NBER WORKING PAPER SERIES

DISCONTINUED POSITIVE FEEDBACK TRADING AND THE DECLINE IN ASSET PRICING FACTOR PROFITABILITY

Itzhak Ben-David
Jiacui Li
Andrea Rossi
Yang Song

Working Paper 28624
http://www.nber.org/papers/w28624

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2021

We thank Nicholas Barberis, Lauren Cohen, Sylvester Flood (Morningstar), Umit Gurun, Paul Kaplan (Morningstar), Dong Lou, Chris Malloy, Andrei Shleifer for helpful comments. We thank seminar participants at The Ohio State University, the University of Utah, the University of Washington, Hong Kong University of Science and Technology, and Arrowstreet Capital, as well as the National Bureau of Economic Research Behavior Finance Workshop for comments and George Aragon for sharing data. Ben-David is with The Ohio State University and the National Bureau of Economic Research, Li is with University of Utah, Rossi is with the University of Arizona, and Song is with the University of Washington. Ben-David is a co-founder and a partner in an investment advisor. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Discontinued Positive Feedback Trading and the Decline in Asset Pricing Factor Profitability
Itzhak Ben-David, Jiacui Li, Andrea Rossi, and Yang Song
NBER Working Paper No. 28624
March 2021
JEL No. G11,G24,G41

ABSTRACT

We show that a June 2002 reform in Morningstar's mutual fund rating methodology led to substantial drop in the profitability of momentum-related asset pricing factors. Before the reform, funds pursuing the same investment style had correlated ratings heavily influenced by recent style performance. Therefore, ratings-chasing flows generated large style-level positive feedback trading. The reform decoupled ratings from style-level performance; consequently, factors that benefited from positive feedback trading experienced a precipitous return decline. The performance decline was limited to the U.S. market where the reform happened. We estimate that the reform explains 25%–50% of the long-term profitability drop in momentum-related factors.

Itzhak Ben-David
The Ohio State University
Fisher College of Business
606A Fisher Hall
Columbus, OH 43210-1144
and NBER
ben-david.1@osu.edu

Jiacui Li
David Eccles School of Business
University of Utah
SFEBB 8123, 1655 Campus Center Dr
Salt Lake City, UT 84112
jiacui.li@eccles.utah.edu

Andrea Rossi
University of Arizona
Eller College of Management
Department of Finance
1130 E. Helen St.
Tucson, AZ 85721
rossi2@email.arizona.edu

Yang Song
Foster School of Business
University of Washington
4273 E Stevens Way NE
Seattle, WA 98195
songy18@uw.edu

1 Introduction

Over the last four decades, asset pricing researchers have identified hundreds of factors (anomalies) in the cross-section of stock returns. However, the profitability of these factors has declined noticeably over time. And perhaps more puzzling, average factor returns *suddenly* dwindled in mid-2002. The kink in profitability is evident in Panel (a) of Figure 1, which plots the average monthly return of 49 popular factors. Visibly, average factor profitability declined sharply after mid-2002, going from 0.43% per month during the earlier period of January 1991 to June 2002 to merely 0.09% after the reform. The drop was particularly sharp for momentum-type factors (e.g., industry momentum), which accounted for a substantial fraction of the overall factor-based profits before 2002.[1]

In this study, we argue that a sizeable part of the post-2002 profitability decline is attributable to a seemingly innocuous institutional change that took place exactly in June 2002: Morningstar's mutual fund rating reform. Before the ratings reform, Morningstar gave mutual funds with similar investment styles highly correlated ratings. Because ratings are a major driver of fund flows, flows to mutual funds with similar styles depended strongly on recent style performance and were highly correlated with each other. As funds scaled up or down their holdings in response to flows (e.g., Lou, 2012), their trading behavior caused substantial positive feedback effects at the style level, contributing to the profits of many factors and especially the momentum-type factors. In June 2002, Morningstar reformed its ratings methodology by ranking funds *within* 3×3 size-value peer groups. By doing so, Morningstar disrupted the ratings-driven positive feedback trading. As a result, factors whose returns benefited from exposure to the pre-June 2002 ratings system experienced a sharp and persistent profitability decline.

In addition to explaining the "2002 kink," our study provides insight into another puzzle. Lou (2012) proposes that investors direct flows towards the best-performing mutual funds,

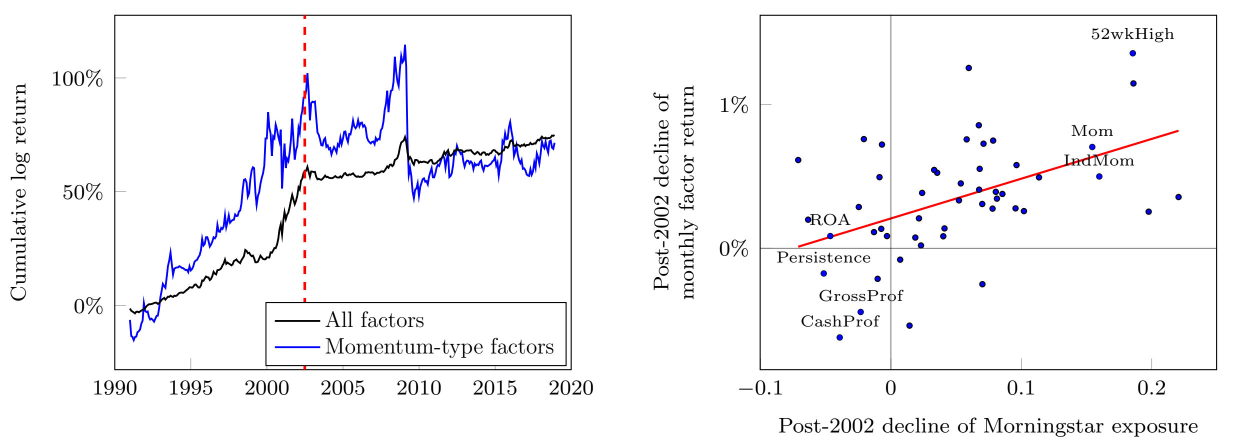
[1] In Appendix B.1, we show that the "2002 kink" is not specific to our choice of factors or our factor construction and is evident in several earlier studies as well.

which creates positive-feedback trading due to the expansion of mutual fund portfolios and thus contributes to momentum profits. However, as the mutual fund sector kept growing over time, one would have expected momentum profits to remain strong, whereas, in fact, momentum profits stalled around mid-2002. Our paper resolves this puzzle. We demonstrate that the positive feedback mechanism between flows and prices is mediated by Morningstar ratings, and that the post-2002 drop is, to a large extent, explained by the Morningstar reform in its rating methodology. In addition, the reform event had a predictably different impact on flows and price pressure across asset pricing factors, allowing sharper identification of flow-induced price effects than is normally possible in other asset pricing settings.

Figure 1. Morningstar Rating Methodology Change and Factor Return Decline

Panel (a) shows the cumulative returns of 49 popular asset pricing factors and momentum-type factors, which include standard ($t-12$, $t-1$) momentum and other factors such as industry momentum. Panel (b) plots the decline of average factor return after June 2002 against the decline in their exposure to Morningstar ratings. From left to right, the labeled factors include return on assets, earnings persistence, gross profitability, cash profitability, momentum, industry momentum, and the 52-week-high strategy. Morningstar exposure measures the degree to which each factor benefits from the positive feedback induced by Morningstar ratings and is defined in Section 3.1 (variable ExpSum(ΔRating)). The red line is the best linear fit.

(a) Asset pricing factor returns　　(b) The cross-section of factor return decline



Our analysis shows that asset pricing factors with high exposure to the change in Morningstar's methodology—i.e., factors that are related to momentum—experienced the largest decline in profitability. Panel (b) of Figure 1 summarizes the relation between factor profitability and ratings across a sample of 49 commonly used asset pricing factors. The plot shows that factors that experienced a larger decline in their exposure to rating-induced positive feedback trading (horizontal axis) also experienced larger profitability declines after 2002 (vertical axis). In the rest of this paper, we seek to establish whether the correlation presented in the plot can be interpreted as causal and to quantify how much of the post-2002 factor profitability decline can be explained by the rating-induced flow mechanism.

We start by documenting strong rating-induced price effects using *stock-level* predictive regressions. Mutual funds recently upgraded (downgraded) by Morningstar attract inflows (outflows), which forces them to scale up (down) their portfolios by increasing (decreasing) their current holdings (Pollet and Wilson, 2008; Lou, 2012). As a result, stocks in mutual fund portfolios experience price pressure in the same direction as the flows. To summarize ratings effects at the stock-level, for each stock, we use a simple measure to capture recent ratings changes experienced by the funds holding that stock, and we then examine whether this measure predicts stock returns. Predictive Fama-MacBeth regressions show that non-microcap stocks with one-standard-deviation higher recent rating changes have 28.7 basis points (bps) higher return in the subsequent month. This predictability is not explained away by controlling for a host of known stock characteristics and, in fact, exhibits a larger magnitude than that of most known return predictors (book-to-market, asset growth, etc.). Further, unlike many other characteristics that have stronger predictive power for small cap stocks (Hou, Xue, and Zhang, 2019), rating-induced price pressure is stronger in large cap stocks, consistent with the fact that mutual funds tend to hold stocks with larger market capitalization. Overall, this evidence suggests that Morningstar ratings, through rating-induced flows, strongly impact prices at the stock level.

To provide more evidence that Morningstar ratings can influence *factor-level* returns, we also conduct an event study at the factor level around the June 2002 Morningstar ratings reform (from January 2002 to December 2002). The event study helps to identify the causal effect of ratings changes on factor returns. Specifically, we construct 49 popular long/short asset pricing factors and, following Hou et al. (2019), we use NYSE break points and value-weighting. The reform caused ratings changes in over 50% of funds, which, through rating-induced price effects, had significant and *heterogeneous* impact on factors. Some factors, such as the size factor, were negatively affected exactly at this event because funds holding their long (short) legs experience reform-induced ratings downgrades (upgrades), while the reverse was true for other factors, such as Ohlson's O-Score. As predicted, factors positively (negatively) affected by the Morningstar ratings reform experienced an increase (reduction) in fund flows and returns right after the June 2002 event. Using all other years other than 2002 as placebo tests, we confirm that the factor-level flow and return patterns documented are unique to 2002. Moreover, proxies for alternative explanations, such as arbitrage activity and liquidity, did not vary materially around the reform event.

Having established that rating-induced flows *can* strongly influence factor returns, we proceed to study the impact of the Morningstar reform on long-term factor profitability. Our hypothesis is that, by disrupting positive feedback trading, the ratings reform diminished the profitability of factors, especially the momentum-type ones. To this end, we first show that the reform indeed caused a disruption in style-level positive feedback trading. Morningstar ratings are primarily based on the ranking of recent fund performance. Before June 2002, mutual funds in the style with top recent performance received flows that were higher *per month* by 1.7% of assets under management (AUM) than funds in the bottom style. Consistent with the flows creating further price impact, stocks held by funds in the top style continue to outperform those in the bottom style by 84 bps per month. After June 2002, the fund flow and return difference between the top and bottom styles became muted.

Naturally, we conjecture that this reform would have the largest negative *long-term* impact on momentum-type factors, which are likely to benefit from positive feedback trading due to how they are constructed (e.g., industry momentum, 52-week-high). The results are consistent with this conjecture. When comparing the period before and after June 2002,[2] we find that momentum-related factors suffered the largest drop in their rating exposure and also the largest profitability drop of approximately 0.8% per month, while other factors

[2]Our sample starts in 1991. This choice is dictated by monthly fund flow data availability in CRSP.

only experienced a profitability drop of 0.29%. To further test the mechanism, we conduct a region-based placebo test using the international versions of momentum factors constructed by other researchers in prior studies. Consistent with the fact that the ratings reform event is unique to the U.S., we do not observe a similar decline in momentum profitability in other countries.

We close our analysis by quantifying the explanatory power of ratings on factor profits through a back-of-envelope calculation. Using a range of estimated price impact parameters, we find that the Morningstar reform accounts for 25% to 50% of the momentum factor profitability decline after June 2002. The reform can also account for a significant fraction of the drop in the other momentum-type factors returns.

This paper is related to studies that aim to explain the decline of factor profitability over time—a phenomenon important not only to academics but also to industry practitioners practicing "smart beta" investing. Existing papers emphasize the role of liquidity (Khandani and Lo, 2011; Chordia, Subrahmanyam, and Tong, 2014; Lee and Ogden, 2015), the entry of arbitrageurs (e.g., hedge funds, see Green, Hand, and Soliman, 2011; Hanson and Sunderam, 2013), and the role of academic publications (Marquering, Nisser, and Valla, 2006; McLean and Pontiff, 2016; Calluzzo, Moneta, and Topaloglu, 2019). It is also possible that some factors reported in the literature were data-mined and stopped working after their "discovery" (Harvey, Liu, and Zhu, 2016; Harvey, 2017; Hou et al., 2019; Huang, Song, and Xiang, 2020b). Specifically to the momentum factor, Daniel and Moskowitz (2016) document that it experienced a crash in the wake of the 2008 financial crisis. These explanations are not mutually exclusive to ours and are likely important contributors to the *secular* factor profitability decline. However, our hypothesis is the only one that can account for the perplexing "2002 kink"—an event which also helps the identification of the Morningstar-based mechanism.

This study is also related to studies of the impact of institutional demand on asset prices. Studies in this area have found price effects due to index composition changes (Harris and

Gurel, 1986; Shleifer, 1986; Wurgler and Zhuravskaya, 2002; Chang, Hong, and Liskovich, 2015), mutual fund flows (Teo and Woo, 2004; Coval and Stafford, 2007; Lou, 2012; Li, 2020; Huang, Song, and Xiang, 2020a), exchange-traded fund flows (Ben-David, Franzoni, and Moussawi, 2018; Brown, Davies, and Ringgenberg, 2021), and other sources of institutional investor demand (Parker, Schoar, and Sun, 2020; Ben-David, Franzoni, Moussawi, and Sedunov, 2021). More recently, Koijen and Yogo (2019) develop a structural methodology to estimate price impact, and Gabaix and Koijen (2020) show that the demand-induced price impact coefficient at the aggregate level is larger than that at the idiosyncratic level, a finding our paper corroborates.

Momentum is perhaps the most prominent "anomaly" and, according to Eugene Fama, "the biggest embarrassment for efficient markets."[3] Scholars seeking to explain momentum profitability have offered myriad explanations, including delayed information diffusion (Hong and Stein, 1999), behavioral biases (Daniel, Hirshleifer, and Subrahmanyam, 1998; Barberis, Shleifer, and Vishny, 1998; George and Hwang, 2004; Grinblatt and Han, 2005), as well as investor attention and media influences (Lee and Swaminathan, 2000; Hou, Xiong, and Peng, 2009; Hillert, Jacobs, and Müller, 2014).[4] Our findings are consistent with the style-investing hypothesis in Barberis and Shleifer (2003). Positive feedback trading has also been identified in Teo and Woo (2004), Lou (2012), and Wahal and Yavuz (2013). Relative to these studies, we contribute by using an exogenous event to identify the price impact of rating-induced trading and by quantifying the explanatory power of our mechanism.

The rest of the paper is organized as follows. Section 2 details the data, factor universe, and variable construction. Section 3 shows the mechanism of rating-induced price pressures using stock-level return predictability regressions and a factor-level event study. Section 4 investigates the reform-induced disruption of style-level positive feedback trading and the resulting impact on asset pricing factors, and Section 5 quantifies the impact of Morningstar

[3] See "Fama on Momentum," AQR 2016, accessible at `https://www.aqr.com/Insights/Perspectives/Fama-on-Momentum`.

[4] The literature on momentum is vast, and we cannot possibly cover all (or even most) of the explanations. Please see Jegadeesh and Titman (2011) for a review.

on the post-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.

2.1 Mutual Fund Data

We obtain monthly fund returns and total net assets (TNA) from the CRSP survivorship bias-free mutual fund data set. Our data are at a monthly frequency and span 1991 to 2018. We start in 1991 because monthly fund flow in CRSP starts in 1990, and some measures require one year of lagged data to construct. We use all U.S. domestic equity mutual funds. While funds are often marketed to different clients through different share classes, they invest in the same fund portfolio and typically only differ in the fee structure. Therefore, we aggregate all share classes at the fund level using Russ Wermers's MFLINKS (Wermers, 2000). We also obtain quarterly fund holdings from Thomson Reuters' S12, which is based on 13F filings.

We obtain Morningstar ratings and fund style categories from Morningstar Direct, and we merge them with the CRSP data using the matching table from Pastor, Stambaugh, and Taylor (2020). Because Morningstar assigns ratings at the share class level, we aggregate ratings at the fund level by TNA-weighting different share classes following Barber, Huang, and Odean (2016). We restrict our analysis to mutual funds with at least $1 million TNA, and we winsorize fund flows at the 0.5% and 99.5% levels. We require the existence of 12 lags of monthly flows, returns, and ratings. The resulting sample comprises a total of 3,305 funds with 454,787 fund-month observations.

2.2 Asset Pricing Factors

We compute 49 popular stock-level characteristics that have been shown to predict returns, mostly following Arnott, Clements, Kalesnik, and Linnainmaa (2019a).[5] Using the classification categories proposed in Hou et al. (2019), these characteristics include 14 in the profitability category (e.g., return on assets), 13 in the investments category (e.g., share issuance), eight in the value/growth category (e.g., book-to-market ratio), 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).

We follow the prescription in Hou et al. (2019) and limit the impact of microcaps in factor construction. Specifically, we use NYSE breakpoints to sort stocks into quintiles and then form factors as long the top quintile and short the bottom quintile. The quintile portfolio returns are value-weighted. Appendix Table A.1 lists all asset pricing factors used in this paper.

2.3 Morningstar Rating and Flow-Induced Trading

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

We define the average Morningstar rating of stock i in month t as the holding-weighted rating of all funds $\mathcal{J}$ that hold the stock:[6]

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

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

[5] We restrict our attention to those that can be constructed using CRSP and Compustat data.

[6] Note that Morningstar assigns ratings at the *fund*-level. The stock-level ratings are computed by us.

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:[7]

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

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

$$\mathrm{Flow}_{j,t} = \frac{\mathrm{TNA}_{j,t}}{\mathrm{TNA}_{j,t-1}} - (1 + \mathrm{Ret}_{j,t}). \tag{4}$$

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

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

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

$$\mathrm{FIT}_{f,t} = \sum_{i \in \text{top quintile}} w^f_{i,t-1} \cdot \mathrm{FIT}^{\mathrm{stock}}_{i,t} - \sum_{i \in \text{bottom quintile}} w^f_{i,t-1} \cdot \mathrm{FIT}^{\mathrm{stock}}_{i,t}, \tag{6}$$

where $w^f_{i,t-1}$ is the lagged market cap weight of stock i in the corresponding quintile portfolio. Summary statistics for the stock-level and factor-level samples both appear in Table 1.

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

[8] Wardlaw (2020) 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.

Table 1. Summary Statistics

Panels A and B present the stock-level and factor-level summary statistics, respectively. Return and FIT (flow-induced trading) are measured monthly. Per Equation (3), FIT is defined as the amount of mutual fund trading induced by fund flows as a fraction of shares held. Morningstar rating is measured in stars (1 to 5), and ExpSum(Δ Rating) is an exponentially weighted sum of the past 12 months of rating changes (defined in Equation (8)). Obs is the average number of observations per month. The last five columns report 1%, 25%, 50%, 75%, and 99% percentile distributions, respectively.

Panel A: Stock-Level								
	Obs	Mean	Std dev	1%	25%	50%	75%	99%
Market cap ($m)	4,405	3,443	16,172	7	110	396	1,587	56,645
Held by num funds	4,405	78.5	104.0	1.0	11.0	36.0	110.0	480.0
Return	4,405	1.10%	15.62%	−38.46%	−6.01%	0.44%	7.04%	52.24%
Rating	4,405	3.369	0.714	1.181	3.000	3.441	3.850	5.000
ExpSum(ΔRating)	4,405	−0.028	0.767	−2.358	−0.306	0.000	0.253	2.268
FIT	4,405	0.55%	2.44%	−4.49%	−0.46%	0.28%	1.22%	8.68%
Panel B: Factor-Level								
	Obs	Mean	Std Dev	1%	25%	50%	75%	99%
Return	49	0.23%	3.69%	−9.94%	−1.60%	0.15%	2.00%	10.71%
Rating	49	0.018	0.207	−0.700	−0.062	0.016	0.103	0.554
ExpSum(ΔRating)	49	0.025	0.256	−0.745	−0.057	0.009	0.089	0.881
FIT	49	0.04%	0.43%	−1.23%	−0.14%	0.03%	0.20%	1.38%

3 Rating-induced Price Effects

In this section, we demonstrate that Morningstar rating changes can indeed influence factor returns significantly by triggering flow-induced trading. To this end, we start with a stock-level exercise, showing that rating-induced return predictability is stronger than that of most stock characteristics. We then conduct an event study at the factor level around the June 2002 Morningstar rating reform, which provides further evidence that rating changes have causal effects on factor-level returns.

Before delving into details, it is worth noting that the mutual fund sector is an important player in the equity market so their trades, collectively, are large enough to generate price impact.[9] When our sample begins in 1991, U.S. equity mutual funds had a total AUM of $326 billion, which was 8.9% of the entire market capitalization. These numbers grew

[9] A number of papers have reached the same conclusion. For instance, see Coval and Stafford (2007), Ben-Rephael, Kandel, and Wohl (2011), and Lou (2012).

steadily over time. By the end of our sample period (2018), equity mutual funds owned \$10,849 billion, which represented 29.3% of overall market capitalization (Figure A.1 in the Appendix).

3.1 Stock-Level Rating-induced Return Predictability

In this subsection, we show that recent rating changes create sizeable return predictability at the stock-level. In fact, the degree of predictability is comparable to that of the strongest known return predictors. Further, consistent the fact that mutual funds seldom hold microcap stocks, the rating-induced predictability is more pronounced in large cap stocks, By contrast, most other stock predictors tend to be stronger in small-cap and microcap stocks (Hou et al., 2019).

Summarizing recent rating changes at the stock level. As a primitive, we validate each link in the rating-induced price impact mechanism. First, funds with positive (negative) rating changes receive in (out) flows, and their managers respond by buying (selling) more of their portfolio holdings. Second, when we aggregate such flow-induced trading (FIT) at the stock level, we find that FIT generates price impact in stocks that revert subsequently. Because the individual links involved in the overall mechanism have been confirmed in earlier studies (e.g., Del Guercio and Tkac, 2008; Lou, 2012; Reuter and Zitzewitz, 2015), we refer the reader to Appendix B.2 for details.

What is an appropriate way to capture the expected stock price impact of rating-induced trading? In Appendix B.2, we show that the price impact primarily comes from recent rating *changes*. This is due to the fact that the FIT-induced price effects revert over time; therefore, the price impact of ratings that remain unchanged over time is canceled out by the reversal of the price impact of earlier flows. *Changes in ratings*, however, create a one-time price impact impulse, relative to the earlier level. We thus use a simple specification to summarize the recent rating changes experienced by funds holding each stock i using exponentially-decaying weights:

$$\text{Ret}_{i,t} = \text{Ret}_{i,t}^{\text{counterfactual}} + \underbrace{\lambda \cdot \text{ExpSum}(\Delta\text{Rating})_{i,t-1}}_{\text{Rating-induced price effects}} \tag{7}$$

$$\text{where ExpSum}(\Delta\text{Rating})_{i,t-1} = \sum_{k=1}^{12} \tau_k \cdot \Delta\text{Rating}_{i,t-k}. \tag{8}$$

In Equation (8), the weights τ_k decay at an exponential rate $\delta = 0.764$, which is estimated from a least-squares fit to the cumulative response of stock returns to past rating changes (see Appendix Figure B.4 for details).[10] We use 12 lags because the impact primarily happens within a year of rating changes, and the estimated decay rate of weights implies a half-life of 2.58 months. We normalize the weights to sum to 12 (i.e., $\sum_{k=1}^{12} \tau_k = 12$), so $\text{ExpSum}(\Delta\text{Rating})_{i,t-1}$ should be interpreted as the rating change over a one-year period. λ is the price impact coefficient.[11] The reasoning behind this specification is explained in detail in Appendix Section B.2. As will be shown momentarily, our results are not sensitive to reasonable variation in the exponential weight parameter δ.

Estimating stock-level return predictability. Thus, we anticipate recent rating change, $\text{ExpSum}(\Delta\text{Rating})_{i,t-1}$, to predict stock returns. How strong is the rating-induced return predictability? To answer this, we estimate stock-level Fama-MacBeth regressions:

$$\text{Ret}_{i,t} = a + b \cdot \text{ExpSum}(\Delta\text{Rating})_{i,t-1} + X_{i,t-1} + \epsilon_{i,t}, \tag{9}$$

where controls $X_{i,t-1}$ always include cumulative fund returns over the past 36 months, aggregated at the stock level, in order to isolate the marginal effect of ratings. The standard errors are calculated using the Newey-West methodology with 12 lags. Table 2 shows the results. The first two columns do not involve more controls, whereas the final two columns

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

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

control for all 49 stock return predictors in Section 2.2. To enable comparison between $\text{ExpSum}(\Delta\text{Rating})_{i,t-1}$ and the other characteristics, we turn all of them into z-scores by month;[12] thus the coefficient can be interpreted as the monthly return associated with a one-standard-deviation increase of the predictor.

Column (1) shows that a one-standard-deviation increase in $\text{ExpSum}(\Delta\text{Rating})_{i,t-1}$ leads to 19.5 bps higher stock return in the subsequent month. When we exclude microcap stocks—defined as those with market capitalization below the 20% NYSE cutoff—the predictive power increases to 28.7 basis points, consistent with the fact that mutual funds tend to hold large cap stocks. This is a key difference between this rating-based predictor and most other predictors that work better on small cap stocks (Hou et al., 2019). Columns (3) and (4) show that, after controlling for characteristics in the Fama-French six factor model (momentum plus FF5 (Fama and French, 2015)), the predictive coefficients only attenuates by 30–40%. Further controlling for all other predictors in Columns (5) and (6) has little effect. All these results are statistically significant at the 1% level. Therefore, we conclude that past Morningstar rating changes are indeed a strong predictor of stock returns.

Horse race against other known return predictors. How does the rating-based return predictability compare with the other 49 predictors? We run a simple horse race. In Figure 2, we compare the return prediction coefficient of $\text{ExpSum}(\Delta\text{Rating})_{i,t-1}$ with all the other ones (also turned into z-scores). Panels (a) and (b) plot the coefficients in univariate regressions while Panels (c) and (d) plot the multivariate regressions where all predictors are used. The regression coefficient for $\text{ExpSum}_{i,t-1}$ is colored red, and the other 49 predictors are colored blue. To ensure that our result is not sensitive to the decay parameter δ, we also run the same estimation using alternative versions of $\text{ExpSum}(\Delta\text{Rating})_{i,t-1}$ with $\delta = 0.3, 0.4, \ldots, 0.9$—which corresponds to half-lives from 0.58 to 6.58 months—and plot those coefficients in green.

[12] We switch the signs, if necessary, so that the predictors are expected to predict returns positively based on the original studies.

Table 2. Predicting Stock Returns using Recent Rating Changes

We use Fama-MacBeth regressions to estimate the stock return predictive power of ExpSum(ΔRating)$_{i,t-1}$—an exponential sum of recent rating changes—after controlling for lagged 36 month cumulative fund return aggregated at the stock level. Columns (1) and (2) do not control for other predictors. Columns (3) and (4) controls for momentum and the Fama-French five-factor characteristics: beta, size, book/market, investment, and profitability (Fama and French, 2015). Columns (5) and (6) also control for all the other 49 stock characteristics in Section 2.2. Columns (1), (3), and (5) use all stocks, while Columns (2), (4), and (6) exclude microcap stocks, defined as those with market cap below the 20% NYSE percentile. Standard errors are calculated using the Newey-West procedure with 12 monthly lags. Coefficients that are statistically significant at the 1%, 5%, and 10% levels are marked by ***, **, and *, respectively.

Dependent variable: monthly stock return (Ret$_{i,t}$)						
	Univariate		FF6 Controls		All Controls	
	All stocks	No microcap	All stocks	No microcap	All stocks	No microcap
	(1)	(2)	(3)	(4)	(5)	(6)
ExpSum(ΔRating)$_{i,t-1}$	0.195***	0.287***	0.144***	0.186***	0.127***	0.174***
	(0.052)	(0.086)	(0.032)	(0.048)	(0.025)	(0.043)
Lagged fund return	Yes	Yes	Yes	Yes	Yes	Yes
FF 6 Predictors	No	No	Yes	Yes	Yes	Yes
All other predictors	No	No	No	No	Yes	Yes
Observations	1,271,496	1,271,496	1,271,496	1,271,496	1,271,496	1,271,496
Adjusted R^2	13.21%	19.45%	19.59%	26.85%	20.96%	30.21%

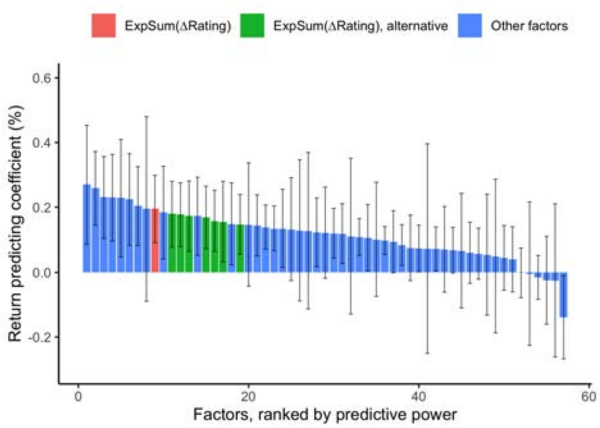
On a univariate basis, while ExpSum(ΔRating)$_{i,t-1}$ only ranks 9$^{\text{th}}$ when all stock are used (Panel (a)), Panel (b) shows that it is the strongest predictor when excluding microcap stocks. Further, this finding is not sensitive to the decay parameter used. On a multivariate basis, when excluding microcaps, ExpSum(ΔRating)$_{i,t-1}$ is the third most powerful predictor following the maximum daily return characteristic in Bali, Cakici, and Whitelaw (2011) and the net operating assets characteristic in Hirshleifer, Hou, Teoh, and Zhang (2004).

It is worth emphasizing that the rating-induced predictability is, unlike other predictors, more powerful in large cap stocks. This fact is consistent with the fact that mutual funds primarily hold large cap stocks. When excluding microcap stocks, ExpSum(ΔRating) ranks first among all predictors on a univariate basis and third on a multivariate basis. The predictor is slightly less powerful when including microcaps, but the degree of return predictability is nonetheless stronger than all the well-known characteristics: size, book-to-market, mo-

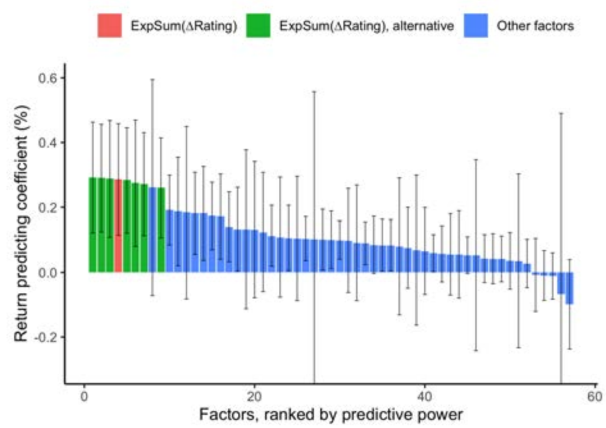
Figure 2. Horse Race Between Rating and Other Stock Return Predictors

We estimate return predictability of ExpSum$(\Delta\text{Rating})_{i,t-1}$ and the other 49 return predictor characteristics using Fama-MacBeth regressions and plot the coefficients. To make coefficients comparable, all predictors are transformed into Z-scores. We switch the signs, if necessary, so that the predictors are expected to predict returns positively based on the original studies. To isolate the marginal predictive power of ratings, all regressions control for the the past 36 month cumulative fund returns aggregated at the stock level. Panels (a) and (b) are results from univariate regressions and Panels (c) and (d) are multivariate regressions where all predictors—one version of ExpSum$(\Delta\text{Rating})_{i,t-1}$ and the other 49 characteristics—are included in the regression. The left panels use all stocks and the right panels use stocks that are above the 20% percentile of NYSE stock market cap. Red bars represent the predictability of ExpSum$(\Delta\text{Rating})_{i,t-1}$. Green bars represent alternative constructions of ExpSum$(\Delta\text{Rating})_{i,t-1}$ with decay parameter $\delta = 0.3$, 0.4, ..., 0.9, which translates to half-lives of 0.58, 0.76, ..., 6.58 months. Blue bars represent the other 49 characteristics (factors) discussed in Section 2.2. The whiskers represent two standard error bands; the standard errors are calculated using the Newey-West procedure with 12 lags.

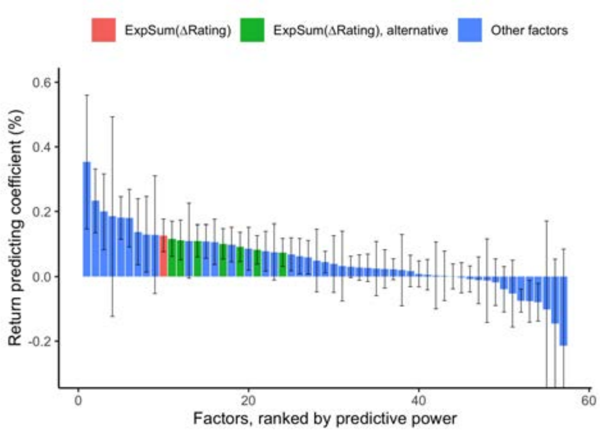
(a) Univariate, all stocks



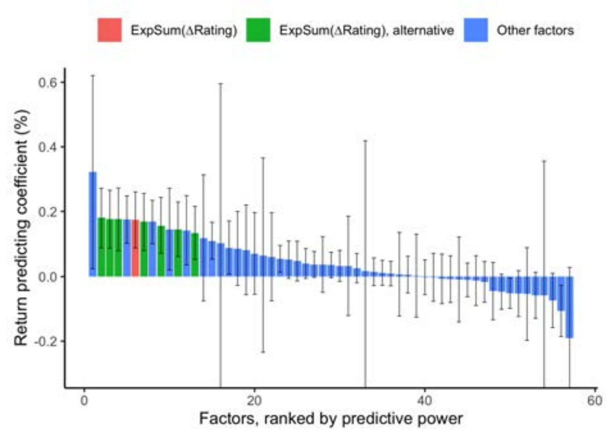
(b) Univariate, no microcap



(c) Multivariate, all stocks



(d) Multivariate, no microcap



mentum, profitability, and investment (asset growth).

3.2 2002 Rating Methodology Reform

We now describe the Morningstar rating methodology reform in June 2002. We will use this reform to further identify the causal effect of rating changes on factor returns. In Sections 4 and 5, we will also study the long-term impact of this rating reform on factor profitability.

After introducing its mutual fund rating system in 1985, Morningstar quickly became the industry leader in providing independent mutual fund ratings. 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 look-back horizon for past performance can be up to ten years, but more weight is applied to more recent periods.[13] 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 A.3.

By 2001, Morningstar received complaints that its ratings failed to correctly measure fund manager skill. 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 were highly dependent on style performance. Following the dotcom crash, many fund managers with heavy technology sector exposure complained that their ratings dropped merely because the sector crashed.

As a result, the research team at Morningstar, spearheaded by the economist Dr. Paul Kaplan, redesigned the rating system.[14] After June 2002, fund ratings were no longer based on how each fund ranked against *all* U.S. equity funds but only on fund rankings *within*

[13]For funds with over 10 years of history, Morningstar computes 3-year, 5-year, and 10-year past returns and combines them. The weights of the three horizons are set at 20%, 30%, and 50%, respectively. Because the three horizons are overlapping, however, the recent years are effectively given much more weight than more distant history.

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

style categories. For diverse U.S. equity funds (87% of all mutual funds in 2002), the style categories are the well-known 3×3 size-value matrix.[15] The change in methodology was announced in February 2002 and was first implemented in Morningstar's monthly ranking of funds at the end of June 2002.

This seemingly innocuous 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 3. For instance, at the height of the dotcom boom in 2000, large-cap growth funds had an average rating of 4 stars, while small-cap value funds only had 1.9 stars. After the change, ratings became uncorrelated with past style performance, and the rating imbalance across styles became negligible. Consistent with flows chasing ratings, Panel (b) shows that style-level fund flow dispersion also became much smaller after the change.

Importantly for our identification purposes, investors continued to chase ratings in a similar manner before and after the reform. To examine this, Panel (c) of Figure 3 simply plots the average monthly mutual fund flows by Morningstar ratings for each year.[16] Throughout our sample period, 5-star funds receive flows that amount to +2.5% of their AUM per month, while 1-star funds experience outflows amounting to −1.5% of their AUM per month. This is a large difference: relative to one-star funds, a five-star fund would *double* its size in approximately one and a half years. The relation between ratings and flows remained remarkably similar throughout the sample period.

In a more formal test, in Panel (d), we estimate the response of fund flows to lagged fund ratings using TNA-weighted Fama-MacBeth regressions that control for 36 lagged months of fund returns. We use three-year rolling windows to examine the time-variation. The coefficient estimate varies only slightly over the sample and has no sharp change around or following the 2002 reform. For example, the average flow-to-rating response was 0.37%

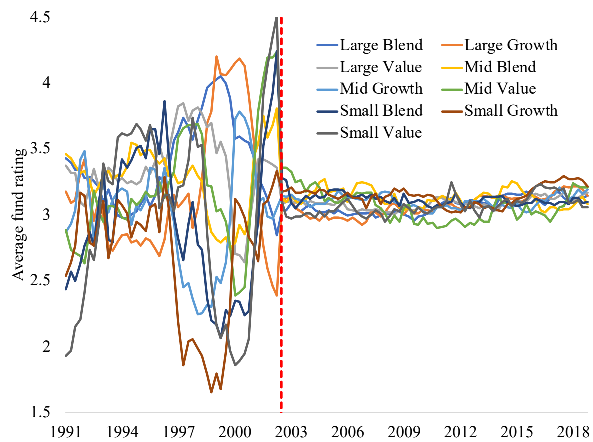
[15] Sector funds—the remaining 13%—were classified into 12 sectors (e.g., financials, utilities).

[16] To sharpen the comparison around the reform, each year is defined to end in June. Therefore, the data point for year 2002 covers the period of July 2001 to June 2002.

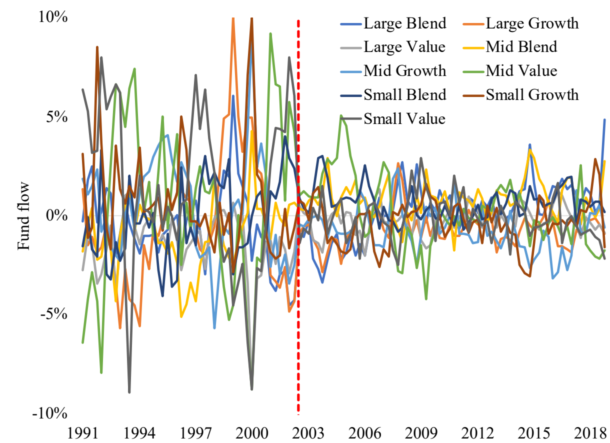
Figure 3. The June 2002 Morningstar Methodology Change

Panel (a) and (b) plot the quarterly average mutual fund ratings and TNA-weighted average fund flows by the 3×3 size-value Morningstar styles. Panel (c) plots the average monthly fund flow by Morningstar rating in each year. In both Panels (b) and (c), fund flows are demeaned by period to focus on the cross-sectional variation. Panel (d) explores the stability of the relation between ratings and flows at the fund level. Specifically, it plots the regression coefficient of fund flows on lagged ratings estimated using three-year rolling windows. Because the regression controls for 36 lags of monthly fund returns, and because it takes three years to compute a rolling average, the graph starts in 1996. (Raw monthly data are available from 1991.) The shaded area indicates the two standard error bands. In all panels, the red dashed vertical line marks the June 2002 Methodology change event.

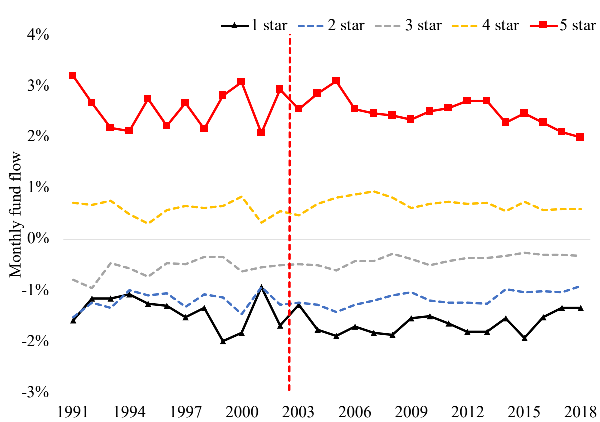
(a) Fund ratings, by style



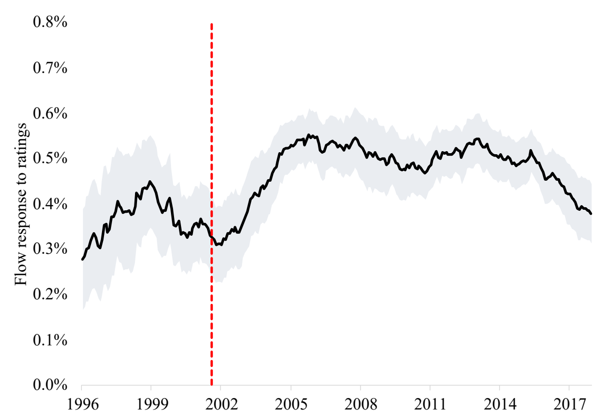
(b) Fund flows, by style



(c) Fund flow by Morningstar rating



(d) Regression coefficient of fund flows on ratings



before June 2002 and 0.48% after June 2002.[17] Therefore, we conclude that mutual fund investors did not change their rating-chasing behavior in response to the reform.

[17] See further analysis indicating that investors did not change their rating-chasing behavior in Ben-David, Li, Rossi, and Song (2019) and Evans and Sun (2021).

3.3 Factor-level Event Study in 2002

Can ratings exert a tangible impact on factor-level returns? To examine this, we conduct an event study using a one-year window (January to December 2002) around the rating reform. There are two benefits to using a short window. First, rating changes in this period are predominantly caused by the rating methodology change. Second, using a short window also reduces the chance that factor returns are impacted by other events such as the NYSE decimalization in early 2001 and the introduction of NYSE auto quoting in 2003 (Hendershott, Jones, and Menkveld, 2011).

As expected, style-level rating, flow, and return changed significantly around the reform, as shown in Panel A of Figure 4. Ratings (and thus flows) were correlated at the style-level before the reform but became evenly spread out afterwards. In the six months before the reform, small-value funds have ratings that are on average 2.07 stars higher than large-growth funds. Correspondingly, they also have higher inflows, and stocks in the small-value style also performed well. In the six months after the reform, ratings, fund flows, and returns became relatively balanced.

How would the reform impact factors? If factors were 'neutral' with respect to Morningstar's style box, then there would be no reason to suspect that the reform would affect them. However, this is not the case for most factors. As an approximate measure of the style exposure of factors, in Panel B of Figure 4, we plot the Fama-French SMB and HML loadings of the factors. Right before the event, some factors such as size, illiquidity, and sales-to-price load onto small and value styles (have high SMB and HML loadings). Those factors, therefore, will experience a sudden rating drop in June 2002. In contrast, factors such as Ohlson's O-score, sales growth, and cash-based profitability load onto large and growth styles (have low SMB and HML loadings), so they are expected to experience a rating increase at the reform event.

We now examine how the reform affected factors in 2002. We sort factors into quintiles based on how their rating is affected by the reform event. To alleviate endogeneity concerns,

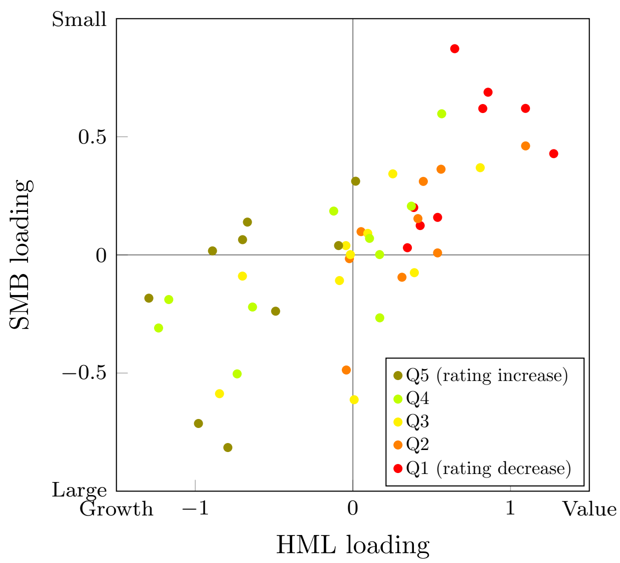
Figure 4. Style-Level Changes Around the June 2002 Event

Panel A shows the average rating, monthly flow, and monthly return of funds by 3×3 fund styles during the 12 months before and after the rating reform event (January–June 2002 and July–December 2002). All variables are demeaned by period to focus on the cross-sectional difference across styles. Larger values are colored in red, and smaller values are colored in green. Panel B plots the Fama-French SMB and HML (size and value) factor loadings of the 49 factors in this study. The loadings are estimated using time-series regressions of daily returns during the six months before the reform date. The factors are grouped into quintiles based on predicted rating changes at the reform event using data in December 2001.

Panel A: Rating, Flow, and Return of 3×3 Mutual Fund Styles Around the Reform

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

Panel B: Factor Style Exposures



we sort the factors using the *predicted* reform-induced rating change computed using data in December 2001, which is *prior to* the event study window. Specifically, using data available up to December 2001, we estimate mutual fund ratings by following the pre-2002 and post-

2002 Morningstar rating methodologies, and then aggregate these ratings up at the factor level. We then predict that each factor f will experience a rating change of

$$\text{PredictedChange}_f = \widehat{\text{Rating}}_{f,\text{Dec 2001}}^{\text{post-2002 methodology}} - \widehat{\text{Rating}}_{f,\text{Dec 2001}}^{\text{pre-2002 methodology}}, \tag{10}$$

where the two terms on the right hand side represent estimated factor-level rating under the two different rating methodologies, respectively.[18] Appendix B.3 explains the prediction process in more detail and verifies that the predictions can accurately forecasting actual factor rating changes at the reform event.

Figure 5 plots what happened to the factors in 2002. Panel (a) plots average ratings of factors and shows a sharp methodology-induced change exactly at the event. Factors in quintile 1 suffer a drop of 0.43 stars, while those in quintile 5 experience a small increase of 0.19 stars. Panels (c) and (e) plot cumulatively monthly factor FIT and returns around the event, respectively. Quintile 1—the factors that benefited from ratings pre-event but suffered post-event—experienced a decline of 1% in monthly FIT and a sharp decline of -3.7% in monthly returns. At the same time, quintile 5 experienced an increase of 0.14% in monthly FIT and a slight increase of 0.75% in monthly returns.[19]

To alleviate the concern that the return and FIT changes could result from other reasons, we also conduct the same exercise in all years other than 2002. The results on rating, FIT, and return changes in other years are shown as the white bars in Panels (b), (d), and (f) with two standard error bands. These panels show that the large change around June is unique to 2002.

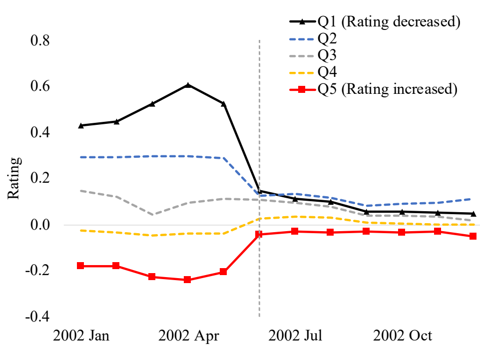
[18] We estimate the pre-2002 ratings under the old methodology, instead of using the actual pre-2002 ratings, to reduce estimation errors. Because we do not have exactly the same data set that Morningstar uses internally, our rating estimation contain errors. However, the same data-induced error is present in both terms in Equation (10), so we are able to difference it out.

[19] In a companion paper, we show that the implied style-level price impact coefficient (the reciprocal of demand elasticity) is approximately 5 (Ben-David, Li, Rossi, and Song, 2020). That is, buying 1% of the market cap outstanding creates a price impact of approximately 5%. This magnitude is consistent with the existing literature that estimates the price impact of undiversifiable demand shocks (e.g., Gabaix and Koijen, 2020).

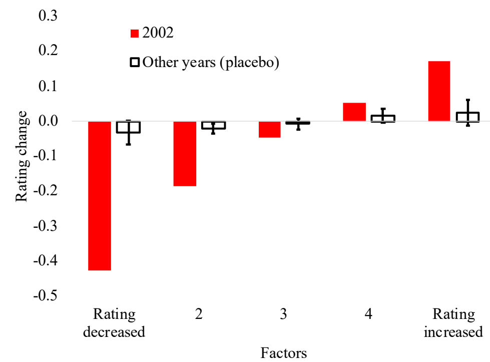
Figure 5. Stock Factors around the June 2002 Event

We perform event studies on the 49 factors using a 12-month window around the reform event (January to December 2002). In the left panels, we sort factors by their *predicted* reform-induced rating change into quintiles, and then plot the evolution of their ratings in Panel (a), cumulative fund flow–induced trading (FIT) in Panel (c), and cumulative returns in Panel (e). To alleviate endogeneity concerns, the rating change prediction only uses data up to December 2001 (prior to the event window). The dashed vertical line is the June 2002 reform event. The right panels conduct the same exercises in years other than 2002 as a placebo test. The red bars plot the average rating, flow-induced trading (FIT), and return changes after June (the average of July to December 2002 minus the average of January to June 2002), while the white bars plot the corresponding results for years other than 2002. The whiskers represent two standard error bands. To focus on cross-sectional dispersion, all variables—ratings, returns, and flows—are demeaned by period.

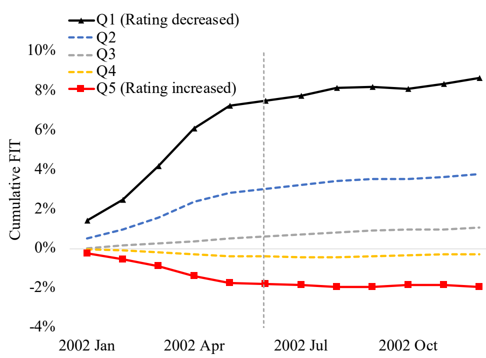
(a) Rating in 2002



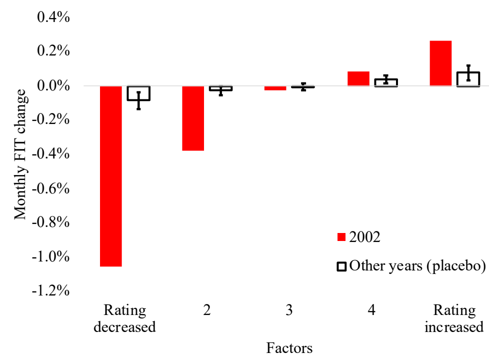
(b) Rating change, 2002 vs placebo



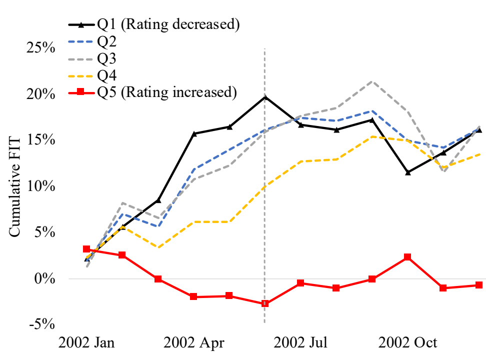
(c) Cumulative FIT in 2002



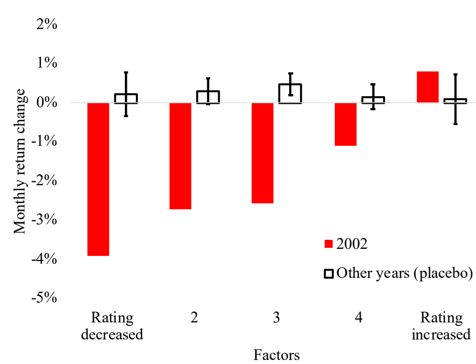
(d) FIT change, 2002 vs placebo



(e) Cumulative return in 2002



(f) Return change, 2002 vs placebo



3.4 Alternative Hypotheses to the Event Study Results

In this subsection, we investigate the concern that the factor price fluctuations around June 2002 may be triggered by changes other than the Morningstar reform.

Arbitrage activity. One natural worry is whether arbitrage forces in these factors have suddenly become stronger in mid 2002. A number of papers present evidence that factor profitability is related to arbitrage activity. For instance, Hanson and Sunderam (2013) argue that value and momentum strategy profits decrease when more capital is devoted to them. McLean and Pontiff (2016) show that factor profitability declines after the strategies were published in academic papers and link it to arbitrage actions. Relatedly, Lou and Polk (2018) show that a return-based measure of arbitrageur activity negatively predict momentum profits.

Did arbitrage activity change in June 2002? We use two measures proposed in the literature to proxy for arbitrage activity in factors. First, we follow Chen, Da, and Huang (2019) to construct a net arbitrage activity (NAT) measure. For each stock, the authors measure the long position of arbitrageurs using aggregate 13F holdings of hedge funds and the short position using aggregate short interest from Compustat.[20] The authors combine the long and short positions into a net position, and subtract the past four-quarter average to arrive at a measure of arbitrageur position changes, which they call NAT. We follow them to compute stock-level NAT and aggregate it at the factor level.

Second, we follow Lou and Polk (2018) to construct a correlation-based measure of arbitrage activity. These authors measure arbitrage activity in the momentum strategy by estimating excess return correlation within the long and short portfolios, which can be gen-

[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. It is worth noting that, while the short side of NAT is updated monthly, the long side relies on 13F holdings and is only updated quarterly.

erated by arbitrageurs trading in the factor.[21] We also compute this measure for all factors.[22]

We plot the evolution of these measures in the 12 month event window in Figure 6. As in Section 3.3, we sort factors into quintiles by their predicted rating change using data up to December 2001. Panel (a) plots the NAT measure, and Panel (b) plots the correlation-based measure. There is no noticeable change in either measure during the event window.

Changes in liquidity. One may also hypothesize that stock market liquidity increased dramatically in June 2002.[23] To examine this possibility, we aggregate the stock-level Corwin and Schultz (2012) bid-ask spread measure for the factors (averaging over the long and short legs) during this period. The results, plotted in Panel (c), show no evidence that liquidity changes account for our findings. Panel (d) shows that monthly trading turnover also had no clear change around the event.

In summary, we do not find any noticeable change in arbitrage trading activity or liquidity — two major forces that could impact factor returns — around June 2002. Thus, this supports the idea that the factor return "kink" in June 2002 is primarily driven by the Morningstar methodology reform.

[21] Specifically, in any given month, they use the previous 52 weeks of data to compute a "comomentum" measure:

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

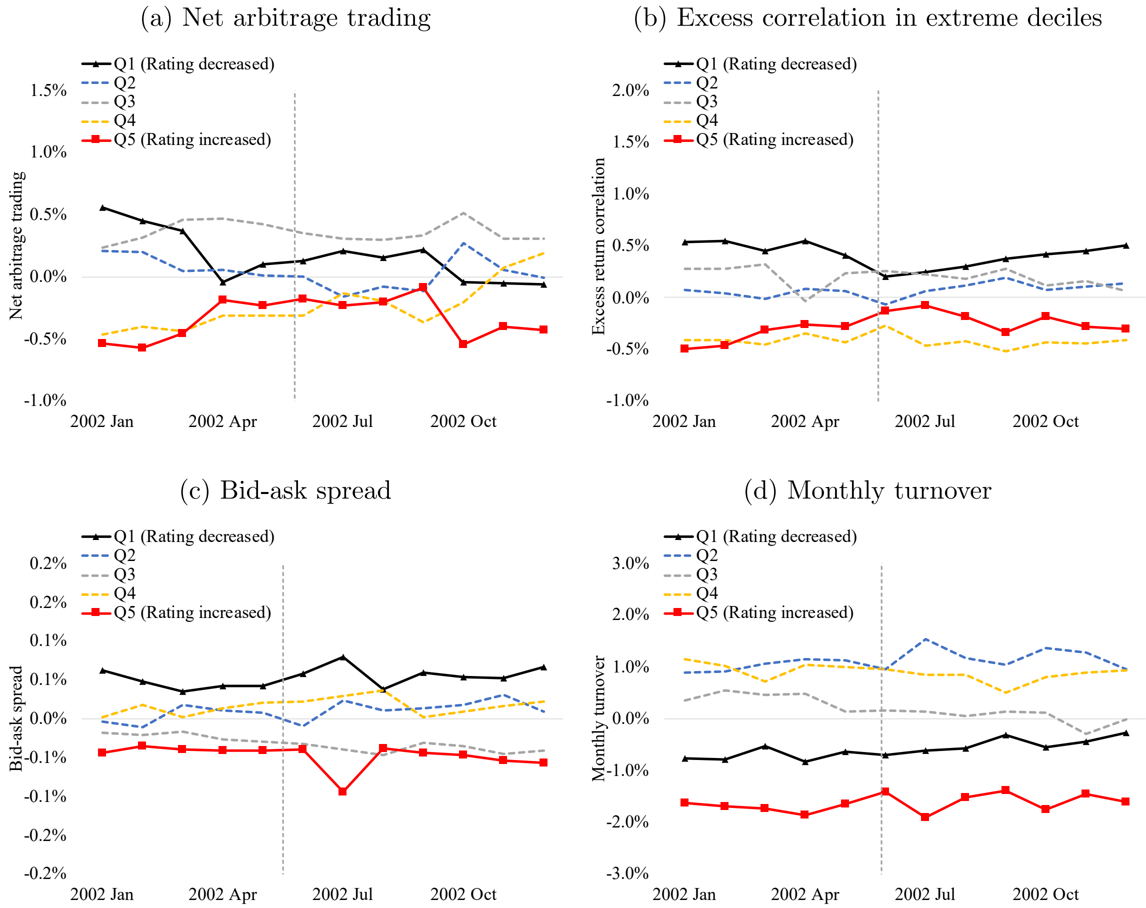
where N^L and N^S are the number of stocks in the long and short leg portfolios, respectively. To compute the partial return correlations, they first subtract Fama-French 30 industry returns from weekly stock returns and then regress the residuals on the Fama-French three factors to obtain alphas. Finally,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

[22] As a sanity check on our replication of their methodology, consistent with Lou and Polk (2018), we find that this measure indeed negatively predicts returns of factors in the momentum category.

[23] Increasing liquidity may explain factor profitability declines through two possibility mechanisms. First, if a factor's profitability comes from demand price pressures, then increasing liquidity will reduce the price impact of such demand shocks. Second, if factor profitability is the result of arbitrageurs not being able to arbitrage away profits, then increasing liquidity may facilitate arbitrage effectiveness and thus reduce residual factor profitability. Of course, the asset pricing literature has also found evidence that illiquidity is a priced risk, so the changes may also come from changes in equilibrium-required rates of return (Amihud, 2002; Pástor and Stambaugh, 2003; Acharya and Pedersen, 2005).

Figure 6. Alternative Explanations: other Influences Around 2002

As in Figure 5, factors are sorted into quintiles by the predicted rating change using data in December 2001. Thus, quintile 1 (or 5) factors are those predicted to experience the largest rating decrease (increase) at the reform event. Panel (a) plots the net arbitrage trading measure in Chen et al. (2019). Panel (b) plots excess return correlation in extreme factor quintiles, a measure of arbitrage activity developed in Lou and Polk (2018). Panel (c) plots average bid–ask spread, measured following Corwin and Schultz (2012), of the long and short factor legs. Panel (d) plots the average monthly trading turnover of the long and short factor legs. To focus on cross-sectional dispersion, all variables are demeaned by month. In all panels, the vertical dashed line marks the methodology change event.



4 The Disruption of Positive-Feedback Trading and Long-run Factor Profitability

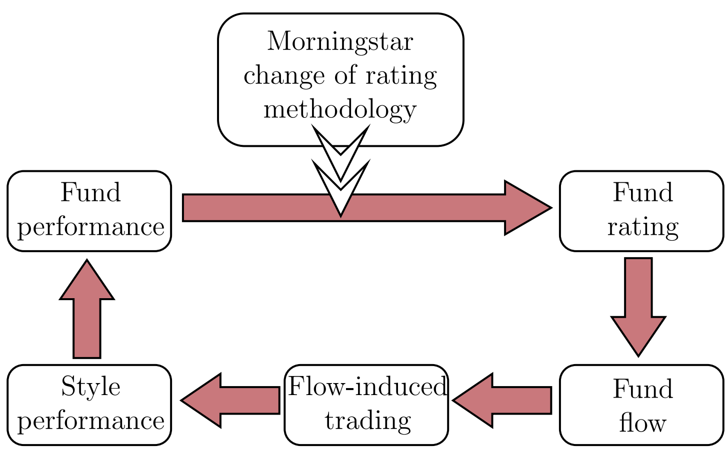
Having established that Morningstar ratings *can* impact factor returns, we now investigate how the 2002 reform influenced long-term factor profitability. We first demonstrate that the 2002 reform disrupted style-level positive feedback trading. Based on this disruption, we conjecture that the reform should reduce the profitability of momentum-type factors which benefit from positive feedback trading. Consistent with this conjecture, these factors indeed suffered the largest profitability decline after June 2002.

4.1 Disruption of Positive-Feedback Trading at the Style Level

The pre-2002 rating methodology generates a positive feedback loop at the style-level. As illustrated in Figure 7, funds in styles that performed well in the recent history get high ratings and attract inflows. They use the new flows to increase their investments in the same style of stocks, so the price of those stocks are pushed up even further. The mechanism also works in the other direction: Funds in underperforming styles experience correlated outflows, resulting in downward price pressure on stocks associated with these styles. The post-June 2002 rating methodology, however, should cause a sudden disruption in this rating-induced positive feedback trading.

Figure 7. Style-Level Rating-Induced Positive Feedback Trading

This flow chart illustrates how the pre-2002 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 inflows (outflows). Third, fund managers buy (sell) fund holdings in response to flows, leading to further stock price pressures.



We confirm the disruption of style-level positive-feedback trading in Figure 8. Specifically, we sort the 3×3 Morningstar fund styles based on past-12-month returns—the typical look-back horizon used in studying momentum. Before June 2002, funds in styles that recently performed well received higher average ratings and higher fund flows. The magnitudes are also large. Panel (a) shows that the average rating spread between funds in the top and bottom styles was about 0.8 stars before 2002 and shrank to almost zero after June 2002.[24] Because rating attracts flows, Panel (b) shows that funds in the top style, on average, received about 1.7% higher flows per month than the bottom style before June 2002, and that difference dropped to around 0.4% after June 2002.

This disruption also has an impact on style returns. In Panel (c), we plot the TNA-weighted style-level fund returns. The top-ranked style exhibit approximately 0.8% higher monthly return than the bottom-ranked style before 2002, and that difference disappeared after 2002, consistent with a reduction of positive-feedback trading after the reform. For robustness, Panel (d) shows similar pattern when using CAPM alpha. The post-2002 change in performance spread (the top style minus bottom style) is statistically significant at the 5% level.[25]

These findings lead us to conjecture that the Morningstar 2002 reform had a negative impact on momentum-type factors. In the 49 factors we study, 5 are classified into the momentum category by Hou et al. (2019): $(t-1,\ t-12)$ and $(t-2,\ t-6)$ momentum (Jegadeesh and Titman, 1993), industry momentum (Moskowitz and Grinblatt, 1999), fifty-two week high (George and Hwang, 2004), and $(t-7,\ t-12)$ intermediate momentum (Novy-Marx, 2012). We predict that they should suffer large profitability declines after June 2002.

4.2 Factor Profitability Decline after June 2002

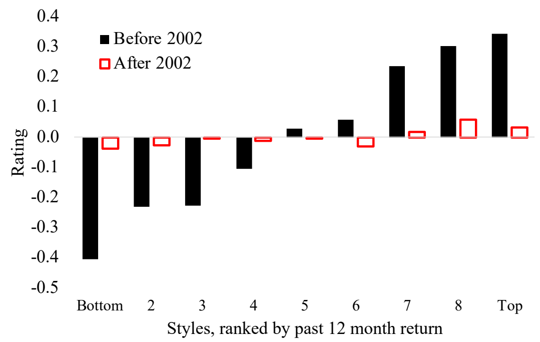
We now examine whether momentum-type factors indeed suffered larger declines in ratings and return profitability after June 2002. It is worth clarifying that there are many other

[24]The data in those graphs are demeaned within month to focus on cross-sectional patterns across styles.
[25]To alleviate the concern that fund returns may also be influenced by transaction costs and fees, we also repeated this exercise using the returns of the stocks held by the funds, rather than the fund returns. The results are unaffected.

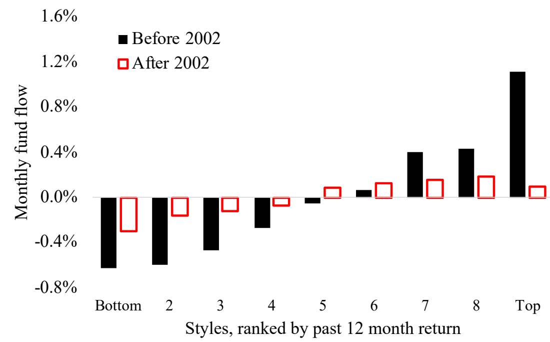
Figure 8. Style-level Positive Feedback Trading, Before and After 2002

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

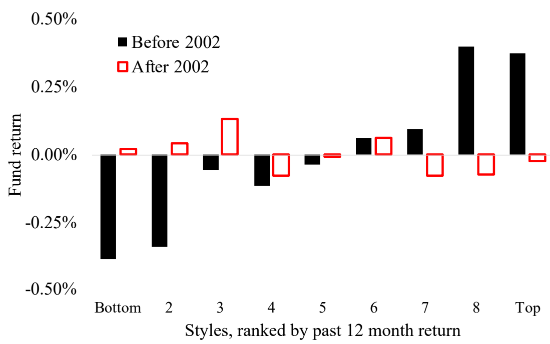
(a) Rating



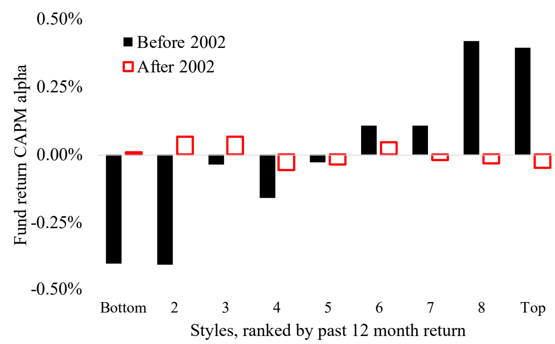
(b) Fund flow



(c) Fund return



(d) CAPM alpha of fund return

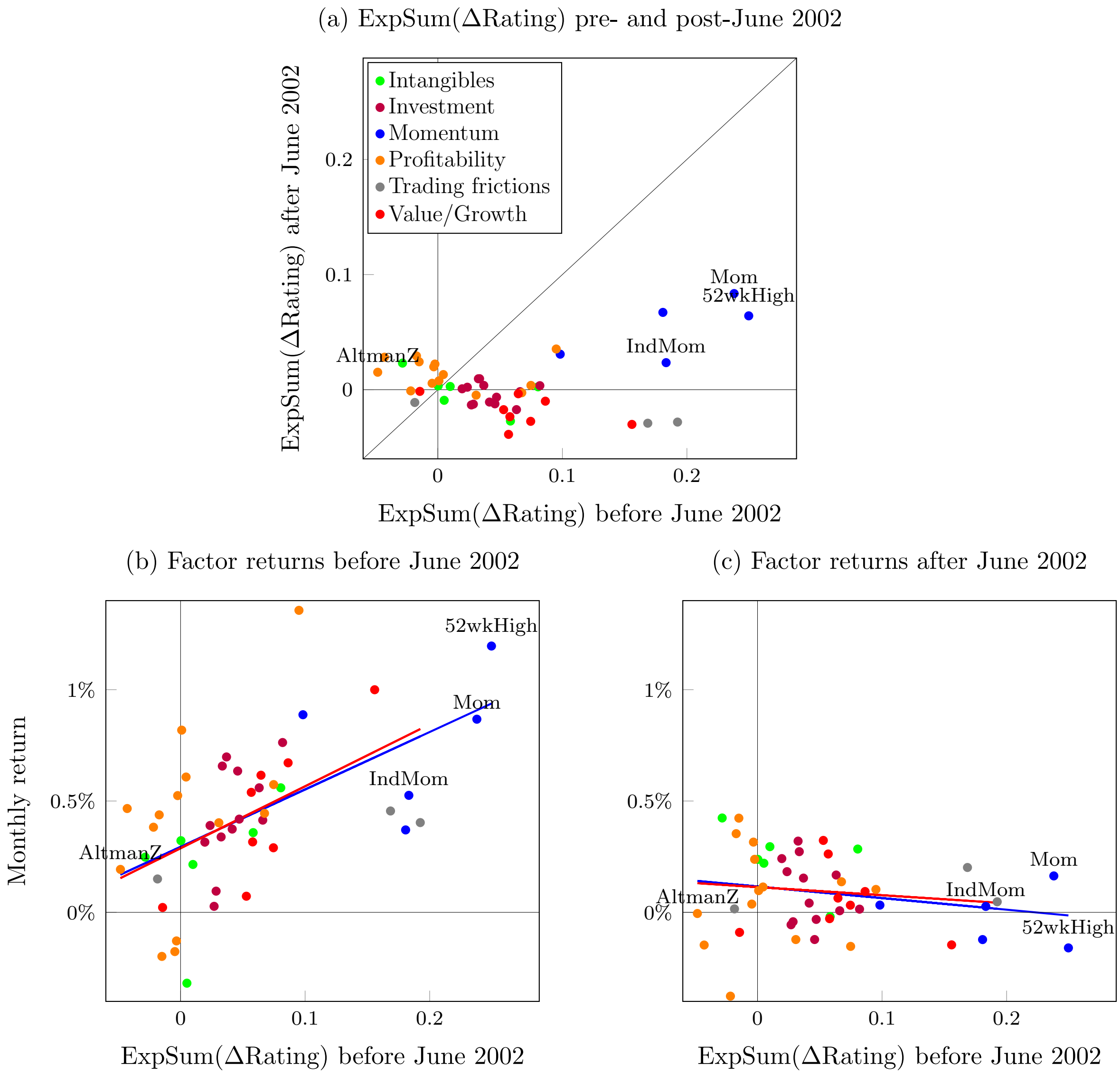


reasons, such as increasing arbitrage activity and liquidity, that could cause secular declines in average factor profitability. However, we are interested in explaining the *cross-sectional difference* of decline across factors.

To measure the impact of recent rating changes on factors, we follow Section 3 to use the exponentially-weighted specification. In Panel (a) of Figure 9, we plot each factor's average post-2002 $\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$ against the pre-2002 values over the sample of 1991 to 2018. We mark factors from different categories using different colors. Clearly, before June 2002, Morningstar served as an important tailwind for many factors, especially those in the momentum categories (colored blue). After June 2002, the $\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$ across factors shrunk. This plot is consistent with our conjecture that momentum-category factors were most positively affected by rating-induced positive-feedback trading before 2002 and suffered the sharpest drop after June 2002.

Figure 9. Factor Ratings and Profitability before versus after June 2002

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



To visualize the cross-sectional differences across factors, in Panels (b) and (c), we plot the pre- and post-2002 average factor return against the pre-2002 ExpSum$(\Delta\text{Rating})_{f,t-1}$. Consistent with our rating-induced price pressure hypothesis, factors that benefit from pre-2002 ratings experienced high returns before the reform but not afterwards. For instance, the profits of the momentum factor was almost 1% each month but became negligible after June 2002. Other momentum-type factors, such as the 52-week-high factor, suffered similar declines in profitability. In Appendix B.4, we provide further details about how momentum-type factors suffered declines in rating, flows, and returns after 2002.

4.3 Placebo Test: Momentum in Other Countries

Because the Morningstar methodology change happened only in the U.S., and because Morningstar is less influential in other countries, we use non-US momentum factor returns in a placebo test of our Morningstar-based explanation.[26] We expect the post-2002 profitability decline to be more pronounced for U.S.-based momentum.[27]

We use the monthly momentum factors for different countries from Ken French's website. To keep factor construction consistent, we also use his construction of the U.S. momentum factor when comparing with momentum factors in Europe, Japan, and Asia Pacific countries other than Japan (APAC ex Japan). His factor construction methodology slightly differs from ours, but the difference is small: the monthly correlation between his and our U.S. momentum factor is 96%.[28] The results are reported in Table 3. Consistent with our hypothesis, only the U.S.-based momentum strategy experienced a large decline in profitability after 2002. In contrast, momentum profits were strong both before and after 2002 in Europe and (to some extent) APAC ex Japan. Momentum profits were low both before and after

[26] Appendix 2 of Morningstar (2016) lists all the historical major Morningstar rating methodology changes. The June 2002 change is unique to the U.S. market.

[27] Asness, Moskowitz, and Pedersen (2013) suggest that arbitrage activity creates spillover across momentum in different markets, leading to higher correlation of momentum profits across markets. To the extent that such spillover is not perfect, however, we still expect to detect a difference between the U.S. and international momentum strategies.

[28] Specifically, the Fama-French construction forms 2×3 size-prior return independent sorts and defines the momentum factor as 1/2 (Small High + Big High) - 1/2 (Small Low + Big Low).

2002 in Japan.[29]

Overall, this section shows that the Morningstar reform — through disrupting style-level positive feedback trading — had the largest negative impact on momentum-type factors. Consistent with the fact that the change only happened in the U.S., we do not see similar return drops in momentum strategies in other markets.

Table 3. Momentum Profitability Decline: The U.S. versus Other Countries

The table reports monthly average return of momentum strategies in different countries from January 1991 to December 2018. The momentum strategy returns are downloaded from Ken French's website. The standard errors are reported in the parenthesis, and returns statistically significant at the 10%, 5%, and 1% levels are denoted with ***, **, and * respectively.

	Monthly momentum return			
	U.S.	Europe	APAC ex Japan	Japan
Before June 2002	1.13%***	1.11%***	0.75%	0.13%
	(0.42%)	(0.33%)	(0.46%)	(0.46%)
After June 2002	0.03%	0.75%***	0.83%***	0.11%
	(0.32%)	(0.28%)	(0.24%)	(0.25%)
After − Before	−1.10%**	−0.36%	0.08%	−0.02%
	(0.53%)	(0.43%)	(0.52%)	(0.52%)

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

5 Quantifying the Effect of the Morningstar Reform

In this section, we quantify the effect of the 2002 reform on factor profitability changes. Like all exercises on counterfactual asset prices, our methodology relies on functional form assumptions and thus should be taken as suggestive.

Following the specification in Section 3.1, we estimate the post-2002 factor profitability change explained by Morningstar as:

$$\lambda \cdot \left(\overline{\text{ExpSum}(\Delta\text{Rating})}_{f,\text{after 2002}} - \overline{\text{ExpSum}(\Delta\text{Rating})}_{f,\text{before 2002}} \right), \tag{11}$$

[29]That momentum strategy return is weak in Japan is a known result (Asness et al., 2013).

where the second term is the change in the average $\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$ before and after 2002. Since ExpSum_{t-1} (the exponentially weighted sum of recent rating changes) can be directly measured, we only need to obtain a well-identified estimate of the price impact parameter λ: the impact on monthly factor return of each star rating change.

5.1 Estimating the Price Impact Parameter

To get a well-identified estimate of λ, we once again make use of the 2002 reform event. Specifically, we estimate a panel regression on factors using the 12-month window around the event:

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

where the control $X_{f,t-1}$ includes factor returns over $t-1$, $t-2$ to $t-6$, and $t-7$ to $t-12$ months as well as factor- and time-fixed effects. These controls are motivated by the recent finding that factor returns exhibit momentum (Gupta and Kelly, 2019; Arnott et al., 2019a). As discussed in Section 3.3, using a short window ensures that we primarily make use of reform-induced rating variation, and thus alleviates endogeneity concerns. To account for the cross-sectional factor return correlation, we adjust the standard errors using a feasible generalized least squares (FGLS) approach.[30]

[30] We use the full sample of factor returns to estimate the covariance matrix C of factor returns and incorporate C into the estimation. Specifically, let y be the vector of factor returns stacked together so that the first 49 entries are the first month, the next 49 entries are the second month, and so forth. Then, we estimate the covariance matrix of y to be

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

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

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

The estimation results are shown in Table 4. For each star rating change, the factor-level price impact is 2.27%, with a t-statistic of 4.28. The result is both statistically and economically significant. Columns (2) to (4) show that the estimates are robust to whether time and factor fixed effects are included.[31]

Table 4. Estimating Price Impact Coefficient (λ) Around the June 2002 Event

We use a panel regression to estimate the predictive relationship between monthly factor returns and the exponentially summed lagged ratings ($\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$). The sample period is the 12 months around the reform (January to December 2002). We control for lagged factor returns in months $t-1$, $t-6$ to $t-2$, and $t-12$ to $t-7$. The four specifications differ in whether factor and month fixed effects are included. The standard errors in parentheses are adjusted for the cross-sectional correlation between factor returns using feasible generalized least squares.

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

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

While the λ estimated based on the event achieves better identification, we also estimate a time-varying λ using five-year windows centered in each year to account for potential non-stationarity.[32] Then, we estimate the amount of post-2002 profitability declined explained by ratings to be:

$$\frac{1}{N_{\text{after}}} \sum_{t \text{ after } 2002.06} \lambda_t \cdot \text{ExpSum}(\Delta\text{Rating})_{f,t-1} - \frac{1}{N_{\text{before}}} \sum_{t \text{ before } 2002.06} \lambda_t \cdot \text{ExpSum}(\Delta\text{Rating})_{f,t-1} \tag{13}$$

[31] As expected in Section 3.3, the implied price impact coefficient in the 2002 event is inline with the estimates obtained by other studies that focus on undiversifiable demand shocks.

[32] That is, for each year s, we use years $(s-2, s-1, s, s+1, s+2)$ to estimate λ using Regression (12). We the apply it to all months in year s in Equation (13).

5.2 Quantification Results

We now apply the estimated λ coefficient to quantify the factor profitability decline that can be explained by Morningstar. Table 5 shows that, after 2002, the momentum factor and other momentum-category factors experienced a drop of monthly return of 70.4 to 80.1 bps, while the other factors experienced a smaller decline averaging 29.4 bps per month. For the momentum factor, quantification based on the two price impact estimates suggests that Morningstar ratings can explain a range of 17.8 to 35 bps of the decline. This amount to $0.178/0.704 \approx 25.3\%$ to $0.350/0.704 \approx 49.7\%$ of the overall decline. For other momentum category factors, ratings can explain 14.2 to 29.9 bps, which is also a sizeable fraction of the decline.

Table 5. Estimated Explanatory Power on Post-2002 Factor Return Decline
We present the change of average monthly factor returns (in percent) from the pre-change period (January 1991 to June 2002) to the post-change period (July 2002 to December 2018). The first row reports in change in monthly return. The next two rows present the estimated component explained by Morningstar rating changes, calculated by multiplying the price impact parameter λ with changes in average $\text{ExpSum}(\Delta\text{Rating})_{f,t-1}$ before and after 2002. The second row uses price impact parameters $\hat{\lambda}$ estimated using using 5-year rolling windows, and the third row uses the $\hat{\lambda}$ estimated using the 12 months around the June 2002 event. Column (1) examines the momentum factor; Column (2) examines other factors in the momentum category; and Column (3) examines all the other factors.

		Factors			Difference	
	Methodology	Momentum Factor	Other Mom-type Factors	Other Factors	(1) − (3)	(2) − (3)
		(1)	(2)	(3)	(4)	(5)
Return change		−0.704	−0.801	−0.294	−0.410	−0.507
Explained	Time-varying λ	−0.178	−0.142	−0.053	−0.125	−0.090
	Event-based λ	−0.350	−0.299	−0.090	−0.261	−0.209

Because other reasons can also contribute to the decline in factor returns over time, our main focus is on comparing the explanatory power on momentum-related factors versus other factors. Column (3) shows that, among all the other factors, Morningstar effects can only explain 5.3 to 9.0 bps of the post-2002 return decline. Relative to those other factors, momentum suffered an additional 41 bps decline in monthly returns, and we estimate that

Morningstar can explain $0.125/0.410 \approx 30.5\%$ to $0.261/0.410 \approx 63.7\%$ of this difference (Column (4)). The last column shows a roughly similar (albeit slightly smaller in magnitude) conclusion when examining other momentum-type factors.

Overall, using a simple quantification approach, we find that the 2002 Morningstar reform can explain approximately one quarter to half of the post-2002 profitability decline of momentum-type factors.

6 Conclusion

Return factors are perhaps the most researched topic in empirical asset pricing. In recent decades, factor investing also became increasingly popular in the investment industry under the name of "smart best investing." However, factor profitability has declined over time. Adding to the puzzle, many factors suffered a sharp and persistent profitability drop starting in mid 2002.

This paper finds that a significant part of the factor profitability decline stems from a seemingly innocuous change in Morningstar's rating methodology. Before June 2002, Morningstar rated funds using their past performance ranking relative to U.S. equity funds. As a consequence, funds pursuing investment strategies associated with recently outperforming styles were rated higher than funds in recently underperforming styles. Investors chasing fund ratings led to significant style-level positive-feedback trading. After the reform, Morningstar rated funds using their past performance ranking against their 3×3 size-value style peers, causing an immediate halt to this positive-feedback trading. Because many factors—and especially the momentum-type ones—benefited from the earlier positive-feedback trading, this halt caused a disruption to their profitability. We estimate that the Morningstar rating reform accounts for approximately 25% to 50% of the post-2002 decline of momentum profits.

Traditionally, researchers believe that factor return predictability arise from compensation for economic risks. However, our finding suggests that a sizeable portion of the profits of momentum-type factors before June 2002 resulted from rating-induced price pressures, which is consistent with recent findings that demand pressures can explain broad market-wide price movements (Gabaix and Koijen, 2020; Li, 2020). For better identification, our paper focuses closely on demand induced by Morningstar ratings. However, it is possible that role of correlated demand, arising from other institutional features or frictions, may be even more consequential for asset pricing than is documented here. Therefore, unlike that assumed in classical "frictionless" asset pricing, demand effects may be a first-order driver of asset price variations (Koijen and Yogo, 2019).

References

Acharya, Viral V, and Lasse Heje Pedersen, 2005, Asset pricing with liquidity risk, *Journal of Financial Economics* 77, 375–410.

Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

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.

Bali, Turan G, Nusret Cakici, and Robert F Whitelaw, 2011, Maxing out: Stocks as lotteries and the cross-section of expected returns, *Journal of Financial Economics* 99, 427–446.

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.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307–343.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, 2018, Do ETFs increase volatility?, *Journal of Finance* 73, 2471–2535.

Ben-David, Itzhak, Francesco Franzoni, Rabih Moussawi, and John Sedunov, 2021, The granular nature of large institutional investors, *Management Science* forthcoming.

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

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, 2020, Advice-driven demand and systematic price fluctuations, Working paper, The Ohio State University.

Ben-Rephael, Azi, Shmuel Kandel, and Avi Wohl, 2011, The price pressure of aggregate mutual fund flows, *Journal of Financial and Quantitative Analysis* 585–603.

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

Brown, David C., Shaun Davies, and Matthew Ringgenberg, 2021, ETF arbitrage and return predictability, *Review of Finance* forthcoming.

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

Chang, Yen-Cheng, Harrison Hong, and Inessa Liskovich, 2015, Regression discontinuity and the price effects of stock market indexing, *Review of Financial Studies* 28, 212–246.

Chen, Andrew Y, and Tom Zimmermann, 2020, Open source cross-sectional asset pricing, *Available at SSRN* .

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.

Corwin, Shane A, and Paul Schultz, 2012, A simple way to estimate bid-ask spreads from daily high and low prices, *Journal of Finance* 67, 719–760.

Coval, Joshua, and Eric 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., and Tobias J. Moskowitz, 2016, Momentum crashes, *Journal of Financial Economics* 122, 221–247.

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

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

Evans, Richard B., and Yang Sun, 2021, Models or stars: The role of asset pricing models and heuristics in investor risk adjustment, *Review of Financial Studies* 34, 67–107.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Gabaix, Xavier, and Ralph S.J. Koijen, 2020, In search of the origins of financial fluctuations: The inelastic markets hypothesis, Working paper, Harvard University.

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

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.

Grinblatt, Mark, and Bing Han, 2005, Prospect theory, mental accounting, and momentum, *Journal of Financial Economics* 78, 311–339.

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

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

Harris, Lawrence, and Eitan Gurel, 1986, Price and volume effects associated with changes in the S&P 500 list: New evidence for the existence of price pressures, *Journal of Finance* 41, 815–829.

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

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

Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld, 2011, Does algorithmic trading improve liquidity?, *Journal of Finance* 66, 1–33.

Hillert, Alexander, Heiko Jacobs, and Sebastian Müller, 2014, Media makes momentum, *The Review of Financial Studies* 27, 3467–3501.

Hirshleifer, David, Kewei Hou, Siew Hong Teoh, and Yinglei Zhang, 2004, Do investors overvalue firms with bloated balance sheets?, *Journal of Accounting and Economics* 38, 297–331.

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, Wei Xiong, and Lin Peng, 2009, A tale of two anomalies: The implications of investor attention for price and earnings momentum, *Available at SSRN 976394* .

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

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

Huang, Shiyang, Yang Song, and Hong Xiang, 2020b, The smart beta mirage, Working paper, University of Washington.

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

Jegadeesh, Narasimhan, and Sheridan Titman, 2011, Momentum, *Annu. Rev. Financ. Econ.* 3, 493–509.

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.

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

Lee, Charles M.C., and Bhaskaran Swaminathan, 2000, Price momentum and trading volume, *Journal of Finance* 55, 2017–2069.

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

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

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.

Morningstar, 2016, The morningstar rating for funds, White paper, Morningstar.

Moskowitz, Tobias J., and Mark Grinblatt, 1999, Do industries explain momentum?, *Journal of Finance* 54, 1249–1290.

Novy-Marx, Robert, 2012, Is momentum really momentum?, *Journal of Financial Economics* 103, 429–453.

Novy-Marx, Robert, and Mihail Velikov, 2016, A taxonomy of anomalies and their trading costs, *The Review of Financial Studies* 29, 104–147.

Parker, Jonathan, Antoinette Schoar, and Yang Sun, 2020, Retail financial innovation and stock market dynamics: The case of target date funds, Working paper, Massachusetts Institute of Technology.

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

Pastor, Lubos, Robert F. Stambaugh, and Lucian A. Taylor, 2020, Fund tradeoffs, *Journal of Financial Economics* 614–634.

Pollet, Joshua M., and Mungo Wilson, 2008, How does size affect mutual fund behavior?, *Journal of Finance* 63, 2941–2969.

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

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

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

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

Wardlaw, Malcolm, 2020, Measuring mutual fund flow pressure as shock to stock returns, *Journal of Finance* 75, 3221–3243.

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

Wurgler, Jeffrey, and Ekaterina Zhuravskaya, 2002, Does arbitrage flatten demand curves for stocks?, *Journal of Business* 75, 583–608.

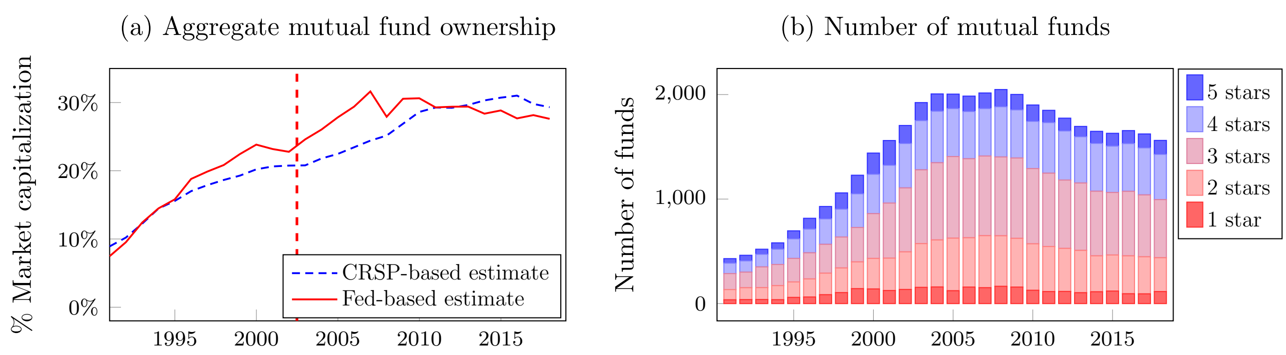
Appendix A Data and Measures

A.1 Mutual fund data

To gauge the importance of the mutual fund sector in the U.S. stock market, Panel (a) of Figure A.1 plots the value-weighted fraction of stocks held by mutual funds from 1991–2018. We present two estimates, one based on the Federal Reserve Flow of Funds and the other on the CRSP mutual fund database. Both estimates show that the aggregate holding of mutual funds increased dramatically over our sample period, starting from constituting approximately 10% of the U.S. stock market in 1991 to approximately 30% by 2018. Panel (b) of Figure A.1 shows the number of mutual funds used in our sample along with the distribution of Morningstar ratings.

Figure A.1. Number of Funds and the Average TNA over Time

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



A.2 Asset pricing factors

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

Table A.1. Asset Pricing Factors

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

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

A.3 Morningstar Methodology

We explain Morningstar's rating methodology and its June 2002 change in detail here. Morningstar ratings are updated every month. There are two steps in Morningstar's rating calculation. First, for each fund with sufficient data, Morningstar calculates performance measures using past returns, with some adjustments based on return volatility and fund loads. Second, it ranks funds by the performance measure and assigns ratings.

Morningstar changed both steps of the methodology in June 2002. The steps are consecutive, though independent. Our analysis shows that the change to the second step (described in Section A.3.2) made the biggest difference to the issues of interest in the study.

A.3.1 Step One: Calculate Performance Measures

The pre-2002 methodology is described in detail in Blume (1998), and we summarize it here. First, Morningstar calculates the cumulative return over the three horizons:

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

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

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

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

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

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

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

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

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

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

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

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

After June 2002, Morningstar began to conduct risk adjustment in a slightly different way.[33] Specifically, Morningstar now summarizes a fund's past performance using the so-called Morningstar risk-adjusted return (MRAR):

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

Here, $r_{i,t} - r_t^f$ is the geometric return in excess of the risk-free rate after adjusting for loads,[34]

[33] 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).

[34] 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.

and γ is the risk aversion coefficient (Morningstar chooses $\gamma = 2$).[35]

The formula indicates that funds with higher return volatility are penalized. To see this, notice that when γ converges to 0, $\mathrm{MRAR}^{\mathrm{T}}(0)$ is equal to the annualized geometric mean of excess returns. When γ is 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. In other words, the risk adjustment can be expressed as $\mathrm{MRAR}^{\mathrm{T}}(0) - \mathrm{MRAR}^{\mathrm{T}}(2)$.

A.3.2 Step Two: Rank Funds and Assign Ratings

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

Before June 2002, Morningstar ranked the past performance of all equity funds together and assigned them ratings with fixed proportions: 10%, 22.5%, 35%, 22.5%, and 10%. 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 was significant rating variation across styles. That variation became negligible after June 2002 (Panel (b) in Figure 3).

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

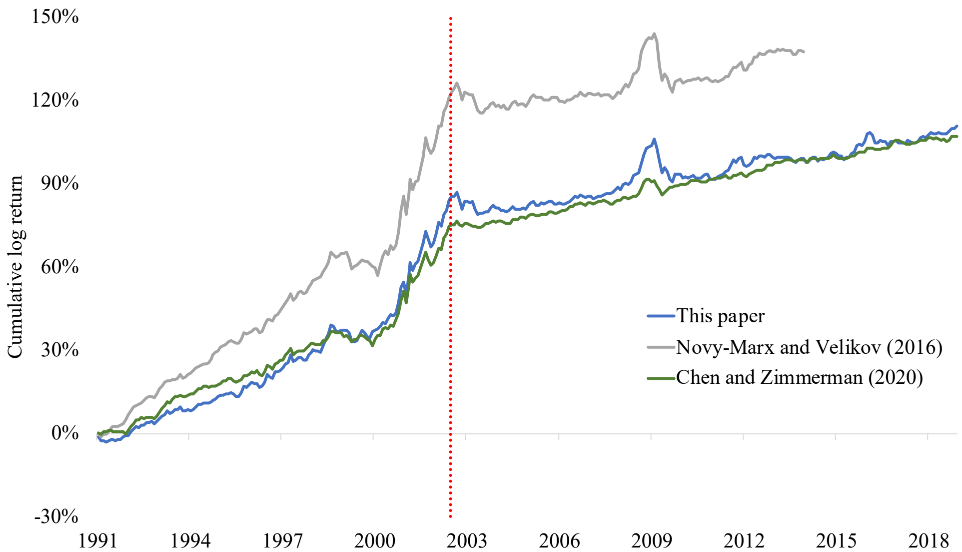
Appendix B Additional Empirical Results

B.1 The Factor Profitability Kink Around June 2002

We first verify that the "2002 kink" of factor profitability is robust to using alternative set of factors and different factor construction methodology. For this purpose, we obtain the 32 factors constructed in Novy-Marx and Velikov (2016). Second, we obtain the 181 factors constructed by Chen and Zimmermann (2020).[36] Figure B.2 plots the cumulative average return of these factors. The same 2002 kink is clearly visible.

Figure B.2. Average Factor Returns in Alternative Data Sets

This figure plots the cumulative log average factor return in our construction (blue line) with returns from alternative data sources. The grey line plots the 32 factors in the data provided by Novy-Marx and Velikov (2016) (gross returns to "simple strategies") and ends in 2013 due to data availability. The dark green line plots the 181 factors (Long/short NYSE-based quintiles, value-weighted) in Chen and Zimmermann (2020). The vertical dashed line marks the June 2002 Morningstar rating reform event.



Second, we note that this "2002 kink" has been documented in some earlier studies such as Daniel and Moskowitz (2016) and Arnott, Harvey, Kalesnik, and Linnainmaa (2019b).

[36]We obtain data in Novy-Marx and Velikov (2016) from Novy-Marx's website (`http://rnm.simon.rochester.edu/data_lib/ToAatTC/index.html`). We use their gross long-short factor returns in "returns to simple strategies". For data in Chen and Zimmermann (2020), we accessed the 0.1.2 version at `https://sites.google.com/site/chenandrewy/open-source-ap`, and used their "test asset portfolios" for NYSE-based value-weighted decile portfolios. We then constructed factors as long long the top decile and short the bottom decile.

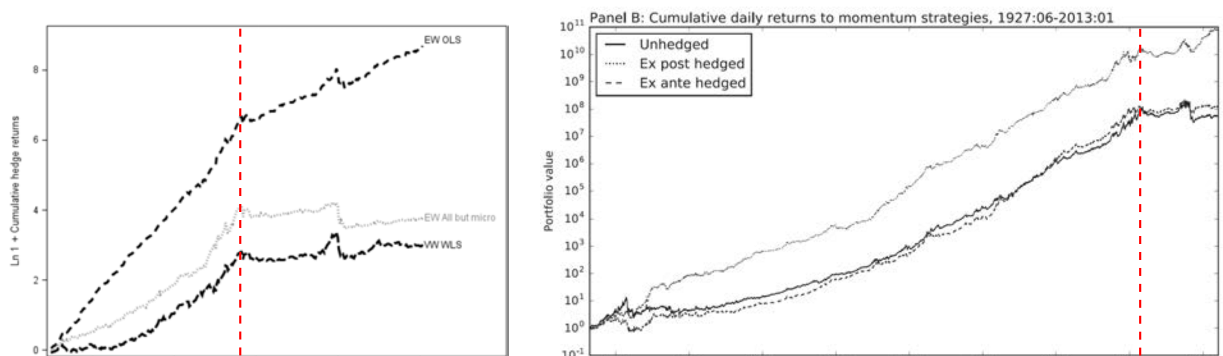
While Green, Hand, and Zhang (2017) investigate return predictability of characteristics using regressions, their study also sheds light on factor portfolios because "portfolio sorts are really the same thing as nonparametric cross-sectional regressions," as explained in Cochrane (2011).

For the reader's convenience, we present screenshots from Green et al. (2017) and Daniel and Moskowitz (2016) in Figure B.3. Panel (a) shows a chart from Green et al. (2017), summarizing the average performance (equally-weighted as well as value-weighted) of 94 characteristics. 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 B.3. Previous Evidence of the Factor Return Kink around June 2002

The figure presents charts in previous studies showing a kink in cumulative factor returns. In both panels, we added a red dashed line to mark the approximate location of June 2002 on the timeline. Panel (a) reproduces Figure 3 of Green et al. (2017). They study a strategy that uses 94 stock characteristics, and the different lines in the Figure represent different portfolio weighting methodologies. "EW OLS" refers to equal-weighting; "EW All but micro" refers to equal-weighting but excluding microcap stocks; "VW WLS" refers to value-weighted strategy. Panel (b) reproduces Figure 4b of Daniel and Moskowitz (2016) which plots the cumulative return to the momentum strategy. The Figures are taken from the latest SSRN versions of each paper: October 2016 version for Green et al. (2017), and July 2015 version of Daniel and Moskowitz (2016), with the authors' permissions.

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

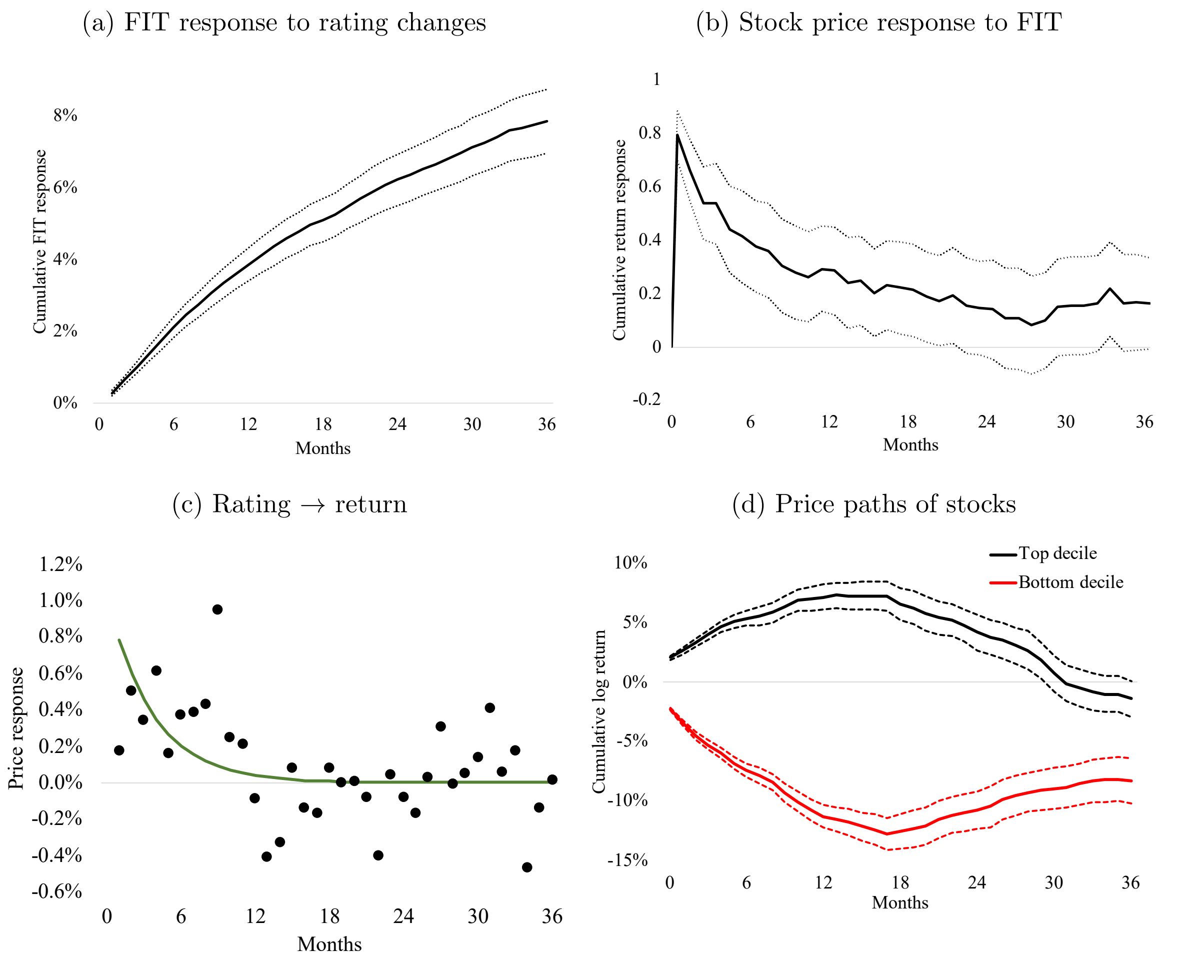


B.2 Estimating Stock-Level Effect of Ratings on Returns

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

Figure B.4. Price Impact of Rating and Flows.

Panel (a) shows the cumulative response of flow-induced trading (FIT) to changes in stock-level ratings. Panel (b) shows the cumulative response of stock returns to FIT. Following Lou (2012), FIT is defined as the nondiscretionary trading induced by mutual fund managers proportionally adjusting existing portfolio holdings in response to fund flows.In both (a) and (b), the dashed lines show two standard error bands. Panel (c) shows the *non*-cumulative response coefficients of stock returns to changes in ratings as well as the fitted exponential response (green line). Panel (d) plots the cumulative value-weighted price path of stocks in the top and bottom deciles of lagged exponential sum of rating changes (ExpSum$(\Delta\text{Rating})_{i,t-1}$). The decile breakpoints are based on NYSE stocks, and the decile portfolio returns are demeaned by period to remove the overall trend of stock value increasing over time. The dashed lines are two standard error bands.



(a) FIT response to rating changes

(b) Stock price response to FIT

(c) Rating → return

(d) Price paths of stocks

We use Fama-MacBeth regressions to estimate the chain of dynamic effects: *i*) the response of fund flow-induced trading (FIT) to Morningstar ratings changes, and *ii*) the response of stock returns to flow-induced trading. The regressions at the the stock level and value-weighted.[37]

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

$$\text{FIT}_{i,t} = a + b_1 \cdot \Delta\text{Rating}_{i,t-1} + \ldots + b_{36} \cdot \Delta\text{Rating}_{i,t-36} + X_{i,t} + u_{j,t}, \tag{21}$$

where $\Delta\text{Rating}_{i,t}$ is the month t rating change of stock i, and controls $X_{i,t}$ include 36 monthly lags of FIT and stock returns. The cumulative response coefficients $(b_1, b_1+b_2, \ldots)$ are plotted in Panel (a) of Figure B.4. In response to a one-star change in rating, stocks experience an average of 8% additional FIT in the subsequent 3 years. This result is consistent with prior research showing that, controlling for past fund performance, discrete changes in ratings create continued fund flow response that last for months (Del Guercio and Tkac, 2008; Reuter and Zitzewitz, 2015; Ben-David et al., 2019).

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

$$\text{Ret}_{i,t} = a + c_0 \cdot \text{FIT}_{i,t} + c_1 \cdot \text{FIT}_{i,t-1} + \ldots + c_{36} \cdot \text{FIT}_{i,t-36} + u_{i,t}. \tag{22}$$

We plot the cumulative response in Panel (b) of Figure B.4. Each 1% increase in mutual fund ownership due to flows leads to immediate price pressures of approximately 0.8% in the contemporaneous month, which gradually reverts in the subsequent one to two years. This result is consistent with the findings related to FIT in Lou (2012).

Combining these two effects, we expect that stock returns will also be affected by rating changes, and particularly by *the most recent* rating changes. The impact of more distant rating changes, such as those 24 months ago, should be weaker. While those rating changes may continue to generate flows, the price pressures generated by their initial price impact are already reverting, so the two effects will partially cancel each other out.

To facilitate our later analysis of rating-induced price impact, it is convenient to sum-

[37] 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.

marize recent rating changes into a weighted average sum where the weights correspond to how much each lagged rating change impacts returns. We obtain such a weighting scheme by directly estimating the response of stock returns on the past 36 lags of stock-level rating changes and plot the coefficients in Panel (c) of Figure B.4. As expected, more recent rating changes are more impactful, and the coefficients on more distant rating changes converge towards zero.

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

$$\text{ExpSum}(\Delta\text{Rating})_{i,t-1} = \sum_{k=1}^{12} \tau_k \cdot \Delta\text{Rating}_{i,t-k}, \tag{23}$$

where $\sum_{k=1}^{12} \tau_k = 12$ and the weights decay with factor $\delta = 0.764$, which is estimated from a least-squares fit to the cumulative response (Panel (c) of Figure B.4).

We end this section with an illustration of price pressures. If the price movements predicted by ExpSum(ΔRating) are truly price pressures, we expect it to eventually revert. In Panel (d) of Figure B.4, we sort stocks in each month t by $\text{ExpSum}(\Delta\text{Rating})_{t-1}$ into NYSE deciles and plot the cumulative returns of the top and bottom deciles. To focus on the cross-sectional dispersion, the returns of the decile portfolios are demeaned by period. As shown in the Figure, the rating-induced price movements in the top and bottom decile portfolios do revert after 3 years.

B.3 Predicting Factor Rating Changes at the Reform Event

In this section, we examine the accuracy of the factor-level rating change-prediction in Equation (10). We first illustrate the prediction method in Panels (a) and (b) of Figure B.5. Those two panels plot the two factors predicted to experience the largest rating decline (size) and increase (O-score). Our estimation matches actual ratings quite well. Before June 2002, the actual ratings closely match the estimated ratings under the old methodology (grey lines), and, after June 2002, the actual ratings closely match the estimated ratings under the new methodology (orange lines). Further, because the changes of factor-level ratings of factors over a few months is small, the predicted rating change using December 2001 data ends up being a reasonable predictor of the actual rating change in June 2002. This is further shown in Panel (c), where we plot the actual June 2002 rating changes of factors against the predicted changes. The latter explains the former with an R^2 of 84%.

B.4 Momentum-type Factors, Before versus After 2002

We now dig deeper into the effect of rating reform on the momentum factor and other momentum-category factors. Momentum is arguably one of the most puzzling factors because of its high profits and the difficulty to rationalize it using risk-based explanations. 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).

Figure B.6 compares ExpSum(ΔRating), FIT, and returns of the five momentum quintile portfolios before and after June 2002.[38] Panel (a) shows that, before the ratings reform, the winner portfolio experienced significant upward rating changes, while the loser portfolio experienced significant downward changes. This pattern became much smaller after June 2002. There is a similar effect on FIT, as shown in Panel (c). Before June 2002, the winner portfolio experienced 0.51% higher monthly flows than the loser portfolio; that difference declined to a meager 0.14% after June 2002.

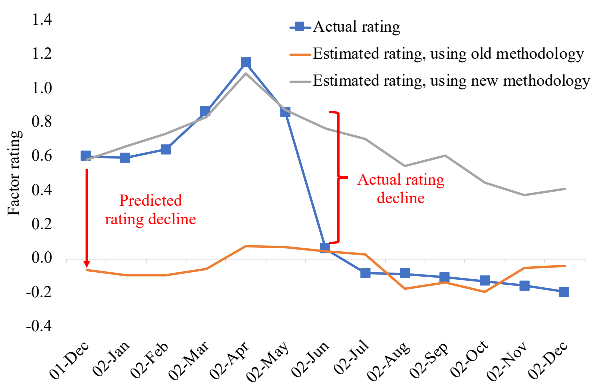
Finally, Panel (e) shows similar patterns in returns. Before June 2002, the winner quintile portfolio enjoyed 0.87% higher monthly return relative to the loser quintile, and that difference declined to 0.16% after June 2002. In Panels (b), (d), and (f), we confirm that similar post-2002 changes also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

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

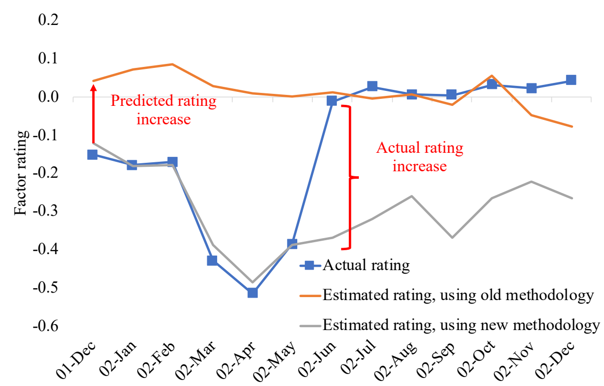
Figure B.5. Predicting Factor-level Rating Changes at the 2002 Reform Event

Panels (a) and (b) illustrate how we predict rating changes of factors at the June 2002 event using data in December 2001. Following Morningstar's rating construction process, we estimate ratings from ground up using fund returns. The grey lines plot the estimated rating under the old (pre-change) methodology, and the orange lines plot the estimated rating under the new (post-change) methodology. We use the difference between the two estimates in December 2001 (marked using red arrows) as the predicted rating change. The blue lines are the actual ratings. Panel (a) and (b) plot the factor with the largest predicted rating decline and increase, respectively (size and O-Score factors). Panel (c) compares the actual rating change in June 2002 against the predicted change using data in December 2001. The factors are sorted into quintiles based on the predicted rating change and colored differently.

(a) Most negatively affected factor



(b) Most positively affected factor



(c) Accuracy of rating change prediction

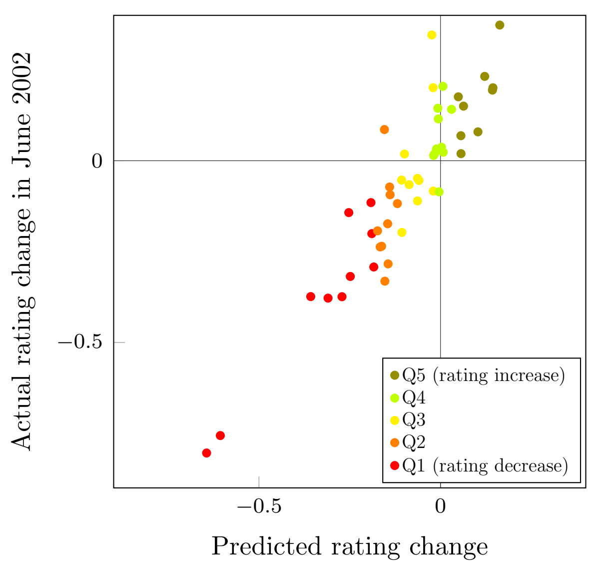
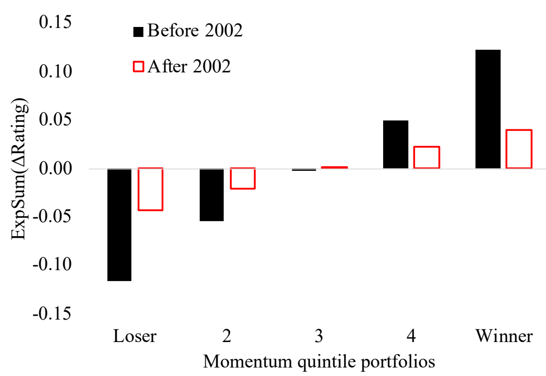


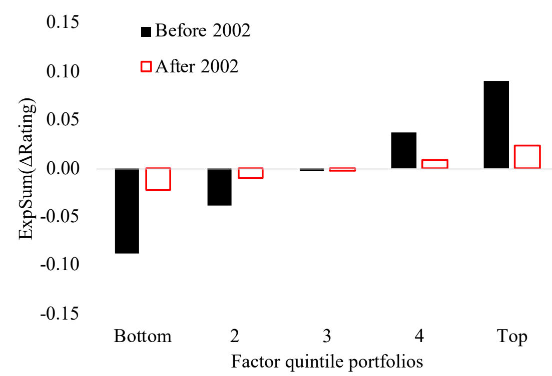
Figure B.6. Momentum-type Factors, Before versus After 2002

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

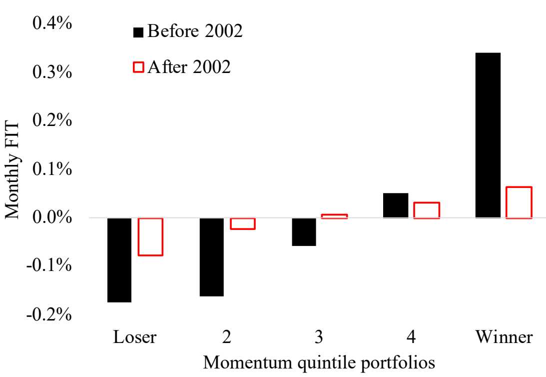
(a) ExpSum(ΔRating), Momentum



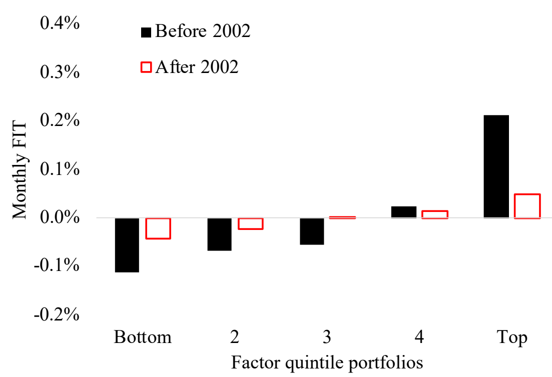
(b) ExpSum(ΔRating), other momentum-category factors



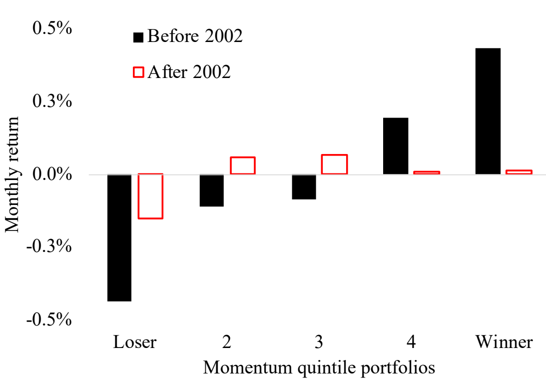
(c) FIT, Momentum



(d) FIT, other momentum-category factors



(e) Return, Momentum



(f) Return, other momentum-category factors