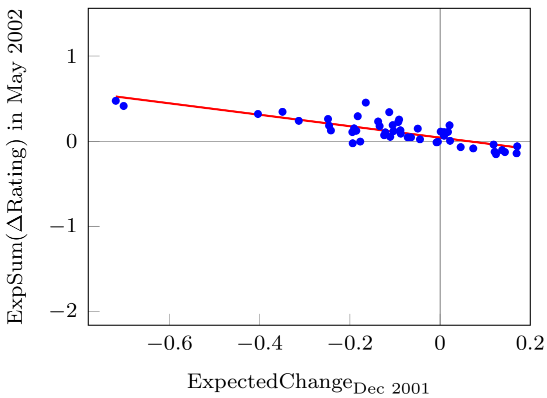
and show the first-stage coefficients in Figure C.2, Panel (c). $\text{ExpectedChange}_{f,\text{Dec 2001}}$ negatively predicts $\text{ExpSum}(\Delta\text{Rating})_{f,t}$ and positively predicts it after the event.

We then use the predicted values from Regression (23) as the independent variable in the 2SLS in the main paper.

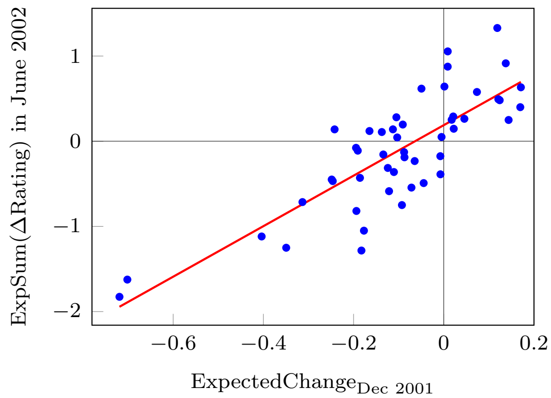
Factor return covariance. Figure C.3 shows the return correlation of factors, clustered by the Hou et al. (2019) categories, in a heat map. High values are shown in red and negative values are shown in blue. There appears to be high correlations within certain categories of

Figure C.2. First Stage of the Regression

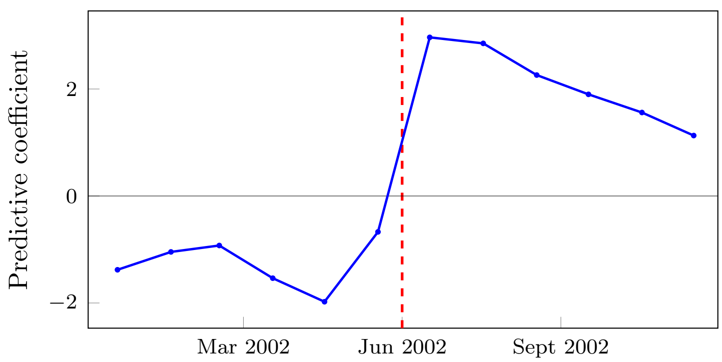
(a) Predicting May 2002 ExpSum(ΔRating)



(b) Predicting June 2002 ExpSum(ΔRating)



(c) Predictive coefficient



factors, such as the momentum category, the trading friction category, and the profitability categories. There are also negative correlations across certain categories.

C.2 Robustness Checks

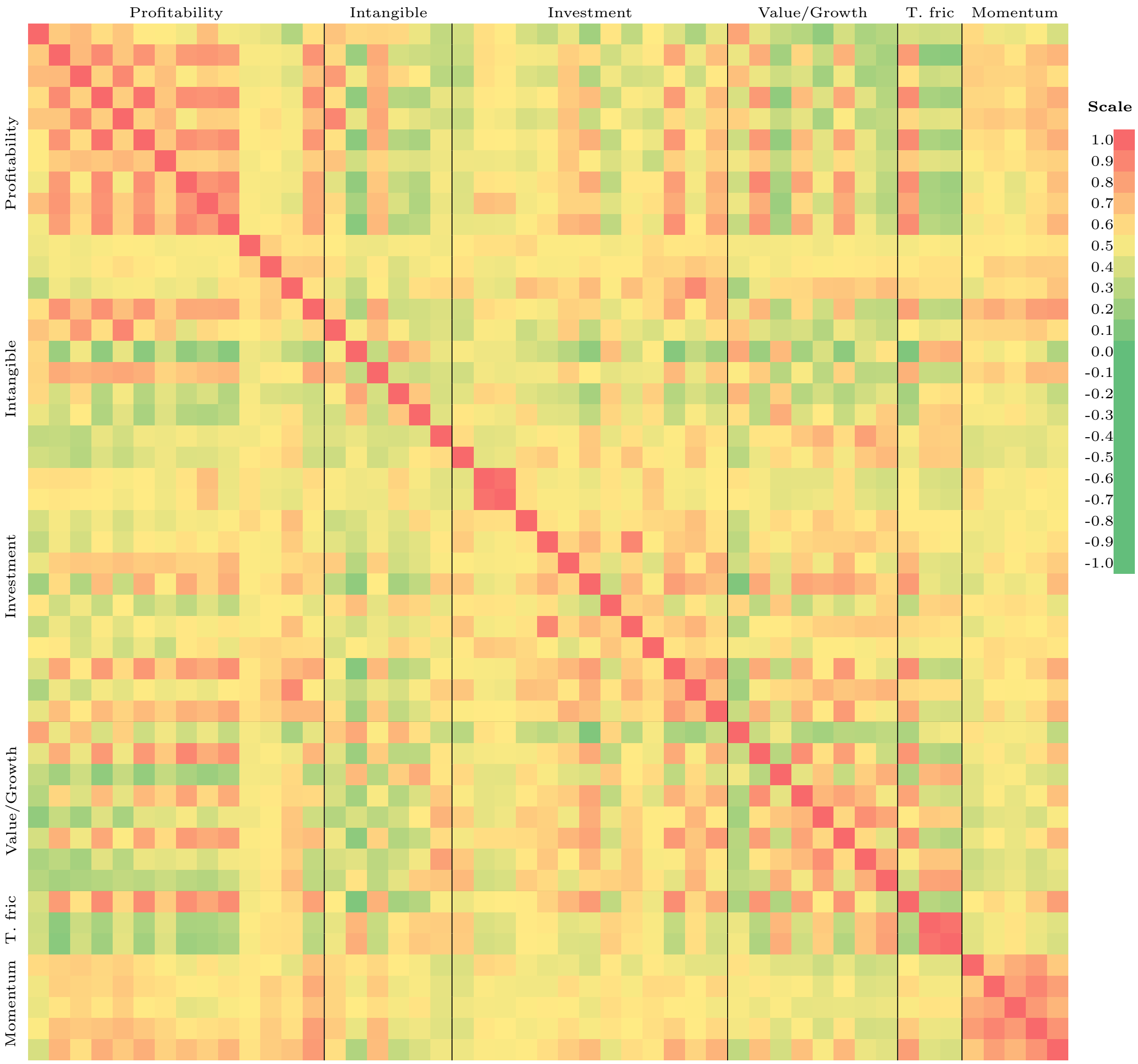
We conduct a number of robustness checks for the regression around the 2002 shock in Section 4.2.

Choice of fixed effects. We include both factor and month fixed effects in the main specification. In Table C.2, we show that including or omitting those fixed effects does not materially change the result.

Choice of return covariance matrix. In the main specification, the FGLS estimator uses a factor return covariance matrix calculated from the full sample of monthly returns. One may worry that covariance is nonstationary over time and thus we should use an estimate

Figure C.3. Correlation Matrix of Factor Returns

High correlations are colored red, and low correlations are colored green. Following Hou et al. (2019), we classify the 49 asset pricing factors into six categories and order them accordingly. The second-to-last category is the trading friction category.



that is more local to 2002. We therefore also estimate a covariance matrix using 36 months or 12 months of data centered around the methodology event. Because there are more factors than months, we use daily factor returns and scale the covariance matrix up to a monthly frequency assuming 21 trading days per month. The regression results in Table C.3 show that the exact covariance matrix choice is not crucial for our results.

Table C.2. Factor Returns around June 2002, Varying Fixed Effects

We regress monthly returns on average ExpSum(ΔRating) during the six months before to six months after the methodology change. Columns (1) to (3) use panel regressions, and Columns (4) to (6) use an instrumented version of the independent variable. The standard errors in parentheses are adjusted for the cross-sectional correlation between factor returns using a feasible generalized least squares approach.

Dependent variable:	Monthly factor return ($\text{Ret}_{f,t}$)(%)					
Regression:	Panel regression			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$	1.909***	1.938***	1.920***	1.210***	1.894***	1.937***
	(0.531)	(0.583)	(0.605)	(0.459)	(0.473)	(0.476)
Factor FE	No	Yes	Yes	No	Yes	Yes
Month FE	No	No	Yes	No	No	Yes
Observations	588	588	588	588	588	588
Adj R^2	1.59%	7.39%	7.82%	0.46%	7.01%	7.71%

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

Table C.3. Factor Returns Around 2002, Varying Covariance Matrix

We regress monthly returns on $\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$ during the six months before versus six months after the methodology change. Columns (1) to (3) use panel regressions, and Columns (4) to (6) use an instrumented version of the independent variable. The standard errors in parentheses are adjusted for the cross-sectional correlation between factor returns using a feasible generalized least squares approach. In Columns (1) and (4), the factor return covariance is estimated using monthly returns throughout the full sample. In Columns (2) and (5), it is estimated using daily returns over 36 months around the methodology change event, and Columns (3) and (6) uses daily returns over 12 months around the event.

Dependent variable:	Monthly factor return ($\text{Ret}_{f,t}$) (%)					
Regression:	Panel regression			2SLS		
Estimation frequency:	Monthly	Daily		Monthly	Daily	
Estimation period:	Full sample	36 month	12 month	Full sample	36 month	12 month
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$	1.920***	1.818***	1.968***	1.937***	1.670***	1.860***
	(0.605)	(0.691)	(0.586)	(0.476)	(0.590)	(0.513)
Factor FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	588	588	588	588	588	588
Adj R^2	7.82%	7.47%	7.54%	7.71%	7.25%	7.16%

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

Appendix D Explanatory Power over the Full Sample

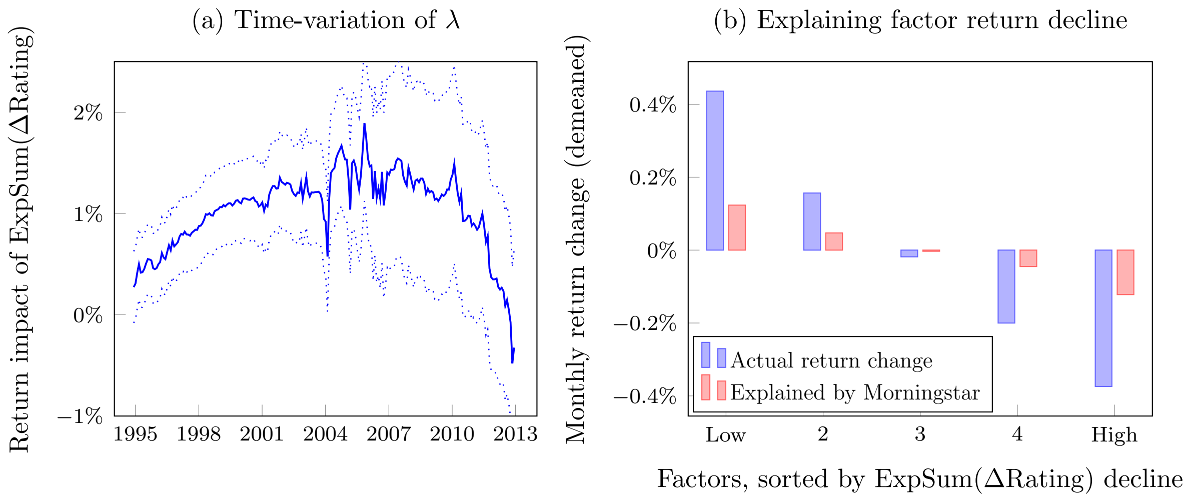
In Section 5.2, we estimate the explanatory power of Morningstar ratings on the post-June 2002 factor profitability decline using the λ estimated from the 2002 shock. We now estimate how the relation varies over time by estimating the factor return predictability (Regression (14)) using 10-year rolling windows. The results are plotted in Panel (a) of Figure D.4. There is indeed variation over the sample period, with λ being higher in the middle part of the sample. We then calculate the return explained by ratings as

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

where $\hat{\lambda}_t$ is estimated using a 10-year window centered around month t.

Figure D.4. Explanatory Power of Morningstar Ratings on Post-June 2002 Profitability Decline, with Time-Varying λ Estimate

Panel (a) plots estimations of λ through predictive regressions of factor returns on $\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$ using rolling 10-year centered windows. The graph starts in 1995 and ends in 2013 due to the need for 10-year windows. The dashed lines represent two standard error bands. Panel (b) plots the equivalent of Figure 14—estimation of the post-June 2002 factor profitability decline explained by Morningstar—using the time-varying estimate in Panel (a). We sort factors into quintiles by their average decline of $\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$ after June 2002. The actual returns changes are plotted in blue, and the part explained by the Morningstar methodology change, estimated using λ times the before-versus-after change in $\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$, is plotted in red. To focus on cross-sectional variation across factors, the data series are demeaned.



We now quantify the explanatory power of Morningstar using this time-varying λ. As in Section 5.2, we sort factors by their post-June 2002 decline of ExpSum(ΔRating) into quintiles so that the top quintile factors experience the largest Morningstar-induced decline. The actual return changes are plotted in blue, and the component explained by Morningstar is plotted in red. To focus on the cross-sectional difference, both series are demeaned. Relative to the bottom quintile, the top quintile factors experience a 0.81% additional decline in factor profitability, out which Morningstar can explain 0.26%, i.e., 30% of the overall

Table D.4. Factors before and after June 2002, Full Sample Statistics

We rank the 49 asset pricing factors in descending order of $\overline{\text{ExpSum}(\Delta\text{Rating})}_{\text{pre 2002}} - \overline{\text{ExpSum}(\Delta\text{Rating})}_{\text{post 2002}}$. Therefore, the first factor is the most negatively affected by the Morningstar methodology change, and the last factor is the least affected.

Factor	Category	ExpSum(ΔRating)		Monthly return	
		Before 6/2002	After 6/2002	Before 6/2002	After 6/2002
52-week high	Momentum	0.310	0.065	1.14%	-0.72%
Size	Trading frictions	0.187	-0.020	0.43%	0.11%
Industry momentum	Momentum	0.235	0.032	0.73%	-0.02%
Long-term reversals	Value/growth	0.151	-0.038	0.90%	-0.02%
Amihud illiquidity	Trading frictions	0.155	-0.022	0.54%	0.25%
Momentum (t-2, t-12)	Momentum	0.268	0.092	1.91%	0.09%
Momentum (t-2, t-6)	Momentum	0.243	0.083	0.92%	-0.13%
Advertising expense	Intangible	0.124	-0.024	0.45%	0.16%
Sales-to-price	Value/growth	0.074	-0.043	0.67%	0.14%
Asset growth	Investment	0.108	-0.007	0.73%	0.07%
Book-to-market	Value/growth	0.081	-0.034	0.53%	0.01%
One-year share issuance	Investment	0.114	0.001	1.35%	0.09%
Abnormal capital investment	Investment	0.094	-0.017	1.13%	0.44%
Enterprise multiple	Value/growth	0.072	-0.022	0.04%	0.13%
Cash flow-to-price	Value/growth	0.063	-0.026	0.49%	0.16%
R&D expense	Intangible	0.096	0.010	1.09%	0.48%
Investment growth	Investment	0.071	-0.013	0.69%	0.03%
Net payout yield	Value/growth	0.085	0.003	1.23%	0.11%
Sustainable growth	Profitability	0.089	0.009	0.62%	0.10%
Distress risk	Profitability	0.121	0.043	1.97%	0.15%
Change in asset turnover	Profitability	0.088	0.010	0.57%	0.01%
Intermediate momentum (t-7,t-12)	Momentum	0.098	0.029	0.90%	0.32%
Total external financing	Investment	0.065	-0.001	0.54%	0.11%
Investment-to-capital	Investment	0.037	-0.026	0.10%	-0.09%
Sales-minus-inventory growth	Profitability	0.049	-0.012	0.42%	-0.14%
Earnings-to-price	Value/growth	0.050	-0.009	0.64%	0.45%
Net operating assets	Investment	0.065	0.011	1.37%	0.47%
Industry adjusted CAPX growth	Investment	0.026	-0.014	0.37%	-0.12%
Investment-to-assets	Investment	0.025	-0.014	0.06%	-0.07%
Maximum daily return	Trading frictions	0.014	-0.015	0.40%	0.13%
Five-year share issuance	Investment	0.038	0.011	0.25%	0.36%
Net working capital changes	Investment	0.026	0.000	0.92%	0.15%
Growth in inventory	Investment	0.030	0.006	0.67%	0.50%
QMJ profitability	Profitability	0.032	0.012	1.11%	0.21%
Profit margin	Profitability	0.004	-0.014	0.72%	-0.18%
Accruals	Investment	0.023	0.005	0.86%	0.19%
Operating profitability	Profitability	0.018	0.007	0.69%	0.03%
Firm age	Intangible	-0.012	-0.010	-0.32%	0.33%
Piotroski's F-score	Profitability	0.001	0.006	-0.03%	0.13%
Return on equity	Profitability	0.013	0.024	0.54%	0.19%
Industry concentration	Intangible	-0.017	0.003	0.12%	0.39%
Sales growth	Value/growth	-0.017	0.007	0.09%	0.09%
Gross profitability	Profitability	0.000	0.026	0.18%	0.36%
Return on assets	Profitability	-0.002	0.025	0.70%	0.06%
Operating Leverage	Intangible	-0.013	0.018	0.01%	0.24%
Earnings persistence	Intangible	-0.019	0.042	0.52%	0.64%
Cash-based profitability	Profitability	-0.036	0.029	-0.06%	0.60%
Ohlson's O-score	Profitability	-0.034	0.033	0.79%	-0.11%
Altman's Z-score	Profitability	-0.071	0.019	0.09%	-0.11%

variation.[24]

[24] Because the λ estimate requires 10 years of data, we extrapolate the earliest estimate to fill the first five years and the latest estimate to fill the last five years.

Appendix E Additional Results

Figure E.5. Long-term Reversal of Rating-Induced Price Pressures.
We sort stocks by NYSE-based ExpSum(Δ Rating) break points into deciles. The two lines plot the cumulative value-weighted log return of the extreme deciles after subtracting market returns. Month 0 is the sorting month.