JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS

doi:10.1017/S0022109023000959

Discontinued Positive Feedback Trading and the Decline of Return Predictability

Itzhak Ben-David
The Ohio State University Fisher College of Business and NBER
ben-david.1@osu.edu (corresponding author)

Jiacui Li
The University of Utah Eccles School of Business
jiacui.li@eccles.utah.edu

Andrea Rossi
The University of Arizona Eller College of Management
rossi2@arizona.edu

Yang Song
University of Washington Foster School of Business
songy18@uw.edu

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.

I. Introduction

Understanding the sources of predictability in securities' returns is a central theme in asset pricing. Traditionally, return predictability in security portfolio returns has been ascribed to risk exposure, and performance predictability in

We thank Nicholas Barberis, Huaizhi Chen, Thummim Cho, Lauren Cohen, Richard Evans (the referee), Thierry Foucault (the editor), Sylvester Flood (Morningstar), Umit Gurun, Bing Han, Paul Kaplan (Morningstar), Dong Lou, Chris Malloy, Christopher Schwarz, John Shim, Andrei Shleifer, Yang Sun, Motohiro 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, 2023 AFA, 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 a co-founder and a partner in an investment advisor.

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 (2022)). Given the advances in the demand-based framework, it is important to assess whether systematic and persistent changes in expected returns can be generated by investors' demand.

This article studies the impact of a significant shift in investor demand on predictability patterns in equity factors and mutual fund returns.[2] 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, that is, predictability based on past performance (Carhart (1997)) and on past flows (the "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 of mutual fund predictability patterns. Overall, our analysis shows that demand effects caused by institutional frictions can be a first-order determinant of the cross section of expected returns.

Our identification strategy builds on the finding that mutual fund investors tend to chase past performance as reflected in Morningstar star ratings (Evans and Sun (2021), Ben-David, Li, Rossi, and Song (2022b)) and exploits a methodological reform in Morningstar's ratings that took place in 2002. Until mid-2002, Morningstar equity fund ratings were based on a universal ranking of past fund performance. Since past performance is highly correlated with investment styles, 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; 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 positive feedback trading caused by Morningstar's reform is an opportunity to study the impact of investors' demand on the predictability of equity factor returns and 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 could also significantly impact the predictability of mutual fund returns.

Our empirical analysis proceeds in two parts. In the first part, we test these predictions for both stock factors and mutual funds. In the U.S. stock market, using

[1]The risk-based perspective for understanding expected security 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]By "factor," we mean a characteristic-sorted long–short portfolio that has a nonzero expected return. It could equivalently be called a "characteristics-based portfolio" or "anomaly portfolio." We do not take a stance on whether covariance with such factors explains the cross section of returns.

https://doi.org/10.1017/S0022109023000959 Published online by Cambridge University Press

either a list of 49 commonly used stock factors we construct or 153 factors from Jensen, Kelly, and Pedersen (2023), we find that momentum-related factors experienced sizeable profitability declines after June 2002. While many other return factors also experienced profitability decay,[3] the decline in momentum is much more pronounced, consistent with our predictions. Across a number of specifications, the most conservative estimate is that the monthly returns of momentum-type factors declined by a statistically significant 23.9 basis points more than the returns of other factors.

Since the reform in Morningstar's rating system was limited to U.S. equity funds, we contrast the decline in U.S. momentum strategies to factors in other regions. We use non-U.S. stock factors from Jensen et al. (2023) to conduct a "triple-difference" regression analysis. Consistent with the proposed mechanism, only momentum-related factors in the U.S. experienced significant profitability declines. Momentum factors in other countries instead showed no such decline. In addition, our results are robust to excluding the "momentum crash" period (Daniel and Moskowitz (2016)). Taken together, the evidence suggests that since June 2002, there has been a *persistent* change in momentum strategy returns confined to the U.S. equity market.

Next, we study the predictability of mutual fund 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 shrank after the reform and the prior finding of long-term style-level flow-induced return reversals (Teo and Woo (2004), Froot and Teo (2008), Wahal and Yavuz (2013), Ben-David, Li, Rossi, and Song (2022a), and Li (2022)), we also expect the Frazzini and Lamont (2008) "dumb money effect" (the finding that funds with high (low) recent 3-year flows subsequently underperform (outperform)) to become weaker after the reform.[4] Our results show that both forms of fund performance predictability declined after the rating reform. Consistent with the proposed mechanism, we do not find statistically significant changes in the predictability of returns in equity funds outside the U.S. or in non-equity U.S. mutual funds.

In the second part of the analysis, we zoom in on a narrow window around the June 2002 reform to provide direct evidence that Morningstar rating changes exert a first-order impact on factor and fund returns. While controlling for all alternative hypotheses is usually not possible when studying expected returns over a long period of time, one can achieve more robust identification when examining the

[3]Prior work has identified many other mechanisms that may cause factor returns to decline across the board, including changes in liquidity (Khandani and Lo (2011), Chordia, Subrahmanyam, and Tong (2014), and 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), and Kim, Ivkovich, and Muravyev (2023)). 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), Falck, Rej, and Thesmar (2022), and Huang, Song, and Xiang (2024)).

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

effect of specific demand shocks on short-term price movements. Because the Morningstar methodology is fully transparent, we can sort stock factors by how much they are expected to be affected by the reform using *pre-event* information. We aggregate fund ratings and fund flows at the factor level to directly measure how each factor is impacted by rating-induced trading. The event study shows that the factors expected to be heavily impacted by the reform experienced sudden drops in ratings, flows, and returns in the 6 months following the reform. Consistent with the rating-induced mechanism, other factors unaffected by the reform did not experience similar effects. Using all years other than 2002 as placebo tests, we confirm that the effects we document are unique to June 2002. Moreover, proxies for other possible influences on return predictability, 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, the mutual funds most exposed to Morningstar's rating reform were impacted the most: they experienced significant changes in ratings, flows, and returns right after the reform. Overall, these event study findings are consistent with the idea that a rating-induced change in style-level demand patterns can strongly impact factor and mutual fund returns.

This article's main contribution is using a natural experiment to identify the effect of predictable correlated investor demand on systematic patterns in expected stock returns and mutual fund performance. Importantly, the rating methodology-induced shift in investor demand we examine is uncorrelated with potential fundamental drivers.

This article 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 or stock-level flows, a point that we elaborate on in Section II.C. Another related article, which can be seen as a stepping stone for this article, is Ben-David et al. (2022a). Specifically, Ben-David et al. (2022a) use the 2002 Morningstar reform to demonstrate that style-level fund flows can cause price pressures in stock prices. This article goes one step further by exploring the full consequences of the Morningstar reform on return predictability patterns in stocks and mutual funds: momentum, mutual fund performance persistence, and the dumb money effect.[6]

More broadly, this article 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), and Pavlova and Sikorskaya (2023)), there is relatively less consensus on whether and how demand

[5]Other studies that use Morningstar ratings as part of their identification strategy include Del Guercio and Tkac (2008), Kim (2020), Evans and Sun (2021), Han, Roussanov, and Ruan (2021), Reuter and Zitzewitz (2021), Ben-David et al. (2022b), and Adelino, Cheong, Choi, and Oh (2023).

[6]In other words, whereas Ben-David et al. (2022a) focuses on establishing that "ratings *can* cause price effects," this article shows what exact effects the Morningstar reform has had since 2002.

https://doi.org/10.1017/S0022109023000959 Published online by Cambridge University Press

can shape *systematic* price movements.[7] The style-level rating-induced positive feedback trading mechanism we identify is related to the "style investing hypothesis" (e.g., Barberis and Shleifer (2003), Teo and Woo (2004), Froot and Teo (2008), Boyer (2011), and Wahal and Yavuz (2013)).

This article also has implications for interpreting performance predictability in mutual funds. Traditional discussions of performance predictability (or lack thereof) often center on managerial skill, such as in the case of Carhart (1997), or investor sophistication in choosing funds, such as in Zheng (1999) and Frazzini and Lamont (2008). We supplement these discussions by showing that demand-based price effects are important components of fund performance. This perspective should be intuitive: mutual funds are, after all, portfolios of securities. If demand effects can systematically influence security returns, they can also impact the returns of security portfolios. Jones and Mo (2021) show that published mutual fund predictability patterns become weaker out-of-sample and propose that 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 play a role. Finally, our article provides a possible explanation for the finding by Choi and Zhao (2021) that the performance persistence results of Carhart (1997) disappear out-of-sample.

The rest of the article is organized as follows: Section II explains how the Morningstar reform disrupts style-level positive feedback trading, and it makes several testable predictions. Section III describes the data. Section IV examines the impact of the reform on asset pricing factors and mutual fund return predictability. Section V performs an event study around the reform date, and Section VI concludes. Robustness checks and additional tests are provided in the Appendix.

II. Morningstar Rating Reform and Predictions

In this section, we describe the Morningstar rating methodology reform that took place at the end of June 2002. We then 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 article.

A. 2002 Rating Methodology Reform

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

[7]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 (2020), and Li (2022)), 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 (Ben-David, Franzoni, Moussawi, and Sedunov (2021), Chen (2024), and Parker, Schoar, and Sun (2023)). More recently, Koijen and Yogo (2019) develop a structural methodology to study price movements arising from demand changes, and this approach is further developed by follow-up work (e.g., Haddad, Huebner, and Loualiche (2021), van der Beck (2022), and Huebner (2023)). Gabaix and Koijen (2022) find evidence for sizeable demand-induced price effects at the aggregate market level.

1. 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 higher weight is applied to more recent periods.[8] 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%).[9]

2. Methodology After the Reform

While the rating methodology has been very consistent over time, Morningstar implemented a major reform in June 2002.[10] After the reform, U.S. equity 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 diversified U.S. equity funds (87% of all mutual funds in 2002), the style categories are the well-known 3×3 size–value Morningstar style box.[11] A notable advantage of Morningstar's proprietary style classification is that it does not rely on funds' self-designated benchmarks, which are often mismatched or changed ex post (e.g., Sensoy (2009), Mullally and Rossi (2022)). The change in methodology was announced in Feb. 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 markedly across styles based on recent style performance, as shown in Graph A of Figure 1, which plots the dispersion of average fund ratings across the 3×3 size–value styles. In the months leading up to the methodology change, the average fund ratings of the top- and bottom-rated styles differed by up to 2 stars, and the standard deviation was up to 1 star. Following the reform, that gap shrank as ratings became uncorrelated with past style performance.[12]

Importantly for our identification purposes, investors continued to chase ratings in a similar manner before and after the reform. This fact has been established in a

[8]For funds with more than 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 overlap, however, the recent years are effectively given much more weight than distant history. If we attribute these horizon weights to each individual past year, the most recent 3 years receive a weight of 53%, and this weight is higher for funds with less history.

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

[10]The change was partially motivated by complaints from fund managers, who argued that they were receiving low ratings when their investment style performed poorly, regardless of how they performed relative to peers. See Section 3 of Ben-David et al. (2022a) for more details.

[11]Sector funds, which make up the majority of the remaining 13%, were classified into 12 sectors (e.g., financials and utilities).

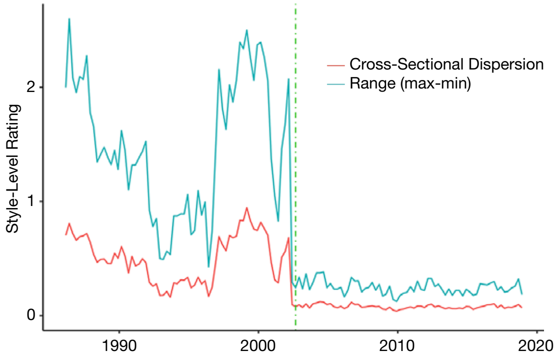
[12]One may wonder why rating dispersion did not drop to exactly 0. A major reason is that Morningstar assigns ratings at the share-class level, so taking an average over share classes would bring the dispersion to 0. Because a fund's share classes have the same underlying portfolio, we compute average ratings at the fund level, following Barber, Huang, and Odean (2016).

FIGURE 1

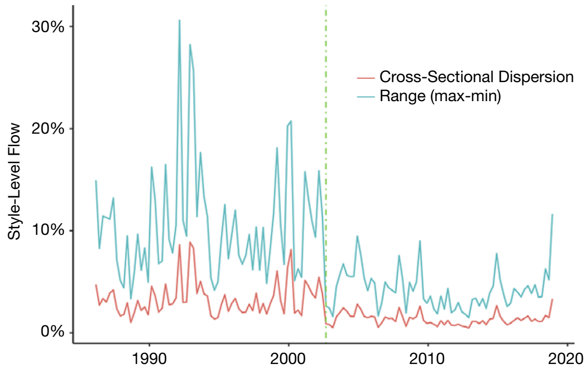
The Morningstar Methodology Reform and Style-Level Flows

We compute total net assets (TNAs)-weighted average quarterly fund ratings and flows for the 3 × 3 size–value Morningstar styles. Graphs A and B of Figure 1 plot the dispersion of these average ratings and flows across styles. Dispersion is measured as the cross-sectional standard deviation (red lines) and the difference between the maximum and minimum values (blue lines). The vertical dashed line marks the June 2002 Morningstar methodology reform event.

Graph A. Dispersion of Average Fund Ratings Across Styles



Graph B. Dispersion of Average Fund Flows Across Styles



number of prior studies (e.g., Evans and Sun (2021), Ben-David et al. (2022a), (2022b)).[13] Most relevant to our analysis, 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.

B. Effects of the Reform on Style-Level Trading

We now demonstrate that the 2002 Morningstar reform had sizeable effects on style-level fund flows: it reduced both the cross-sectional dispersion of style flows and the magnitude of style-level positive feedback flows. As mutual fund managers responded to inflows and outflows by buying or selling stocks they held (e.g., Lou

[13]See, e.g., Figures 1b and 4b in Ben-David et al. (2022a).

https://doi.org/10.1017/S0022109023000959 Published online by Cambridge University Press

(2012)), these fund flow changes impacted the style-level trading imbalances in stocks.

1. 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 in style-level rating dispersion naturally led to a reduction in style-level flow dispersion. This effect is shown in Graph 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. After the reform, the average cross-sectional dispersion (standard deviation) of quarterly style-level fund flows declined from 3.39% to 1.35%, and the average range (maximum minus minimum) declined from 10.66% to 4.23%, respectively.[14]

2. 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 Graph A of Figure 2: funds in styles that performed well in the recent past receive high ratings and attract inflows. By mandate, funds use the new flows to primarily increase their investments in stocks with the same style, 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, caused a sudden disruption to this rating-induced style-level positive feedback trading. We present evidence of disruption in Graphs B and C of Figure 2. Specifically, we sort the 3 × 3 Morningstar fund styles based on lagged-12-month returns, the typical look-back horizon used when studying momentum. Before the reform, funds in styles that recently outperformed received higher average ratings and higher fund flows. The magnitudes are large. Graph B shows that the average rating spread between funds in the top and bottom styles was about 0.8 stars before the reform, shrinking to almost 0 after the reform. Because high ratings attract flows, Graph 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. For expositional clarity, the data in these graphs are demeaned within each month to focus on cross-sectional patterns.

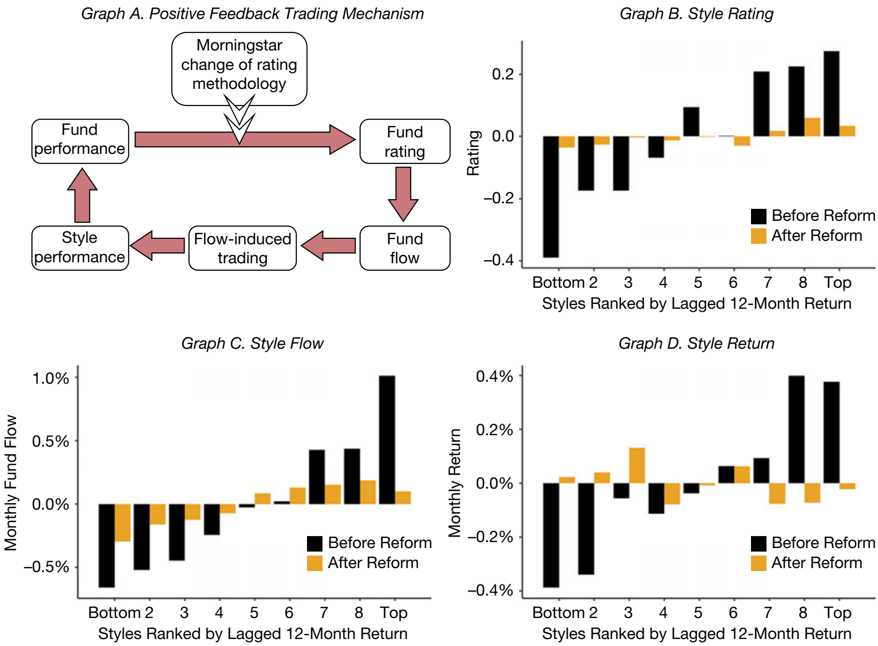
We conjecture that this disruption could have an impact on style returns. As an illustrative exercise, in Graph D of Figure 2, we plot total net assets (TNAs)-weighted style-level fund returns. Prior to the reform, the performance spread between the top- and bottom-ranked styles was approximately 80 basis points per month. The performance difference disappears after the reform. In unreported robustness checks, we find similar patterns when measuring returns using the CAPM alpha, and the post-reform change in the alpha spread is statistically

[14]More formally, in a regression framework, Ben-David et al. (2022b) estimate that a simple rating fixed effects model can explain about 73% of the post-2002 decline in within-month average category flow dispersion (see their Table 4).

FIGURE 2

Style-Level Positive Feedback Trading Before and After Reform

Figure 2 shows that the style-level positive feedback trading largely halted after the Morningstar methodology change in June 2002. The flow chart in Graph A illustrates how pre-2002 ratings generate style-level positive feedback trading. In Graphs B–D, we sort the 3 × 3 Morningstar styles by their lagged 12-month returns. Graphs B and C plot the TNA-weighted average rating and fund flows of the sorted styles. Graph 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 1991–2018, and the start date is dictated by the need for monthly fund flow data from CRSP.



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.

Of course, this illustrative exercise does not control for other reasons why style-level momentum returns could have declined. Section IV provides a formal analysis.

C. On the Price Effect Mechanism

This section further discusses the mechanism underpinning our return predictions.

1. Why Do We Expect Sizeable Price Effects?

Conceptually, the price effects of trading can be modeled as follows (e.g., Gabaix and Koijen (2022)):

$$\text{PRICE\_EFFECT} = \text{PRICE\_MULTIPLIER} \times \text{TRADING\_QUANTITY}, \tag{1}$$

where the PRICE_MULTIPLIER, defined as in Gabaix and Koijen (2022), is a unitless variable that measures the impact on the dollar value of assets per dollar of trading. As such, price effects can be large only if both trading quantity and price multipliers are large. We argue that both are satisfied in our setting.

Style-level trading imbalance is large. As discussed in Section II.B, rating-induced style-level fund flows are sizeable. Combined with the fact that the average mutual funds holding exceeded 20% of the U.S. stock market during our sample period (see Figure 2 of Ben-David et al. (2022a)), this implies sizeable stock trading at the style level due to rating-induced flows before June 2002.

Style-level price multipliers are large. Asset pricing theory predicts that price multipliers are substantially higher at more aggregate levels than at the stock level. Johnson (2006) shows that, in a standard Lucas endowment economy model, the price multiplier of the market factor is equal to the representative investor's relative risk aversion parameter, which is around 2 under conventional calibration (e.g., He and Krishnamurthy (2013)). Therefore, buying 1% of the aggregate stock market leads to a 2% price impact, which is in sharp contrast to tiny theory-implied price multipliers at the stock level.[15] This sharp contrast between aggregate- and stock-level price multipliers is also true when we consider factors other than the market, as An (2022) shows in a no-arbitrage framework. Consistent with this view, empirical studies generally find style- and factor-level price multipliers to be large (e.g., Ben-David et al. (2022a), Gabaix and Koijen (2022), Li and Lin (2022), and Peng and Wang (2022)).[16]

2. Additional Remarks

Next, we clarify other aspects related to the economic mechanism in our framework.

Idiosyncratic positive-feedback fund flows do not contribute to momentum. This article emphasizes style-level positive feedback trading. One may naturally ask: what about *fund-specific* flows and their positive feedback effect? Since Chevalier and Ellison (1997), it is well known that fund flows chase past fund performance. One may naturally suspect that uncorrelated *fund-level* positive feedback trading may also contribute to momentum-strategy profitability. By definition, the Morningstar reform left return chasing of *within-style* fund rankings relatively unchanged; based on this argument, one may expect the flow-momentum channel to have continued to be strong even after the rating reform.[17]

We argue that return chasing of (uncorrelated) fund-specific returns is unlikely to be related to momentum. In fact, idiosyncratic fund flows that are not correlated in the cross section (e.g., at the style or factor levels) lead to little trading in stocks, so idiosyncratic fund-level flows cannot be significant contributors to momentum factor returns.

To see this, consider a stock that recently experienced high idiosyncratic returns. How will these returns impact future mutual fund demand for this stock?

[15]For instance, Petajisto (2009) uses a mean–variance calibration to show that the stock-level price multiplier is on the order of 0.0001.

[16]We further confirm this fact in our fund flow-based setting. 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 of Fama–MacBeth and panel regressions show that style-level price multipliers are two to three times larger than stock-level multipliers.

[17]We thank our discussant Huaizhi Chen for suggesting that we clarify this point.

https://doi.org/10.1017/S0022109023000959 Published online by Cambridge University Press

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, and thus will not create significant additional demand for the stock via the fund flow channel.[18]

Fund flows likely do not carry cash flow information. An important aspect of our mechanism is that rating-induced fund flows do not represent informed trading that conveys cash flow information. Consistent with this interpretation, prior literature has found that fund flow-induced price effects are transient and tend to revert over roughly a year. This tendency has been found at the stock level (Lou (2012)), the style level (Li (2022)), and also when focusing on rating-induced fund flows (Ben-David et al. (2022a)).[19] Further, Ben-David et al. (2022b) document that rating-chasing investors are primarily household investors who exhibit less-than-fully-sophisticated behavior, making it even less likely that fund flows carry superior information about cash flows.

D. Testable Predictions

The discussion in the previous subsection leads to several testable predictions about post-reform changes in expected returns in stock factors and mutual funds. The predictions are summarized in column 1 of Table 1.

First, we anticipate that a disruption in positive feedback trading would reduce the profitability of momentum-type stock factors. In addition to the typical

TABLE 1

Testable Predictions

Table 1 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 the U.S. stock market, we expect the effects to be specific to the U.S. equity market. The first row considers the impact on asset pricing factor profitability, and the next 2 rows consider the impact on mutual fund return predictability.

	2002 Morningstar Reform	
	Impacted 1	Not Impacted 2
Asset pricing factor profitability	Momentum-type factors	Non-momentum-type or international factors
Mutual fund return predictability:		
1) Performance persistence (Carhart (1997))	U.S. equity funds	International equity funds
2) Flow-based predictability (dumb money)		Non-equity U.S. funds

[18]Conversely, fund flows chasing *common* return components can lead to significant positive feedback trading in stocks. For instance, suppose that, over the past few months, small-value stocks performed well on average. Then, ceteris paribus, all funds in the small-value category would see their returns rise. Under the pre-reform Morningstar rating scheme, these funds would receive high ratings and large inflows, leading to higher aggregate demand for small-value stocks. Appendix A.1 provides further evidence that, when examining *stock* trading (rather than fund flows), the Morningstar reform altered positive feedback trading at the style level but not at the stock level.

[19]When we focus on short-term patterns such as stock momentum or mutual fund return persistence, the reversal effect has yet to kick in. We expect the reversal effect to manifest in the dumb money effect, which sorts funds based on longer-term past returns. See Section II.D for details.

momentum factors based on past stock returns (Jegadeesh and Titman (1993), Novy-Marx (2012)), we also expect similar effects for industry momentum (Moskowitz and Grinblatt (1999)) and the 52-week-high strategy (George and Hwang (2004)).

Second, we anticipate disruption to the performance persistence of U.S. equity mutual funds. The mechanism is as follows: most mutual funds have persistent style tilts by mandate. Before the Morningstar reform in 2002, funds in styles that recently performed well (poorly) were assigned high (low) ratings and thus experienced large inflows (outflows). As they buy (sell) stocks in their own styles, this creates positive (negative) price pressures, which further positively (negatively) impact these funds' future performance. This feedback loop contributes to the persistent performance effect documented by Carhart (1997). After the Morningstar reform, the average ratings are equalized across styles; thus, this effect should be muted.

Finally, we expect to see a reduction in the "dumb money effect," which refers to the empirical finding that mutual funds with high (low) recent 3-year fund flows have low (high) subsequent returns (Frazzini and Lamont (2008)). Our reasoning is as follows: several articles have shown that fund flow-induced price pressure reverts over time (Coval and Stafford (2007), Lou (2012)), particularly at the style level (Li (2022)), which generates a negative association between longer-term past flows and subsequent fund returns. Before the reform, this mechanism could directly contribute to the dumb money effect. Since style-level flow dispersion was significantly reduced after the reform, we would expect the dumb money effect to weaken.[20]

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 momentum factors, it should not affect other types of factors. Further, because the Morningstar reform is specific to the U.S. stock market,[21] we also use international equity funds, U.S. non-equity funds, and international factor returns as controls.

When testing the profitability decline of momentum-type factors, comparing against non-momentum factors and momentum factors in other countries is essential because our prediction should be seen as ceteris paribus. As discussed in the introduction, several other mechanisms can also lead to declines in factor profitability over time. These mechanisms impact all factors, but the Morningstar reform's impact is limited to momentum-type factors in the U.S. stock market only and was enacted on a specific date. In other words, we are interested in the *incremental* impact on the momentum-type factors above and beyond other factors.

[20]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). We note 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 meaningfully affect stock reversals.

[21]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.

https://doi.org/10.1017/S0022109023000959 Published online by Cambridge University Press

III. Data and Variable Construction

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

A. Asset Pricing Factors

Panel A of Table 2 summarizes the U.S. and international factors data. For international factors, we rely on the data from Jensen et al. (2023). For U.S. factors, in addition to their factors, we also construct our own factors (described momentarily) for two reasons. First, Jensen et al. (2023) only provide factor returns but not the underlying stock characteristics. Therefore, we compute our own stock characteristics, which enables us to determine factor-level Morningstar rating exposure,

TABLE 2

Summary Statistics

Panel A of Table 2 summarizes the monthly stock factor return data. Columns 1–3 report the average monthly return, the number of factors, and the number of momentum-type factors using our data. Columns 4–6 report equivalent statistics for U.S. factors in Jensen et al. (2023), and columns 7–10 report equivalent statistics for the international factors in Jensen et al. (2023). Panel B summarizes the distribution of each variable in the quarterly mutual fund samples; the first column reports the average number of unique funds. Returns are reported in percentages in both panels.

Panel A. Stock Factors

	U.S. Factors						International Factors			
	Our Data			Jensen et al. (2023) Data			Jensen et al. (2023) Data			
		No. of Factors			No. of Factors			No. of Factors		
	Return	All	Mom	Return	All	Mom	Return	All	Mom	No. of 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	No. of Obs.	Mean	Std. Dev.	1%	5%	25%	50%	75%	95%	99%
	1	2	3	4	5	6	7	8	9	10
U.S. 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	−0.24	−13.97	−7.12	−2.17	−0.24	1.67	7.00	13.97
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-U.S. Equity Mutual Funds										
RETURN (%)	1,417	1.44	9.70	−23.94	−16.21	−3.65	2.19	6.96	17.01	24.75
CAPM_ALPHA (%)	1,417	−0.75	6.32	−16.03	−10.31	−4.28	−0.97	2.39	9.89	17.61
PREV_1Y_RETURN (%)	1,417	4.78	21.67	−59.10	−36.78	−5.50	8.14	17.95	35.21	51.04
PREV_3Y_FLOW	1,417	−0.05	0.86	−3.22	−1.68	−0.32	0.12	0.44	0.97	1.49
Non-Equity U.S. 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
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

https://doi.org/10.1017/S0022109023000959 Published online by Cambridge University Press

a necessary input for the tests in Section V. Second, we want to ensure our results are not sensitive to factor construction methodologies. Specifically, in response to the criticism in Hou et al. (2020) that many factors' returns are concentrated in microcaps, we follow their recommendation in constructing factors to alleviate this concern.

1. U.S. Factors

We examine monthly returns of long–short factors for 1987 to 2018. The start date is guided by the launch of Morningstar fund ratings.[22] To reduce concerns about result sensitivity to the choice and construction of asset pricing factors, we use 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 (2023), 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), 8 in the value/growth category (e.g., BM), 6 in the intangibles category (e.g., industry concentration), 5 in the momentum category (e.g., momentum of Jegadeesh and Titman (1993)), and 3 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. (2023).* Jensen et al. (2023) constructed many 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.[23]

We use their "value-weighted capped factors," which are based on value-weighted tercile portfolios. In addition, they also cap the market weight of each stock at the 80th NYSE percentile, a practice intended to ensure one mega-stock does not dominate a portfolio. For brevity, we refer readers to the description in Jensen et al. (2023) for more details.

2. International Factors

As explained in Section II.D, 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.

[22]Morningstar began providing ratings in 1985, and we control for 1 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.

[23]These include the 52-week-high strategy in George and Hwang (2004), $(t-6, t-1)$ and $(t-12, t-1)$ residual momentum in Blitz et al. (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 $h = 1,3,6,12$).

(2023). We include all factors that are available starting from 1991.[24] After imposing this requirement, the sample includes 1,337 factors from 30 countries, out of which 172 are momentum-type factors.[25]

B. Mutual Fund Data

We obtain quarterly mutual fund data from the CRSP survivorship bias-free mutual fund data starting from 1980[26] and from Morningstar Direct starting from 1990. Summary statistics appear in Panel B of Table 2.

1. U.S. Equity Mutual Funds

We use CRSP objective codes starting with "ED" to identify U.S. domestic equity funds and restrict attention to those with net asset values above $10 million. We use MFLINKS to map the share classes to fund identifiers and aggregate data at the fund level. To investigate the Carhart and dumb money effects, we also require 3 years of data history so we can compute previous 1-year returns and previous 3-year flows. We also compute CAPM alphas, which require 36 months of trailing returns for estimating market betas.[27]

We download Morningstar ratings and fund style categories from Morningstar Direct and merge them with the CRSP fund flow 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). Overall, the sample contains 4,567 unique funds and 173,189 fund-quarters, with an average of 1,110 funds in each period.

In Section V.A, we also use fund holdings data to aggregate the effect of fund ratings and flows at the stock level. For that analysis, we obtain quarterly fund holdings from Thomson Reuters S12, which is based on 13F filings.

2. Other Mutual Funds Used as Controls

We obtain the following two control groups:

1. *Non-U.S. equity mutual funds.* We download returns and assets under management (AUMs) of equity mutual funds domiciled in the European Union, the U.K., and Japan from Morningstar Direct. We filter out all funds that may hold U.S. equities using the information in Morningstar category names.[28] We aggregate the share-class-level data to the fund level. This placebo sample contains 5,176 unique funds and 147,390 fund-quarters, with an average of 1,417 funds in each period.

---

[24] Data availability for many countries starts around 1990.

[25] The Jensen et al. (2023) data set 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.

[26] The MFLINKS mapping of mutual fund share classes to funds is unavailable before 1980.

[27] 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, the month t CAPM alpha is computed as $\alpha_{i,t} = \mathrm{RET}_{i,t} - \mathrm{RF}_t - \hat{\beta}_{t-1} \cdot (\mathrm{MKT}_t - \mathrm{RF}_t)$.

[28] For instance, we filter out funds in the category "Japan Fund – World ex-Japan Equity" because they may hold U.S. stocks. In fact, due to the large size of the U.S. stock market, many funds with an international focus devote a substantial fraction of their portfolio to U.S. equities. In contrast, "Japan Fund – Greater China Equity" is included because funds in this category cannot hold U.S. stocks.

https://doi.org/10.1017/S0022109023000959 Published online by Cambridge University Press

2. *Non-equity U.S. mutual funds.* This control group is composed of 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. The resulting sample contains 8,689 unique funds and 310,154 fund-quarters, with an average of 1,988 unique funds in each period.[29]

IV. Return Predictability Before and After the Reform

In this section, we present stock factor return and mutual fund predictability patterns that are consistent with our predictions in Section II.D.

A. Stock Factor Returns

1. U.S. Factors

We start by testing our prediction about U.S. factor returns by estimating the following difference-in-differences panel regression specification:

$$\text{(2)} \quad \text{RET}_{f,t} = a\text{MOM\_TYPE}_f + b\text{POST2002}_t + c\text{MOM\_TYPE}_f \times \text{POST2002}_t + \text{CONTROLS}_{f,t} + \epsilon_{f,t},$$

where $\text{RET}_{f,t}$ is the return of factor f in month t, MOM_TYPE_f is an indicator of whether the factor f is of momentum-type (defined in Section III.A), and POST2002_t is an indicator that equals 1 after the Morningstar reform in June 2002.

Regression results are presented in Table 3. In columns 1–3, we use the 49 factors we constructed. Column 1 has no additional controls. The results indicate

TABLE 3

Post-Reform Profitability Decline: U.S. Factors

In Table 3, we estimate panel regressions of monthly long–short stock factor returns on the interaction of an indicator for whether a factor is of the momentum type (MOM_TYPE) and an indicator that equals 1 after the June 2002 Morningstar reform (POST2002). All regressions cluster standard errors by factor. Columns 1–3 use the 49 long–short quintile factors we construct, and columns 4–6 use the 153 long–short tercile factors from Jensen et al. (2023). 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 Jan. 2008 to June 2009 from the sample. Standard errors are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Dependent Variable: MONTHLY_FACTOR_RETURN (%)					
	Our Constructed Factors			Jensen et al. (2023) Factors		
	1	2	3	4	5	6
MOM_TYPE × 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)
MOM_TYPE	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
No. of 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

[29]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.

that, relative to other factors, the profitability of momentum-type factors declined by 53.4 basis points per month after the Morningstar reform. Motivated by the finding that factors themselves can exhibit momentum (Gupta and Kelly (2019), Ehsani and Linnainmaa (2022)), in column 2, we also control for lagged factor returns over the months of $t-1$, $t-6$ to $t-2$, and $t-12$ to $t-7$. The effect slightly weakens to 46.1 basis points per month. In columns 4 and 5, we repeat the same regressions using the U.S. factors from Jensen et al. (2023) and find broadly similar results with slightly smaller magnitudes.

One concern is that our results could reflect the short-term severe crash that momentum strategies experienced during the Global Financial Crisis of 2008–2009 (Daniel and Moskowitz (2016)), rather than a persistent decline in momentum profits over the post-Morningstar reform period. 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 Jan. 2008 to June 2009 by the NBER recession dating committee. As expected, omitting these observations reduces the size of the coefficient (to 28.5 basis points using our factors and 23.9 basis points using the Jensen et al. (2023) factors) but the effect remains statistically significant at the 5% level, indicating that the decline in momentum profitability exists *outside* the Global Financial Crisis and is not solely driven by the "momentum crash."

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

2. International Factors

We next test whether the post-reform profitability decline is specific to momentum-type strategies in the U.S.[30] As mentioned in Section III.A, we use the capped value-weighted factors from Jensen et al. (2023).

In the first 3 columns of Table 4, we focus on momentum-type factors across different countries and estimate the following difference-in-differences panel regression:

$$\text{(3)}\quad \text{RET}_{f,c,t} = a\text{US}_c + b\text{POST2002}_t + c\text{US}_c \times \text{POST2002}_t + \text{CONTROLS}_{f,c,t} + \epsilon_{f,c,t},$$

where $\text{RET}_{f,c,t}$ is the monthly return of factor f from country c in month t; US_c is an indicator for whether a factor is based on U.S. stocks; and POST2002_t is an indicator that equals 1 after the Morningstar reform. 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

[30]We thank James Choi for this suggestion.

TABLE 4

Post-Reform Profitability Decline: International Factors

In Table 4, we estimate panel regressions on monthly returns of long–short stock factors from all countries in Jensen et al. (2023). In columns 1–3, we focus on momentum-type factors and regress their returns on a U.S. indicator (US) and an indicator that equals 1 after the June 2002 Morningstar reform (POST2002). In columns 4–6, we use all factors and also add a third interaction with an indicator of whether a factor is of the momentum type (MOM_TYPE). All regressions cluster standard errors by factors. 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 Jan. 2008 to June 2009 from the sample. Standard errors are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	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 × MOM_TYPE × 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)
MOM_TYPE				0.353***	0.321***	0.316***
				(0.041)	(0.037)	(0.037)
MOM_TYPE × US				0.006	0.009	0.011
				(0.112)	(0.102)	(0.100)
MOM_TYPE × POST2002				−0.024	−0.022	−0.100**
				(0.046)	(0.042)	(0.043)
Lagged factor returns	No	Yes	Yes	No	Yes	Yes
Omit momentum crash	No	No	Yes	No	No	Yes
No. of 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

the reform, but U.S. momentum-type factors experienced an additional 54.3 basis point decline, a magnitude similar to that found in column 1 of Table 3. In column 2 of Table 4, we also control 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.

In columns 4–6 of Table 4, we estimate the following triple-difference panel regression using all the factors and including all the countries:

$$
\begin{aligned}
(4)\quad \mathrm{RET}_{f,c,t} = {} & a\mathrm{US}_c + b\mathrm{POST2002}_t + c\mathrm{MOM\_TYPE}_f + d\mathrm{US}_c \times \mathrm{POST2002}_t \\
& + e\mathrm{US}_c \times \mathrm{MOM\_TYPE}_f + f\mathrm{MOM\_TYPE}_f \times \mathrm{POST2002}_t \\
& + g\mathrm{US}_c \times \mathrm{MOM\_TYPE}_f \times \mathrm{POST2002}_t + \mathrm{CONTROLS}_{f,c,t} + \epsilon_{f,c,t}.
\end{aligned}
$$

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 × MOM_TYPE × POST2002 is −35.9 to −39.5 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.

We again caution that the reduction in style-level positive feedback trading we study is likely only one of the causes of the drop in the profitability of U.S. momentum.

By controlling for various other factors, the results in this section are an imperfect attempt to gauge economic magnitudes. Taking the most conservative estimate across Tables 3 and 4, we conclude that the Morningstar reform could explain a monthly decline of 23.9 basis points for momentum strategies after 2002.

3. Factor Momentum

Prior research shows that factors that have performed well (poorly) in the recent past tend to continue to outperform (underperform) subsequently, a finding called "factor momentum" (e.g., Gupta and Kelly (2019), Ehsani and Linnainmaa (2022)). The mechanism in this article also has implications for factor momentum. Specifically, to the extent that factors are correlated with styles as defined in Morningstar's 3×3 size–value matrix, we expect the Morningstar reform to weaken factor momentum strategies after 2002 and that impact to be confined to the U.S.

Our results are broadly consistent with this prediction.[31] We use the international factor returns in Jensen et al. (2023) and follow Ehsani and Linnainmaa (2022) to construct both cross-sectional and time-series factor momentum strategies by country. Indeed, relative to other countries, factor momentum profits in the U.S. declined more after the Morningstar reform. When using factors constructed using equal-weighted stock returns, the incremental profitability decline is approximately 0.77% a month and statistically significant at the 5% level. When using factors constructed using value-weighted returns, the incremental decline is 0.42% to 0.52% a month and no longer statistically significant at the 5% level (t-stat in a range of 1.21 to 1.55).

B. Mutual Fund Return Predictability

We next evaluate the prediction that the mutual fund performance persistence effect of Carhart (1997) and the dumb money effect of Frazzini and Lamont (2008) may attenuate after June 2002. We sort funds into deciles by the corresponding sorting variable: previous 12-month returns for the former effect and previous 3-year fund flows (times -1) for the latter. Therefore, the 10th and 1st deciles represent funds predicted to have the highest and lowest performance according to the original papers, respectively.

We then estimate the following panel regression:

$$\alpha_{i,t}^{\text{CAPM}} = a\text{BOTTOM}_{i,t} + b\text{TOP}_{i,t} + c\text{BOTTOM}_{i,t} \times \text{POST2002}_t + d\text{TOP}_{i,t} \times \text{POST2002}_t + \text{CONTROLS}_{i,t} + \epsilon_{i,t}, \quad (5)$$

where the dependent variable is the quarterly CAPM alpha of fund i in quarter t;[32] $\text{BOTTOM}_{i,t}$ and $\text{TOP}_{i,t}$ are indicators that equal 1 if the fund belongs to the top or bottom deciles; and POST2002_t is an indicator that equals 1 after the Morningstar reform. We control for time and fund fixed effects and cluster standard errors by

[31] These untabulated results are available from the authors.

[32] We examine CAPM alpha, rather than the Fama–French 3-factor alpha, because the price effects we examine are primarily at the size-value style levels, so it is not appropriate to control for size and value effects.

https://doi.org/10.1017/S0022109023000959 Published online by Cambridge University Press

TABLE 5

Post-Reform Performance Predictability Decline: Mutual Funds

In Table 5, we estimate panel regressions of quarterly mutual fund CAPM alphas on indicator variables and their interactions. BOTTOM and TOP refer to funds in the bottom or top decile when ranked based on past 1-year returns or past 3-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 1 after the Morningstar reform. Columns 1 and 4 are estimated using U.S. equity funds; columns 2 and 5 use non-U.S. equity funds; and columns 3 and 6 use U.S. non-equity funds. All regressions include quarter and fund fixed effects; standard errors are 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. The standard errors of these differences are estimated using the delta method. Standard errors are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Dependent Variable: QUARTERLY_FUND_ALPHA (%)					
	Carhart			Dumb Money		
Fund Sample:	U.S.	Non-U.S.	U.S. Non-Equity	U.S.	Non-U.S.	U.S. Non-Equity
	1	2	3	4	5	6
Panel A. Regressions						
TOP	1.276*	0.629	−0.132	0.708***	0.218	0.077
	(0.716)	(0.720)	(0.181)	(0.200)	(0.243)	(0.070)
BOTTOM	−1.584**	−0.306	−0.066	−0.901***	−0.903**	0.115
	(0.636)	(0.677)	(0.176)	(0.283)	(0.425)	(0.070)
TOP × POST2002	−0.710**	−0.882	0.146	−0.670***	−0.041	−0.062
	(0.805)	(0.755)	(0.257)	(0.230)	(0.268)	(0.087)
BOTTOM × POST2002	1.929***	0.262	0.269	0.955***	0.539	−0.098
	(0.732)	(0.761)	(0.281)	(0.299)	(0.440)	(0.102)
No. of obs.	173,189	147,390	310,153	173,189	147,390	310,153
Adj. R^2	0.089	0.395	0.373	0.085	0.396	0.372
Panel B. Estimated Differences						
TOP − BOTTOM	2.860**	0.935	−0.066	1.609***	1.121***	−0.037
	(1.221)	(1.259)	(0.332)	(0.385)	(0.427)	(0.119)
(TOP − BOTTOM) × POST2002	−0.369***	−1.144	−0.123	−1.624***	−0.581	0.036
	(1.373)	(1.348)	(0.499)	(0.413)	(0.453)	(0.166)

quarter and fund. Our main interest is the difference between coefficients *d* and *c*, which captures the post-reform change in the return difference between top- and bottom-ranked funds.

Regression results using U.S. domestic equity funds are shown in columns 1 and 4 in Panel A of Table 5. 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 was nonexistent 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.

Are these results unique to U.S. equity funds? As placebo tests, columns 2 and 4 of Table 5 show estimates of the same regressions using non-U.S. equity funds.[33] While we do find slight decreases in the Carhart and dumb money effects in these funds, the results are not statistically distinguishable from 0. In columns 3 and 6, we estimate the same regressions on non-equity U.S. funds. We find no Carhart or dumb money effects before the reform, nor do we find any significant decline after the reform.

[33]When sorting on Carhart or dumb money effects into deciles, we conduct the sorting within fund domicile regions (European Union, U.K., and Japan).

https://doi.org/10.1017/S0022109023000959 Published online by Cambridge University Press

Overall, the results are consistent with the idea that the Morningstar reform impacted these two forms of mutual fund performance predictability patterns exclusively in the U.S. equity market. However, we caution that the available control-based falsification tests are far from perfect. In particular, international equity mutual funds may differ from U.S. equity funds in ways other than their exposure to the rating reform (e.g., regulation).[34]

Our finding of a decline in U.S. equity mutual fund performance persistence is consistent with Choi and Zhao (2021), who found that the original Carhart (1997) results did not persist out-of-sample. Our mechanism provides a possible explanation for their empirical finding. Specifically, for each year *t*, Figure 1 in Choi and Zhao (2021) plots the rolling average CAPM alphas of the long–short return spread of mutual funds sorted using the Carhart criterion over a 10-year window that ends in year *t*. Their plotted line dropped to close to 0 around 2011 (which reports the average alpha over 2002 and 2011), suggesting that the Carhart (1997) effect started declining in 2002, which is consistent with our rating-based mechanism.

V. Event Study Around the Rating Reform

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

While obtaining definitive identification in asset pricing tests over long periods of time is probably impossible, 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 1-year window (Jan. to Dec. 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, partially alleviating the concern 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)), both of which fall outside our event study window.

Which factors and funds should be most affected by the Morningstar reform in this short window? As discussed in Section II.D, over the long run, we expect

[34]We also use separate accounts in two exercises to further test our mechanism. In the first exercise, we follow Evans and Fahlenbrach (2012) and Evans and Sun (2021), who find that mutual funds often have multiple twin share classes sold to different investors. We verify that the 2002 reform event-induced fund flow changes are specific to regular mutual fund share classes with ratings, consistent with the findings in Evans and Sun (2021). In the second exercise, we confirm that there is no dumb money effect in institutional accounts.

momentum-type factors to be most affected. However, the impact of rating-induced trading is time-varying, so the impact in this short window depends on which factors and funds happened to load on styles suffering the largest reform-induced rating drops by the end of June 2002.

In Section V.A, we describe how we first map ratings and flows into stocks and then aggregate them at the factor level. In Section V.B, we devise a method to measure the "reform exposure" of factors and funds; we then conduct the event study in Section V.C. The fact that the factors and funds with the highest short- 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 full sample panel regressions in the previous section.

A. Mapping Ratings and Flows into Factors

Our mechanism focuses on the impact of ratings and fund flows on factors, so we start by summarizing these variables at the factor level. As explained in Section III.A, because these calculations require that we have data on the stock characteristics that underlie factor construction, we use our U.S. factors for this exercise as the Jensen et al. (2023) data only provide factor returns, not the stock characteristics.

1. Measuring Factor-Level Ratings and Flow-Induced Trading

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 $\mathcal{J}(i)$ that hold the stock:

$$\text{(6)} \qquad \text{RATING}_{i,t} = \frac{\sum_{\text{fund}\, j \in \mathcal{J}(i)} \text{SHARES\_HELD}_{i,j,t-1} \cdot \text{RATING}_{j,t}}{\sum_{\text{fund}\, j \in \mathcal{J}(i)} \text{SHARES\_HELD}^{\text{fund}}_{i,j,t-1}}.$$

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

$$\text{(7)} \qquad \text{FIT}_{i,t} = \frac{\sum_{\text{fund}\, j \in \mathcal{J}(i)} \text{SHARES\_HELD}_{i,j,t-1} \cdot \text{FLOW}_{j,t}}{\sum_{\text{fund}\, j \in \mathcal{J}(i)} \text{SHARES\_HELD}_{i,j,t-1}}.$$

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)).[35] In short, FIT is the total amount of nondiscretionary mutual fund trading in stock i caused by fund flows. As argued in Lou (2012), whereas discretionary trading is more likely to be related to fundamentals, FIT isolates the nondiscretionary trading that is only attributable to fund flows.

[35]Specifically, $\text{FLOW}_{j,t} = \frac{\text{TNA}_{j,t}}{\text{TNA}_{j,t-1}} - \left(1 + \text{RET}_{j,t}\right)$. For simplicity, our construction in equation (7) assumes a one-to-one pass-through of fund flows to stock trading, which is slightly different from Lou (2012), which finds a slightly higher pass-through rate for outflows than inflows. All subsequent analyses are qualitatively unaffected by this simplification. The results in Li (2022) are similar; see his footnote 6 and Appendix A.I for details.

We next aggregate stock-level ratings and FIT at the factor level. For each factor f in each month t, we compute the following:

$$\text{(8)}\quad \text{RATING}_{f,t} = \sum_{i \in \text{top quintile}} w^{f}_{i,t-1}\text{RATING}_{i,t} - \sum_{i \in \text{bottom quintile}} w^{f}_{i,t-1}\text{RATING}_{i,t},$$

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

where $w^{f}_{i,t-1}$ is the lagged weight of stock i in the corresponding factor portfolio.

B. Predicting the Reform's Impact on Factors in the Event Study

We next use data from Dec. 2001, which is the last month *prior to* the 12-month event study window, to predict how each factor's rating (equation (8)) will be affected by the reform. Specifically, for each fund j, we estimate how its rating will change due to the reform:

$$\text{(10)}\quad \text{PREDICTED\_CHANGE}_j = \widehat{\text{RATING}}^{\text{post2002 methodology}}_{j,\text{Dec 2001}} - \widehat{\text{RATING}}^{\text{pre2002 methodology}}_{j,\text{Dec 2001}},$$

where the two terms on the right are our estimates of fund ratings using Dec. 2001 data under the two different Morningstar methodologies. We then aggregate these predicted fund-level rating changes from equation (10) 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 a similar fashion.[36]

Appendix Section A.3 shows further details on predicting factor-level rating changes. The Dec. 2001-based factor-level rating change prediction can explain most variation in the actual factor-level rating change in June 2002 with an R^2 of 84%.

C. Event Study

1. Stock Factors

We start by focusing on stock factors. After computing predicted factor-level rating changes as described in Section V.B, we use them to sort the 49 factors into quintiles. Figure 3 presents the event study results. Graph A plots the average ratings of factors and shows a sharp methodology-induced change exactly at the event. Factors in quintile 1 suffered a drop of 0.43 stars, while those in quintile 5 experienced an increase of 0.19 stars.

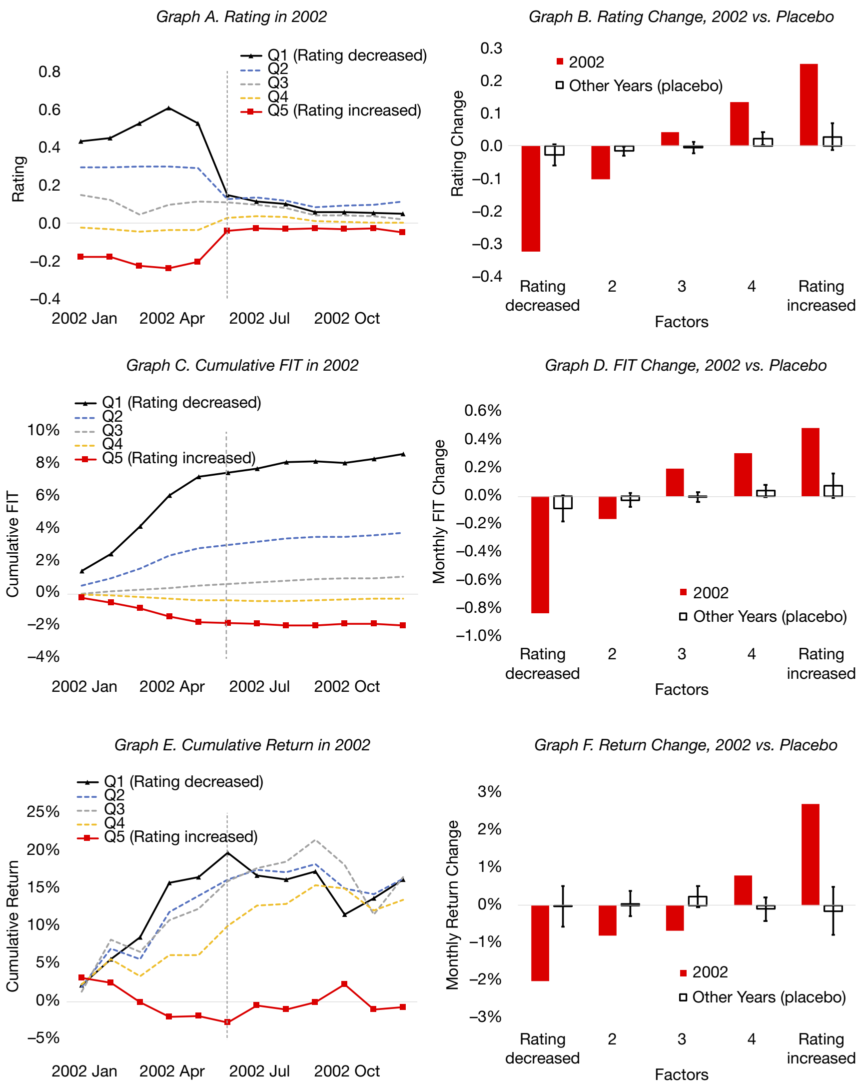
Graphs C and E of Figure 3 plot cumulative monthly factor FIT and returns around the event, respectively. Quintile 1 (the factors that benefited the most from rating-induced flows in the months leading to the reform) experienced a decline of approximately 1% in monthly FIT and a sharp decline of −3.7% in monthly returns.

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

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 (Jan. to Dec. 2002). In Graphs A, C, and E of Figure 3, we sort factors by their *predicted* reform-induced rating change into quintiles and then plot the evolution of their ratings in Graph A, cumulative fund flow-induced trading (FIT) in Graph C, and cumulative returns in Graph E. To alleviate endogeneity concerns, the rating change prediction only uses data up to Dec. 2001 (prior to the event window). The dashed vertical line is the June 2002 reform event. Graphs B, D, and F conduct the same exercises in years other than 2002 as a placebo test. The red solid bars plot the average rating, FIT, and return changes after June (the average of July to Dec. 2002 minus the average of Jan. 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.



https://doi.org/10.1017/S0022109023000959 Published online by Cambridge University Press

Conversely, quintile 5 experienced an increase of 0.14% in monthly FIT and a slight increase of 0.75% in monthly returns.[37]

To alleviate the concern that the return and FIT changes may 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 Graphs B, D, and F of Figure 3, respectively. The 95% confidence intervals are also shown in whiskers. The results show that the patterns found around the Morningstar reform are unique to the reform year of 2002, suggesting that the results are not mechanical.

2. 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.

Figure 4 shows that mutual funds that were predicted to suffer rating decreases indeed saw declines in ratings, FIT, and returns. Graphs B, D, and F 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. Further, Appendix Section A.5 finds that results are qualitatively similar when using 6-month instead of 12-month event windows.

D. 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. Naturally, we cannot rule out all alternative hypotheses, but we confirm that there is no evidence of any systematic changes in arbitrage activity or liquidity patterns around the event. The subsequent results on factors and mutual funds are similar, so we only report the former for brevity.

1. Arbitrage Activity

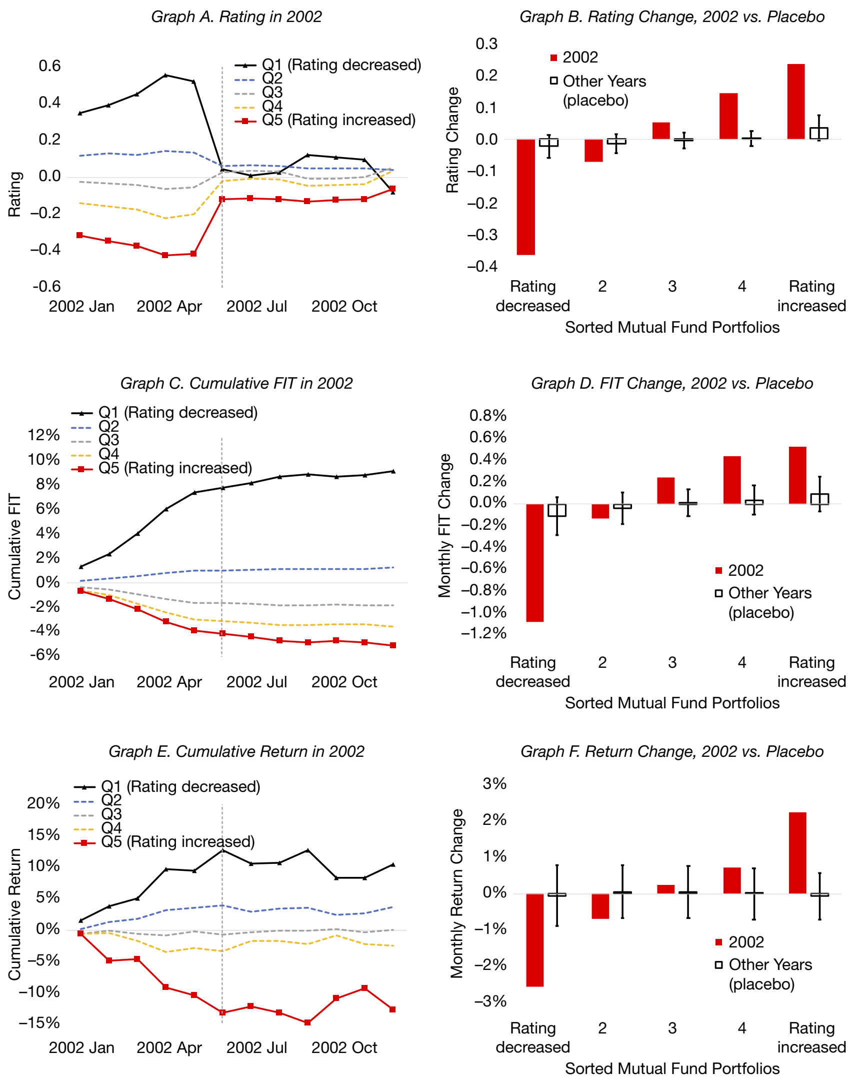
One natural concern is whether arbitrageurs abruptly changed their factor-level trading intensity in mid-2002. A number of articles present evidence that factor profitability is related to arbitrage activity. For instance, Hanson and Sunderam (2013) argue that value and momentum strategy profits decrease when more capital is devoted to them. McLean and Pontiff (2016) show that factor profitability declined after the strategies were published in academic articles and link it to arbitrageur actions. Relatedly, Lou and Polk (2022) show that a return-based measure of arbitrageur activity negatively predicts momentum profits.

[37]In a companion paper, we show that the implied style-level price multiplier (the reciprocal of investor demand elasticity) is approximately 5 (Ben-David et al. (2022a)). That is, buying 1% of the shares outstanding creates a price impact of approximately 5%.

FIGURE 4

Domestic Equity Mutual Funds Around the June 2002 Event

Figure 4 is the equivalent of Figure 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 (Jan. to Dec. 2002). In Graphs A, C, and E, we sort mutual funds by their *predicted* reform-induced rating change into quintiles and then plot the evolution of their ratings in Graph A, cumulative fund flow-induced trading (FIT) in Graph C, and cumulative returns in Graph E. To alleviate endogeneity concerns, the rating change prediction only uses data up to Dec. 2001 (prior to the event window). The dashed vertical line is the June 2002 reform event. Graphs B, D, and F conduct the same exercises in years other than 2002 as a placebo test. The red solid bars plot the average rating, FIT, and return changes after June (the average of July to Dec. 2002 minus the average of Jan. 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.



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.[38] The authors combine the long and short positions into a net position and subtract the past-4-quarter average 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.[39] We compute this measure for all factors.[40]

We plot the evolution of these measures in the 12-month event window in Figure 5. As in Section V.B, we sort factors into quintiles by their predicted rating change using data in Dec. 2001. Graph A plots the NAT measure, and Graph B plots the correlation-based measure. We detect no noticeable change in either measure during the event window.

2. Changes in Liquidity

One may also hypothesize that stock market liquidity increased dramatically in June 2002.[41] 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 Graph C of Figure 5, show no

[38]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.

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

$$\mathrm{CO\_MOMENTUM}_t = \frac{1}{2}\cdot\left[\frac{1}{N^L\left(N^L-1\right)}\sum_i\sum_{j\neq i}\mathrm{PARTIAL\_CORR}\left(\mathrm{RET}_i,\mathrm{RET}_j\right)\right.$$

$$\left.+\frac{1}{N^S\left(N^S-1\right)}\sum_i\sum_{j\neq i}\mathrm{PARTIAL\_CORR}\left(\mathrm{RET}_i,\mathrm{RET}_j\right)\right],$$

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

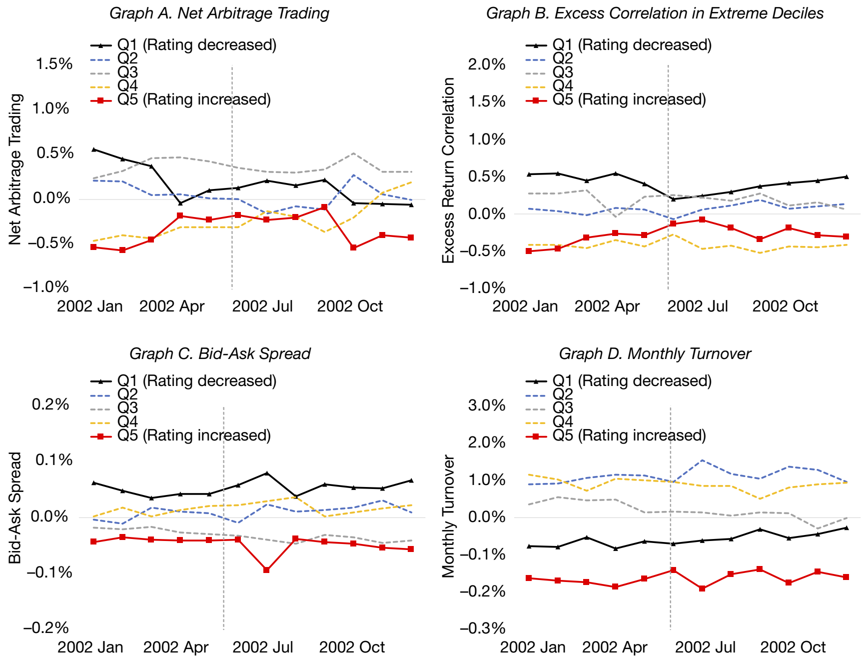
[40]As a sanity check on our replication of their methodology, consistent with Lou and Polk (2022), we find that this measure negatively predicts the returns of factors in the momentum category. Admittedly, this correlation-based measure may be less useful for measuring arbitrage activity in factor strategies with slower turnover such as value, but it could be useful for high-turnover strategies such as momentum.

[41]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.

FIGURE 5

Other Potential Influences of Factor Returns Around 2002

As in Figure 3, factors in Figure 5 are sorted into quintiles by the predicted rating change using data from Dec. 2001. Thus, quintile 1 (or 5) factors are those predicted to experience the largest rating decrease (increase) at the reform event. Graph A plots the net arbitrage trading measure in Chen et al. (2019). Graph B plots excess return correlations in extreme factor quintiles, a measure of arbitrage activity developed in Lou and Polk (2022). Graph C plots the average bid–ask spread of the long and short factor legs, measured following Corwin and Schultz (2012). Graph 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 graphs, the vertical dashed line marks the Morningstar methodology change event.



evidence that liquidity changes account for our findings. Graph D shows that monthly trading turnover also had no clear change around the event.

In summary, around June 2002, we do not find 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.

VI. 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 article, we use a natural experiment to demonstrate that demand movements caused by institutional features can have a first-order impact on the expected returns of 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 decline in profitability that was above and beyond that experienced by non-momentum-type factors and international 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. We further show that these changes were specific to the U.S. stock market, which is consistent with the fact that the Morningstar reform only impacted U.S. equity mutual funds.

More broadly, our findings join a growing number of studies indicating that demand effects can drive systematic price movements (see the literature review in Gabaix and Koijen (2022)). Our article focuses on the role of the Morningstar rating reform because it allows for sharp 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.

Appendix A. Additional Empirical Results

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

A.1. The Morningstar Reform Did Not Alter Idiosyncratic Stock-Level Positive Feedback Trading

Section II.B explains that the Morningstar reform impacted style-level positive feedback trading. A seemingly natural question follows: given that fund flows continue to chase fund ratings, which are based on past fund performance, should we not continue to see flow-induced positive feedback stock trading even after the reform?

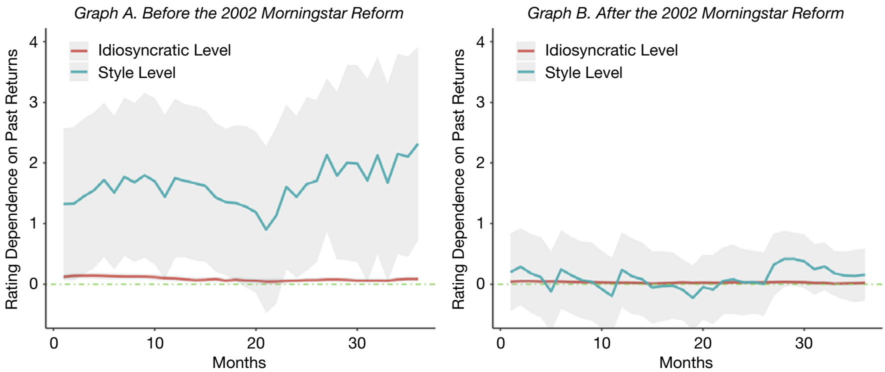
In this section, we demonstrate that fund-level positive feedback flows (without correlated style-level structures) do not lead to stock-level positive feedback trading. Why? Well, the average stock is held by 78.5 funds, so for any given stock, there has to be a *correlated* change in the performance and ratings of funds holding that stock in order to generate sufficiently large trading 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, its return is unlikely to have a large effect on fund performance or ratings. In contrast, suppose the style-level (small-cap growth) return was high in the recent past. Under the pre-reform Morningstar methodology, this means that all small-cap funds would have

FIGURE A.1

Morningstar Reform Only Impacted Style-Level Positive Feedback Trading in Stocks

Figure A.1 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). Graphs 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.



performed well and thus received higher ratings, leading to more flows into all small-cap growth funds, which would then lead to buying pressure in all small-cap growth stocks. After the methodology reform, this style-level positive feedback trading became muted.

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

$$\text{RET}_{i,t} = \text{STYLE\_RET}_{i,t} + \text{IDIOSYNCRATIC\_RET}_{i,t}, \tag{11}$$

where $\text{STYLE\_RET}_{i,t}$ is defined as the value-weighted return of the corresponding 3×3 size–book/market style portfolio, and $\text{IDIOSYNCRATIC\_RET}_{i,t}$ 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. Graph A shows that, before the reform, stock ratings heavily depended on past style-level returns but not idiosyncratic returns. This result confirms that Morningstar-induced positive feedback trading happens exclusively at the style level. Graph B shows that, after the reform, the rating dependence on past style returns also becomes muted. We note that the rating dependence on past idiosyncratic stock returns is close to 0 both before and after the reform.

A.2. *Price Multipliers Are Larger at the Style Level*

As discussed in Section II.C, a number of studies have shown that style-level price multipliers are larger than those at the idiosyncratic level. We also examine this relation 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 V.A. 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:

$$\text{FIT}_{i,t} = \text{STYLE\_FIT}_{i,t} + \text{IDIOSYNCRATIC\_FIT}_{i,t}, \tag{12}$$

where $\text{STYLE\_FIT}_{i,t}$ is the value-weighted average FIT of the 3×3 size–book/market style that the stock belongs to, and $\text{IDIOSYNCRATIC\_FIT}_{i,t}$ 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 a 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:

$$\text{RET}_{i,t} = a + b_{\text{STYLE}} \cdot \text{STYLE\_FIT}_{i,t} + b_{\text{IDIOSYNCRATIC}} \cdot \text{IDIOSYNCRATIC\_FIT}_{i,t} + \epsilon_{i,t} \tag{13}$$

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 that are available from 1980, while columns 2 and 4 use monthly data that are available from 1991. Panel B conducts a t-test between the two coefficients.

TABLE A.1

Estimates of Fund Flow-Induced Price Multipliers

In Table A.1, 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 value-weighted average at the 3×3 size–book/market portfolio level, and 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 the 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. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Dependent Variable: STOCK_RETURN (%)			
	Fama–MacBeth Regression		Panel Regression	
	Quarterly	Monthly	Quarterly	Monthly
	1	2	3	4
Panel A. Regressions				
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	NA	NA	Yes	Yes
Sample period	1980–2018	1991–2018	1980–2018	1991–2018
No. of 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 coefficient difference	2.53	2.56**	8.16***	6.96***
	(2.45)	(1.26)	(1.79)	(1.92)

While we see some variation in the coefficient estimates, a clear pattern emerges: the style-level multipliers are significantly larger than the idiosyncratic-level multipliers. Their differences are statistically significant at the 1% level for all specifications (see Panel B). As pointed out by Schmickler (2020), these price multiplier estimates may be biased upward 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 the idea that style-level multipliers are larger than at the idiosyncratic level (Ben-David et al. (2022a), Gabaix and Koijen (2022), Li and Lin (2022), and Peng and Wang (2022)), we argue that our findings indicate that the same is likely true in the context of flow-induced price effects.

A.3. Accuracy of the Factor Rating Change Prediction

In this section, we examine the accuracy of the factor-level rating change prediction in Section V.B. We first illustrate the predictions in Graphs A and B of Figure A.2. The two graphs plot the two factors predicted to experience the largest rating decline (size factor) and increase (*O*-score factor). Our estimation matches actual ratings quite well. Before June 2002, the actual ratings closely match the estimated ratings under the old methodology (grey lines); 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 over a few months are small, the predicted rating change using Dec. 2001 data ends up being a reasonable predictor of the actual change in June 2002. This is further shown in Graph C, where we plot the actual June 2002 factor rating changes 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 articles in Figure A.3.

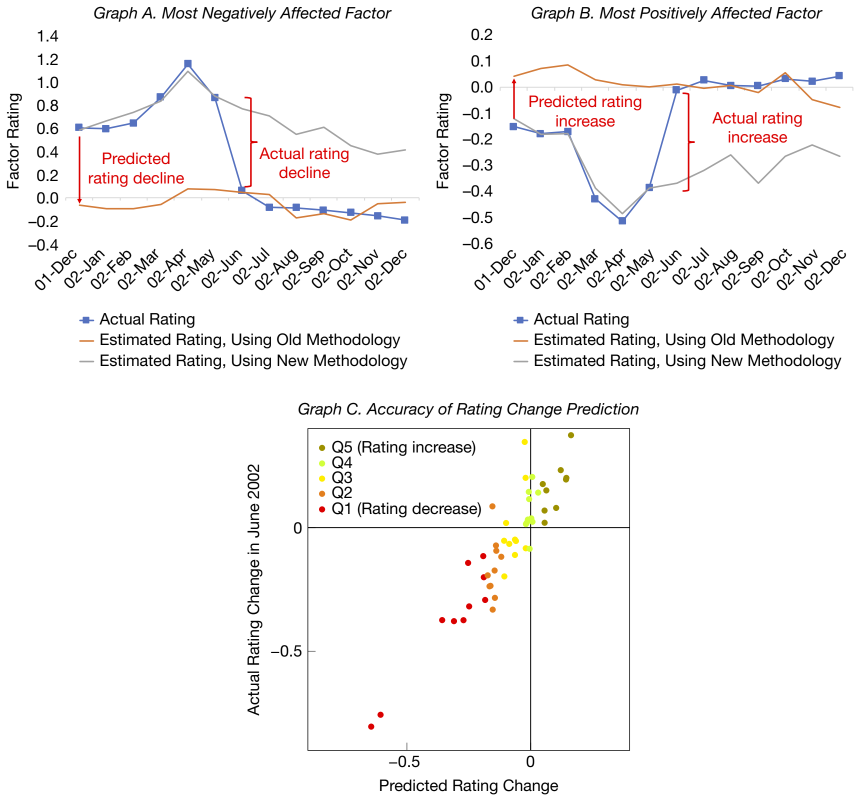
Graph A of Figure A.3 shows a chart from Green et al. (2017) summarizing the average performance (equal-weighted as well as value-weighted) of 94 characteristics. Methodologically speaking, their result is closer to the factor momentum strategy discussed in Arnott et al. (2023), which Ehsani and Linnainmaa (2022) show to be highly related to stock momentum.[42] Graph B shows a chart from Daniel and Moskowitz (2016) summarizing the performance of the stock momentum strategy.

[42]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.

FIGURE A.2

Predicting Factor Rating Changes Around the 2002 Reform Event

Graphs A and B of Figure A.2 illustrate how we predict rating changes of factors at the June 2002 event using data from Dec. 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 Dec. 2001 (marked using red brackets) as the predicted rating change. The blue lines are the actual ratings. Graphs A and B plot the factor with the largest predicted rating decline and increase, respectively (size and *O*-score factors). Graph C compares the actual rating change in June 2002 against the predicted change using data from Dec. 2001. The factors are sorted into quintiles based on the predicted rating change and colored differently.



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.

A.5. *Robustness Check for Event Studies*

The event studies in Section V.C use 12-month event windows. Given the very precise nature of the rating shock, we also perform a robustness check by using a shorter 6-month event window. Instead of using holdings from Dec. 2001, we implement the same event studies using Mar. 2002 holdings and a 6-month window surrounding the shock (Apr. 2002 to Sept. 2002).

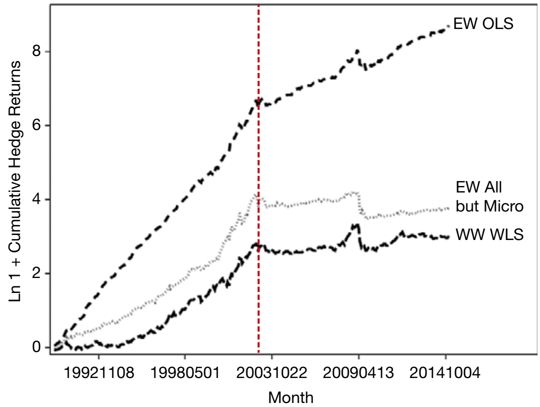
The results are plotted in Figure A.4. Graphs A, C, and E show results for stock factors and are the equivalent of Graphs B, D, and F in Figure 3. Graphs B, D, and F of Figure A.4 show results for mutual funds and are the equivalent of Graphs B, D,

FIGURE A.3

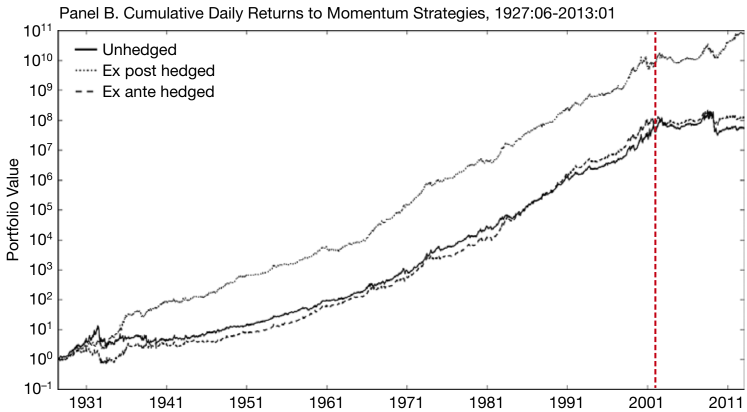
Prior Evidence of Momentum-Type Strategy Profitability Decline

Figure A.3 presents charts used in previous studies that show a kink in cumulative factor returns. In both graphs, we added a red dashed line to mark the approximate location of June 2002 on the timeline. Graph 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 a value-weighted strategy. Graph B reproduces Figure 4b of Daniel and Moskowitz (2016), which plots the cumulative return to the momentum strategy. The screenshots are taken from the latest SSRN version of each article: the Oct. 2016 version for Green et al. (2017), and the July 2015 version of Daniel and Moskowitz (2016), with the authors' permissions.

Graph A. Green, Hand, and Zhang (2017, Fig 3)



Graph B. Daniel and Moskowitz (2016, Fig 4b)

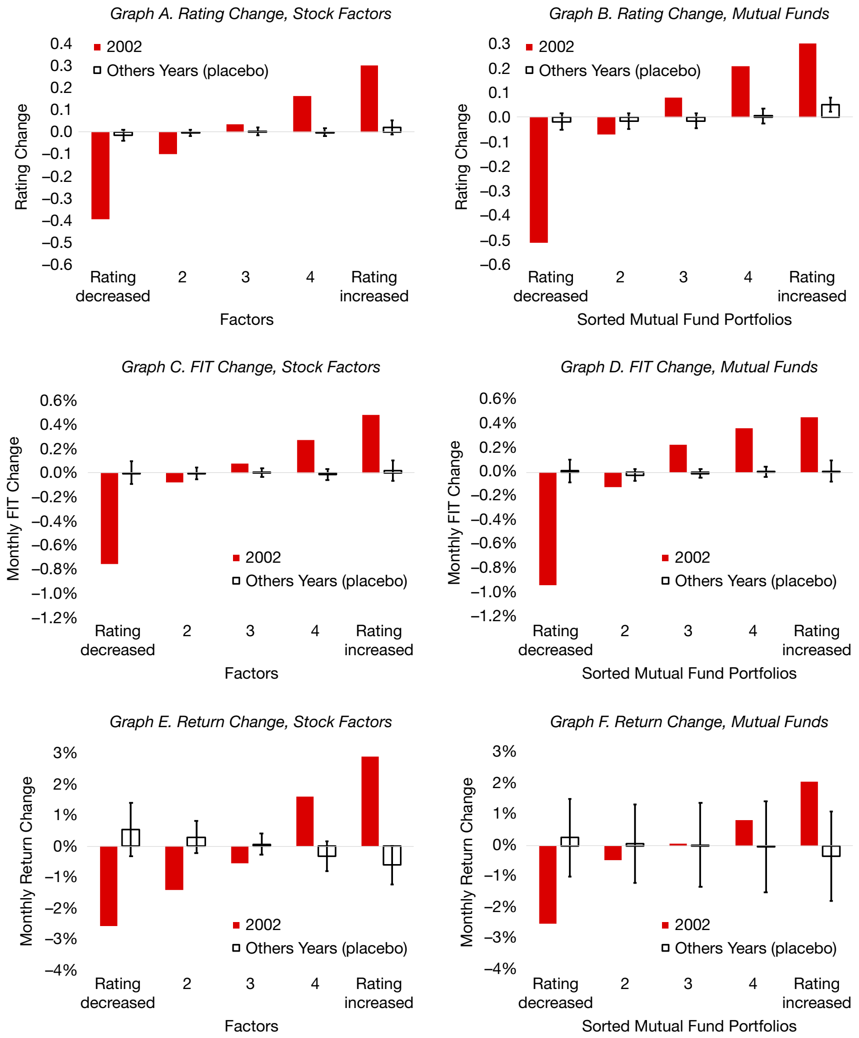


and F in Figure 4. These plots show that our earlier conclusion based on 12-month event windows is qualitatively similar when performed using 6-month event windows. As plotted by the red solid bars, factors and mutual funds predicted to face rating decreases (quintile 1) indeed have lower ratings, FIT, and returns after the event; the reverse is true for factors and mutual funds predicted to experience rating increases (quintile 5). The 2002 event period results are statistically different from the other placebo years, which are plotted in the white bars.

FIGURE A.4

Event Study Using 6-Month Event Windows

Figure A.4 is the equivalent of Figures 3 and 4 using 6-month event windows (3 months before to 3 months after the event). Graphs A, C, and E examine stock factors and Graphs B, D, and F examine mutual funds. We sort them by their reform-induced rating changes predicted using Mar. 2001 data (prior to the event window) into quintiles. The red solid bars plot the average rating, FIT, and return changes after the event (the average of July to Sept. 2002 minus the average of Apr. 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 each month.



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

Table B.1 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 article 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)
	5-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 Moskowitz (JF 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 abbreviation: 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; RF, Review of Finance; RFS, Review of Financial Studies.

References

Adelino, M.; C. Cheong; J. Choi; and J. Y. J. Oh. "Mutual Fund Flows and the Supply of Capital in Municipal Financing." NBER Working Paper No. 30980 (2023).

An, Y. "Flow-Based Arbitrage Pricing Theory." Working Paper, Johns Hopkins University (2022).

Aragon, G. O.; E. Li; and L. A. Lindsey. "Exploration or Exploitation? Hedge Funds in Venture Capital." Working Paper, Arizona State University (2018).

Arnott, R. D.; M. Clements; V. Kalesnik; and J. T. Linnainmaa. "Factor Momentum." *Review of Financial Studies*, 36 (2023), 3034–3070.

Barber, B. M.; X. Huang; and T. Odean. "Which Factors Matter to Investors? Evidence from Mutual Fund Flows." *Review of Financial Studies*, 29 (2016), 2600–2642.

Barberis, N., and A. Shleifer. "Style Investing." *Journal of Financial Economics*, 68 (2003), 161–199.

Ben-David, I.; F. Franzoni; and R. Moussawi. "Do ETFs Increase Volatility?" *Journal of Finance*, 73 (2018), 2471–2535.

Ben-David, I.; F. Franzoni; R. Moussawi; and J. Sedunov. "The Granular Nature of Large Institutional Investors." *Management Science*, 67 (2021), 6629–6659.

Ben-David, I.; J. Li; A. Rossi; and Y. Song. "Ratings-Driven Demand and Systematic Price Fluctuations." *Review of Financial Studies*, 35 (2022a), 2790–2838.

Ben-David, I.; J. Li; A. Rossi; and Y. Song. "What Do Mutual Fund Investors Really Care About?" *Review of Financial Studies*, 35 (2022b), 1723–1774.

Berk, J. B., and R. C. Green. "Mutual Fund Flows and Performance in Rational Markets." *Journal of Political Economy*, 112 (2004), 1269–1295.

Blitz, D.; J. Huij; and M. Martens. "Residual Momentum." *Journal of Empirical Finance*, 18 (2011), 506–521.

Boyer, B. H. "Style-Related Comovement: Fundamentals or Labels?" *Journal of Finance*, 66 (2011), 307–332.

Brown, D. C.; S. Davies; and M. Ringgenberg. "ETF Arbitrage, Non-Fundamental Demand, and Return Predictability." *Review of Finance*, 25 (2021), 937–972.

Calluzzo, P.; F. Moneta; and S. Topaloglu. "When Anomalies Are Publicized Broadly, Do Institutions Trade Accordingly?" *Management Science*, 65 (2019), 4555–4574.

Carhart, M. M. "On Persistence in Mutual Fund Performance." *Journal of Finance*, 52 (1997), 57–82.

Chang, Y.-C.; H. Hong; and I. Liskovich. "Regression Discontinuity and the Price Effects of Stock Market Indexing." *Review of Financial Studies*, 28 (2015), 212–246.

Chen, A. Y., and T. Zimmermann. "Open Source Cross-Sectional Asset Pricing." *Critical Finance Review*, 27 (2022), 207–264.

Chen, H. "Cash Induced Demand." *Journal of Financial and Quantitative Analysis*, 59 (2024), 195–220.

Chen, Y.; Z. Da; and D. Huang. "Arbitrage Trading: The Long and the Short of It." *Review of Financial Studies*, 32 (2019), 1608–1646.

Chevalier, J., and G. Ellison. "Risk Taking by Mutual Funds as a Response to Incentives." *Journal of Political Economy*, 105 (1997), 1167–1200.

Cho, T. "Turning Alphas into Betas: Arbitrage and Endogenous Risk." *Journal of Financial Economics*, 137 (2020), 550–570.

Choi, J. J., and K. Zhao. "Carhart (1997) Mutual Fund Performance Persistence Disappears Out of Sample." *Critical Finance Review*, 10 (2021), 263–270.

Chordia, T.; A. Subrahmanyam; and Q. Tong. "Have Capital Market Anomalies Attenuated in the Recent Era of High Liquidity and Trading Activity?" *Journal of Accounting and Economics*, 58 (2014), 41–58.

Cochrane, J. H. "Presidential Address: Discount Rates." *Journal of Finance*, 66 (2011), 1047–1108.

Corwin, S. A., and P. Schultz. "A Simple Way to Estimate Bid–Ask Spreads from Daily High and Low Prices." *Journal of Finance*, 67 (2012), 719–760.

Coval, J., and E. Stafford. "Asset Fire Sales (and Purchases) in Equity Markets." *Journal of Financial Economics*, 86 (2007), 479–512.

Daniel, K. D., and T. J. Moskowitz. "Momentum Crashes." *Journal of Financial Economics*, 122 (2016), 221–247.

De Bondt, W. F., and R. Thaler. "Does the Stock Market Overreact?" *Journal of Finance*, 40 (1985), 793–805.

Del Guercio, D., and P. A. Tkac. "Star Power: The Effect of Morningstar Ratings on Mutual Fund Flow." *Journal of Financial and Quantitative Analysis*, 43 (2008), 907–936.

Ehsani, S., and J. T. Linnainmaa. "Factor Momentum and the Momentum Factor." *Journal of Finance*, 77 (2022), 1877–1919.

Evans, R. B., and R. Fahlenbrach. "Institutional Investors and Mutual Fund Governance: Evidence from Retail–Institutional Fund Twins." *Review of Financial Studies*, 25 (2012), 3530–3571.

Evans, R. B., and Y. Sun. "Models or Stars: The Role of Asset Pricing Models and Heuristics in Investor Risk Adjustment." *Review of Financial Studies*, 34 (2021), 67–107.

Falck, A.; A. Rej; and D. Thesmar. "When Do Systematic Strategies Decay?" *Quantitative Finance*, 22 (2022), 1955–1969.

Frazzini, A., and O. A. Lamont. "Dumb Money: Mutual Fund Flows and the Cross-Section of Stock Returns." *Journal of Financial Economics*, 88 (2008), 299–322.

Froot, K., and M. Teo. "Style Investing and Institutional Investors." *Journal of Financial and Quantitative Analysis*, 43 (2008), 883–906.
Gabaix, X., and R. S. Koijen. "In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis." Working Paper, Harvard University (2022).
George, T. J., and C.-Y. Hwang. "The 52-Week High and Momentum Investing." *Journal of Finance*, 59 (2004), 2145–2176.
Green, J.; J. R. Hand; and M. T. Soliman. "Going, Going, Gone? The Apparent Demise of the Accruals Anomaly." *Management Science*, 57 (2011), 797–816.
Green, J.; J. R. Hand; and X. F. Zhang. "The Characteristics that Provide Independent Information About Average US Monthly Stock Returns." *Review of Financial Studies*, 30 (2017), 4389–4436.
Gupta, T., and B. T. Kelly. "Factor Momentum Everywhere." *Journal of Portfolio Management*, 45 (2019), 13–36.
Haddad, V.; P. Huebner; and E. Loualiche. "How Competitive is the Stock Market? Theory, Evidence from Portfolios, and Implications for the Rise of Passive Investing." Working Paper, SSRN (2021).
Han, X.; N. L. Roussanov; and H. Ruan. "Mutual Fund Risk Shifting and Risk Anomalies." Working Paper, University of Pennsylvania (2021).
Hanson, S. G., and A. Sunderam. "The Growth and Limits of Arbitrage: Evidence from Short Interest." *Review of Financial Studies*, 27 (2013), 1238–1286.
Harris, L., and E. Gurel. "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 (1986), 815–829.
Harvey, C. R. "Presidential Address: The Scientific Outlook in Financial Economics." *Journal of Finance*, 72 (2017), 1399–1440.
Harvey, C. R.; Y. Liu; and H. Zhu. "… and the Cross-Section of Expected Returns." *Review of Financial Studies*, 29 (2016), 5–68.
He, Z., and A. Krishnamurthy. "Intermediary Asset Pricing." *American Economic Review*, 103 (2013), 732–770.
Hendershott, T.; C. M. Jones; and A. J. Menkveld. "Does Algorithmic Trading Improve Liquidity?" *Journal of Finance*, 66 (2011), 1–33.
Heston, S. L., and R. Sadka. "Seasonality in the Cross-Section of Stock Returns." *Journal of Financial Economics*, 87 (2008), 418–445.
Hou, K.; C. Xue; and L. Zhang. "Replicating Anomalies." *Review of Financial Studies*, 33 (2020), 2019–2133.
Huang, S.; Y. Song; and H. Xiang. "Noise Trading and Asset Pricing Factors." Working Paper, University of Washington (2020).
Huang, S.; Y. Song; and H. Xiang. "The Smart Beta Mirage." *Journal of Financial and Quantitative Analysis*, forthcoming (2024).
Huebner, P. "The Making of Momentum: A Demand-System Perspective." Working Paper, University of California at Los Angeles (2023).
Jegadeesh, N., and S. Titman. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance*, 48 (1993), 65–91.
Jensen, T. I.; B. T. Kelly; and L. H. Pedersen. "Is There a Replication Crisis in Finance?" *Journal of Finance*, 78 (2023), 2465–2518.
Johnson, T. C. "Dynamic Liquidity in Endowment Economies." *Journal of Financial Economics*, 80 (2006), 531–562.
Jones, C. S., and H. Mo. "Out-of-Sample Performance of Mutual Fund Predictors." *Review of Financial Studies*, 34 (2021), 149–193.
Khandani, A. E., and A. W. Lo. "Illiquidity Premia in Asset Returns: An Empirical Analysis of Hedge Funds, Mutual Funds, and US Equity Portfolios." *Quarterly Journal of Finance*, 1 (2011), 205–264.
Kim, S. H. "Do Mutual Funds Manipulate Star Ratings? Evidence from Portfolio Pumping." Working Paper, University of Texas at Dallas (2020).
Kim, Y. H.; Z. Ivkovich; and D. Muravyev. "Causal Effect of Information Costs on Asset Pricing Anomalies." Working Paper, Michigan State University (2023).
Koijen, R. S., and M. Yogo. "A Demand System Approach to Asset Pricing." *Journal of Political Economy*, 127 (2019), 1475–1515.
Lee, J., and J. P. Ogden. "Did the Profitability of Momentum and Reversal Strategies Decline with Arbitrage Costs After the Turn of the Millennium?" *Journal of Portfolio Management*, 41 (2015), 70–83.
Lewellen, J. "The Cross-Section of Expected Stock Returns." *Critical Finance Review*, 4 (2015), 1–44.
Li, J. "What Drives the Size and Value Factors?" *Review of Asset Pricing Studies*, 12 (2022), 845–885.
Li, J., and Z. Lin. "Prices Are Less Elastic at More Aggregate Levels." Working Paper, University of Utah (2022).

Lou, D. "A Flow-Based Explanation for Return Predictability." *Review of Financial Studies*, 25 (2012), 3457–3489.
Lou, D., and C. Polk. "Comomentum: Inferring Arbitrage Activity from Return Correlations." *Review of Financial Studies*, 35 (2022), 3272–3302.
Marquering, W.; J. Nisser; and T. Valla. "Disappearing Anomalies: A Dynamic Analysis of the Persistence of Anomalies." *Applied Financial Economics*, 16 (2006), 291–302.
McLean, R. D., and J. Pontiff. "Does Academic Research Destroy Stock Return Predictability?" *Journal of Finance*, 71 (2016), 5–32.
Morningstar. "The Morningstar Rating for Funds." White Paper, Morningstar (2016).
Moskowitz, T. J., and M. Grinblatt. "Do Industries Explain Momentum?" *Journal of Finance*, 54 (1999), 1249–1290.
Mullally, K., and A. Rossi. "Moving the Goalposts? Mutual Fund Benchmark Changes and Performance Manipulation." Working Paper, University of Arizona (2022).
Novy-Marx, R. "Is Momentum Really Momentum?" *Journal of Financial Economics*, 103 (2012), 429–453.
Parker, J.; A. Schoar; and Y. Sun. "Retail Financial Innovation and Stock Market Dynamics: The Case of Target Date Funds." *Journal of Finance*, 78 (2023), 2673–2723.
Pástor, L.; R. F. Stambaugh; and L. A. Taylor. "Fund Tradeoffs." *Journal of Financial Economics*, 138 (2020), 614–634.
Pavlova, A., and T. Sikorskaya. "Benchmarking Intensity." *Review of Financial Studies*, 36 (2023), 859–903.
Peng, C., and C. Wang. "Factor Demand and Factor Returns." Working Paper, University of Notre Dame (2022).
Petajisto, A. "Why Do Demand Curves for Stocks Slope Down?" *Journal of Financial and Quantitative Analysis*, 44 (2009), 1013–1044.
Reuter, J., and E. Zitzewitz. "How Much Does Size Erode Mutual Fund Performance? A Regression Discontinuity Approach." *Review of Finance*, 25 (2021), 1395–1432.
Schmickler, S. "Identifying the Price Impact of Fire Sales Using High-Frequency Surprise Mutual Fund Flows." Working Paper, Princeton University (2020).
Sensoy, B. A. "Performance Evaluation and Self-Designated Benchmark Indexes in the Mutual Fund Industry." *Journal of Financial Economics*, 92 (2009), 25–39.
Shleifer, A. "Do Demand Curves for Stocks Slope Down?" *Journal of Finance*, 41 (1986), 579–590.
Teo, M., and S.-J. Woo. "Style Effects in the Cross-Section of Stock Returns." *Journal of Financial Economics*, 74 (2004), 367–398.
van der Beck, P. "On the Estimation of Demand-Based Asset Pricing Models." Working Paper, Ecole Polytechnique Fédérale de Lausanne (2022).
Wahal, S., and M. D. Yavuz. "Style Investing, Comovement and Return Predictability." *Journal of Financial Economics*, 107 (2013), 136–154.
Wurgler, J., and E. Zhuravskaya. "Does Arbitrage Flatten Demand Curves for Stocks?" *Journal of Business*, 75 (2002), 583–608.
Zheng, L. "Is Money Smart? A Study of Mutual Fund Investors' Fund Selection Ability." *Journal of Finance*, 54 (1999), 901–933.