Discontinued Positive Feedback Trading and the Decline of Return Predictability

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

December 1, 2022

Abstract

We show that demand effects generated by institutional frictions can influence systematic return predictability patterns in stocks and mutual funds. Identification relies on a reform to the Morningstar rating system, which we show caused a structural break in style-level positive feedback trading by mutual funds. As a result, momentum-related factors in stocks, as well as performance persistence and the "dumb money effect" in mutual funds, experienced a sharp decline. Consistent with the proposed channel, return predictability declined right after the reform, was limited to the U.S. market, and was concentrated in factors and mutual funds most exposed to the mechanism.

Keywords: Positive feedback trading, mutual funds, anomalies, momentum, performance persistence
JEL Classification: G11, G12, G24, G41

*We thank Nicholas Barberis, Huaizhi Chen, Thummim Cho, Lauren Cohen, Sylvester Flood (Morningstar), Umit Gurun, Bing Han, Paul Kaplan (Morningstar), Dong Lou, Chris Malloy, Christopher Schwarz, John Shim, Andrei Shleifer, Moto Yogo, and Lu Zheng for helpful comments. We thank seminar participants at The Ohio State University, the University of Utah, the University of Washington, the University of California-Irvine, Cornell University, Hong Kong University of Science and Technology, and Arrowstreet Capital, as well as 2021 WFA, 2021 EFA, 2022 MFA, Pacific Northwest Finance Conference, and the National Bureau of Economic Research Behavior Finance Workshop participants 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 the 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. Emails: ben-david.1@osu.edu, jiacui.li@eccles.utah.edu, rossi2@arizona.edu, and songy18@uw.edu.

1 Introduction

Understanding the sources of predictability in securities' returns is a central theme in asset pricing. Traditionally, return predictability in stock portfolios has been ascribed to risk exposure. And performance predictability in mutual funds is usually credited to managerial skill.[1] However, recent studies highlight the importance of investor demand—even if unrelated to cash flow expectations or hedging motives—in explaining assets' return patterns (e.g., Koijen and Yogo, 2019; Gabaix and Koijen, 2021). Given the advances in the demand-based framework, it is important to assess whether systematic and persistent return patterns can be generated by investors' demand.

This paper studies the impact of a significant shift in investor demand on predictability patterns in equity factors and mutual fund returns. We show that a major mid-2002 reform to the Morningstar mutual fund rating methodology caused an exogenous decline in positive feedback trading by mutual funds at the style level. Based on this mechanism, we explore the impact of this reform on the profitability of momentum-related factors and known predictability patterns in mutual fund performance, i.e., predictability based on past performance (Carhart, 1997) and on past flows ("dumb money effect" of Frazzini and Lamont, 2008). Indeed, we document that this institutional change contributed to the decline of momentum-related factor profits and mutual fund predictability patterns. Overall, our analysis shows that demand effects caused by institutional frictions can be a first-order determinant of long-term expected returns.[2]

Our identification strategy builds on the findings that mutual fund investors tend to chase past performance, as reflected in Morningstar star ratings (Ben-David, Li, Rossi, and Song, 2022b), and a methodological reform in Morningstar's ratings, which took place in

[1] The risk-based perspective for understanding expected stock returns is articulated in Cochrane (2011). Managerial skill is central in interpreting mutual fund performance and capital flows in Berk and Green (2004) and follow up work.

[2] Also using the 2002 Morningstar reform event, Ben-David, Li, Rossi, and Song (2022a) show that correlated demand can exert large influence on *short-run price fluctuations* of style portfolios. This paper focuses on the *long-term* impact on *expected returns* of stock factors and mutual fund return predictability.

2002 (Ben-David et al., 2022a). Until mid-2002, Morningstar equity fund ratings were based on a universal return ranking. Since past performance is highly correlated with investment style, flows were directed to funds in the best-performing styles, putting price pressure on the underlying stocks and leading to further outperformance in the following months. In June 2002, Morningstar revised its methodology and began ranking funds *within* style. After the reform, top-ranked funds exist in similar proportion in every style, and hence rating-chasing flows are distributed much more equally across styles. This seemingly innocuous institutional change led to a sudden decline in positive feedback trading and return persistence at the style level.

The disruption in the positive feedback trading caused by Morningstar's reform is an opportunity to study the impact of investors' demand on the predictability of equity factors and of mutual fund returns. We argue that if positive feedback trading were an important contributing driver of momentum-related factors, we should observe a sharp decline in their profitability after the reform. Furthermore, since mutual funds tend to pursue strategies related to investment styles, the Morningstar methodology-induced changes in style-level stock returns can also have a major impact on the predictability of mutual fund returns.

Our empirical analysis proceeds in two parts. In the first part, we test these predictions in stock factors and mutual funds. In the U.S. stock market, using either a list of 49 commonly used stock factors we construct or 153 factors from Jensen, Kelly, and Pedersen (2022), we find that momentum-related factors experienced large profitability declines after June 2002. Based on our constructed factors, the monthly return of momentum-related factors declined from 77 basis points to approximately zero. Based on the Jensen et al. (2022) factors, the decline is from 67 basis points to 4 basis points. While many other factors also experienced profitability declines, we find that the profitability decline in momentum is much more severe, consistent with our predictions. It is likely that other market-wide mechanisms proposed in prior work[3] may have improved market efficiency or risk sharing. The mechanism we identify is non-mutually exclusive of other channels. Importantly, the mutual fund rating reform was unique to the U.S. market and was implemented in a particular date, allowing us to identify the effect more precisely than it is usually possible in similar asset pricing settings.

Since the reform in Morningstar's rating system was limited to U.S. funds, we contrast the decline in performance predictability in the U.S. to that in other regions. We use international stock factors from Jensen et al. (2022) to conduct "triple-difference" regression analysis. Consistent with the proposed mechanism, only momentum-related factors in the U.S. experienced significant profitability declines. The same is not true about momentum factors in other countries. In addition, our results are robust to excluding the "momentum crash" period (Daniel and Moskowitz, 2016).

Next, we study the predictability of mutual funds' performance. Based on the post-2002 disruption of style-level positive feedback trading, we predict that the Carhart (1997) performance persistence will decline. Further, combining the fact that style-level flows become much muted after the reform and the prior finding of style-level flow-induced return reversals (Ben-David et al., 2022a; Li and Lin, 2022), we also expect the Frazzini and Lamont (2008) "dumb money effect"—the finding that funds with high (low) recent three-year flows subsequently underperform (outperform)—to become weaker after the reform.[4] Our results show that both forms of fund performance predictability declined precipitously after the rating reform. Consistent with the proposed mechanism, we do not find similar patterns in equity funds outside the U.S. nor in non-equity U.S. mutual funds.

In the second part of the analysis, we zoom in on a narrow window around the June

[3]Prior work has identified many other mechanisms that may cause factor returns to decline: changes in liquidity (Khandani and Lo, 2011; Chordia, Subrahmanyam, and Tong, 2014; Lee and Ogden, 2015), and increased arbitrage activity (Marquering, Nisser, and Valla, 2006; Green, Hand, and Soliman, 2011; Hanson and Sunderam, 2013; McLean and Pontiff, 2016; Calluzzo, Moneta, and Topaloglu, 2019; Cho, 2020). Several studies propose that some factors may result from possible data-mining or overfitting (Harvey, Liu, and Zhu, 2016; Harvey, 2017; Hou, Xue, and Zhang, 2020; Huang, Song, and Xiang, 2020b; Falck, Rej, and Thesmar, 2021).

[4]Specifically, before the Morningstar rating reform, volatile style-level flows lead to large style-level price fluctuations and subsequent reversals, and the reversals contribute to the dumb money effect. Section 2.4 explains this mechanism in detail.

2002 reform to provide direct evidence that the rating reform had a first-order impact on factor returns. While it is usually not possible to control for all alternative hypotheses when studying long-term expected returns, one can achieve more robust identification when examining the effect of specific demand shocks on short-term price movements. Making use of the transparency of the Morningstar rating methodology, we sort stock factors by how much they are expected to be affected by the reform using *ex-ante* information. To directly measure how each factor is impacted by rating-induced trading, we aggregate fund ratings and fund flows at the factor level. The event study shows that stock factors that were expected to be impacted by the reform (based on pre-event information) experienced sudden drops in ratings, flows, and returns in the six months following the reform. Consistent with the rating-induced mechanism, other factors that were not affected by the reform did not experience similar "kinks." Using all years other than 2002 as placebo tests, we confirm that the effects we document were unique to 2002. Moreover, proxies for other possible influences, such as arbitrage activity and liquidity, did not vary materially around the reform event.

We also examine the predictability of mutual fund returns and find similar effects around mid-2002. In particular, mutual funds that were most exposed to Morningstar's rating reform were impacted the most: they experienced large changes in ratings, flows, and returns right after the reform. Overall, these findings from the event study are consistent with the idea that the rating-induced change in style-level demand patterns can strongly impact factor and mutual fund returns.

This paper's main contribution is using a natural experiment to identify the effect of correlated investor demand on systematic patterns in expected stock returns and mutual fund returns. Importantly, the shift in investor demand that we use to identify our effects are uncorrelated with potential fundamental drivers of the performance predictability of stocks and mutual funds.

This paper is most related to Lou (2012), which argues that return-chasing mutual fund flows can impact expected returns. The main innovation lies in the use of the Morningstar reform for identification.[5] Relative to Lou (2012), we also further clarify the mechanism: the effect primarily comes from style-level correlated flows, rather than idiosyncratic fund-level flows, a point that we elaborate in Section 2.3. Another related paper is Ben-David et al. (2022a) which also uses the 2002 Morningstar reform to demonstrate that style-level fund flows can cause short-term fluctuations in stock prices. The current study differs by exploring the impact of positive feedback trading on *long-term* patterns in *expected returns* in stocks and mutual funds.

More broadly, this paper contributes to the literature studying the impact of demand on *systematic* components of asset prices. While the earlier work on index composition changes convincingly showed that demand could impact the prices of individual stocks (Harris and Gurel, 1986; Shleifer, 1986; Wurgler and Zhuravskaya, 2002; Chang, Hong, and Liskovich, 2015; Pavlova and Sikorskaya, 2021), there is relatively less consensus on whether and how demand can shape *systematic* price movements.[6] The style-level rating-induced positive feedback trading mechanism we identify is also consistent with the "style investing hypothesis" which has been examined in previous work (e.g., Barberis and Shleifer, 2003; Teo and Woo, 2004; Wahal and Yavuz, 2013).

This paper also has implications for interpreting performance predictability in mutual funds. Traditional discussions of performance predictability—or lack thereof—often center on the managerial skill, such as in the case of Carhart (1997), or investor sophistication in choosing funds, such as in the case of Zheng (1999) and Frazzini and Lamont (2008). In contrast, we show that demand-based price effects are also important. This perspective should be intuitive: mutual funds are, after all, portfolios of securities. If demand effects

[5] Other studies that use Morningstar ratings as part of their identification strategy include Del Guercio and Tkac (2008), Kim (2020), Reuter and Zitzewitz (2021), Evans and Sun (2021), Han, Roussanov, and Ruan (2021), Ben-David et al. (2022b), and Adelino, Cheong, Choi, and Oh (2022).

[6] Other studies on demand-based price effects use mutual fund flows (Teo and Woo, 2004; Coval and Stafford, 2007; Lou, 2012; Huang, Song, and Xiang, 2020a; Li, 2021), exchange-traded fund flows (Ben-David, Franzoni, and Moussawi, 2018; Brown, Davies, and Ringgenberg, 2021), microstructure measures of order flow imbalance (Li and Lin, 2022), and other sources of institutional investor demand (Parker, Schoar, and Sun, 2020; Ben-David, Franzoni, Moussawi, and Sedunov, 2021; Hartzmark and Solomon, 2021). More recently, Koijen and Yogo (2019) develop a structural methodology to estimate price impact, and Gabaix and Koijen (2021) show that demand-induced price impact at the aggregate market level is considerable.

can systematically influence security returns, they can also impact the returns of security portfolios. Jones and Mo (2021) show that discovered mutual fund predictability patterns become weaker out-of-sample and propose the decline is related to arbitrage activity and mutual fund competition. While applicable only to a subset of known fund return predictors, our findings suggest that changes in demand patterns can also be an important contributor. Finally, our paper provides a possible explanation for Choi and Zhao (2022), which finds that the performance persistence results of Carhart (1997) disappear out of sample.

The rest of the paper is organized as follows. Section 2 explains how the Morningstar reform disrupts style-level positive feedback trading and makes several testable predictions. Section 3 describes the data. Section 4 examines the impact of the reform on asset pricing factors and mutual fund return predictability. Section 5 performs an event study around the reform date, and Section 6 concludes. Robustness checks and additional tests are provided in the Appendix.

2 Morningstar Rating Reform and Predictions

In this section, we describe the Morningstar rating methodology reform that was implemented at the end of June 2002. Then, we explain why it led to a disruption in style-level positive feedback trading. Based on this mechanism, we make testable predictions to be examined throughout the rest of the paper.

2.1 2002 Rating Methodology Reform

We now describe the Morningstar rating methodology reform in June 2002.

Methodology before the reform. 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 10 years, but more weight is applied to more recent periods.[7] 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%).[8]

Methodology after the reform. While the rating methodology has been very stable over time, Morningstar implemented a major reform in June 2002.[9] After the reform, fund ratings were no longer based on how each fund ranked against *all* U.S. equity funds but only on fund rankings *within* 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.[10] 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 1, which plots the rating dispersion of 3×3 size–value fund styles. In the months leading to the methodology change, the average fund ratings of the top- and bottom-rated styles differed by up to 2 stars. Following the reform, that gap dropped dramatically and ratings also became uncorrelated with past style performance.[11]

[7]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.

[8]The Morningstar methodology is fully transparent. Appendix B of Ben-David et al. (2022a) provides further detail on the exact rating computation.

[9]The change was partially motivated by complaints from fund managers, who argued that they were receiving low ratings simply because their investment style performed poorly, but not because of how they managed the funds. See Section 3 of Ben-David et al. (2022a) for more details.

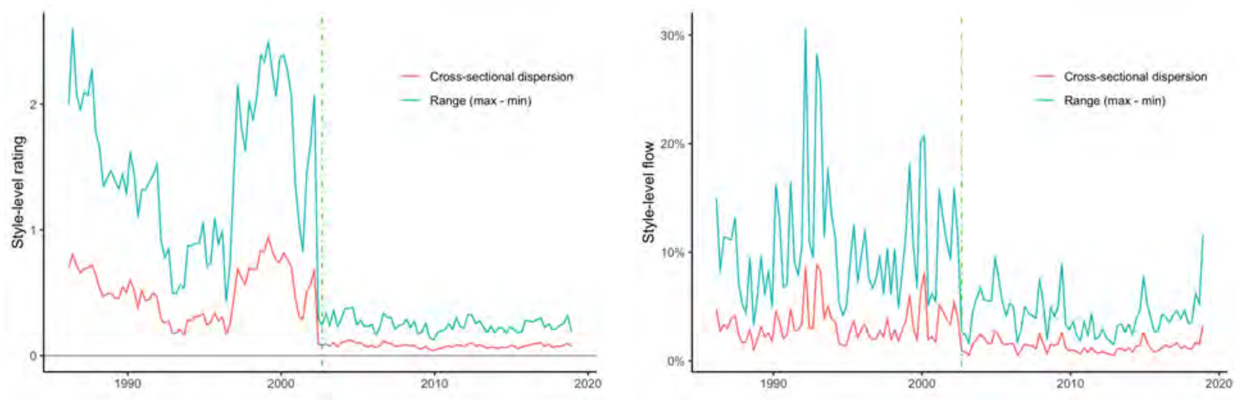
[10]Sector funds—the remaining 13%—were classified into 12 sectors (e.g., financials, utilities).

[11]One may wonder why rating dispersion did not drop to exactly zero. A major reason is that Morningstar assigns ratings at the share-class level, so taking an average over share classes would bring the dispersion to zero. Because a fund's share classes have the same underlying portfolio, we compute average ratings at the fund-level following Barber, Huang, and Odean (2016).

Figure 1. The Morningstar Methodology Reform and Style-Level Flows

Panels (a) and (b) plot the dispersion of quarterly fund ratings and total net assets (TNA)-weighted average fund flows by the 3 × 3 size–value Morningstar styles. Dispersion is measured either as the cross-sectional standard deviation (red lines) or the difference between maximum and minimum values (blue lines). The vertical dashed line marks the June 2002 Morningstar methodology reform event.

(a) Dispersion of style-level ratings (b) Dispersion of style-level fund flows



Importantly for our identification purposes, investors continued to chase ratings in a similar manner before and after the reform. This has been shown by a number of prior studies (Evans and Sun, 2021; Ben-David et al., 2022a,b).[12] Therefore, by equalizing the distribution of ratings across investment styles, Morningstar's reform effectively *redirected* fund flows to be more equally distributed over styles. Consequently, rating-chasing flows stopped chasing style-level returns.

2.2 Effects of Reform on Positive Feedback Trading

We now demonstrate that the 2002 Morningstar reform had significant effects on style-level trading. In addition to reducing cross-sectional dispersion of style-level flows, it also reduced style-level positive feedback trading.

Reduction of style-level flow dispersion. Because Morningstar ratings are a major driver of fund flows (e.g., Reuter and Zitzewitz, 2021; Ben-David et al., 2022b), a reduction of style-level rating dispersion naturally led to the a reduction of style-level flow dispersion.

[12] See, for example, Figure 1(b) and Figure 4(b) in Ben-David et al. (2022a).

This is shown in Panel (b) of Figure 1 where we plot the cross-sectional dispersion (standard deviation) and range (maximum minus minimum) of style-level fund flows across the 3×3 Morningstar styles. Before the reform, the average cross-sectional dispersion (standard deviation) of style-level ratings was 0.50 and the average maximum-minus-minimum range was 1.46, both of which shrunk to effectively zero after the reform. Similarly, the dispersion of quarterly style-level fund flows declined from 3.39% to 1.35%, and the range declined from 10.66% to 4.23%, respectively.

Disruption of style-level positive feedback trading. The pre-reform rating methodology generated a positive feedback loop at the style level that was disrupted after the reform.

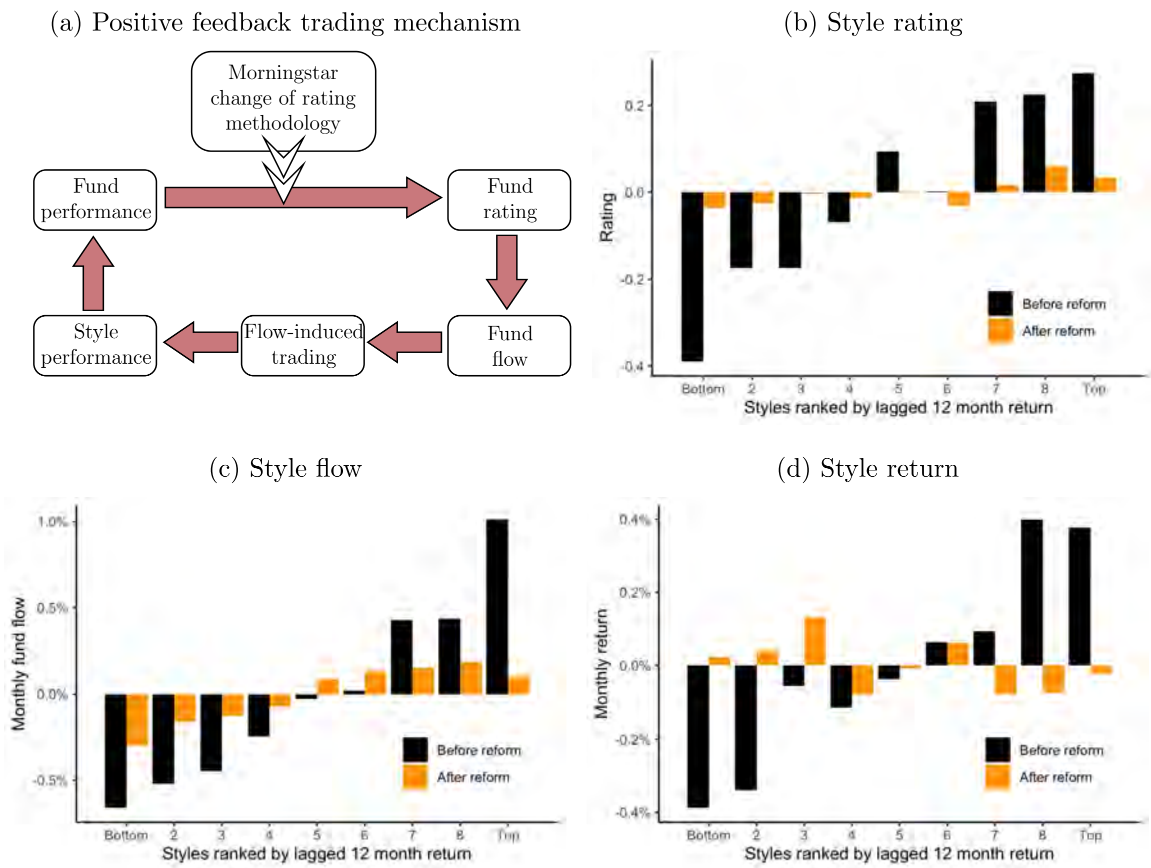
The pre-reform mechanism is illustrated in Panel (a) of Figure 2: Funds in styles that performed well in the recent past receive high ratings and attract inflows. Funds use the new flows to increase their investments in the same style of stocks, so the prices of those stocks are pushed up even further. 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, causes a sudden disruption in this rating-induced style-level positive feedback trading. We confirm this style-level disruption in Panels (b) and (c) of Figure 2. 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 the reform, funds in styles that recently performed well-received higher average ratings and higher fund flows. The magnitudes are also large. Panel (b) shows that the average rating spread between funds in the top and bottom styles was about 0.8 stars before reform and shrank to almost zero after the reform. Because high ratings attract flows, Panel (c) shows that funds in the top style received about 1.7% higher flows per month than the bottom style before the reform, and that difference dropped to around 0.4% after the reform.[13]

[13] The data in these graphs are demeaned within each month to focus on cross-sectional patterns across styles.

Figure 2. Style-Level Positive Feedback Trading Before and After Reform

This figure shows that the style-level positive feedback trading largely halted after the Morningstar methodology change in June 2002. The flow chart in Panel (a) illustrates how pre-2002 ratings generate positive style-level positive feedback trading. In Panels (b) to (d), we sort the 3×3 Morningstar styles by their lagged 12-month returns. Panels (b) and (c) plot the TNA-weighted average rating and fund flows of the sorted styles. Panel (d) plots the return of funds in those styles. All variables are demeaned to focus on the cross-sectional difference across styles. The sample years include 1991 to 2018, and the start date is dictated by the need for monthly fund flow data from CRSP.



This disruption also has a significant impact on style returns. In Panel (d), we plot total net assets (TNA)-weighted style-level fund returns. Prior to the reform, the performance spread between the top- and bottom-ranked styles was approximately 80 basis points per month. The performance difference disappears after the reform. In unreported robustness checks, we find similar patterns when measuring returns using the capital asset pricing model (CAPM) alpha, and the post-reform change in the alpha spread is statistically significant at the 5% level. To alleviate the concern that fund returns may also be influenced by transaction costs and fees, we also repeat this exercise using the returns of the stocks held by the funds, rather than fund returns. The results are unaffected.

In summary, the 2002 Morningstar rating reform led to a disruption in the flow and return dynamics at style level.

2.3 What is Special About Style-level Demand?

This section explains why changes in demand at the style level can have a sizeable impact on stock returns, and why demand at the stock (idiosyncratic) level has an only a minute impact on returns. Conceptually, the price effect of demand can be modeled as follows:

$$\text{Price Effect} = \text{Price Multiplier} \times \text{Demand} \tag{1}$$

where the price multiplier refers to the dollar value of total price impact for each dollar of demand (Gabaix and Koijen, 2021).[14] Thus, the price effect can be large if both demand and price multipliers are large. The effect of the Morningstar reform on style-level fund flows satisfies both conditions. Appendix A.1 provides further evidence that the Morningstar rating reform altered positive feedback trading at the style level but not at the stock level.

1. **Price multipliers are larger at the style-level.** A series of papers have estimated market-level, style-level, or factor-level price multipliers and generally find that systematic-level price multipliers are larger than at the stock-level (e.g., Ben-David et al., 2022a; Li and Lin, 2022). See Gabaix and Koijen (2021) for a survey.

   We further confirm this fact in our specific setting by examining fund flow-induced trading. In Appendix A.2, we estimate the price multipliers associated with fund flow-induced trading at both the style and stock (idiosyncratic) levels. The results

[14] It can also be interpreted as an elasticity: Price Multiplier $= \frac{dP/P}{dQ/Q}$, where P is the per-share stock price, Q is the number of shares outstanding, and dQ is the change in demand. To see the equivalence, note that $\frac{\text{Total dollar price impact}}{\text{Total dollar demand change}} = \frac{dP \cdot Q}{dQ \cdot P}$.

of Fama-MacBeth and panel regressions show that style-level price multipliers are 2-3 times larger than the stock-level multipliers.

2. **Only return chasing of *common* return components impacts momentum strategy profitability.** One may be confused by our emphasis on style-level positive feedback trading: in addition to style-level positive feedback trading, there is also strong evidence of fund-level positive feedback trading which, intuitively, might also impact momentum strategy profitability. If fund-level positive feedback trading is not disrupted by the Morningstar reform, one may anticipate that it would continue to contribute to momentum profitability after the reform.[15]

   In fact, most of *fund-level* return-chasing should not impact momentum defined at the *stock-level*. To see why imagine that a stock experienced high idiosyncratic returns in a recent period. How will this impact future mutual fund demand for this stock? The answer is *very little*: because most funds have diversified holdings, the return of a single stock will barely impact the returns of the funds that hold that stock.

   Conversely, fund flows chasing *common* return components will lead to significant positive feedback trading in stocks. For instance, suppose that, over the past few months, small-value stocks performed well on average. Under the pre-reform Morningstar rating scheme, small-value funds would have received high ratings and large inflows, which would then lead to higher demand for small-value stocks.[16]

2.4 Testable Predictions

The discussion in the previous subsection leads to a number of testable predictions on post-reform changes in expected returns in stock factors and mutual funds. The predictions are summarized in Column (1) of Table 1.

[15] We thank Huaizhi Chen for suggesting that we clarify this point.

[16] This discussion also clarifies the mechanism in the seminal findings of Lou (2012). Lou finds evidence that return-chasing in mutual funds can partially account for stock momentum effects.

First, we anticipate that a disruption in positive feedback trading would reduce the profitability of momentum-type stock factors. In addition to the standard momentum factors based on past stock returns (Jegadeesh and Titman, 1993; Novy-Marx, 2012), we also expect similar findings in other factors, which are defined based on past performance, e.g., industry momentum (Moskowitz and Grinblatt, 1999) and the 52-week high strategy (George and Hwang, 2004).

Second, we also anticipate disruption to the performance persistence of U.S. equity mutual funds. Relative to stock momentum, mutual fund performance persistence is even more directly connected to style-level positive feedback trading: funds are diversified portfolios with wide dispersion in size and value style tilts (by mandate and/or due to active management choices).

Finally, we anticipate a reduction in the "dumb money effect" which refers to the empirical finding that mutual funds with high (low) recent three-year fund flows have low (high) subsequent returns (Frazzini and Lamont, 2008). The reasoning is as follows. A number of papers have shown that fund flow-induced price pressure reverts over time (Coval and Stafford, 2007; Lou, 2012; Li, 2021), a fact that would directly contribute to the dumb money effect. Since style-level flow dispersion reduced significantly after the reform, we would expect the dumb money effect to weaken.[17]

Factors and funds unaffected by the reform. Our mechanism also provides natural "control groups" that would not be affected by the Morningstar rating reform, which are listed in Column (2) of Table 1. Because the Morningstar reform is specific to the U.S. stock market,[18] we use non-equity funds and—subject to data availability—non-U.S. factor returns as controls.

[17]One might ask why we do not also predict a decline in the long-run stock return reversals finding of De Bondt and Thaler (1985). It is worth noting that the dumb money effect is defined by sorting on past *flows*, while stock reversals are defined by sorting on past stock *returns*. Because flows can only explain a small fraction of past return variation, the "signal to noise ratio" is low, so we do not expect the fund flow-based mechanism to have a large effect on stock reversals.

[18]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.

Table 1. Testable Predictions

This table summarizes the main testable predictions based on the Morningstar reform in June 2002. Column (1) lists return predictability patterns that should decline after the reform, while Column (2) lists "control group" assets and portfolios that should not be affected. Because the Morningstar reform only impacted U.S. stock market, we expect the effects to be specific to the U.S. equity market. The first row considers impact on asset pricing factor profitability and the next two rows consider impact on mutual fund return predictability.

	2002 Morningstar reform	
	Impacted	Not impacted
	(1)	(2)
Asset pricing factor profitability	Momentum-type factors	Non-Momentum-type/ Non-U.S. factors
Mutual fund return predictability: 1) Performance persistence 2) Flow-based predictability (dumb money)	U.S. equity mutual funds	Non-equity funds

When testing the decline of profitability of momentum-type factors, comparing against non-momentum factors and momentum factors in other countries is important because our prediction should be seen as *ceteris paribus*. As discussed in the introduction, there are several other mechanisms that can lead to long-term declines in factor profitability. These mechanisms impact all factors, but the Morningstar reform has an impact that is limited to momentum-type factors in the U.S. stock market and was enacted at a specific date. In other words, we are interested in the *incremental* impact of the disruption in feedback trading due to the Morningstar reform.

3 Data and Variable Construction

This section describes the data for stock factors and mutual funds. Summary statistics appear in Table 2.

3.1 Asset Pricing Factors

Panel A of Table 2 summarizes the U.S. and international factors data.

U.S. factors. We examine monthly returns of long-short factors over the period of 1987 to 2018. The start date is guided by the launch of Morningstar fund ratings.[19] To minimize the sensitivity of the results to the choice and construction of asset pricing factors, we use the two sets of factors described below.

1. **Our constructed factors.** We construct 49 popular stock characteristics-based long-short factors that have been shown to predict returns; our choice of factors mostly follows Arnott, Clements, Kalesnik, and Linnainmaa (2021), and we restrict our attention to those that can be constructed using CRSP and Compustat data. Using the classification categories proposed in Hou et al. (2020), these 49 characteristics-based factors 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. (2020) to limit the impact of microcaps in factor construction. Specifically, we use NYSE breakpoints to sort stocks into characteristics-based quintiles and then form value-weighted long-short factors. Appendix Table B.1 lists all the factors we construct.

2. **U.S. factors from Jensen et al. (2022)**. Jensen et al. (2022) constructed a large number of long-short factors and made their returns publicly available on Professor Bryan Kelly's website. This data set includes 153 U.S.-based factors that are available since 1987. Of these, eight are momentum-related factors.[20]

[19] Morningstar began providing ratings in 1985, and we control for one year of lagged factors returns, motivated by the finding that factor returns can exhibit momentum (Gupta and Kelly, 2019; Ehsani and Linnainmaa, 2022). Our results are not sensitive to changes in the start date.

[20] These include the fifty-two week high strategy in George and Hwang (2004), $(t-6, t-1)$ and $(t-12, t-1)$ residual momentum in Blitz, Huij, and Martens (2011), $(t-12, t-7)$ "intermediate momentum" in Novy-Marx (2012), lagged return in Heston and Sadka (2008), and four different forms of stock momentum from Jegadeesh and Titman (1993) ($(t-h, t-1)$ where $h = 1, 3, 6, 12$).

We use their "value-weighted capped factors" which are value-weighted long-short tercile portfolios. In addition, they also cap the market weight of each stock at the 80^{th} NYSE percentile, a practice that is intended to make sure one mega-stock does not dominate a portfolio. This concern is particularly relevant for less developed markets with fewer stocks. For brevity, we refer readers to the description in Jensen et al. (2022) for more details.

International factors. As we explain in Section 2, we expect rating-induced demand effects to impact only U.S.-based factors, so non-U.S. factors can be used as placebo assets. For this purpose, we use the monthly international factor returns in Jensen et al. (2022). We include all factors that are available starting from 1991.[21] After imposing this requirement, the sample includes 1,337 factors from 30 countries, out of which 172 are momentum-type factors.[22]

3.2 Mutual Fund Data

We obtain quarterly mutual fund data from the CRSP survivorship bias-free mutual fund data from 1980.[23] Summary statistics appear in Panel B of Table 2.

Domestic equity mutual funds. We use CRSP objective code starting with "ED" to identify domestic equity funds and restrict attention to those with net asset values above ten million. We use MFLINKS to map the share classes to fund identifiers and aggregate at the fund level (Wermers, 2000). To investigate the Carhart and dumb money effects, we also require having past three years of data so we can compute previous one-year return and three-year flows. We also compute CAPM alphas, which require 36 months of trailing

[21] Data availability for many countries starts around 1990, so requiring data to start from 1987 would significantly reduce the sample size.

[22] The Jensen et al. (2022) dataset applies the same factor construction, when applicable, to all countries. For instance, this means that there will be a Jegadeesh and Titman (1993) $(t-12, t-1)$ standard momentum factor for each country.

[23] The MFLINKS mapping of mutual fund share classes to funds is not available before 1980.

Table 2. Summary Statistics

Panel A reports the monthly stock factor return data. Columns (1) to (3) report the average monthly return, the number of factors, and the number of momentum-type factors using our data. Columns (4) to (6) report equivalent statistics for U.S. factors in Jensen et al. (2022), and Columns (7) to (10) report equivalent statistics for the international factors in Jensen et al. (2022). Panel B summarizes the distribution of each variable in the CRSP quarterly mutual fund sample. The sample includes domestic equity mutual funds and non-equity mutual funds. The first column reports the average number of unique funds. Returns are reported in percent in both Panels A and B.

Panel A: Stock factors

	U.S. factors						International factors			
	Our data			Jensen et al. (2022) data			Jensen et al. (2022) data			
	Return	Num factors		Return	Num factors		Return	Num factors		Num
		All	Mom		All	Mom		All	Mom	countries
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1987–1990	0.29	49	5	0.22	153	8	0.10	1,311	170	30
1991–1994	0.17	49	5	0.14	153	8	0.11	1,334	172	30
1995–1998	0.30	49	5	0.25	153	8	0.19	1,336	172	30
1999–2002	0.79	49	5	0.71	153	8	0.51	1,335	172	30
2003–2006	0.00	49	5	0.00	153	8	0.08	1,337	172	30
2007–2010	0.09	49	5	0.10	153	8	0.09	1,337	172	30
2011–2014	0.12	49	5	0.16	153	8	0.27	1,337	172	30
2015–2018	0.14	49	5	0.12	153	8	0.16	1,337	172	30

Panel B: Mutual funds

Variable	Obs	Mean	StdDev	1%	5%	25%	50%	75%	95%	99%
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Domestic equity mutual funds										
Return	1,110	2.14	9.49	−25.44	−15.75	−1.90	2.95	7.32	15.89	24.36
CAPM alpha	1,110	−0.20	4.75	−13.97	−7.12	−2.17	−0.24	1.67	7.00	13.97
TNA ($M)	1,110	1,859	9,381	15	27	105	333	1,133	6,645	24,365
Prev 1y return	1,110	10.10	20.39	−42.66	−26.95	−0.23	11.40	21.32	40.37	64.18
Prev 3y flow	1,110	0.19	0.88	−1.34	−0.86	−0.34	−0.01	0.53	1.93	3.32
Non-equity mutual funds										
Return	1,988	0.81	2.05	−5.07	−1.93	0.00	0.64	1.50	3.84	6.68
CAPM alpha	1,988	0.15	2.02	−5.70	−2.77	−0.41	0.00	0.85	3.18	5.67
TNA ($M)	1,988	1,010	3,968	12	18	57	171	578	4,037	15,264
Prev 1y return	1,988	3.52	4.74	−7.50	−1.89	0.49	2.99	5.58	11.31	17.25
Prev 3y flow	1,988	0.16	0.82	−1.35	−0.87	−0.32	0.00	0.48	1.74	3.05

returns for estimating betas.[24] Overall, the sample contains 4,567 unique funds and 173,189 fund-quarters, with an average of 1,110 funds in each period.

[24]Specifically, for each fund i in each month t, we use a time-series regression over the previous 36 months to estimate beta ($\hat{\beta}_{t-1}$). Then, month t CAPM alpha is estimated as $\alpha_{i,t} = \text{Ret}_{i,t} - \text{rf}_t - \hat{\beta}_{t-1} \cdot (\text{Mkt}_t - \text{rf}_t)$.

Non-equity mutual funds. The control group of non-equity funds is composed of the CRSP mutual funds with objectives that do not start with "E" (equity). We also filter out funds with more than 10% of their holdings in common stock. This placebo sample contains 8,689 unique funds and 310,154 fund-quarters, and there are an average of 1,988 unique funds in each period.[25]

4 Return Predictability Before and After the Reform

In this section, we test our predictions on stock factor returns and mutual fund performance predictability before and after the Morningstar reform.

4.1 Stock Factor Returns

U.S. factors. We start by testing the prediction on U.S. factor returns by estimating a panel regression:

$$\text{Ret}_{f,t} = a \text{ MomType}_f + b \text{ Post2002}_t + c \text{ MomType}_f \times \text{Post2002}_t + \text{Controls}_{f,t} + \epsilon_{f,t} \quad (2)$$

where $\text{Ret}_{f,t}$ is the return of factor f in month t, MomType_f is an indicator of whether the factor f is a momentum-type factor (defined in Section 3.1), and Post2002_t is an indicator that equals one after the Morningstar reform in June 2002.

Regression results are presented in Table 3. In Columns (1) to (3), we use the 49 factors that we constructed. Column (1) contains no additional controls. The results indicate that, relative to other factors, the profitability of momentum-type factors declined by 53 basis points per month after the Morningstar reform. Motivated by the finding that factors can exhibit momentum (Gupta and Kelly, 2019; Ehsani and Linnainmaa, 2022), Column (2) also controls for lagged factor returns over the months of $t-1$, $t-6$ to $t-2$, and $t-12$ to $t-7$.

[25] Because fund-level identifiers for the non-equity sample are not available for most of the sample period, we treat share classes as funds in this sample.

The effect only slightly weakens to 46 basis points per month. In Columns (4) and (5), we repeat the same regressions using the U.S. factors from Jensen et al. (2022) and find broadly similar results.

One concern might be that our results reflect the severe crash that momentum strategies experienced during the 2008–2009 financial crisis (Daniel and Moskowitz, 2016), and not necessarily a persistent decline of momentum profits over the entire post-reform period, as the Morningstar reform implies. To alleviate the concern that our results may be driven by this crash, Columns (3) and (6) of Table 3 report results after excluding the recession period around the financial crisis, defined as January 2008 to June 2009 by the National Bureau of Economic Research (NBER) recession dating committee. As expected, this reduces the size of the coefficient but the effect remains statistically significant at the 5% level, indicating that the decline of momentum profitability exists outside the 2008–2009 financial crisis period, hence is not solely driven by the "momentum crash."

Interestingly, with hindsight, it appears that some results reported in two existing papers (Green, Hand, and Zhang, 2017; Daniel and Moskowitz, 2016) are also consistent with our prediction. Specifically, these papers report results consistent with the fact that momentum-type strategies experienced profitability declines starting in mid-2002, even though testing for this change was not the objective of those papers. Appendix A.4 provides further details. Their results further show that the finding of post-2002 momentum factor return is robust to alternative factor construction methodologies.

International factors. We now test whether the post-reform profitability decline is specific to momentum-type strategies in the U.S.[26] As mentioned in Section 3.1, we use the capped-valued weighted factors data in Jensen et al. (2022).

In the first three columns of Table 4, we focus on momentum-type factors across different

[26] We thank James Choi for this suggestion.

Table 3. Post-Reform Profitability Decline: U.S. Factors

We estimate panel regressions of monthly long-short stock factor returns on the interaction of an indicator for whether a factor is of momentum type (MomType) and an indicator that equals one after the June 2002 Morningstar reform (Post2002). All regressions cluster standard errors by factor. Columns (1) to (3) use the 49 long-short decile factors we construct and Columns (4) to (6) use the 153 long-short tercile factors from Jensen et al. (2022). Columns (1) and (4) do not include additional controls. Columns (2) and (5) also control for lagged factor returns over the months of $t-1$, $t-6$ to $t-2$, and $t-12$ to $t-7$. Columns (3) and (6) further exclude the momentum crash period of January 2008 to June 2009 from the sample. Standard errors are reported in parentheses. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

Dependent variable:	Monthly factor return (%)					
	Our constructed factors			Jensen et al. (2022) factors		
	(1)	(2)	(3)	(4)	(5)	(6)
MomType × Post2002	−0.534***	−0.461***	−0.285**	−0.419***	−0.370***	−0.239**
	(0.137)	(0.118)	(0.135)	(0.098)	(0.086)	(0.096)
MomType	0.423***	0.363***	0.364***	0.359***	0.314***	0.313***
	(0.123)	(0.104)	(0.104)	(0.104)	(0.089)	(0.088)
Post2002	−0.252***	−0.221***	−0.249***	−0.208***	−0.185***	−0.205***
	(0.051)	(0.044)	(0.045)	(0.022)	(0.020)	(0.020)
Lagged factor returns	No	Yes	Yes	No	Yes	Yes
Omit momentum crash	No	No	Yes	No	No	Yes
Obs	18,816	18,816	17,934	58,752	58,752	55,680
Adj R^2	0.008	0.008	0.006	0.002	0.007	0.006

countries and estimate a panel regression:

$$\text{Ret}_{f,c,t} = a\ \text{US}_c + b\ \text{Post2002}_t + c\ \text{US}_c \times \text{Post2002}_t + \text{Controls}_{f,c,t} + \epsilon_{f,c,t} \qquad (3)$$

where $\text{Ret}_{f,c,t}$ is the monthly return of factor f from country c in month t, and US_c is an indicator for whether a factor is based on U.S. stocks. Column (1) does not include any controls, and the results show that the sharp post-2002 decline in momentum profitability is specific to the United States. Momentum-type factors in other countries only experienced a decline of 9 basis points in monthly returns after the reform, but U.S. momentum-type factors experienced an additional 54 basis point decline. Column (2) also controls for lagged factor returns over the months of $t-1$, $t-6$ to $t-2$, and $t-12$ to $t-7$. Column (3) further omits the momentum crash period. The inference is largely unaffected across these specifications.

Table 4. Post-Reform Profitability Decline: International Factors

We estimate panel regressions on monthly returns of long-short stock factors from all countries in Jensen et al. (2022). In Columns (1) to (3), we focus on momentum-type factors and regress their returns on an indicator of the country U.S. (US) and an indicator that equals one after the June 2002 Morningstar reform (Post2002). In Columns (4) to (6), we use all factors and also add a third interaction with an indicator of whether a factor is of the momentum type (MomType). All regressions cluster standard errors by factors. Columns (1) and (4) do not include additional controls. Columns (2) and (5) also controls for lagged factor returns over the months of $t-1$, $t-6$ to $t-2$, and $t-12$ to $t-7$. Columns (3) and (6) further exclude the momentum crash period of January 2008 to June 2009 from the sample. Standard errors are reported in parentheses. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

Dependent variable:	Monthly factor return (%)					
	Momentum-type factors			All factors		
	(1)	(2)	(3)	(4)	(5)	(6)
US × Post2002	-0.543^{***}	-0.542^{***}	-0.561^{***}	-0.148^{***}	-0.136^{***}	-0.150^{***}
	(0.105)	(0.105)	(0.115)	(0.026)	(0.024)	(0.024)
US × MomType × Post2002				-0.395^{***}	-0.364^{***}	-0.359^{***}
				(0.108)	(0.099)	(0.108)
US	0.134	0.130	0.131	0.128^{***}	0.117^{***}	0.115^{***}
	(0.108)	(0.110)	(0.109)	(0.028)	(0.026)	(0.025)
Post2002	-0.085^{*}	-0.081^{*}	0.031	-0.061^{***}	-0.057^{***}	-0.064^{***}
	(0.045)	(0.045)	(0.046)	(0.013)	(0.012)	(0.012)
MomType				0.353^{***}	0.321^{***}	0.316^{***}
				(0.041)	(0.037)	(0.037)
MomType × US				0.006	0.009	0.011
				(0.112)	(0.102)	(0.100)
MomType × Post2002				−0.024	−0.022	0.100^{**}
				(0.046)	(0.042)	(0.043)
Lagged factor returns	No	Yes	Yes	No	Yes	Yes
Omit mom crash	No	No	Yes	No	No	Yes
Obs	65,730	65,730	62,490	542,474	542,474	515,807
Adj R^2	0.000	0.007	0.003	0.005	0.005	0.005

In Columns (4) to (6) of Table 4, we estimate a triple-interaction panel regression using all the factors and including all the countries:

$$\begin{aligned}\text{Ret}_{f,c,t} = {} & a\ \text{US}_c + b\ \text{Post2002}_t + c\ \text{MomType}_f + d\ \text{US}_c \times \text{Post2002}_t + e\ \text{US}_c \times \text{MomType}_f \\ & + f\ \text{MomType}_f \times \text{Post2002}_t + g\ \text{US}_c \times \text{MomType}_f \times \text{Post2002}_t \\ & + \text{Controls}_{f,c,t} + \epsilon_{f,c,t} \end{aligned} \tag{4}$$

Similar to earlier regressions, Column (4) does not include additional controls; Column (5) controls for lagged factor returns, and Column (6) omits the momentum crash period. Across specifications, the triple-interaction coefficient for US $\times$ MomType $\times$ Post2002 is -36 to -40 basis points, indicating that the post-reform profitability decline is concentrated in U.S. momentum-type factors, as predicted. Overall, the results are consistent with the fact that the Morningstar reform only impacted U.S. stock markets and only impacted momentum-type factors. Therefore, the momentum-type factors in the U.S. suffered stronger declines than non-momentum-type factors or factors outside of the U.S.

4.2 Mutual Fund Return Predictability

We now evaluate the prediction that mutual fund performance persistence and dumb money effects should attenuate after June 2002. We sort funds into ten deciles by the corresponding sorting variable—previous twelve-month returns in the case of Carhart (1997) and previous three-year fund flows (times -1) in the case of Frazzini and Lamont (2008). Therefore, the 10$^{\text{th}}$ and 1$^{\text{st}}$ deciles represent funds that are predicted to have the highest and lowest returns according to the original papers, respectively.

We then estimate a panel regression:

$$\begin{aligned}\alpha_{i,t}^{\text{CAPM}} &= a\ \text{Bottom}_{i,t} + b\ \text{Top}_{i,t} + c\ \text{Bottom}_{i,t} \times \text{Post2002}_t \\ &+ d\ \text{Top}_{i,t} \times \text{Post2002}_t + \text{Controls}_{i,t} + \epsilon_{i,t} \qquad (5)\end{aligned}$$

where the dependent variable is the quarterly CAPM alpha of fund i in quarter t,[27] $\text{Bottom}_{i,t}$ and $\text{Top}_{i,t}$ are indicators that equal one if the fund belongs to the top or bottom deciles, and Post2002_t is an indicator that equals one after the Morningstar reform. We control for time and fund fixed effects and cluster standard errors by time and fund. Our main object

[27] We examine CAPM alpha, rather than Fama-French three factor-alpha, because the price effects we examine in this paper are primarily at the size–value style levels, so it is not appropriate to control for size and value effects.

of interest is the difference between coefficients d and c, which represents the post-reform change in the return difference between top- and bottom-ranked funds.

Table 5. Post-Reform Predictability Decline: Mutual Funds

We estimate panel regressions of quarterly mutual fund CAPM alphas on indicator variables and their interactions. "Bottom" and "Top" refers to funds being in the bottom or top decile when ranked based on past one-year return or past three-year flows (times −1) which are, respectively, the sorting variables for Carhart (1997) and Frazzini and Lamont (2008) ("dumb money"). "Post2002" is an indicator that equals one after the Morningstar reform. Columns (1) and (3) are estimated using domestic equity funds while Columns (2) and (4) use non-equity funds. All regressions include quarter and fund fixed effects, and standard errors are also clustered by quarter and fund. Panel A reports regression results. Panel B reports the differences between the "Top" and "Bottom" funds before and after 2002, respectively. The standard errors of these differences are estimated using the delta method. Standard errors are reported in parentheses. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

Panel A: Regressions

Dependent variable:	Fund return (%)			
	Carhart		Dumb money	
Fund sample:	Domestic equity	Non-equity	Domestic equity	Non-equity
	(1)	(2)	(3)	(4)
Top	1.276*	−0.132	0.708***	0.077
	(0.716)	(0.181)	(0.200)	(0.070)
Bottom	−1.584**	−0.066	−0.901***	0.115
	(0.636)	(0.176)	(0.283)	(0.070)
Top × Post2002	−1.710**	0.146	−0.670***	−0.062
	(0.805)	(0.257)	(0.230)	(0.087)
Bottom × Post2002	1.929***	0.269	0.955***	−0.098
	(0.732)	(0.281)	(0.299)	(0.102)
Fixed effects	Yes	Yes	Yes	Yes
Obs	173,189	310,153	173,189	310,153
Adj R^2	0.089	0.373	0.085	0.372

Panel B: Estimated differences

Top − Bottom	2.860**	−0.066	1.609***	−0.037
	(1.221)	(0.332)	(0.385)	(0.119)
(Top − Bottom) × Post2002	−3.639***	−0.123	−1.624***	0.036
	(1.373)	(0.499)	(0.413)	(0.166)

Regression results using U.S. domestic equity funds are shown in Columns (1) and (3) in Panel A of Table 5. In the table, the sorting (into "Bottom" and "Top") of funds are done according to the Carhart and dumb money strategies, respectively. The relevant differences between coefficients are reported in Panel B. Before the Morningstar reform, funds ranked in the top decile based on the Carhart method outperformed those in the bottom decile by 2.86% per quarter, and that outperformance entirely disappeared after the Morningstar reform. Similarly, before the reform, funds ranked in the top decile based on the dumb money method outperformed the bottom-ranked funds by 1.61% per quarter, and that effect also disappeared after the reform. As a placebo test, in Columns (2) and (4), we also estimate the same regressions on non-equity funds and find no such result. Overall, the results are consistent with the idea that the Morningstar reform reduced these two forms of mutual fund return predictability patterns.

Our results on the decline of performance persistence are consistent with Choi and Zhao (2022). The authors of that paper found that the original Carhart (1997) results did not persist out of sample; our findings provide a possible explanation for their findings. In fact, Figure 1 of Choi and Zhao (2022) suggests that the disappearance of the Carhart (1997) effect started around 2002. Specifically, their Figure 1 plots a ten-year rolling window of CAPM alpha of the long-short return spread of mutual funds sorted using the Carhart criterion. The positive alpha declined around 2005 and became close to zero around 2011, which is approximately ten years after the Morningstar reform.

5 Event Study Around the Rating Reform

So far, we have focused on testing predictions for changes in long-term expected returns. While the results are consistent with our proposed mechanism, definitive causal identification is difficult to achieve over long sample periods because one cannot control for all the possible determinants of factor and fund returns. Naturally, the same criticism applies to all existing attempts to explain long-term expected returns using preferences, sentiment, or other mechanisms.

While, in asset pricing, it is probably impossible to obtain definitive identification over long samples, it may be possible to isolate a causal effect in quasi-natural experiments spanning short samples. In this section, we use an event-study approach to zoom in on a short one-year window (January to December 2002) around the reform event to examine whether style-level rating changes *can* have a first-order impact on factor returns and fund returns. Over this short period, fund rating changes are predominantly caused by the rating reform itself. Further, we can reduce the chance that returns are impacted by other events such as the NYSE decimalization in early 2001 or the introduction of NYSE auto-quoting in 2003 (Hendershott, Jones, and Menkveld, 2011)—all of which may plausibly impact factor returns in the long run. None of these other events fall within our event study window.

Which factors and funds are most affected by the rating reform in this short window? As discussed in Section 2.4, over the long run, we expect momentum-type factors to be most affected. However, the impact of rating-induced trading is time-varying, so the impact of the reform in the short term depends on whichever factors and funds happened to "load" on styles that experience the largest reform-induced rating drops at the end of June 2002.

In Section 5.1 we describe the mapping of ratings and flows into stocks and then aggregating at the factor level. In Section 5.2, we devise a method to measure the "reform exposure" of factors and funds; we then conduct the event study in Section 5.3. The fact that the factors and funds with the highest short-term and long-term exposures are not the same subsets is empirically useful: it means this event study is an independent test, rather than just a derivative of the long-term panel regressions in the previous section.

5.1 Mapping Ratings and Flows into Factors

Our mechanism focuses on the impact of ratings and fund flows on factors, so we start by measuring these variables at the factor level. Because these calculations require having stock characteristics that underlie factor construction, we use our U.S. factors for this exercise, as the Jensen et al. (2022) data only provide factor returns but not the stock characteristics used for forming factor portfolios.

Matching fund ratings to stocks via mutual fund holdings. Recall that ratings are assigned to mutual funds, and factors are exposed to ratings through the mutual funds that hold stocks in the factors. Therefore, our calculation needs to go through mutual fund holdings.

We obtain quarterly fund holdings from Thomson Reuters' S12 which is based on 13F filings. We download Morningstar ratings and fund style categories from Morningstar Direct, and we merge them with the CRSP fund flows data using the matching table from Pástor, 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 et al. (2016).

Measuring factor-level ratings and flow-induced trading (FIT). Because factors are defined as long-short stock portfolios, we first measure ratings and flows at the stock level and then aggregate them up to the factor level.

For each stock i in month t, we define its rating as the holding-weighted rating of all funds $\mathcal{J}$ that hold the stock:

$$\text{Rating}_{i,t} = \frac{\sum_{\text{fund}\, j \in \mathcal{J}} \text{SharesHeld}_{i,j,t-1} \cdot \text{Rating}_{j,t}}{\sum_{\text{fund}\, j \in \mathcal{J}} \text{SharesHeld}^{\text{fund}}_{i,j,t-1}} \tag{6}$$

Similarly, we follow Lou (2012) to calculate flow-induced trading (FIT) for each stock i in each month t:

$$\text{FIT}_{i,t} = \frac{\sum_{\text{fund}\, j \in \mathcal{J}} \text{SharesHeld}_{i,j,t-1} \cdot \text{Flow}_{j,t}}{\sum_{\text{fund}\, j \in \mathcal{J}} \text{SharesHeld}_{i,j,t-1}}. \tag{7}$$

Here, the 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).[28] In short, FIT is the total amount of nondiscretionary mutual fund trading in stock i caused by fund flows. As

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

argued 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.

We then aggregate rating FIT at the factor level. For each factor f in each month t, we compute:

$$\text{Rating}_{f,t} = \sum_{i \in \text{top quintile}} w^f_{i,t-1} \text{Rating}_{i,t} - \sum_{i \in \text{bottom quintile}} w^f_{i,t-1} \text{Rating}_{i,t} \tag{8}$$

$$\text{FIT}_{f,t} = \sum_{i \in \text{top quintile}} w^f_{i,t-1} \text{FIT}_{i,t} - \sum_{i \in \text{bottom quintile}} w^f_{i,t-1} \text{FIT}_{i,t} \tag{9}$$

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

5.2 Predicting Reform Impact on Factors in the Event Study

We now use data in December 2001, which is the last month *prior to* the event study window, to predict how each factor's rating will be affected by the reform. Specifically, for each fund j, we estimate how its rating will change due to the reform:

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

where the two terms on the right are our estimates of fund ratings using December 2001 data under the two different Morningstar methodologies. We then aggregate the fund-level predicted rating changes to the stock-level (Equation (6)), and then aggregate the stock-level ratings to the factor-level (Equation (8)) or the mutual fund portfolio-level in the equivalent fashion.[29]

Appendix A.3 shows further details on predicting factor-level rating changes. The prediction is rather accurate, as the predicted factor-level rating change explains the actual factor-level rating change in the month of June 2002 with an R^2 of 84%.

[29]Because we know the holdings of each mutual fund, we can simply treat each mutual fund as a factor portfolio and compute holding value-weighted average stock ratings.

5.3 Event Study

Stock factors. We start the event study by focusing on stock factors. Using the method explained above, the 49 factors are sorted into quintiles based on their predicted rating changes. Figure 3 presents the results. Panel (a) plots the 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 an increase of 0.19 stars. The rating changes happened abruptly in June 2002, a fact that corroborates the accuracy of the predicted rating changes based on December 2001 data (Appendix A.3).

Panels (c) and (e) plot cumulative monthly factor FIT and returns around the event, respectively. Quintile 1—the factors that benefited the most from rating-induced flows in the months leading to the reform—experienced a decline of 1% in monthly FIT and a sharp decline of −3.7% in monthly returns. Conversely, quintile 5 experienced an increase of 0.14% in monthly FIT and a slight increase of 0.75% in monthly returns.[30]

To alleviate the concern that the return and FIT changes could be mechanical, we perform placebo tests by conducting the same exercise in all years other than 2002. The placebo results for ratings, FIT, and return changes are shown as the white bars in Panels (b), (d), and (f), respectively. 95% confidence intervals are also shown. The results show that the patterns found around the reform month of June 2002 are unique to that year, suggesting that the results are not mechanical.

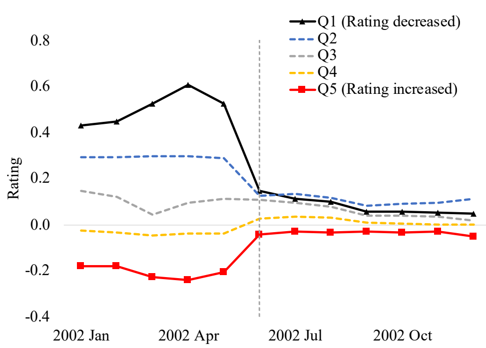
Mutual funds. To examine how the event study impacted mutual funds, we also perform a similar event study using fund returns as opposed to factor returns. Specifically, we first compute the holding-weighted predicted rating changes for all U.S. domestic equity mutual funds and sort them into quintiles, similar to how we sorted factors. Then, we examine the behavior of ratings, FIT, and returns of the sorted mutual fund portfolios in Figure 4.

[30] In a companion paper, we show that the implied style-level price impact coefficient (the reciprocal of demand elasticity) is approximately 5 (Ben-David et al., 2022a). 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, 2021).

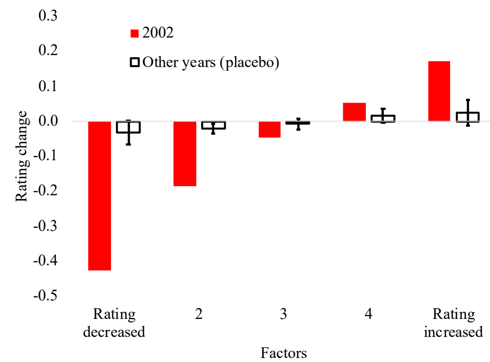
Figure 3. U.S. Stock Factors around the June 2002 Event

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, FIT, and return changes after June (the average of July to December 2002 minus the average of January to June 2002), and the white bars plot the corresponding results for years other than 2002. The whiskers represent 95% confidence intervals. To focus on cross-sectional dispersion, all variables—ratings, returns, and flows—are demeaned within month.

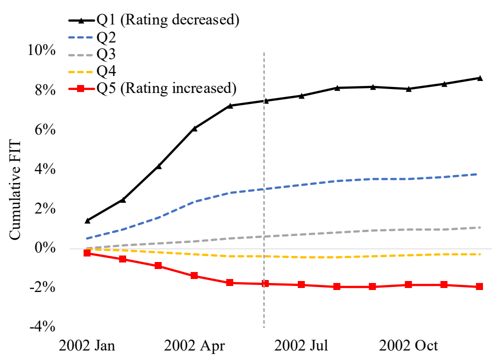
(a) Rating in 2002



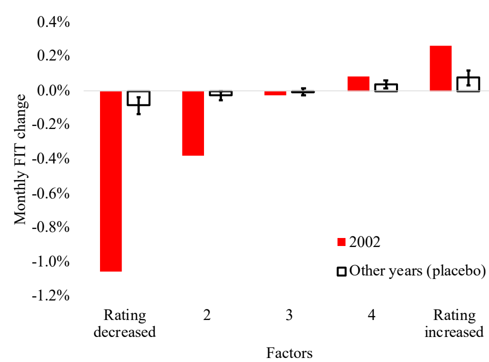
(b) Rating change, 2002 vs. placebo



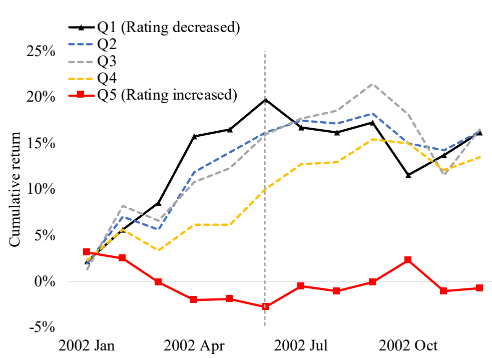
(c) Cumulative FIT in 2002



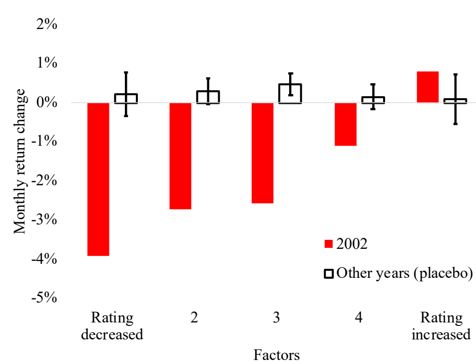
(d) FIT change, 2002 vs. placebo



(e) Cumulative return in 2002



(f) Return change, 2002 vs. placebo



The figure shows that mutual funds that were predicted to suffer rating decreases indeed saw declines in ratings, FIT, and returns. The right panels of the figure also compare the 2002 changes against other years. The placebo tests show that the patterns observed around 2002 are indeed unique to that year. Overall, the results presented in these two event studies suggest that the rating reform causally impacted factor returns as well as mutual fund returns in a significant and predictable way.

5.4 Alternative Explanations for the Event Study Results

We now discuss the concern that the factor and mutual fund return fluctuations around June 2002 may have been triggered by changes other than the Morningstar reform.[31]

Arbitrage activity. One natural worry is whether arbitrage forces in these factors suddenly became 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 declined after the strategies were published in academic papers and link it to arbitrage actions. Relatedly, Lou and Polk (2022) show that a return-based measure of arbitrageur activity negatively predicts 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.[32] The authors combine the long and short positions into a net position and subtract the past four-quarter average

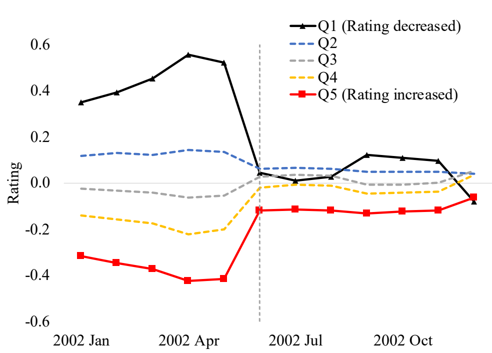
[31] The results on factors and mutual funds are similar so we only report the former for brevity.

[32] 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. Note that, while the short side of NAT is updated monthly, the long side relies on 13F holdings and is only updated quarterly.

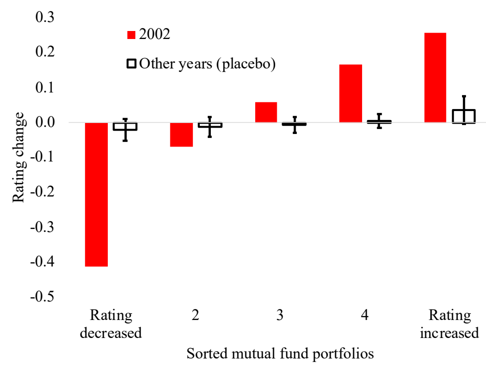
Figure 4. Domestic Equity Mutual Funds around the June 2002 Event

This figure is the equivalent of 3 using mutual funds. We perform event studies on all U.S. domestic equity mutual funds using a 12-month window around the reform event (January to December 2002). In the left panels, we sort mutual funds 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, FIT, and return changes after June (the average of July to December 2002 minus the average of January to June 2002), and the white bars plot the corresponding results for years other than 2002. The whiskers represent 95% confidence intervals. To focus on cross-sectional dispersion, all variables—ratings, returns, and flows—are demeaned within month.

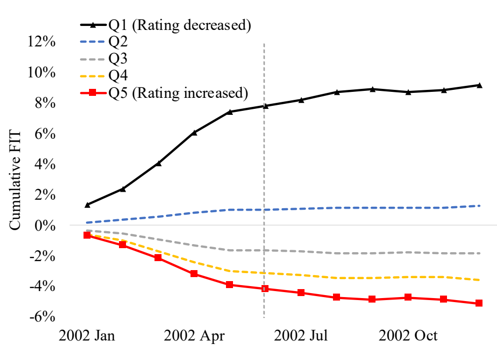
(a) Rating in 2002



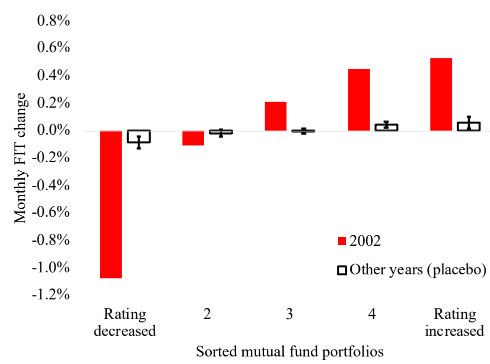
(b) Rating change, 2002 vs. placebo



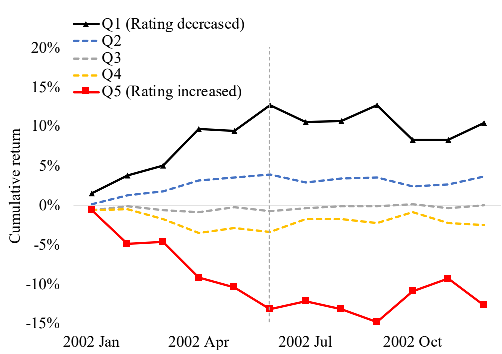
(c) Cumulative FIT in 2002



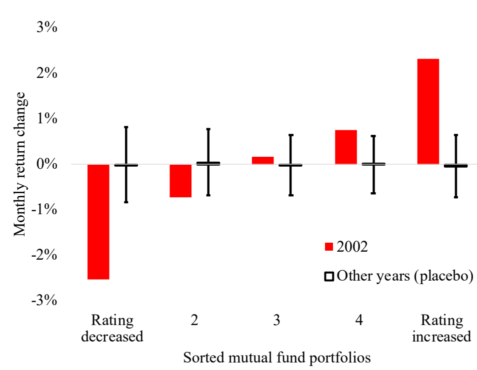
(d) FIT change, 2002 vs. placebo



(e) Cumulative return in 2002



(f) Return change, 2002 vs. placebo



to arrive at a measure of arbitrageur position changes, which they call NAT. We follow this methodology in computing stock-level NAT and aggregate it at the factor level.

Second, we follow Lou and Polk (2022) 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 generated by arbitrageurs trading in the factor.[33] We compute this measure for all factors.[34]

We plot the evolution of these measures in the 12-month event window in Figure 5. As in Section 5, 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.[35] 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

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

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

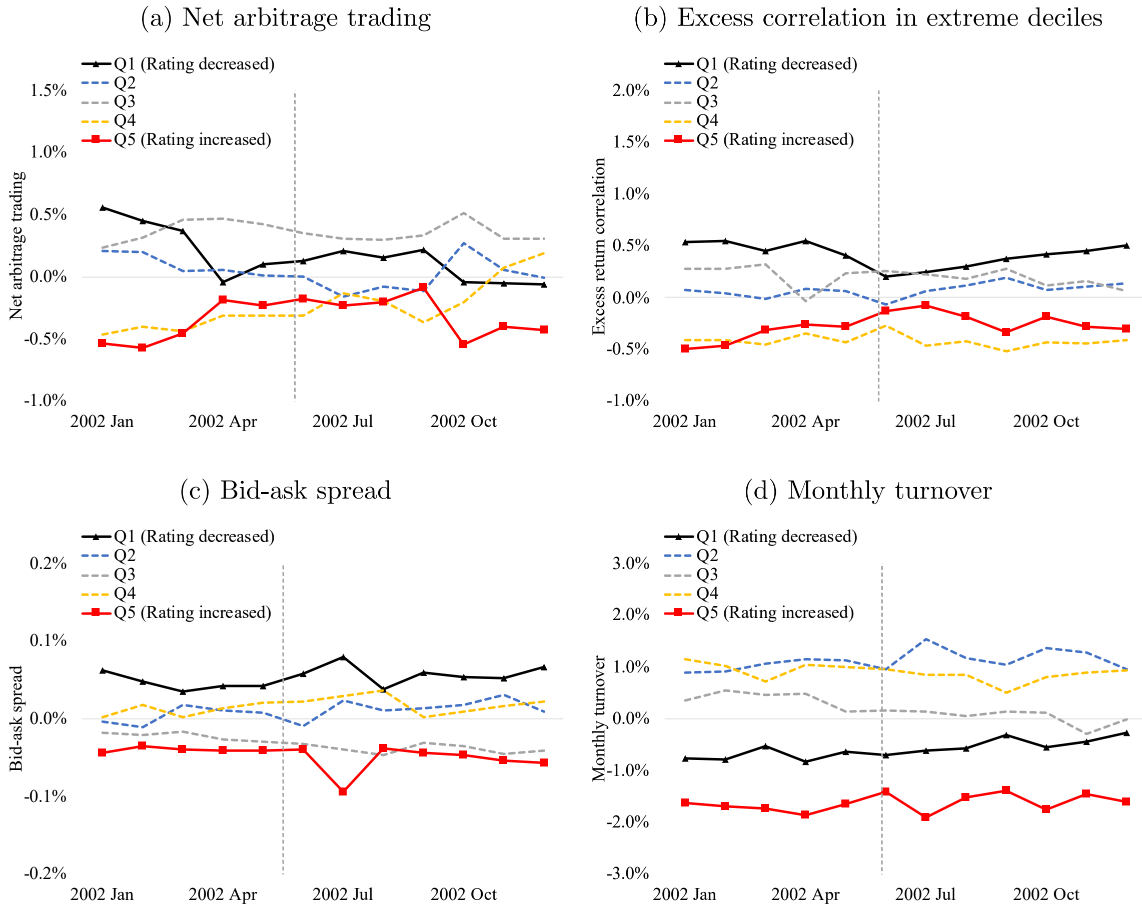
where N^L and N^S are the number of stocks in the long and short leg portfolios, respectively. To compute the partial return correlations, they first subtract Fama-French 30 industry returns from weekly stock returns and then regress the residuals on the Fama-French three factors to obtain alphas. Finally,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

[34]As a sanity check on our replication of their methodology, consistent with Lou and Polk (2022), we find that this measure indeed negatively predicts returns of factors in the momentum category.

[35]Increasing liquidity may explain factor profitability declines through two possible 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 5. Alternative Explanations: Other Influences around 2002

As in Figure 3, 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 (2022). Panel (c) plots the 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 Morningstar methodology change event.



no clear change around the event.

In summary, we do not find around June 2002 any noticeable change in arbitrage trading activity or liquidity, two major forces that could impact factor returns. Thus, the event study supports the idea that Morningstar rating changes can exert a tangible price impact on factor returns.

6 Conclusion

While asset pricing researchers generally agree that demand shocks can impact asset prices, it is less clear whether demand matters for systematic patterns in expected returns. In this paper, we use a natural experiment to demonstrate that a demand effect caused by institutional frictions can have a first-order impact on expected returns in stock factors and mutual funds. Specifically, we show that a seemingly innocuous change in Morningstar's rating methodology led to a disruption of mutual fund positive feedback trading at the style level. After the reform, momentum-type stock factors—which benefit from positive feedback trading—experienced a significant decline in profitability, and the observed decline is above and beyond that experienced by non-momentum-type factors. In mutual funds, we also find that the Carhart (1997) performance persistence and Frazzini and Lamont (2008) dumb money effect weakened after the reform, both of which are expected consequences of the Morningstar reform-induced demand changes. We further show that these changes in expected stock and mutual fund returns are specific to U.S. stocks, which is consistent with the fact that the Morningstar reform only impacted the U.S. stock market.

More broadly, our findings join a growing number of studies indicating that demand effects can drive systematic price movements (see literature review in Gabaix and Koijen, 2021). Our paper focuses on the role of the Morningstar ratings reform since it allows sharper inference. However, it is possible, and even likely, that the role of correlated demand and positive feedback trading, arising from other institutional features or frictions, may be even more consequential for asset pricing than documented here and previously believed. Therefore, unlike the assumption embedded in classical "frictionless" asset pricing, demand effects may be a first-order driver of asset prices (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.

Adelino, Manuel, Sophia Chiyoung Cheong, Jaewon Choi, and Ji Yeol Jimmy Oh, 2022, Mutual fund flows and capital supply in municipal financing, Working paper, National Bureau of Economic Research.

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, Working paper, Arizona State University.

Arnott, Robert D., Mark Clements, Vitali Kalesnik, and Juhani T. Linnainmaa, 2021, Factor momentum, Working paper, Dartmouth College.

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

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

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* 67, 6629–6659.

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, 2022a, Ratings-driven demand and systematic price fluctuations, *Review of Financial Studies* forthcoming.

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, 2022b, What do mutual fund investors really care about?, *Review of Financial Studies* 35, 1723–1774.

Berk, Jonathan .B, and Richard C. Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269–1295.

Blitz, David, Joop Huij, and Martin Martens, 2011, Residual momentum, *Journal of Empirical Finance* 18, 506–521.

Brown, David C., Shaun Davies, and Matthew Ringgenberg, 2021, ETF arbitrage and return predictability, *Review of Finance* 25, 937–972.

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

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

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, 2022, Open source cross-sectional asset pricing, *Critical Finance Review* 27, 207–264.

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

Cho, Thummim, 2020, Turning alphas into betas: Arbitrage and endogenous risk, *Journal of Financial Economics* 137, 550–570.

Choi, James J., and Kevin Zhao, 2022, Carhart (1997) mutual fund performance persistence disappears out of sample, *Critical Finance Review* forthcoming.

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

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

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

Ehsani, Sina, and Juhani T. Linnainmaa, 2022, Factor momentum and the momentum factor, *Journal of Finance* forthcoming.

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.

Falck, Antoine, Adam Rej, and David Thesmar, 2021, Why and how systematic strategies decay?, Working paper, Massachusetts Institute of Technology.

Frazzini, Andrea, and Owen A Lamont, 2008, Dumb money: Mutual fund flows and the cross-section of stock returns, *Journal of Financial Economics* 88, 299–322.

Gabaix, Xavier, and Ralph S.J. Koijen, 2021, 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.

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

Han, Xiao, Nikolai L Roussanov, and Hongxun Ruan, 2021, Mutual fund risk shifting and risk anomalies, Working paper, University of Pennsylvania.

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.

Hartzmark, Samuel M., and David H. Solomon, 2021, Predictable price pressure, Working paper, University of Chicago.

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.

Heston, Steven L, and Ronnie Sadka, 2008, Seasonality in the cross-section of stock returns, *Journal of Financial Economics* 87, 418–445.

Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating anomalies, *Review of Financial Studies* 33, 2019–2133.

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.

Jensen, Theis Ingerslev, Bryan T. Kelly, and Lasse Heje Pedersen, 2022, Is there a replication crisis in finance?, *Journal of Finance* forthcoming.

Jones, Christopher S, and Haitao Mo, 2021, Out-of-sample performance of mutual fund predictors, *Review of Financial Studies* 34, 149–193.

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.

Kim, Sanghyun Hugh, 2020, Do mutual funds manipulate star ratings? Evidence from portfolio pumping, Working paper, University of Texas at Dallas.

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

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.

Lewellen, Jonathan, 2015, The cross-section of expected stock returns, *Critical Finance Review* 4, 1–44.

Li, Jiacui, 2021, What drives the size and value factors?, Working paper, University of Utah.

Li, Jiacui, and Zihan Lin, 2022, Prices are less elastic at more aggregate levels, Working paper, Univesity of Utah.

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

Lou, Dong, and Christopher Polk, 2022, Comomentum: Inferring arbitrage activity from return correlations, *Review of Financial Studies* forthcoming.

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.

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.

Pástor, Ľuboš, Robert F. Stambaugh, and Lucian A. Taylor, 2020, Fund tradeoffs, *Journal of Financial Economics* 138, 614–634.

Pavlova, Anna, and Taisiya Sikorskaya, 2021, Benchmarking intensity, Working paper, London Business School.

Peng, Cameron, and Chen Wang, 2021, Factor demand and factor returns, Working paper, University of Notre Dame.

Reuter, Jonathan, and Eric Zitzewitz, 2021, How much does size erode mutual fund performance? A regression discontinuity approach, *Review of Finance* 25, 1395–1432.

Schmickler, Simon, 2020, Identifying the price impact of fire sales using high-frequency surprise mutual fund flows, Working paper, Princeton University.

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.

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.

Zheng, Lu, 1999, Is money smart? A study of mutual fund investors' fund selection ability, *Journal of Finance* 54, 901–933.

Appendix A Additional Empirical Results

Appendix A.1 and Appendix A.2 explore the mechanism of how the Morningstar reform impacts style-level positive feedback trading. Appendix A.3 provides further details about predicting rating changes in the event study of Section 5. Appendix A.4 shows corroborating results from previous studies on the decline of momentum-type factor profits after mid-2002.

A.1 The Morningstar Reform Only Altered Style-Level Positive Feedback Trading

Section 2.2 explains that the Morningstar reform impacts style-level positive feedback trading. A natural question is: what about the effects of fund-level ratings? Shouldn't that continue to impact idiosyncratic-level positive feedback trading?

In this section, we show that the reform does not impact idiosyncratic-level positive feedback trading. To influence how fund trading depends on past returns, the reform needs to create a tangible impact in fund ratings to cause changes in flows. The average stock is held by 78.5 funds (see Table 2), so for any given stock, there has to be a *correlated* change in the ratings of funds holding that stock in order to generate sufficiently large rating-induced flow pressure. Therefore, while past style-level returns—which can induce correlated fund return changes—can have a large impact on a stock's rating, past idiosyncratic stock returns do not.

For a concrete example, consider a small-cap growth stock that is held by many small-cap growth funds. Suppose the stock's idiosyncratic return was high in the recent past. Because that stock is only a small part of each fund's portfolio, this shock is unlikely to have a sufficiently large effect on fund ratings. In contrast, suppose the style-level (small-cap growth) return was high in the recent past. Under the pre-reform methodology, this means that all small-cap funds would have performed well and thus receive higher ratings, leading to more positive feedback fund flows across small-cap growth stocks. After the methodology reform, this style-level positive feedback trading became muted by design.

Figure A.1. Morningstar Reform Only Impacted Style-Level Positive Feedback

This figure plots the panel regression coefficients of stock-level ratings (Equation (6)) on the past 36 lags of monthly stock returns, which have been decomposed into style-level returns (3 × 3 Fama-French size–book/market styles) and idiosyncratic-level returns (the residual). Panels (a) and (b) plot the regression coefficients, and the shaded areas represent 95% confidence intervals. The regressions control for month fixed effects and cluster standard errors by month.

(a) Before the 2002 Morningstar reform

(b) After the 2002 Morningstar reform

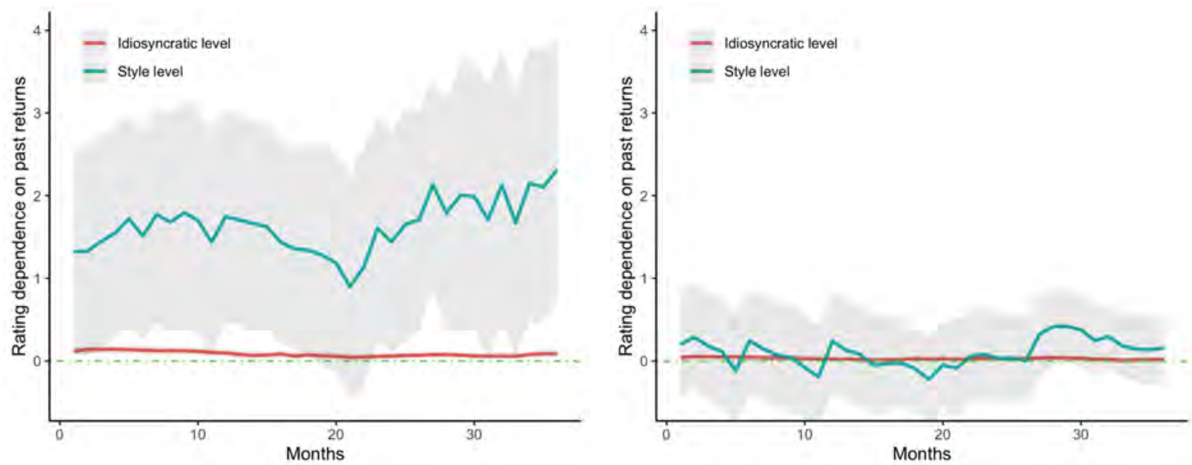


Figure A.1 illustrates these points using panel regressions of stock-level ratings (Equation (6))—defined as the holdings-weighted average rating of all funds that hold the stock—on the past 36 monthly lags of stock returns. To separately estimate the impact of different return components, we decompose each stock's return into

$$\text{Ret}_{i,t} = \text{StyleRet}_{i,t} + \text{IdiosyncraticRet}_{i,t}, \tag{11}$$

where $\text{StyleRet}_{i,t}$ is defined as the value-weighted return of the corresponding 3 × 3 size–book/market style portfolio, and $\text{IdiosyncraticRet}_{i,t}$ is the residual. We regress stock ratings on 36 lags of each of these two components, controlling for month-fixed effects, and plot the coefficients in Figure A.1. Panel (a) shows that before the reform, stock ratings heavily depended on past style-level returns but not idiosyncratic returns. This confirms that the Morningstar-induced positive feedback trading happens exclusively at the style level.

Panel (b) shows that after the reform, the rating dependence on past style returns becomes muted. It is worth noting that the rating dependence on past stock returns is close to zero both before and after the reform.

A.2 Price Multipliers are Larger at the Style-Level

As discussed in Section 2.3, a number of studies have shown that style-level price multipliers are larger than that at the idiosyncratic level. We also examine this in the context of fund flow-induced price effects.

We follow Lou (2012) to compute flow-induced trading (FIT) at the stock-level, as described in Section 5.1. In order to measure price multipliers, instead of normalizing FIT by the number of shares held, we normalize it here by the number of shares outstanding. Then, we decompose stock-level FIT into two components:

$$\text{FIT}_{i,t} = \text{StyleFIT}_{i,t} + \text{IdiosyncraticFIT}_{i,t} \tag{12}$$

where $\text{StyleFIT}_{i,t}$ is the value-weighted average FIT of the 3×3 size–book/market style that the stock belongs to, and $\text{IdiosyncraticFIT}_{i,t}$ is defined as a residual. We construct the 3×3 portfolios using NYSE break points in the stock characteristics from Chen and Zimmermann (2022). To avoid microcap stocks, we filter out stocks with market capitalization below the 20th NYSE percentile, following Lewellen (2015) and Hou et al. (2020).

To estimate price multipliers, we estimate regressions on stock returns:

$$\text{Ret}_{i,t} = a + b_{\text{style}} \cdot \text{StyleFIT}_{i,t} + b_{\text{idiosyncratic}} \cdot \text{IdiosyncraticFIT}_{i,t} + \epsilon_{i,t} \tag{13}$$

and compare the multiplier estimates b_{style} and $b_{\text{idiosyncratic}}$. The results are reported in Table A.1. In Columns (1) and (2), we estimate Fama-MacBeth regressions. In Columns (3) and (4), we estimate panel regressions with time and stock fixed effects, and cluster standard errors by time and stock. Columns (1) and (3) use quarterly data which are available from

1980, while Columns (2) and (4) use monthly data which are available from 1991. The last row of the table conducts a t-test between the two coefficients.

Table A.1. Estimates of fund flow-induced price multipliers

We estimate the price multipliers associated with fund flow-induced trading. We first follow Lou (2012) to compute stock-level flow-induced trading (FIT), defined as the amount of aggregate mutual fund trading due to mutual fund managers adjusting their holdings in response to fund flows. We then separate FIT into two components, the first being the value-weighted average at the 3×3 size–book/market portfolio level and the second being an idiosyncratic residual. To estimate price multipliers, we regress contemporaneous stock returns on style and idiosyncratic FIT using Fama-MacBeth regressions in Columns (1) and (2), as well as panel regressions with time and stock fixed effects in Columns (3) and (4). Columns (1) and (3) use quarterly data which are available since 1980. Columns (2) and (4) use monthly data which are available from 1991. We cluster standard errors by time and stock for panel regressions. Panel A reports regression results. Panel B reports the difference between the style- and idiosyncratic-level coefficient estimates. Standard errors are reported in parentheses. Levels of significance are presented as follows: *p<0.1; **p<0.05; ***p<0.01.

Panel A: Regressions

Dependent variable:	Stock return (%)			
	Fama-MacBeth regression		Panel regression	
	Quarterly	Monthly	Quarterly	Monthly
	(1)	(2)	(3)	(4)
Style FIT	6.01**	6.29***	11.83***	10.97***
	(2.44)	(1.24)	(1.75)	(1.88)
Idiosyncratic FIT	3.48***	3.72***	3.67***	4.01***
	(0.29)	(0.22)	(0.38)	(0.40)
Time and stock FE	N/A	N/A	Yes	Yes
Sample period	1980–2018	1990–2018	1980–2018	1990–2018
Obs	333,114	832,478	333,114	832,478
R^2	0.229	0.178	0.225	0.173

Panel B: Estimated differences

Style − Idiosyncratic	2.53	2.56**	8.16***	6.96***
coefficient difference	(2.45)	(1.26)	(1.79)	(1.92)

While there is some variation in the coefficient estimates, a clear pattern emerges: the style-level multipliers appear significantly larger than the idiosyncratic-level multipliers. Their differences are statistically significant at 1% level for all specifications. As pointed out by Schmickler (2020), these price multiplier estimates may be biased upwards due to possible reverse causality concerns. However, to the extent that reverse causality does not differ significantly between the style and idiosyncratic levels, it would be sensible to *compare* these coefficients. Combined with the finding in the existing literature that supports larger style-level than idiosyncratic multipliers (Gabaix and Koijen, 2021; Li and Lin, 2022; Ben-David et al., 2022a; Peng and Wang, 2021), we argue that our findings indicate that the same is likely true in the context of flow-induced price effects.

A.3 Accuracy of Factor Rating Change Prediction

In this section, we examine the accuracy of the factor-level rating change prediction in Section 5.2.

We first illustrate the predictions in Panels (a) and (b) of Figure A.2. The 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 in factor-level ratings of factors over a few months are small, the predicted rating change using December 2001 data ends up being a reasonable predictor of the actual rating change that occurred 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%.

A.4 Previous Studies Related to Momentum Profitability Decline

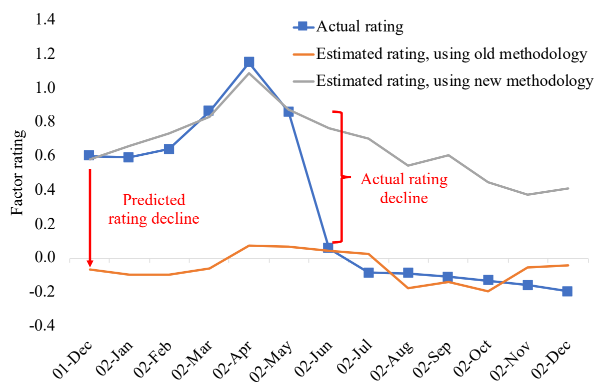
We note that earlier studies have also shown evidence that suggests post-2002 return declines, even though detecting profitability changes is not their objective. For the reader's convenience, we present screenshots from those papers in Figure A.3.

Panel (a) shows a chart from Green et al. (2017) summarizing the average performance (equally-weighted as well as value-weighted) of 94 characteristics. Methodologically speaking, their result is closer to the factor momentum strategy discussed in Arnott et al. (2021), which

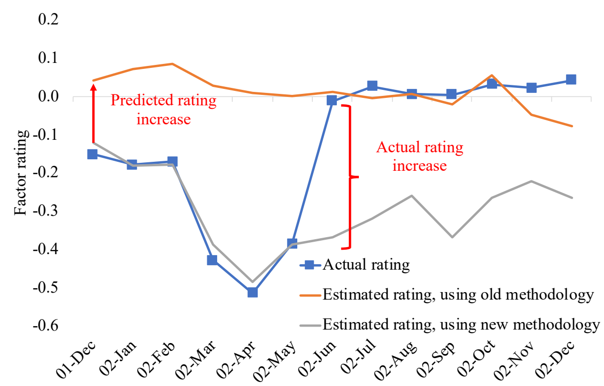
Figure A.2. Predicting Factor Rating Changes at the 2002 Reform Event

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 the 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. Panels (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.

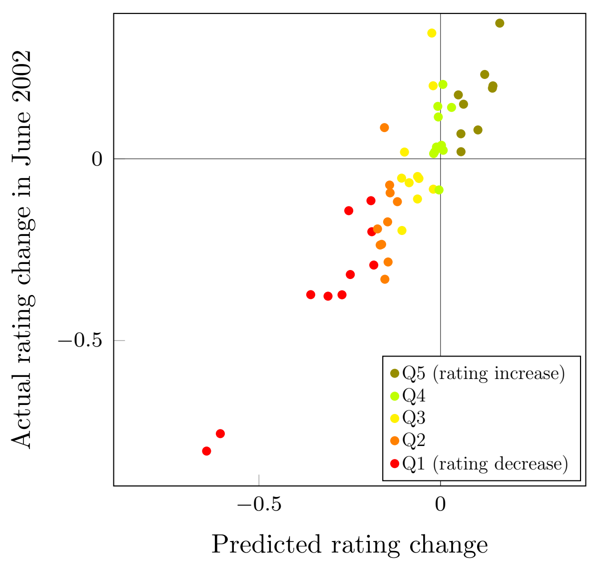
(a) Most negatively affected factor



(b) Most positively affected factor



(c) Accuracy of rating change prediction



Ehsani and Linnainmaa (2022) show to be highly related to stock momentum.[36] Panel (b)

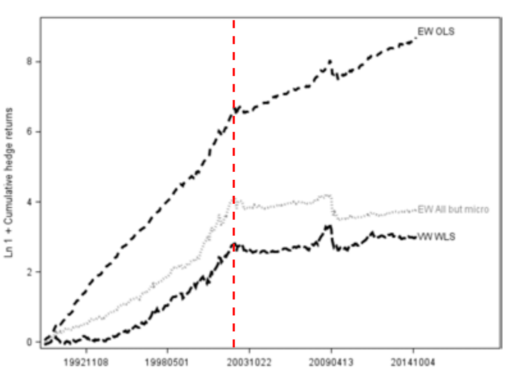
[36]Specifically, they investigate the profits to predicting stock returns based on rolling multivariate Fama-MacBeth regressions with many stock characteristics. Therefore, their strategy ends up going long on characteristics that recently performed well and short on those that performed poorly—which is similar to how the factor momentum strategy is constructed. Even though they investigate characteristics and do not form factors, Cochrane (2011) notes that"portfolio sorts are really the same thing as nonparametric cross-sectional regressions," so the Green et al. (2017) findings also shed light on factor-based results.

shows a chart from Daniel and Moskowitz (2016) summarizing the performance of the stock momentum strategy. In both charts, we added a dashed line for June 2002. Also, in both cases, we see a clear change in the profitability of the strategies after the reform.

Figure A.3. Prior Evidence of Momentum-Type Strategy Profitability Decline

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)

Appendix B Data and Measures

Table B.1 shows the list of 49 U.S. asset pricing factors we construct. Following Hou et al. (2020), we classify them into six categories: intangibles, investment, momentum, profitability, trading frictions, and value/growth.

Table B.1. Our U.S. Stock Factors

The table lists the 49 U.S. stock factors we construct in this study. The first column classifies the factors into six categories, based on Hou et al. (2020). The second column is the factor name and the third column lists the first academic paper published on the factor.

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.