What Do Mutual Fund Investors Really Care About?

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

September 24, 2019

Abstract

Two recent influential studies document that mutual fund flows respond to past returns that are adjusted for market exposure, leading one paper to conclude that the Capital Asset Pricing Model (CAPM) is the "closest to the true asset pricing model". We re-examine these tests and find that they cannot reject the null hypothesis that investors make no adjustment for any known systematic risk exposure when allocating capital to funds. Instead, investors rely on simple signals (unadjusted past returns and Morningstar ratings). Furthermore, we present evidence consistent with investors following Morningstar blindly, regardless of the way the ratings are constructed.

Keywords: Mutual funds, retail investors, fund flows, Morningstar, CAPM
JEL Classification: G11, G24, G41

*We thank Brad Barber, Xing Huang, and Terry Odean for kindly sharing their data. We thank Jonathan Berk, Justin Birru, Scott Cederburg, John Cochrane, Sylvester Flood (Morningstar), Francesco Franzoni, John Graham, Cam Harvey, David Hirshleifer, Ron Kaniel, Paul Kaplan (Morningstar), Steve Kaplan, Adam Reed, Martin Schmalz, Mike Schwert, Berk Sensoy, Andrei Shleifer, Rick Sias, Nancy Xu, Deniz Yavuz, and seminar participants at The Ohio State University, Boston College, Southern Methodist University, the University of Texas at Dallas, and the University of Arizona, as well as participants at the Pacific Northwest Finance Conference, the Financial Research Association conference, the Fourth Annual Young Scholars Finance Consortium, and SFS Cavalcade for helpful comments. Ben-David is with The Ohio State University and the National Bureau of Economic Research, Li is with University of Utah, Rossi is with the University of Arizona, and Song is with the University of Washington. Email: ben-david.1@osu.edu, jiacui.li@eccles.utah.edu, rossi2@email.arizona.edu, and songy18@uw.edu.

1 Introduction

Whether and the extent to which investors perceive risk in financial markets is a fundamental research question in asset pricing. Several studies have attempted to tackle this question in the mutual funds space, where both the past performance of fund managers and investors' capital allocation are observable to researchers. Two celebrated studies—Barber, Huang, and Odean (2016) (BHO) and Berk and van Binsbergen (2016) (BvB)[1]—set out to address the issue using different empirical techniques. Despite the difference in methodologies, both studies find evidence that when allocating capital according to past fund performance, investors appear to discount performance attributable to exposure to the market factor, consistent with investors using the Capital Asset Pricing Model (CAPM) as their benchmark model.[2]

The results of BHO and BvB were very influential on the finance research community, immediately spawning a slew of follow-up research. Blocher and Molyboga (2017) and Agarwal, Green, and Ren (2018) applied BHO and BvB's empirical methods in the hedge fund space and found that hedge fund investors also appear to account for exposure to market risk. Several studies accept BHO and BvB's results and propose rationalizations for why investors only attend to the market risk factor, but not to other factors such as size and value. Chakraborty, Kumar, Muhlhofer, and Sastry (2018) argue that the results of BvB and BHO might be driven by investors' limited attention. Specifically, they posit that investors adjust for market returns but not for other factors because only market returns are readily available to investors. Evans and Sun (2018) argue that the results of BHO and BvB are partially explained by the fact that investors use Morningstar ratings as their main signal for investment, and that ratings and CAPM alphas are correlated. Jegadeesh and Mangipudi (2017) propose that the BHO and BvB tests may be too restrictive and therefore they do not necessarily reveal the true (complex) asset pricing that investors use. Further, they claim that the tests put the CAPM at an empirical advantage over more elaborated asset pricing models because parameter estimations are inherently more precise for a simple model (like the CAPM) than for a multifactor model such as the Fama-French three-factor model.

Our study is also a contemporaneous follow-up on BHO and BvB. However, unlike the other follow-up studies, we do not attempt to rationalize their results. Rather, we challenge

[1] To demonstrate their prominence and contribution to the debate: compared to all other studies published in 2016 in their respective journals, both studies are in the top decile of citation counts (as of May 2019).

[2] BHO interpret their result as indicating that retail investors are unsophisticated to not account for other known risk factors such as size and value. BvB conclude that the CAPM is the "closest to the asset pricing model investors are actually using" (p. 2).

BHO and BvB to reject the obvious null hypothesis arising from past literature. Our null hypothesis is based on the large body of mutual fund literature showing that mutual fund investors are not sophisticated and respond to simple signals, such as unadjusted returns and external ratings. We show that neither BHO and BvB can reject this null hypothesis. Moreover, we show that investors blindly follow external ratings, which are widely used by investors and explain a large variation of fund flows, suggesting that investors have little understanding of the methodology and risk adjustment of ratings.

At the core of our study is the paradigm of the scientific approach to financial economics, that a new finding should be contrasted with the existing prior (Harvey, 2017). As such, the new tests and evidence of BHO and BvB should be contrasted with the existing evidence and consensus about the behavior of mutual fund investors. Our null hypothesis is that mutual fund investors are not particularly sophisticated investors. Specifically, they rely on simple and readily-available signals (unadjusted past returns and/or external signals) when allocating capital to funds. First, this null hypothesis is supported by the characteristics of retail investors and by previous empirical evidence. According to the 2011 ICI Fact Book, for example, mutual fund investors in the U.S. are primarily households: 93.7% of long-term mutual fund assets were held by households. These investors exhibit behavior that is generally considered unsophisticated.[3] Prior studies show the mutual fund investors invest based on attention-grabbing and easy-to-process signals, often based on unadjusted past performance.[4] Second, the information available to mutual fund investors is limited and simple, often including items like the name of the fund, size, past performance, and—at times—external rating (like Morningstar or Lipper). In contrast, these investors do not have easy access to data about fund exposure to risk factors (e.g., market beta), and it is questionable whether they would know how to apply this information if it were available. Further casting doubt on the idea that individual savers would account for a fund's market beta when making investment decisions, a content analysis of mutual fund advertisements in the most circulated personal finance magazine performed by Jones and Smythe (2003) finds

[3]For example, they invest in high-fee funds (Barber, Odean, and Zheng, 2005; Choi and Robertson, 2018), time the market poorly (Frazzini and Lamont, 2008; Akbas, Armstrong, Sorescu, and Subrahmanyam, 2015; Friesen and Nguyen, 2018), display a naive understanding of diversification (Benartzi and Thaler, 2001), value recent performance the same as very old performance when making capital allocation decisions (Phillips, Pukthuanthong, and Rau, 2016), and prefer funds that recently experienced an extremely positive monthly return (Akbas and Genc, 2019).

[4]For example, investors are more likely to invest in funds with higher Morningstar ratings (Del Guercio and Tkac, 2008; Reuter and Zitzewitz, 2015) and sustainability ratings (Hartzmark and Sussman, 2019), and in those mentioned in Wall Street Journal rankings (Kaniel and Parham, 2017). Furthermore, investors allocate more capital to funds that advertise (Jain and Wu, 2000; Reuter and Zitzewitz, 2006) and whose stock holdings appear in the media (Solomon, Soltes, and Sosyura, 2014), and seem to fail to account for the fact that fund families selectively advertise only their best-performing funds (Koehler and Mercer, 2009).

that advertisements never report a fund's beta or similar measures of risk. Indeed, when discussing their own results, BHO admit that how retail investors adjust for funds' market beta when assessing performance is "a mystery" (p. 5).

To examine the explanatory power of asset pricing models versus simple signals, we begin by eyeballing the flows of investors to active mutual funds based on various signals. We find that fund flows are best explained by Morningstar ratings, with 5-star funds (7.4% of fund-months) attracting a total of $656 billion over the sample period of 1997 to 2011 (see Figure 1 in Section 3). In contrast, the funds that performed best according to unadjusted returns or factor model-based alphas receive significantly less flows. Moreover, the amount of flows they attract are almost indistinguishable from each other ($274 billion to $326 billion), so controlling for commonly used factors does not improve the power to predict flows. We also compare the explanatory power of unadjusted past returns and Morningstar ratings. While both signals independently have strong explanatory power for fund flows, Morningstar's explanatory power is greater. Overall, these findings further support the null hypothesis and are inconsistent with the idea that investors use the CAPM to direct their mutual fund investments.

We then perform two exercises to examine whether the specific tests proposed by BHO and BvB can reject the null hypothesis. The BHO test examines the response of fund flows to different components of returns—alpha and returns from exposure to different risk factors—and find evidence that investors respond less to market-related returns. For this exercise, we simulate a data set of mutual fund flows under the null. Specifically, the simulated flows chase past unadjusted fund returns and ratings, and the flow-performance sensitivities are estimated from the actual data using monthly cross-sectional regressions of flows on past returns. By construction, the simulated flows do not discount any factors. However, when feeding these data into the BHO test, the resulting coefficients are statistically indistinguishable from the coefficients estimated in the actual data, indicating that the test proposed by BHO cannot reject the null. Further investigation shows that the BHO finding is a statistical artifact due to time-varying sensitivity of flows to performance.[5]

BvB perform a nonparametric test that examines the correlation between the signal (e.g., positive alpha) and an indicator of positive flows. To compare asset pricing models against our null, we generate a binary data set in which we record an investment signal based on Morningstar (e.g., buy all funds with 5-star ratings). When examined in the BvB test, the Morningstar signal performs significantly better than all signals based on asset pricing models. In contrast, alphas from all asset pricing models barely and inconsistently

[5]The time-varying nature of the mutual fund flow-performance sensitivity has been documented by Starks and Sun (2016), Franzoni and Schmalz (2017), and Harvey and Liu (2019).

outperform unadjusted returns. In sum, the BvB test also does not reject the null hypothesis that investors rely on simple signals.

The fact that investors do not adjust their flows to academics' risk factors, but rather follow Morningstar raises the question of what is the role of Morningstar in investors' decision making process. Morningstar ratings are based on past returns (mildly adjusted for past volatility) and, in the later part of our sample, fund performance rankings within size and value buckets. Thus, there are two possible reasons for why investors rely on Morningstar's ratings. First, investors care about dimensions like value and size benchmarking, and rely on Morningstar to advise them about risk-adjusted performance. Second, investors have little understanding of (or even little interest in) risk adjustment, and simply rely on Morningstar's expertise because of its reputation as an independent agency. This explanation is in the spirit of Mullainathan, Schwartzstein, and Shleifer (2008) and Gennaioli, Shleifer, and Vishny (2015), who argue that financial advisors instill trust in investors. As such, investors do not have a specific risk adjustment procedure in mind, but rather take Morningstar's advice at face value.

To understand better the way investors use Morningstar ratings in their decision making, we examine investor reaction to a significant methodological change that Morningstar implemented in 2002. In June 2002, Morningstar changed how ratings are constructed. Before the change, ratings were based on fund performance ranking among *all* U.S. equity funds; beginning in June 2002, Morningstar started ranking funds *within* size and value categories. The change in methodology led to a major reshuffling of Morningstar ratings where more than 50% of funds experienced a change in ratings in that month.

Our tests show that, despite the extensive change in methodology, investors' flows responded with equal intensity to ratings before and after the methodological change. In other words, investors seem to have followed Morningstar's recommendations regardless of whether ratings reflected style benchmarking. This result is consistent with the second hypothesis, that investors do not have a pre-specified risk adjustment in mind, but rather trust Morningstar to pick the best investments for them, in line with the theory of Gennaioli et al. (2015).

Overall, our evidence shows that mutual investors do not perform any meaningful risk adjustment when allocating capital to mutual funds. These results invalidate the previous findings of BHO and BvB. Furthermore, it appears that investors follow Morningstar's external ratings blindly. For investors, Morningstar might be providing investors peace of mind, à-la money doctors (Gennaioli et al., 2015).

The rest of this study is organized as follows. Section 2 presents the null hypothesis: Investors chase unadjusted past returns and attention-grabbing signals. Section 3 describes the data we use. Section 4 shows that the BHO and BvB tests, when conducted properly, cannot reject the null hypothesis that investors chase past returns and follow Morningstar ratings. Section 5 explores the way in which investors use Morningstar ratings. Section 6 provides concluding remarks.

2 The Null Hypothesis

In his 2017 presidential address at the American Finance Association meeting, Campbell Harvey urged empiricists to specify the null hypothesis and test their models against it, a fundamental tenet of modern scientific empirical inquiry (Harvey, 2017). BHO and BvB propose tests that attempt to determine which asset pricing models are used by investors. We argue, however, that before taking asset pricing models to the data, one needs to examine whether the proposed tests can reject the null hypothesis. In other words, one needs to determine whether the proposed asset pricing models explain the data better than known investor behavior.

What is the null hypothesis in this case? There is a decades-long literature showing that mutual fund investors follow two types of signals when allocating capital to funds: unadjusted past returns, and simple and attention-grabbing signals.

Since the late 1990s, researchers have documented that retail mutual fund investors chase past returns (Chevalier and Ellison, 1997; Sirri and Tufano, 1998; Karceski, 2002; Sapp and Tiwari, 2004; Friesen and Sapp, 2007; Bailey, Kumar, and Ng, 2011; Christoffersen, Musto, and Wermers, 2014; Akbas and Genc, 2019, among many others). These findings are consistent with survey-based evidence that investors form beliefs about future performance by extrapolating from past performance (Vissing-Jorgensen, 2003; Bacchetta, Mertens, and Van Wincoop, 2009; Adam, Marcet, and Beutel, 2017; Greenwood and Shleifer, 2014) and do not take risk adjustment into account in the traditional asset pricing sense (Heuer, Merkle, and Weber, 2017; Choi and Robertson, 2018; Adam, Matveev, and Nagel, 2018).

In addition, many studies document that investors use external ratings and attention-grabbing signals to guide their capital allocation. For example, investors rely on Morningstar ratings (Del Guercio and Tkac, 2008; Reuter and Zitzewitz, 2015), Wall Street Journal rankings (Kaniel and Parham, 2017), and sustainability rankings (Hartzmark and Sussman, 2019). Furthermore, investors allocate more capital to funds that advertise (Jain and Wu, 2000; Reuter and Zitzewitz, 2006) and to those whose stock holdings are covered in the media (Solomon et al., 2014). Among all the different signals that investors are exposed to, we focus on Morningstar ratings due to their prominence in the mutual fund industry. In the United States, Morningstar is the undisputed leader in this industry (Del Guercio and Tkac, 2008). Its most well-known product, the 5-star rating system, was introduced in 1985 and is widely used by financial professionals and advisors. Ratings are also used by asset management companies in advertising (Blake and Morey, 2000; Morey, 2003). Finally, Morningstar ratings have been shown to have a strong independent influence on investor flows (Del Guercio and Tkac, 2008; Reuter and Zitzewitz, 2015).[6]

To further increase our confidence that past returns and Morningstar ratings are good candidates for the null hypothesis, Figure 1 plots net flows to the funds that are top-ranked by Morningstar (5 stars) or top-ranked (same proportion as 5-star) by the following alternative performance measures: unadjusted returns,[7] the CAPM alpha, the Fama-French three-factor alpha, and the Fama-French-Carhart four-factor alpha. (Further details about the construction of these variables are reported in the next section.) Panel A shows net flows by year, and Panel B plots the same flows cumulated over the sample period. The results support our choice of the null. First, funds with top Morningstar ratings receive more inflows than funds that are deemed best-performing according to any of the asset pricing models considered, and the difference is economically large ($22 billion per year, on average). Second, based on this initial test, there is no clear evidence that alphas from any asset pricing model dominate unadjusted returns.[8]

To sum up, in light of the previous literature and the summary statistics presented, both unadjusted past returns and external ratings should be part of the null hypothesis. Thus, any test designed to assess which asset pricing model best describes mutual fund investors' behavior should simultaneously be able to clearly reject the null that investors chase unadjusted past returns and ratings.

3 Data and Measures

To make our results more directly comparable to prior literature, we use the same sample of funds used by BHO, which spans January 1991 to December 2011.[9] To limit discrepancies driven by variable construction or other methodological choices, we take the fund flow

[6] According to Morningstar's company statistics dated December 2017 (kindly provided by Morningstar), 11.9 million individual investors, 255,000 financial advisors, 2,700 institutional clients, 1,500 asset management firms, 31 retirement-plan providers, and 285,000 plan sponsors subscribe to its ratings.

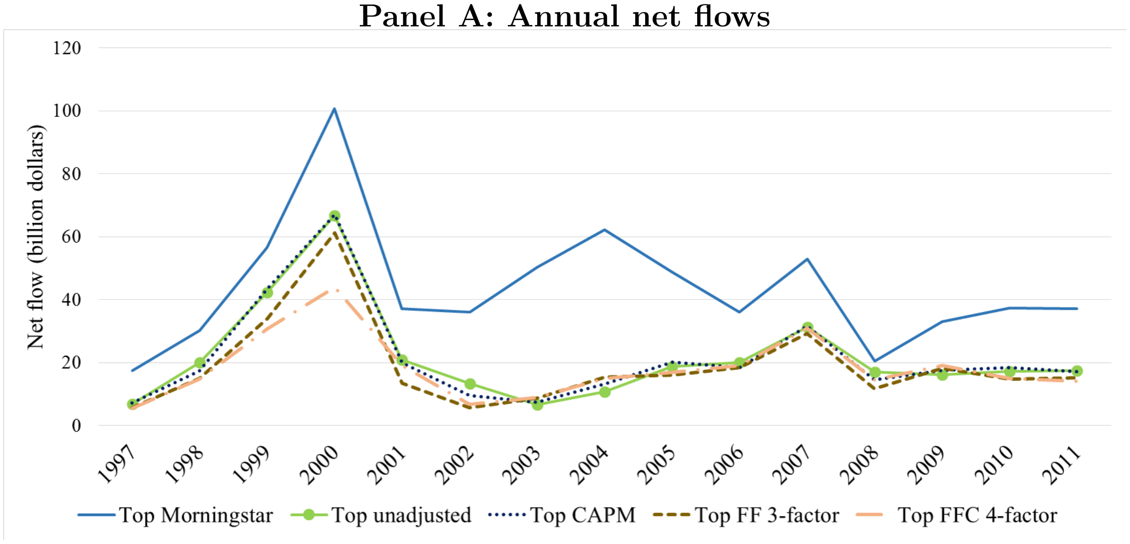
[7] In Figure 1, "top unadjusted" are funds that rank highest when using unadjusted returns. Because we are sorting funds within each month, this is equivalent to using returns in excess of the risk-free rate or of the market as the sorting criterion.

[8] Further analysis in Appendix A.1 confirms that while both unadjusted past returns and ratings both explain flows, ratings explain fund flows better.

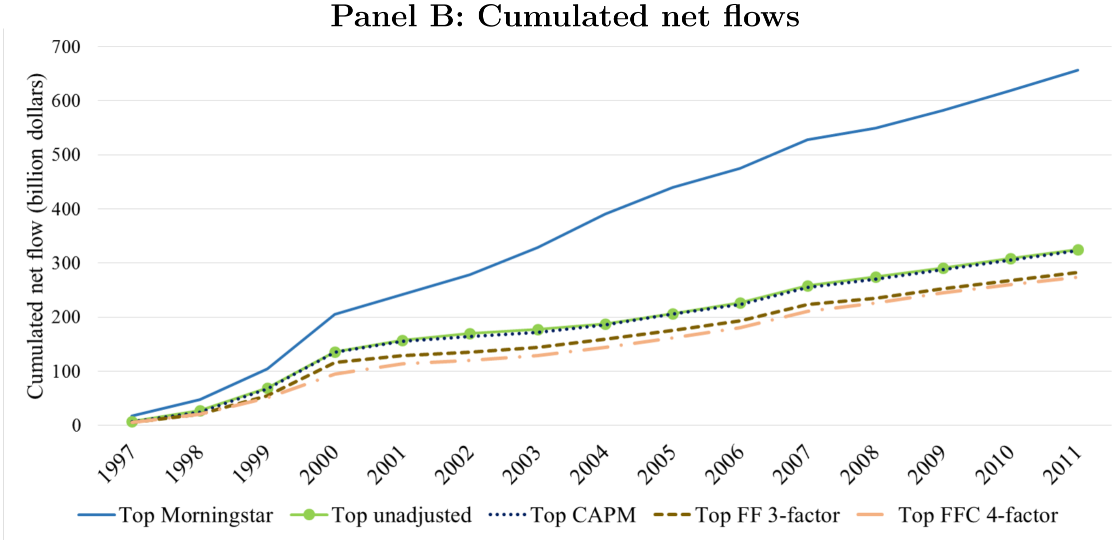
[9] We thank the authors for generously sharing their data. They restrict the sample to mutual funds that start in 1991 because the Center for Research in Security Prices (CRSP) database contains monthly total net assets beginning in 1991.

Figure 1. Flows to top-ranked funds. Panel A presents annual aggregate net flows to top-ranked funds when ranked according to four return-based performance measures and Morningstar. Each month, funds with a 5-star rating and an equivalent number of funds with the highest past performance according to each model are considered top-ranked. Because funds are ranked within each month, rankings based on unadjusted returns, returns in excess of the risk-free rate, and returns in excess of the market are the same. We thus report the results for these ranking rules under the same label "Top unadjusted." Panel B shows the accumulated value of the annual flows in Panel A.

Panel A: Annual net flows



Panel B: Cumulated net flows



variable and several other variables (expense ratios, fund style assignments, ratings, etc.)

directly from the BHO dataset. Extending the BHO dataset to include observations up to the end of 2017 does not materially alter our conclusions.

We briefly explain how BHO constructed their dataset for the reader's convenience. The BHO dataset, spanning from 1991 to 2011, is based on the standard CRSP survivor-bias-free mutual fund database. BHO focus on actively managed equity mutual funds. They eliminate index funds, balanced funds, and exchange-traded funds (ETFs). While funds are often marketed to different clients through different share classes, they invest in the same portfolio, and typically the only difference is the fee structure. Therefore, all share classes are aggregated at the fund level.

Table 1. Descriptive statistics for the mutual fund sample. Our sample includes actively managed US-focused equity mutual funds with a Morningstar rating. Variables are computed using monthly data from January 1991 to December 2011, and the main sample starts in July 1997 and ends in December 2011. Fund-month observations are grouped based on their Morningstar Star rating at the beginning of that month. Fund flows are net flows in month t divided by month $t-1$ assets under management. Weighted past return and weighted past FFC-alpha are, respectively, the past 18-month unadjusted return and the Fama-French-Carhart (FFC) four-factor alpha (Carhart, 1997), both weighted using an exponential decay model estimated in the data (see Equation (12) for details). Return volatility is calculated as the standard deviation of fund returns in excess of the risk-free rate.

	Morningstar Rating					
	1 Star	2 Stars	3 Stars	4 Stars	5 Stars	All
Fund-month observations	17,024	60,416	92,131	60,613	18,279	248,463
Fund size (avg; \$million)	501	752	1,294	2,136	3,460	1,444
Fund age (avg; years)	16.2	16.7	16.9	17.4	16.5	16.9
Fund flow (avg)	−1.54%	−1.23%	−0.69%	0.17%	1.14%	−0.53%
Weighted past return (avg)	−0.08%	0.18%	0.36%	0.55%	0.78%	0.37%
Weighted past FFC-alpha (avg)	−0.43%	−0.18%	−0.027%	0.12%	0.36%	−0.03%
Return volatility (avg; 1 year)	5.51%	5.05%	4.85%	4.81%	4.89%	4.93%
Return volatility (avg; 5 years)	6.44%	5.61%	5.27%	4.99%	4.96%	5.33%
Market beta (avg)	1.11	1.06	1.00	0.95	0.91	1.00
Size beta (avg)	0.26	0.19	0.16	0.15	0.16	0.18
Value beta (avg)	−0.06	0.00	0.05	0.09	0.10	0.04
Momentum beta (avg)	−0.01	0.01	0.02	0.03	0.05	0.02
Fraction of positive flows	16.8%	20.0%	31.0%	50.1%	67.3%	35.1%

Following the fund flow literature, the investment flow for fund p in month t is defined as the net flow into the fund divided by the lagged total net assets (TNA). Formally, the flow is calculated as

$$\mathrm{F}_{p,t} = \frac{\mathrm{TNA}_{p,t}}{\mathrm{TNA}_{p,t-1}} - (1 + R_{p,t}). \tag{1}$$

Here, $\mathrm{TNA}_{p,t}$ is fund p's total net assets at the end of month t, and $R_{p,t}$ is the fund return in month t. The analysis is restricted to mutual funds with at least \$10 million TNA and flows between -90% and $1,000\%$. The CRSP mutual fund dataset is then merged with Morningstar data to obtain each fund's Morningstar rating and investment style. Funds for which a Morningstar rating or investment style is not available are dropped from the sample. The resulting sample comprises a total of 3,432 funds.

Table 1 provides descriptive statistics for the final sample consisting of nearly 250,000 fund-month observations. During our sample period, the average fund has a modestly negative monthly flow of -0.53%, manages $\$1,444$ million, and has an average age of 16.9 years. Funds with higher Morningstar ratings tend to be larger and receive higher investor flows. Consistent with the algorithm that Morningstar uses to assign ratings (Appendix B), higher rated funds also tend to have higher past returns and lower return volatility. Table 1 also presents fund factor loadings on the Fama-French-Carhart (FFC) four factors (Carhart, 1997) when estimated using rolling 60-month regressions. Higher rated funds have higher value and momentum betas, on average.

The tests in our paper, as well as those in BvB and BHO, are designed to compare the ability of different signals—alphas from asset pricing models and Morningstar ratings—to explain fund flows. For each asset pricing model, we estimate the corresponding alpha from past fund returns following BHO. We provide the exact details of the derivation of the alphas in Appendix A.2.

4 Can BHO and BvB Reject the Null Hypothesis?

In this section, we show that the tests proposed by BHO and BvB cannot reject the null hypothesis that investors chase unadjusted fund returns and Morningstar ratings.

4.1 BHO Cannot Reject the Null

4.1.1 Explanation of BHO Methodology

BHO examine how fund flows respond to different components of fund returns using a panel regression framework. They find that flows appear to respond less strongly to the market component of returns than to the alpha component and other factor-related returns, and they interpret this finding as evidence that investors use a model akin to the CAPM when evaluating mutual funds.

We start by replicating BHO's exercise. For each month t, we decompose a fund's return into four factor-related components—market, size, value, and momentum factors—and alpha.[10] For each factor, the factor-related component is calculated as the fund's factor loading (estimated using the most recent 60 months of returns) multiplied by the factor's realized return in month t. Following BHO, these components are accumulated over the prior 18 months using an exponential decay function (see Appendix A.2 for details). The return components related to the fund's market, size, value and momentum exposures are labeled MKTRET, SIZRET, VALRET, and MOMRET, respectively. The residual, i.e., the fund's alpha, is labeled $\text{ALPHA}^{\text{FFC}}$.

To infer how investors respond to different components of a fund's return, we follow BHO and estimate the following panel regression with time fixed effects (FEs):

$$\begin{aligned} F_{p,t} = b_0 + \gamma X_{p,t} + b_{\text{ALPHA}}\text{ALPHA}^{\text{FFC}}_{p,t} + b_{\text{MKTRET}}\text{MKTRET}_{p,t} + b_{\text{SIZRET}}\text{SIZRET}_{p,t} \\ + b_{\text{VALRET}}\text{VALRET}_{p,t} + b_{\text{MOMRET}}\text{MOMRET}_{p,t} + \mu_t + e_{p,t}. \end{aligned} \quad (2)$$

Here, $F_{p,t}$ is the percentage of fund flows in month t, μ_t is the time fixed effect for month t, and $X_{p,t}$ is a vector of control variables. The controls include the total expense ratio, a dummy variable for no-load, a fund's return standard deviation over the prior five years, the log of fund size in month $t-1$, the log of fund age, and lagged fund flows from month $t-19$.[11] The coefficients $b_{\text{ALPHA}}, b_{\text{MKTRET}}, \ldots$, measure how fund flows respond to different return components. Standard errors are double-clustered by fund and month.

We report the estimates for the flow response to each of the five components of fund returns in Column (1) of Table 2. Taken at face value, these estimates suggest that investors treat fund return components differently. Specifically, the response appears to be stronger for alpha but weaker for returns attributable to factor exposures, and especially so for the market factor. These results are very similar to the results presented by BHO in Column (1) of their Table 5. Our estimated coefficient for the response to market-related returns is 72% smaller than that for the response to alpha and is almost identical to that of BHO (they find the MKTRET coefficient to be 71% smaller than the coefficient on the Alpha component).

[10] Barber et al. (2016) also consider three industry factors constructed following Pástor and Stambaugh (2002). They do not find evidence that investors attend to the industry factors. For consistency with the factors used by BvB and in the relevant extant literature (e.g., Del Guercio and Reuter, 2014), we only consider the four factors of the Carhart (1997) model.

[11] These control variables are the same as those used by BHO.

Table 2. Response of fund flows to components of fund returns: Testing the null hypothesis. This table presents coefficient estimates from panel regressions of the percentage of fund flow (the dependent variable) on the components of a fund's return in Equation (2). Column (1) uses actual fund flows, while Columns (2) and (3) use flows generated under the null hypothesis that investors respond equally to each return component. Specifically, in Model 1, flows respond only to cross-sectional differences in total returns across funds. In Model 2, in addition to total returns, flows also respond to Morningstar ratings. Columns (4) and (5) present the differences between the coefficients observed in the data and under the null hypotheses. The controls include the total expense ratio, a dummy variable for no-load, a fund's return standard deviation over the prior five years, the log of fund size in month $t-1$, the log of fund age, and lagged fund flows from month $t-19$. Standard errors are presented in parentheses. Standard errors in Column (1) are panel regression standard errors double-clustered by fund and month. Standard errors in the other columns are bootstrapped empirical standard errors. **, and *** indicate significance at the 5% and 1% level, respectively. Please refer to the text for additional details.

	Observed in the data	**Observed under null hypothesis**		**Difference: data - null hypothesis**	
		Model 1	**Model 2**	**Model 1**	**Model 2**
	(1)	(2)	(3)	(4)	(5)
$\text{ALPHA}^{\text{FFC}}$	0.84***	0.80***	0.81***	0.05	0.03
	(0.030)	(0.009)	(0.008)	(0.031)	(0.031)
MKTRET	0.24***	0.33***	0.29***	-0.10**	-0.06
	(0.042)	(0.018)	(0.018)	(0.046)	(0.045)
SIZERET	0.60***	0.66***	0.64***	-0.06	-0.04
	(0.060)	(0.020)	(0.019)	(0.063)	(0.063)
VALRET	0.52***	0.55***	0.54***	-0.02	-0.01
	(0.063)	(0.014)	(0.014)	(0.065)	(0.065)
MOMRET	0.71***	0.77***	0.78**	-0.06	-0.07
	(0.051)	(0.021)	(0.020)	(0.055)	(0.054)
Month FE	Yes	Yes	Yes	-	-
Controls	Yes	Yes	Yes	-	-
Observations	248,463	248,463	248,463	-	-
Adjusted R^2	0.181	0.179	0.177	-	-

4.1.2 BHO Test Results with Null-Consistent Data

To explore whether BHO can reject our null hypothesis, we generate flows data under the null hypothesis that investors respond equally to all components of fund returns. By construction, these simulated flows *do not* differentiate between return components. We then feed the generated data into the BHO panel regression to see whether the BHO test can make the correct inference about the fact that investors in the simulated data indeed do not care about the market factor.

The simulated data are generated in the following way. For each month t, we run cross-sectional regressions of fund flows on fund returns and controls. The predicted fund flows obtained from these regressions represent the flows that we would observe under the null. We consider two regression models to generate the counterfactual flows. In the first, we run

$$F_{p,t} = b_0 + \gamma X_{p,t} + b_1 R_{p,t} + e_{p,t}. \tag{3}$$

Here, $R_{p,t}$ is the past 18-month weighted fund return for fund p, and $X_{p,t}$ contains the same set of control variables used to estimate Equation (2). In light of our subsequent finding that investors heavily chase Morningstar ratings (Section 4.2), we also include indicators for Morningstar ratings in the second and more realistic model:

$$F_{p,t} = \gamma X_{p,t} + \sum_{k=1}^{5} \gamma_{p,t}^{k} \mathrm{I}_{(\mathrm{star=k})} + b_1 R_{p,t} + e_{p,t}, \tag{4}$$

where $\mathrm{I}_{(\mathrm{star=k})}$ is the Morningstar rating indicator variable for a fund with k stars. Then, for each period, we add bootstrapped fund flow residuals to the predicted flows to obtain simulated flows, and repeat this procedure 1,000 times for each of the two models.[12]

We then re-estimate BHO's main panel regression (Equation (2)) using the simulated flows, and report the results in Columns (2) and (3) of Table 2. The coefficients and empirical standard errors reported are, respectively, the average and the standard deviation of each coefficient across the 1,000 simulations.

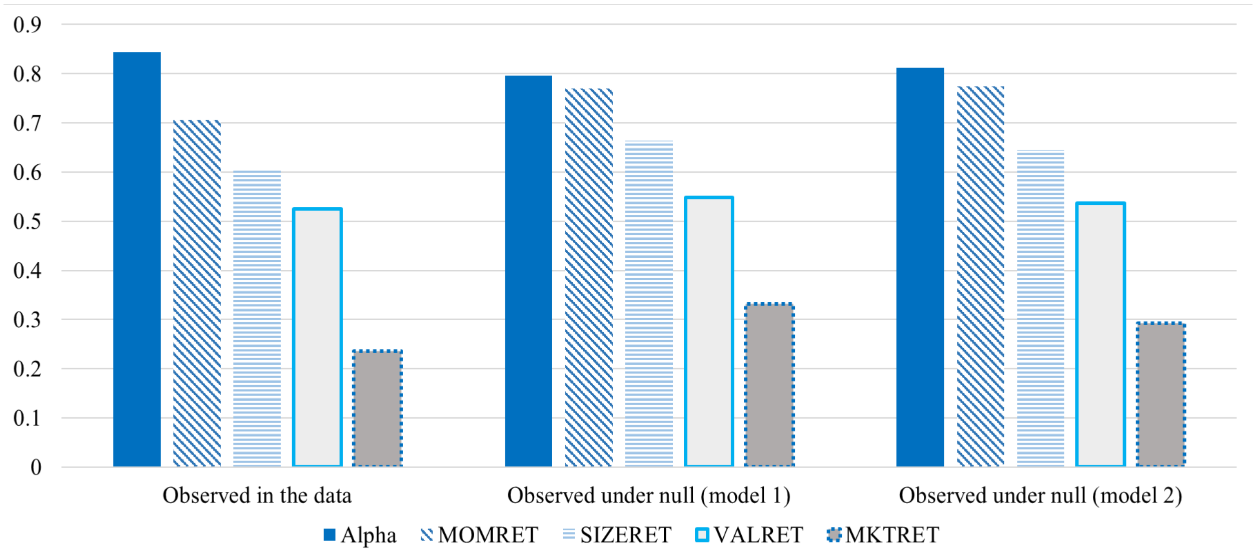
The results are striking. Applying the BHO methodology to simulated flows produces essentially identical results to the original BHO findings using actual flows. In Columns (4) and (5) of Table 2, we compare the coefficient estimates using the simulated flows and find

[12] The R^2s in the cross-sectional regressions are around 20%. In order to preserve the fund-level time-series correlation of fund flow residuals and the variation in their cross-sectional dispersion over time, we randomly select blocks of 18 consecutive residuals at the fund level, and then bootstrap these blocks across funds. Compared to simple bootstrapping, this method leads to average standard errors on estimated coefficients that are closer to those observed in the data, but does not materially alter the economic or statistical inference from this test.

that except for one coefficient they are statistically indistinguishable from the results found using actual flows. Therefore, we conclude that the BHO findings cannot reject the null hypothesis that investors chase unadjusted returns.

Our result is unlikely to be a coincidence. Not only do we reproduce BHO's "discounting for market" finding using simulated flows, we also reproduce the degree of "discounting" in the BHO regression for all of the other factors. This is also graphically shown in Figure 2. For instance, the BHO regression indicates that VALRET, the return related to the value factor, is also somewhat discounted. Using our test, we also observe this same result under the null.

Figure 2. Flow-performance sensitivity test. This figure displays the coefficient estimates for the response of fund flows to various components of fund returns as illustrated in Equation (2), ranked from larger to smaller. The same coefficients are also reported in Table 2. The first set of coefficients from the left is estimated using actual fund flows, and the second and third sets are estimated using flows generated under the null hypothesis that flows respond equally to all cross-sectional differences in fund returns, regardless of their source. Please refer to the text for additional details on the modelling of the null hypotheses.



4.1.3 Why Does the BHO Test Produce Misleading Results?

Why does the BHO test mistakenly conclude that investors adjust for market-related returns? We hypothesize that this is due to time-varying flow-performance sensitivity (FPS), through the following mechanism. During periods of extreme market returns, the cross-sectional dispersion in funds' market-related returns (MKTRET) is particularly large.[13] However, in periods when the market has extreme returns, investors respond the least to past performance (low flow-performance sensitivity; FPS), a fact that has been highlighted by Starks and Sun (2016), Franzoni and Schmalz (2017), and Harvey and Liu (2019). Thus, BHO's panel regressions place more weight on more volatile observations. As a result, the panel regression coefficient on MKTRET will overweight periods with low FPS relative to the coefficient on ALPHA.[14] This is a time-series effect, and it takes place regardless of whether investors care about exposure to the market factor in the cross-section or not.

The fact that flows simulated under the null—which has time-varying FPS but does not incorporate factor-discounting—can explain the findings of BHO is consistent with our hypothesis. We further test our hypothesis and find support for this mechanism. To keep the discussion brief, we present the results of the analysis in Appendix A.3.

4.2 BvB Cannot Reject the Null

4.2.1 Explanation of BvB's Methodology

BvB's approach is to examine the performance of an asset pricing model by computing how frequently the signs of the alpha from the model match the signs of net flows to the fund. We now explain the methodology used by BvB for readers' convenience. For each fund p in each month t, let $F_{p,t}$ denote the fund flow and let $\mathrm{ALPHA}^{\mathcal{M}}_{\mathrm{p,t}}$ denote the alpha estimated using the asset pricing model $\mathcal{M}$. Notice that $\mathrm{ALPHA}^{\mathcal{M}}_{\mathrm{p,t}}$ is calculated using historical returns prior to t. Following the method of BvB,[15] for each asset pricing model $\mathcal{M}$, we run the following regression:

$$\mathrm{sign}(F_{p,t}) = \beta^{\mathcal{M}}_0 + \beta^{\mathcal{M}}_1 \mathrm{sign}(\mathrm{ALPHA}^{\mathcal{M}}_{\mathrm{p,t}}) + \epsilon_{\mathrm{p,t}}, \tag{5}$$

where $\mathrm{sign}(F_{p,t})$ and $\mathrm{sign}(\mathrm{ALPHA}^{\mathcal{M}}_{\mathrm{p,t}})$ take on values in $\{-1, 1\}$. Lemma 2 of BvB shows that a linear transformation of the regression slope is directly related to the frequency at which

[13] The reason why MKTRET is widely dispersed is because it is computed as the fund's beta (which does not vary much over time) multiplied by the market return. During months of extreme returns, by construction, the market return is either a large positive or negative number, causing the distribution of MKTRET to be more dispersed.

[14] See Pástor, Stambaugh, and Taylor (2017) for a formal discussion of the relation between the coefficients from period-by-period regressions and panel regressions with fixed effects.

[15] Our implementation thus differs slightly from BvB, who use alphas that are contemporaneous with the flows. We lag the alphas by one month to avoid look-ahead bias and to be more consistent with the flow-performance literature.

the sign of the alpha and of the flow agree with each other:

$$\frac{\beta_1^{\mathcal{M}}+1}{2} = \frac{\Pr(\text{sign}(F_{p,t})=1|\text{sign}(\text{ALPHA}_{p,t}^{\mathcal{M}})=1)}{2} + \frac{\Pr(\text{sign}(F_{p,t})=-1|\text{sign}(\text{ALPHA}_{\text{p,t}}^{\mathcal{M}})=-1)}{2}. \quad (6)$$

In their Table 2, BvB find that the signs of the CAPM alpha match the signs of fund flows better than other commonly used risk models. The CAPM alpha also does better than the "market-adjusted" benchmark, defined as the fund return minus the market return. Thus, they conclude that the CAPM is closest to the "true" model used by investors.

4.2.2 BvB Performance with Null-Consistent Data

BvB's test differs from BHO's methodology on two dimensions. First, BvB use a non-parametric test. The test imposes no functional form on the relation between flows and alphas. The only restriction is that positive flows are expected to go to funds with positive signals. In their study, BvB already explore the performance of unadjusted returns against asset pricing models. Their test shows that unadjusted returns slightly underperform other asset pricing models. They do not, however, test the performance of Morningstar ratings. As seen in Section 2, there is good reason to believe that Morningstar ratings should be part of the null hypothesis, as ratings explain flows materially better than other signals.

Our test examines whether the BvB test can reject the null hypothesis that investors allocate money in accordance with Morningstar ratings as opposed to with asset pricing models. We assign a fund an indicator equal to 1 if its rating is $\geq i$ where, $i = 3, 4$, and 5. In our sample, funds with ratings ≥ 3, ≥ 4, and $= 5$ comprise, respectively, 68.9%, 31.8%, and 7.4% of fund-month observations. We estimate Equation (5) for each of the asset pricing models and for our rating-based heuristic models. Following BvB, we double-cluster standard errors by fund and month. The results are shown in the first two columns of Table 3.

Consistent with the findings of BvB, we also find that the CAPM performs better than the market-adjusted model, the FF three-factor model, and the FFC four-factor model.[16] However, the rating-based heuristic significantly outperforms the CAPM and the other models, and the degree of outperformance is larger than the maximum difference among all other models. The best-performing heuristic, which has investors reallocating money into 5-star funds, gets the sign of the flows right 67.9% of the time, while the CAPM gets the flow signs

[16] BvB also include some dynamic equilibrium models in their tests. In their study, these models are generally dominated by the CAPM and by multifactor models. Therefore, we do not include them in our tests.

Table 3. Horse race of different models. The first two columns are estimates of Equation (5) for each model considered. For ease of interpretation, the table reports $(\beta_1^{\mathcal{M}}+1)/2$ as a percentage, and the models are presented in decreasing order of the point estimate of $\beta_1^{\mathcal{M}}$. The remaining columns provide statistical significance tests of the pairwise model horse races based on Equation (7). Each cell reports the t-statistic of the hypothesis that $\beta^{\text{row}} > \beta^{\text{column}}$. For both univariate and pairwise tests, standard errors are double-clustered by fund and month.

Model	Estimate of $(\beta_1^{\mathcal{M}}+1)/2$	Univariate t-stat	Rating ≥ 4	Rating ≥ 3	CAPM	Market-adjusted	FF 3-factor	FFC 4-factor	Excess return
Rating = 5	67.90	29.17	5.47	9.38	9.55	10.44	11.88	12.20	11.03
Rating ≥ 4	64.41	36.20	-	9.94	7.93	8.95	11.08	11.56	8.41
Rating ≥ 3	61.04	32.28	-	-	1.02	2.50	4.16	4.86	5.06
CAPM	60.58	25.80	-	-	-	3.39	5.42	5.67	4.65
Market-adjusted	59.87	24.72	-	-	-	-	2.21	3.02	4.15
FF 3-factor	59.21	24.89	-	-	-	-	-	1.89	2.84
FFC 4-factor	58.94	25.85	-	-	-	-	-	-	2.59
Excess return	57.03	12.09	-	-	-	-	-	-	-

right 60.6% of the time. The difference is approximately 7.3%, which, for comparison, is twice as large as the 3.6% difference between the CAPM (60.6%) and the worst-performing model (excess returns, 57.0%).

To assess the statistical significance of these results, we follow BvB in conducting pairwise model horse races. For any two models $\mathcal{M}1$ and $\mathcal{M}2$, we run the following regression:

$$\text{sign}(F_{p,t}) = \gamma_0 + \gamma_1 \left(\frac{\text{sign}(\text{ALPHA}_{\text{p,t}}^{\mathcal{M}1})}{\widehat{\text{var}}(\text{ALPHA}_{\text{p,t}}^{\mathcal{M}1})} - \frac{\text{sign}(\text{ALPHA}_{\text{p,t}}^{\mathcal{M}2})}{\widehat{\text{var}}(\text{ALPHA}_{\text{p,t}}^{\mathcal{M}2})} \right) + \xi_{p,t}, \tag{7}$$

where $\widehat{\text{var}}(\text{ALPHA}_{\text{p,t}}^{\mathcal{M}1})$ and $\widehat{\text{var}}(\text{ALPHA}_{\text{p,t}}^{\mathcal{M}2})$ are the sample variance of the alpha measures. Following BvB, we consider $\mathcal{M}1$ to be a better model of investor behavior if $\gamma_1 > 0$ with statistical significance. We double-cluster standard errors by fund and month. The results are reported in the remaining columns in Table 3. The first two ratings-based models both outperform the CAPM with strong statistical significance, with t-statistics of 9.55 and 7.93, respectively.[17]

Based on the diagnostic test devised by BvB, the conclusion is that Morningstar ratings

[17]In an untabulated exercise, we show that this finding is robust to using different past return windows ranging from one month to ten years.

explain investors' capital reallocation better than the CAPM and all other asset pricing models considered.

4.3 Economic Significance of Morningstar's Outperformance

The BvB test is a theoretically grounded application of the idea that investors channel flows to funds with positive NPV. However, by only using signs of alphas and flows, it disregards more granular variation in alpha and is not designed to shed light on economic magnitudes. In this section, we carry out additional tests to address these concerns.

We rank funds using different measures and examine the net flow difference between top- and bottom-ranked funds. Because Morningstar ratings are discrete (1 to 5 stars), we also discretize the alphas of asset pricing models to place the different measures on equal footing. Specifically, we sort funds in each month by each performance measure and use the number of 5-star and 1-star funds to classify top- and bottom-ranked funds, respectively. For instance, if there were 150 funds with a 5-star rating, then the 150 funds with the highest CAPM alpha are defined as top-ranked by CAPM. On average, 7.4% and 6.9% of fund-month observations are classified as top- and bottom-ranked, respectively. Then, for top- and bottom-ranked funds, we calculate the fraction of funds with positive flows, the average flow as a fraction of TNA, and the average dollar flow. Table 4 reports the results.[18]

When classifying fund-months based on Morningstar ratings, 67.3% of top-ranked funds receive positive flows whereas only 16.8% of bottom-ranked funds receive positive flows, a spread of 50.6%. This is significantly higher than all other measures, which generate spreads in the 40.6% to 45.8% range. Morningstar also outperforms by a sizable margin in predicting dollar flow measures, indicating that the outperformance is economically meaningful.[19] Note that, in these tests, the CAPM model no longer outperforms rankings based on unadjusted returns.

In sum, the test proposed by BvB cannot reject the null hypothesis that investors use Morningstar to guide their investments. In fact, Morningstar performs significantly better relative to all asset pricing models tested, and the rankings of asset pricing models vary based on the method and measure of flows used.

[18] Because we rank funds within each month, rankings based on unadjusted returns, returns in excess of the risk-free rate, and returns in excess of the market return are all the same. Therefore, we report the results for these measures only once under the label "market-adjusted."

[19] When classifying both 4- and 5-star funds to be top-ranked and 1- and 2-star funds to be bottom-ranked, we still find that Morningstar outperforms all asset pricing models. The results are available upon request.

Table 4. Flows to top- and bottom-ranked funds. The top-ranked funds are the 5-star funds and the best 7.4% of funds for each model. The bottom-ranked funds are the 1-star funds and the worst 6.9% of funds for each model. Because funds are ranked within each month, rankings based on unadjusted returns, returns in excess of the risk-free rate, and returns in excess of the market are the same. We thus report the results for these ranking rules under the same label "Unadjusted."

	Positive flow (%)			**Fund flow (%)**			**Fund flow ($ Mn)**		
	Top	**Bottom**	**Diff**	**Top**	**Bottom**	**Diff**	**Top**	**Bottom**	**Diff**
Morningstar	67.3	16.8	50.6	1.15	−1.53	2.68	37.3	−8.0	45.3
Unadjusted	61.7	15.9	45.8	0.99	−1.85	2.85	18.8	−15.9	34.7
CAPM	60.3	15.2	45.2	0.93	−1.91	2.84	18.0	−15.2	33.2
FF 3-factor	57.8	16.1	41.6	0.80	−1.81	2.61	15.6	−13.8	29.4
FFC 4-factor	57.5	17.0	40.6	0.78	−1.76	2.54	14.8	−12.3	27.1

5 How Do Investors Utilize Morningstar Ratings?

Our results so far show that fund flows do not appear to adjust for risk exposure in a way that is consistent with commonly-used asset pricing models. At the same time, we find that a significant portion of flows respond to Morningstar ratings. One may still hypothesize that the investors who rely on Morningstar intend to perform risk adjustment, but use the conveniently available ratings as a proxy for risk adjustment, especially because Morningstar ranks funds within size and value categories in recent years.

An alternative hypothesis is that investors use Morningstar ratings naïvely without understanding the process underlying the ratings. In this sense, Morningstar may have a comforting role for investors (Gennaioli et al., 2015). Investors may not have specific risk adjustment in mind, but rather trust Morningstar to pick the best investments for them. This explanation is consistent with Mullainathan et al. (2008) who document that mutual fund advertisements emphasize trust and expertise, rather than performance metrics. Morningstar may be viewed by investors as an independent expert in evaluation of mutual funds.[20]

To investigate the role that Morningstar plays in investors' decision making process, we exploit a major change in the methodology Morningstar uses to assign ratings. Prior to June 2002, U.S. equity funds were ranked against all other funds (despite the fact that

[20]Consistent with this explanation, Morningstar sends similar messages to investors in its website. As of May 2019, Morningstar's website have headers like "[Morningstar provides] Independent investment research" as well as links to the "Best investments".

Morningstar style categories existed before that date). Starting from June 2002, star ratings are assigned to funds based on their performance relative to other funds in their assigned 3×3 size-value style category.[21] Therefore, before June 2002, mediocre fund managers whose benchmark's style happened to outperform other styles were mechanically assigned high ratings. If investors understand how ratings are constructed and intend to use them to adjust for size and value exposure, then they should not have relied on these ratings before June 2002.

Figure 3, Panel A, illustrates the effect of the methodological change and the extent to which ratings explained flows before and after this event. For each month in the sample, the bars show the fraction of funds with rating changes. In most months, an average of 11% of funds experience rating changes,[22] while 54% of funds changed ratings in June 2002. Using data before versus after the change, we regress fund flows on rating and style indicator variables while simultaneously controlling for past fund returns:

$$F_{p,t} = b_0 + \gamma X_{p,t} + b_1 R_{p,t} + \sum_{k=1}^{5} b_k \mathbf{I}_{(\text{rating}=k)} + \sum_{s} c_s \mathbf{I}_{(\text{style category}=s)} + \epsilon_{p,t}.$$

We then examine the marginal R^2 of rating and style indicators.[23] These figures are respectively plotted as blue solid lines and red dashed lines in Figure 3, Panel A, (right axis). We find that the marginal R^2 of ratings is almost identical before and after the change (4.2% vs 4.5%, respectively). Thus, before June 2002, ratings explained flows about as much as

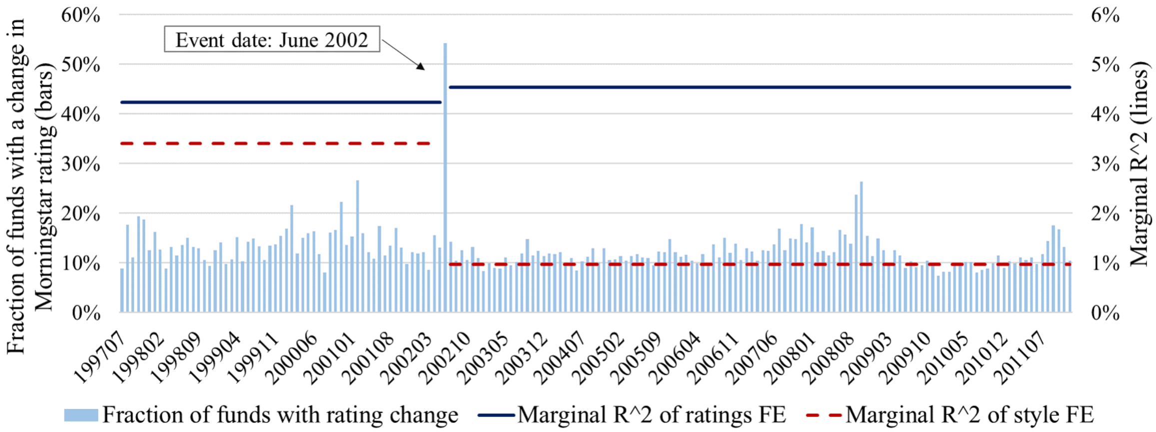
[21] Morningstar classifies diversified U.S. equity mutual funds into style categories based on the so-called Morningstar style box. Each fund is assigned to one of nine style categories based on its size tilt (small, mid-cap, or large) and value tilt (value, blend, or growth). This information is usually presented together with a fund's star rating in a fund summary and in marketing materials (e.g., `http://corporate.morningstar.com/us/documents/MethodologyDocuments/MethodologyPapers/MorningstarCategory_Classifications.pdf`). When assigning a fund to a given style group, Morningstar uses the fund's actual stock holdings. The fact that Morningstar provides an independent style categorization can potentially be useful to investors because fund managers sometimes choose inappropriate self-specified benchmarks (Sensoy, 2009). See `https://corporate.morningstar.com/US/documents/MethodologyDocuments/FactSheets/MorningstarRatingForFunds_FactSheet.pdf` for details.

[22] This is not surprising because, given that funds are assigned to one of five rating categories and therefore there are four thresholds, at any given time a large number of funds are just above or just below one of the thresholds. Consequently, a small difference in performance in the previous month can make the rating change at the beginning of the current month. Moreover, a number of other methodological details increase the frequency at which funds just above or below the threshold tend to cross the threshold. For instance, when a new fund enters or leaves the sample of funds that are rated by Morningstar (e.g., if a fund is liquidated or merges with another fund due to poor performance), the relative ranking of other funds will change, leading some to cross one of the thresholds even if their relative performance does not change. In unreported analysis, we find that about 40% of the changes in ratings are reversed within two months (except for the June 2002 changes, of which only 15% reverted within two months).

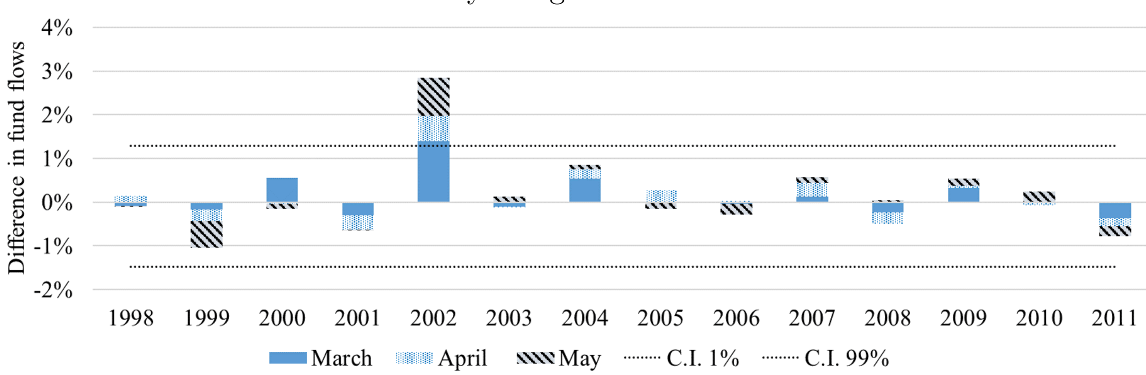
[23] In the full sample, they are equal to 4.5% and 1.5%, respectively.

Figure 3. Event study: Change in the Morningstar methodology. In June 2002, Morningstar started to rank funds within style categories as opposed to across all U.S. equity funds. In Panel A, we show the fraction of funds with a change in rating as well as the marginal R^2 of ratings and style fixed effects in flow-performance regressions before and after the event. In Panels B and C, we present the results of an event study based on the change in methodology. Years other than 2002 serve as placebo tests. Bootstrapped confidence intervals at the 1% and 99% levels are also presented.

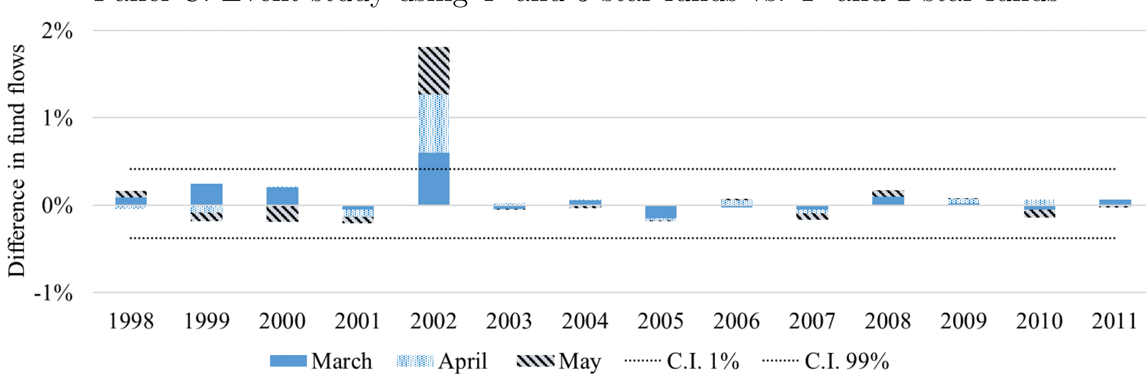
Panel A: Rating changes by month



Panel B: Event study using 5-star funds vs. 1-star funds



Panel C: Event study using 4- and 5-star funds vs. 1- and 2-star funds



after June 2002, even though pre-June 2002 ratings did not adjust for style. In fact, the explanatory power of style fixed effects drops from 3.4% before the change in methodology to 1.0% after.

We present evidence that further strengthens the conclusion that investors did not perform style adjustments before Morningstar started doing so in June 2002. In the months of March, April, and May of each year, we calculate the average fund flow based on the current rating versus the counterfactual rating in June. We then calculate the following measure:

$$(\overline{\text{flow}}_{\text{current high}} - \overline{\text{flow}}_{\text{current low}}) - (\overline{\text{flow}}_{\text{June high}} - \overline{\text{flow}}_{\text{June low}}), \tag{8}$$

where the first term is the spread in flows between high-rated and low-rated funds based on the actual current rating, and the second term is the spread based on the June rating.[24] In 2002, if this measure is negative, it means that investors independently adjusted for style returns in fund performance. In Panel B, 5-star funds are considered high-rated and 1-star funds are considered low-rated. In Panel C, 4- and 5-star funds are considered high-rated and 1- and 2-star funds are considered low-rated. We calculate the same measure for other years as placebo tests and use the results to generate confidence intervals.[25]

The calculated measure for the year 2002 is positive and statistically different from zero in both versions of the test. This result implies that, on average, investors did not account for style benchmarks before June 2002, but rather simply moved money into funds with high absolute performance and out of funds with low absolute performance. This evidence suggests that investors followed the actual rating and failed to distinguish between ratings that were high because of relative outperformance and ratings that were high because of high average style returns.

The event study carried out above highlights two clear facts. First, controlling for past fund returns, Morningstar ratings had the same strong influence on investors' allocation

[24] One might be concerned that this test could be subject to look-ahead bias. In March, April, and May, we sort funds based on their (future) June rating, which partially depends on future information about fund performance. If fund flows predict the cross-section of future fund performance (at the one-, two-, and three-month horizons), then the measure we compute using Equation (8) might be biased. However, we view this as improbable for several reasons. First, it seems unlikely that flows would systematically predict changes in performance that are large enough to lead to significant changes in future ratings within three months. Second, we calculate the same measure in all other years as a placebo test to verify whether this potential look-ahead bias is indeed an issue. As Figure 3 shows, there is no bias, i.e., in the placebo years, the measure is positive 11 times, negative 11 times, and virtually zero 4 times.

[25] We do so using a bootstrap procedure in which we first re-sample the focal year and then randomly draw three months within that year (with replacement). Based on this nonparametric test, we can conclude not only that the direction of the flows in the months preceding June 2002 was the opposite of what we would expect if investors independently adjusted for style exposure, but also that the observed flow pattern was not random.

decisions both before and after Morningstar introduced peer-benchmarking based on size and value categories. Second, investors do not appear to independently adjust for style benchmarks by themselves.

Our results indicate that investors rely on Morningstar's ratings blindly, irrespective of its ranking methodology. These findings are most consistent with the explanation that investors view Morningstar's ratings as a recommendation about the best funds from an independent expert. As in Gennaioli et al. (2015), Morningstar's is a trusted advisor rather than a vehicle for risk adjustment.

6 Conclusion

Understanding how investors allocate capital is an important question in the study of financial markets. Because of its wealth of data on both fund performance and investor reaction to this performance, the actively managed mutual fund industry has been used extensively as a window to understanding investor behavior. Two influential studies, Berk and van Binsbergen (2016) (BvB) and Barber et al. (2016) (BHO), find that investors appear to behave as if they adjust for fund performance explained by market exposure. This conclusion is in sharp contrast to the prior literature that finds that investors chase unadjusted fund returns and external ratings—which, we argue, is the sensible null hypothesis to compare against.

In this paper, we find that neither the BHO nor the BvB tests can reject our basic null hypothesis. The results of BHO are mostly spurious. Once the methodological issues are addressed, we find that their results are consistent with the null hypothesis that investors do not adjust for any factor exposure. When we use the BvB test, we also find that rating-chasing explains flows much better than any asset pricing model, so the BvB test also cannot reject the null. Finally, we show that investors appear to follow ratings blindly, not likely understanding the way in which Morningstar constructs its ratings. Morningstar may make investors feel better about their investment decisions à-la money doctors (Gennaioli et al., 2015).

Given that a slew of papers have either adopted the methodology of BvB and BHO, or accepted their finding that "investors use CAPM", we believe it is important to clarify what we truly can learn from the behavior of mutual fund investors. Overall, our results show that retail investors do not appear to incorporate risk adjustment into their investments, but rather pursue easy-to-follow signals and advice that they trust.

References

Adam, Klaus, Albert Marcet, and Johannes Beutel, 2017, Stock price booms and expected capital gains, *American Economic Review* 107, 2352–2408.

Adam, Klaus, Dmitry Matveev, and Stefan Nagel, 2018, Do survey expectations of stock returns reflect risk-adjustments?, Working paper, University of Chicago.

Agarwal, Vikas, Clifton Green, and Honglin Ren, 2018, Alpha or beta in the eye of the beholder: What drives hedge fund flows, *Journal of Financial Economics* 127, 417–434.

Akbas, Ferhat, Will J. Armstrong, Sorin Sorescu, and Avanidhar Subrahmanyam, 2015, Smart money, dumb money, and capital market anomalies, *Journal of Financial Economics* 118, 355–382.

Akbas, Ferhat, and Egemen Genc, 2019, Do mutual fund investors overweight the probability of extreme payoffs in the return distribution?, *Journal of Financial and Quantitative Analysis* forthcoming.

Bacchetta, Philippe, Elmar Mertens, and Eric Van Wincoop, 2009, Predictability in financial markets: What do survey expectations tell us?, *Journal of International Money and Finance* 28, 406–426.

Bailey, Warren, Alok Kumar, and David Ng, 2011, Behavioral biases of mutual fund investors, *Journal of Financial Economics* 102, 1–27.

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

Barber, Brad M., Terrance Odean, and Lu Zheng, 2005, Out of sight, out of mind: The effects of expenses on mutual fund flows, *Journal of Business* 78, 2095–2120.

Benartzi, Shlomo, and Richard H. Thaler, 2001, Naive diversification strategies in retirement saving plans, *American Economic Review* 91, 79–98.

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

Blake, Christopher R., and Matthew R. Morey, 2000, Morningstar ratings and mutual fund performance, *Journal of Financial and Quantitative Analysis* 35, 451–483.

Blocher, Jesse, and Marat Molyboga, 2017, The revealed preference of sophisticated investors, *European Financial Management* 23, 839–872.

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

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

Chakraborty, Indraneel, Alok Kumar, Tobias Muhlhofer, and Ravi Sastry, 2018, Does limited investor attention explain mutual fund flows? Evidence from sector funds, Working paper, University of Miami.

Chevalier, Judith, and Glenn Ellison, 1997, Risk taking by mutual funds as a response to incentives, *Journal of Political Economy* 105, 1167–1200.

Choi, James J., and Adriana Z. Robertson, 2018, What matters to individual investors? Evidence from the horse's mouth, Working paper, Yale University.

Christoffersen, Susan E. K., David K. Musto, and Russ Wermers, 2014, Investor flows to asset managers: Causes and consequences, *Annual Review of Financial Economics* 6, 289–310.

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

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

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

Evans, Richard B., and Yang Sun, 2018, Models or stars: The role of asset pricing models and heuristics in investor risk adjustment, Working paper, University of Virginia.

Franzoni, Francesco, and Martin C. Schmalz, 2017, Fund flows and market states, *Review of Financial Studies* 30, 2621–2673.

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

French, Kenneth R., 2008, Presidential address: The cost of active investing, *Journal of Finance* 63, 1537–1573.

Friesen, Geoffrey C., and Viet Nguyen, 2018, The economic impact of mutual fund investor behaviors, Working paper, University of Nebraska-Lincoln.

Friesen, Geoffrey C., and Travis R.A. Sapp, 2007, Mutual fund flows and investor returns: An empirical examination of fund investor timing ability, *Journal of Banking & Finance* 31, 2796–2816.

Gennaioli, Nicola, Andrei Shleifer, and Robert Vishny, 2015, Money doctors, *Journal of Finance* 70, 91–114.

Greenwood, Robin, and Andrei Shleifer, 2014, Expectations of returns and expected returns, *Review of Financial Studies* 27, 714–746.

Hartzmark, Samuel M., and Abigail Sussman, 2019, Do investors value sustainability? A natural experiment examining ranking and fund flows, *Journal of Finance* forthcoming.

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

Harvey, Campbell R., and Yan Liu, 2019, Cross-sectional alpha dispersion and performance evaluation, *Journal of Financial Economics* fortchcoming.

Heuer, Justus, Christoph Merkle, and Martin Weber, 2017, Fooled by randomness: Investor perception of fund manager skill, *Review of Finance* 21, 605–635.

Jain, Prem C., and Joanna S. Wu, 2000, Truth to mutual fund advertising: Evidence on future performance and fund flows, *Journal of Finance* 55, 937–958.

Jegadeesh, Narasimhan, and Chandra Sekhar Mangipudi, 2017, What do fund flows reveal about asset pricing models?, Working paper, Emory University.

Jones, A. Michael, and Tom Smythe, 2003, The information content of mutual fund print advertising, *Journal of Consumer Affairs* 37, 22–41.

Kaniel, Ron, and Robert Parham, 2017, WSJ category kings: The impact of media attention on consumer and mutual fund investment decisions, *Journal of Financial Economics* 123, 337–356.

Karceski, Jason, 2002, Returns-chasing behavior, mutual funds, and beta's death, *Journal of Financial and Quantitative Analysis* 37, 559–594.

Koehler, Jonathan J., and Molly Mercer, 2009, Selection neglet in mutual fund advertisements, *Management Science* 55, 1107–1121.

Morey, Matthew R., 2003, The kiss of death: A 5-star Morningstar mutual fund rating?, Working paper, Pace University.

Mullainathan, Sendhil, Joshua Schwartzstein, and Andrei Shleifer, 2008, Coarse thinking and persuasion, *Quarterly Journal of Economics* 123, 577–619.

Pástor, L'uboš, and Robert F. Stambaugh, 2002, Mutual fund performance and seemingly unrelated assets, *Journal of Financial Economics* 63, 315–349.

Pástor, Ľuboš, Robert F. Stambaugh, and Lucian A Taylor, 2017, Do funds make more when they trade more?, *Journal of Finance* 72, 1483–1528.

Phillips, Blake, Kuntara Pukthuanthong, and P. Raghavendra Rau, 2016, Past performance may be an illusion: Performance, flows, and fees in mutual funds, *Critical Finance Review* 5, 351–298.

Reuter, Jonathan, and Eric Zitzewitz, 2006, Do ads influence editors? Advertising and bias in the financial media, *Quarterly Journal of Economics* 121, 197–227.

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

Sapp, Travis, and Ashish Tiwari, 2004, Does stock return momentum explain the "smart money" effect?, *Journal of Finance* 59, 2605–2622.

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

Sharpe, William F., 1998, Morningstar's risk-adjusted ratings, *Financial Analyst Journal* 54, 21–33.

Sirri, Erik R., and Peter Tufano, 1998, Costly search and mutual fund flows, *Journal of Finance* 53, 1589–1622.

Solomon, David H., Eugene Soltes, and Denis Sosyura, 2014, Winners in the spotlight: Media coverage of fund holdings as a driver of flows, *Journal of Financial Economics* 113, 53–72.

Starks, Laura, and Sophia Sun, 2016, Economic policy uncertainty, learning and incentives: Theory and evidence on mutual funds, Working paper, University of Texas.

Vissing-Jorgensen, Annette, 2003, Perspectives on behavioral finance: Does "irrationality" disappear with wealth? Evidence from expectations and actions, *NBER Macroeconomics Annual* 18, 139–194.

Appendix A Additional Results

A.1 The Explanatory Power of Past Returns versus Ratings

Figure 1 and Table 4 show that Morningstar ratings explain flows better than past returns. In those tests, we have followed BHO and measured past performance as a weighted average of monthly returns over the past 18 months. In this section, we find that even when we allow fund flows to depend on past returns in more flexible ways, ratings still explain more flow variation than return chasing—the most cited and studied determinant of fund flows (Christoffersen et al., 2014).

To allow for flexible flow dependence on returns, we regress fund flows on lagged ratings and up to $H = 120$ lags (ten years) of past monthly returns:

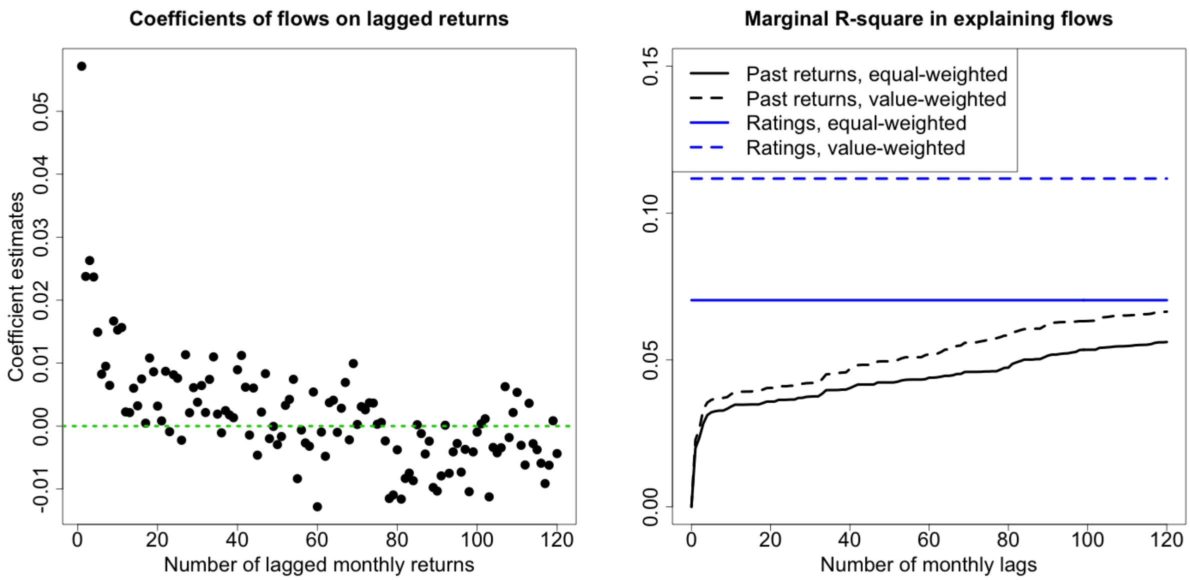
$$F_{p,t} = b_0 + b_1 \sum_{k=1}^{5} \mathbf{1}_{\text{rating}=k} + c_1 R_{p,t-1} + c_2 R_{p,t-2} ... + c_H \cdot R_{p,t-H} + \epsilon_{p,t}. \tag{9}$$

The left panel of Figure A.I plots the regression coefficients on lagged returns. Consistent with BHO, the flow response coefficients to past flows becomes smaller as we extend the horizon, and becomes indistinguishable from zero after roughly 20 lags.

The right panel of Figure A.I plots the marginal adjusted R^2 of ratings versus past returns as we increase the number of lagged returns. Strikingly, even if we allow for 120 lags, the marginal R^2 of past returns can only reach 6.6% when weighting funds by assets under management (AUM) and 5.6% when equal-weighting. In contrast, even after controlling for all 120 lags of returns, lagged ratings explain 11.2% and 7.0% of flow variation in AUM-weighted and equal-weighted regressions, respectively. Therefore, we conclude that rating-chasing is quantitatively more important than the well-documented return-chasing phenomenon.

In sum, we quantify the explanatory power of rating-chasing and find that it explains even more flow variation than return-chasing, a fact that has not yet been documented in the literature.

Figure A.I. Explanatory power of past returns and ratings. We regress monthly fund flows on up to 120 lagged monthly returns and lagged Morningstar ratings (Regression (9)). The left panel plots the regression coefficients on lagged monthly returns (equal-weighted regression). The right panel plots the marginal R^2 of past returns (black) versus ratings (blue). The horizon axis denotes the number of lags of past returns included. The marginal R^2 for ratings plotted is when we have controlled for all 120 past monthly returns.



A.2 Estimating Alphas Using Asset Pricing Models

In this section, we describe how we calculate the alphas for the various models. Our methodology tracks very closely the procedure used by BHO, and it is also similar to the method used in BvB.

Consider the Fama-French-Carhart (FFC) four-factor model as an example. For each fund p in month t, we estimate the following time-series regression using the 60 months of returns from month $t-60$ to month $t-1$:

$$\begin{aligned} R_{p,\tau} - RF_{\tau} &= a_{p,t}^{\text{FFC}} + b_{p,t}(\text{MKT}_{\tau} - RF_{\tau}) + s_{p,t}\text{SMB}_{\tau} + h_{p,t}\text{HML}_{\tau} \\ &\quad + u_{p,t}\text{UMD}_{\tau} + \epsilon_{p,\tau}, \qquad \tau = t-60, \ldots, t-1. \end{aligned} \tag{10}$$

Here, $R_{p,\tau}$ is the fund return net of fees in month τ, and RF_{τ} is the one-month Treasury bill rate. $\text{MKT}, \text{SMB}, \text{HML}$, and UMD are the market, size, value, and momentum factors in Carhart (1997), respectively.[26] The regression intercept $a_{p,t}^{\text{FFC}}$ is the four-factor-adjusted average return, while regression coefficients $b_{p,t}$, $s_{p,t}$, $h_{p,t}$, and $u_{p,t}$ capture fund exposures to the four factors, respectively. Following BHO, we next estimate alpha as the realized return not explained by lagged factor exposures:

$$\hat{\alpha}_{p,t}^{\text{FFC}} = R_{p,t} - RF_t - \left[\hat{b}_{p,t}(\text{MKT}_t - RF_t) + \hat{s}_{p,t}\text{SMB}_t + \hat{h}_{p,t}\text{HML}_t + \hat{u}_{p,t}\text{UMD}_{\text{t}}\right], \quad (11)$$

where $\hat{b}_{p,t}$, $\hat{s}_{p,t}$, $\hat{h}_{p,t}$, and $\hat{u}_{p,t}$ are the estimated regression coefficients in Equation (10).

Investors often respond to fund performance slowly (Coval and Stafford, 2007). Therefore, we follow BHO and model the response of flows to past returns using an exponential decay model over the past 18 months. For instance, we calculate the four-factor alpha using

$$\text{ALPHA}_{p,t}^{\text{FFC}} = \frac{\sum_{s=1}^{18} e^{-\lambda(s-1)}\hat{\alpha}_{p,t-s}^{\text{FFC}}}{\sum_{s=1}^{18} e^{-\lambda(s-1)}}, \quad (12)$$

where $\hat{\alpha}_{p,t}^{\text{FFC}}$ is from Equation (11) and the decay parameter $\lambda = 0.20551497$ is the same used in BHO. λ is estimated from the empirical relationship between flows and past returns at different lags. The advantage of this weighting method is that it does not require researchers to arbitrarily assume that investors respond to performance over a specific horizon.

Similarly, we calculate the CAPM alpha as

$$\text{ALPHA}_{p,t}^{\text{CAPM}} = \frac{\sum_{s=1}^{18} e^{-\lambda(s-1)}\hat{\alpha}_{p,t-s}^{\text{CAPM}}}{\sum_{s=1}^{18} e^{-\lambda(s-1)}}, \quad (13)$$

$$\text{where } \hat{\alpha}_{p,t}^{\text{CAPM}} = R_{p,t} - RF_t - \hat{\beta}_{p,t}(\text{MKT}_t - RF_t), \quad (14)$$

and $\hat{\beta}_{p,t}$ is estimated using univariate regressions of fund returns on market returns in the 60 months prior to t. Finally, using the same weighting scheme, we also calculate alphas relative to the Fama-French three-factor model ($\text{ALPHA}_{p,t}^{\text{FF}}$).

[26] We download Treasury bill rates and factor returns from Kenneth French's website (`http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`).

A.3 A Deeper Dive into BHO's Tests

A.3.1 Does Time-Varying FPS Distort BHO's Results?

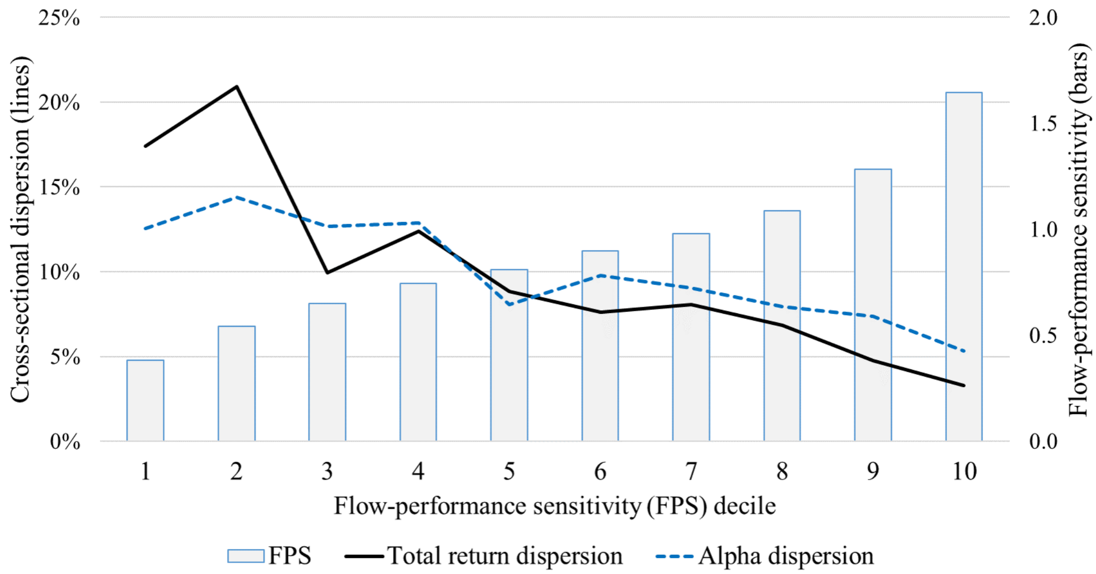
Our hypothesis is that time-varying flow-performance sensitivity (FPS) can mechanically generate the market-discounting results in BHO's panel regression with time FEs. To examine our hypothesis, we first verify the empirical relationship between FPS and the dispersion of different return components in our data. In Figure A.II, we sort the 175 calendar months in the sample into ten groups based on FPS in that month. We measure each month's FPS as the slope in the monthly cross-sectional regressions of fund flows on prior 18-month weighted total fund returns. In each of the five plots, the bars represent the average FPS in each FPS-sorted decile. The lines represent the average dispersion (cross-sectional variance) in fund return components in each decile. For ease of discussion, the dispersion is presented as the average variance in each decile relative to the sum of the average variance across all deciles.

As hypothesized, Figure A.II shows that cross-sectional dispersion in factor-related returns is particularly high in periods with low FPS. In Panel A, we plot the dispersion in total fund returns and in the four-factor alphas. In the two deciles with the lowest FPS, total fund returns have higher dispersion than alphas, while the opposite is true in the five deciles with high FPS. Panel B shows the average dispersion of each factor-related return in a given FPS decile against that of the alpha. Consistent with Starks and Sun (2016), Franzoni and Schmalz (2017), and Harvey and Liu (2019), the dispersion in the market-related return is particularly high when the FPS is low. We also observe similar patterns for the value factor and, to a lesser extent, for the size and momentum factors.

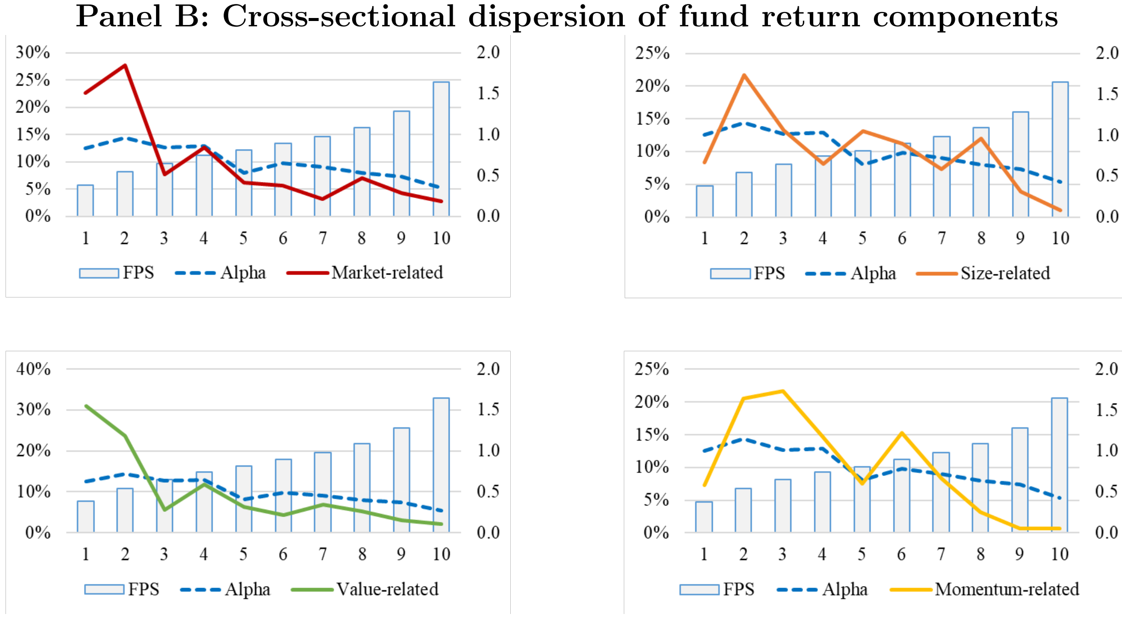
Note that the patterns discussed here are economically large. For instance, the average FPS in the top FPS decile is more than four times as large as it is in the bottom decile. The dispersion of the market-related return in the bottom FPS decile is eight times as large as in the top decile. Therefore, this pattern can generate significant downward bias in the

Figure A.II. Return dispersion and time-varying flow-performance sensitivity. We split the entire sample period into ten subperiods based on the strength of the cross-sectional flow-performance sensitivity (FPS) in each month. In Panel A, we report the fraction of cross-sectional dispersion in total fund returns and four-factor alphas in each of the ten subperiods. In Panel B, we report the fraction of cross-sectional dispersion in the four factor-related return components and four-factor alphas in each of the ten subperiods. In Panel B, the axes represent the same variables as in Panel A, but their labels are not reported in the interest of space.

Panel A: Cross-sectional dispersion of fund returns and alphas



Panel B: Cross-sectional dispersion of fund return components



panel regression estimate of the market-related coefficient. This evidence serves to support and explain the results we presented in Section 4.1 suggesting that the main test of BHO produces spurious results.

A.3.2 Flow-Performance Horse-Race Regressions Are Not Reliable

In Section 4.1, we took a closer look at the test specification used by BHO that appeared to provide the strongest evidence of market-related return discounting. We found that the data are actually consistent with investors treating all sources of fund returns equally. Here, we present additional evidence that other specifications of BHO can also lead to misleading results that favor the CAPM.

In particular, an intuitive way to infer which model investors use to allocate capital across mutual funds is to regress flows on two different model-based measures of performance and see which one correlates with flows the most. BHO found that CAPM alphas correlate with flows more than Fama-French three-factor alphas and alphas from all models that include additional factors. Moreover, they also found that market-adjusted returns correlate with flows less than CAPM alphas but more than alphas from models with additional factors.

These "horse-race" tests are also likely to be biased and misleading, for the same reasons we outlined throughout Section 4.1. To verify this hypothesis, following the logic of the test carried out in Table 2, we regress actual fund flows and counterfactual fund flows (again generated under the null that flows respond equally to all components of fund returns as per Equations (3) and (4)) onto pairs of different measures of fund performance. We further control for fund characteristics (size, age, etc.) and time FE, as did BHO. The results are presented in Figure A.III.

Based on (unreported) bootstrapped standard errors, in all horse races, the coefficient on each performance measure obtained using true fund flows is not significantly statistically different from those obtained under the null hypotheses. Figure A.III shows that these alternative specifications also tend to produce spurious results. For instance, CAPM alphas *appear* to predict actual flows *and* counterfactual flows better than unadjusted returns, even though counterfactual flows are generated under the null that flows respond equally to all components of fund returns. Hence, we conclude that these additional tests cannot reject the hypothesis that investors' flows do not distinguish between different sources of fund returns.

Appendix B Overview of Morningstar Ratings

Mutual funds have become increasingly popular over the last 35 years as a way to own stocks (French, 2008). The increasing demand has led to an explosion in the number of funds offered, and currently, the number of existing U.S. equity funds exceeds the number of publicly traded firms. The large number of available products created the need to classify and rate these funds. The fund rating industry emerged to satisfy this need.

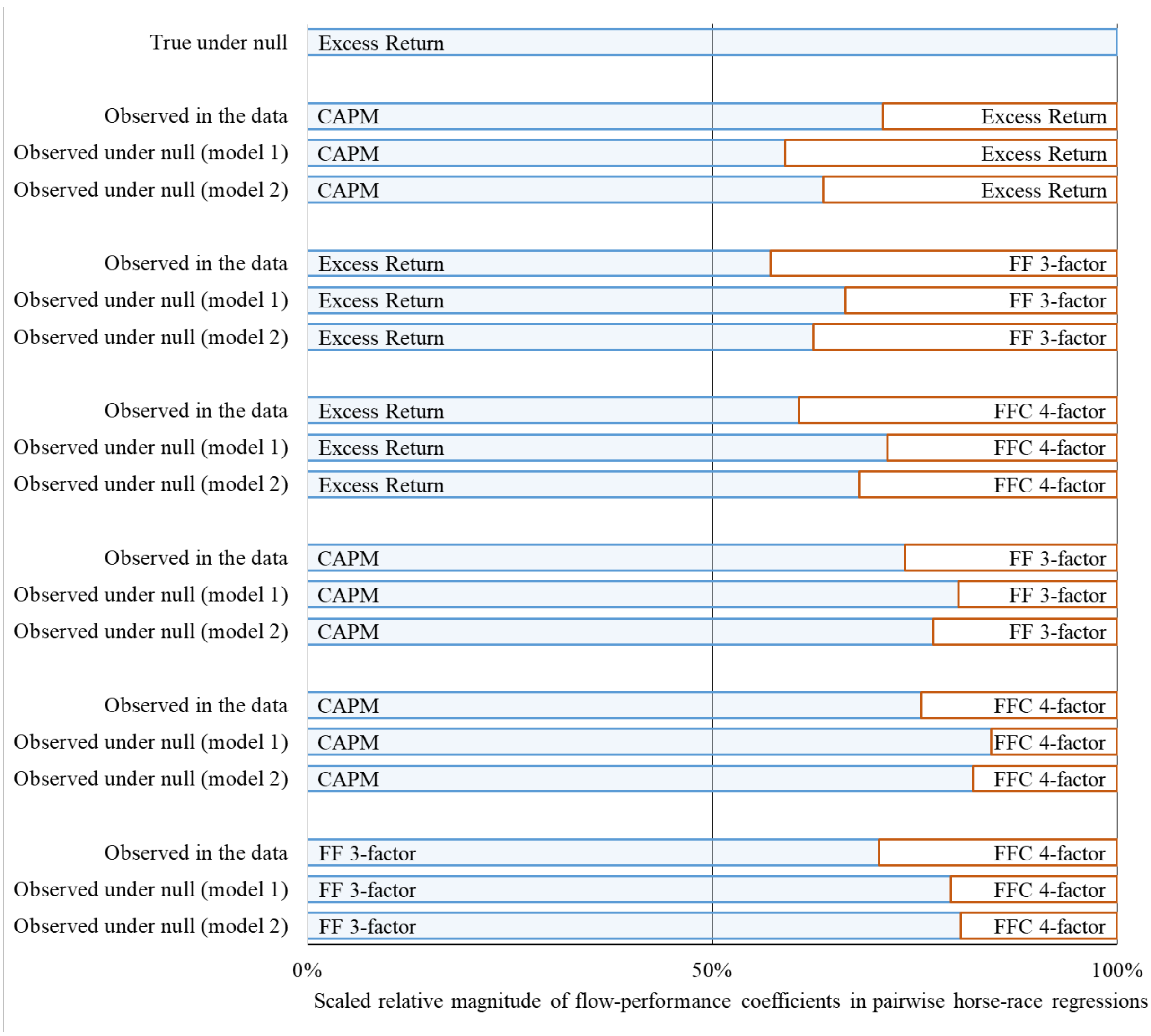
Morningstar explains its rating methodology in a publicly available manual.[27] Ratings are assigned using a relative ranking system and are updated each month. Mutual funds are ranked against funds in their peer group using past volatility-adjusted returns, and peer groups are defined as style categories (e.g., foreign large value) within broadly defined groups (e.g., international equities). Consistent with the relevant literature, our study focuses on U.S. equities funds that Morningstar assigns to one of nine (3×3) styles based on their size tilt (small, mid-cap, or large) and value tilt (value, blend, or growth).[28] Ranking within styles was introduced in June 2002; before then, all U.S. equity funds were ranked against each other without regard for their investment style (see Section 5). The top 10% of funds within each style category are assigned five stars. The subsequent 22.5%, 35%, 22.5%, and 10% of funds are assigned four, three, two, and one stars, respectively.

Morningstar summarizes a fund's past performance using the so-called Morningstar risk-

[27] The Morningstar manual is available at `https://corporate.morningstar.com/US/documents/MethodologyDocuments/FactSheets/MorningstarRatingForFunds_FactSheet.pdf`. See also Blume (1998).

[28] An additional category, called "Leveraged Net Long," was introduced in the U.S. equities group as of September 30, 2007. We do not include these funds in our sample.

Figure A.III. Spurious flow-performance correlations in pairwise model horse-race regressions. This figure displays the coefficient estimates for pairwise model horse races in explaining the percentage of fund flows in a panel regression framework with time fixed-effects. The sum of the coefficients is scaled to 100% for convenience of exposition. For each pair of models, the first set of coefficients is estimated using actual fund flows, and the second and third sets are estimated using flows generated under the null hypothesis that flows respond equally to all cross-sectional differences in fund returns, regardless of their source. Notice that, due to the cross-sectional nature of the test, unadjusted returns and returns in excess of the risk-free rate or in excess of the market return yield the same results. Please refer to Section 4 for additional details on the modelling of the null hypotheses.



adjusted return (MRAR):

$$\text{MRAR}(\gamma) = \left[\frac{1}{T}\sum_{t=1}^{T}(1 + R_t - RF_t)^{-\gamma}\right]^{-\frac{12}{\gamma}} - 1, \tag{15}$$

where $R_t - RF_t$ is the geometric return in excess of the risk-free rate in month t, $\gamma = 2$ is the risk aversion coefficient, and T is the number of past monthly returns used. The formula penalizes funds with higher return volatility. No other adjustment is carried out. For example, exposure to risk factors is not taken into account.

To see how MRAR penalizes for volatility, notice that when γ converges to 0, MRAR(0) is equal to the annualized geometric mean of excess returns.[29] When γ is set to be greater than 0, holding the geometric mean return constant, the formula yields a lower MRAR value for funds whose monthly returns deviate more from their mean. Specifically, the risk adjustment can be expressed as MRAR(0) – MRAR(2).[30]

Depending on the age of the fund, separate MRAR measures are calculated using the past three, five, and ten years of monthly excess returns. Each MRAR measure is further adjusted for sales charges, loads, and redemption fees. Because these costs can vary across different share classes of the same fund, Morningstar ratings are assigned at the share class level rather than at the fund level. We follow BHO in calculating the fund star rating as the total net asset-weighted star rating across all share classes.

Morningstar rates share classes for multiple time horizons—three years, five years, and ten years—when data availability permits. Share classes with a history shorter than three years

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

[30]In general, ranking funds based on their MRAR(2) is similar to ranking them based on their Sharpe ratio calculated over the same period. For instance, Sharpe (1998) reports that in an earlier sample, the correlation between Morningstars risk-adjusted return percentile (within category) and the Sharpe ratio percentile was 0.986. See also `https://web.stanford.edu/~wfsharpe/art/stars/stars8.htm` and Del Guercio and Tkac (2008) for additional evidence that Morningstar's risk adjustment leads to rankings that are highly correlated with Sharpe ratio rankings.

are not rated.[31] These horizon-specific ratings, subject to availability, are then consolidated into an overall rating, which is the most salient and influential one. Specifically, if a fund is less than five years old, its overall rating equals the three-year rating. If a fund is between five and ten years old, the overall rating equals the weighted average of the five-year and the three-year ratings, with weights of 60% and 40%, respectively. If the track record is longer than ten years, the overall rating is a weighted average of the ten-year rating (50% weight), the five-year rating (30% weight), and the three-year rating (20% weight).[32]

[31] See `https://corporate.morningstar.com/US/documents/MethodologyDocuments/FactSheets/MorningstarRatingForFunds_FactSheet.pdf` for details.

[32] The weighted horizon-dependent ratings are then rounded to the nearest integer when producing the overall rating.