What Do Mutual Fund Investors Really Care About?

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

November 18, 2018

Abstract

Recent studies use mutual fund flows to infer which asset pricing model investors use. Among the tested models, the Capital Asset Pricing Model (CAPM) was found to be "closest to the true asset pricing model." We show that, in fact, fund flow data is most consistent with investors relying on fund rankings (Morningstar ratings) and chasing recent returns. We find no evidence that investors account for market beta or exposures to other risk factors when allocating capital among mutual funds. Flows are weaker for high-volatility funds only because Morningstar penalizes funds for high total volatility.

*We thank Brad Barber, Xing Huang, and Terrance Odean for kindly sharing their data. We thank Jonathan Berk, Justin Birru, Scott Cederburg, Sylvester Flood (Morningstar), Francesco Franzoni, John Graham, Cam Harvey, David Hirshleifer, Paul Kaplan (Morningstar), Steve Kaplan, Martin Schmalz, Berk Sensoy, Rick Sias, Nancy Xu, and seminar participants at The Ohio State University, Boston College and the University of Arizona for helpful comments. Ben-David is with The Ohio State University and NBER, Li is with Stanford University, Rossi is with the University of Arizona, and Song is with the University of Washington. Email: ben-david.1@osu.edu, jiacui@stanford.edu, rossi2@email.arizona.edu, and songy18@uw.edu.

1 Introduction

How investors allocate capital across mutual funds has been the focus of academic debate in recent years. Some financial economists argue that investors' mutual fund choices provide a lens to how investors perceive risk in financial markets. Two celebrated studies, Barber, Huang, and Odean (2016) (henceforth BHO) and Berk and van Binsbergen (2016) (henceforth BvB), study mutual fund flows using different empirical techniques.[1] Both reach the same conclusion: among the asset pricing models tested, investors appear to use the Capital Asset Pricing Model (CAPM). BvB conclude that the CAPM is the "closest to the asset pricing model investors are actually using" (p.2). While the idea that investors allocate capital using risk-adjusted returns is appealing, it is potentially at odds with other empirical findings documenting that investors display behaviors that may be considered suboptimal or unsophisticated. For instance, mutual fund investors respond to external rankings (Morningstar: Del Guercio and Tkac (2008), Reuter and Zitzewitz (2015), Wall Street Journal: Kaniel and Parham (2017), sustainability: Hartzmark and Sussman (2018)), chase past returns (Chevalier and Ellison (1997), Choi and Robertson (2018)), and so forth.[2]

In this study, we reconcile the results from the two streams of the literature. Motivated by the fact that households hold the vast majority of mutual fund assets,[3] we test whether simple and readily-available signals explain investors behavior better than asset pricing models. Specifically, we test whether Morningstar's star ratings explain mutual fund flows better than risk-adjusted returns. Morningstar ratings are the ideal candidate for our tests for several reasons. First, Morningstar is the leader of the US fund rating industry and its star ratings are often provided to investors by financial advisors, brokers, defined-contribution retirement plan sponsors, and by fund companies themselves through marketing material. The ratings are also available for free on Morningstar's website. Second, Morningstar ratings do not adjust for fund exposure to any systematic risk factor (see section 2 for an in-depth discussion). Third, these ratings are available for most U.S. equity mutual funds.

Our results show that ratings are the main determinant of capital allocation across mutual funds, followed by past returns. We find no evidence that investors account for mutual fund

[1] Agarwal, Green, and Ren (2018a) and Blocher and Molyboga (2018) applied these empirical methods in the hedge fund space.

[2] Academic studies have found that mutual fund investors prefer funds that report holdings of recent winners and lottery stocks (Solomon, Soltes, and Sosyura (2014), Agarwal, Jiang, and Wen (2018b), Chuprining and Ruf (2018)), react to advertisements and media coverage that do not signal skill (Jain and Wu (2000) and Solomon et al. (2014)), generate 'dumb money' flows (Frazzini and Lamont (2008), Akbas, Armstrong, Sorescu, and Subrahmanyam (2015), Friesen and Nguyen (2018)) and make suboptimal retirement planning choices (Xiao, Zhang, and Kalra (2018)).

[3] According to the 2011 ICI Fact Book, at the end of 2010, 93.7% of long-term mutual fund assets (equity and bond funds) in the US were held by households, consisting of 90.2 million individuals.

exposure to the market or to other risk factors. We also show that fund flows are weaker for high-volatility funds only because Morningstar ratings penalize funds for high volatility.

In the first part of this article, we adopt the diagnostic test proposed by BvB and compare the performance of Morningstar ratings to alphas from asset pricing models in predicting mutual fund flows. BvB's test measures the degree of agreement between the direction of net fund flows (inflows or outflows) and different signals (e.g., the sign of a fund's alpha using different asset pricing models, or Morningstar ratings, in our case). We first replicate BvB's main finding. Consistent with their results, the sign of alphas from common asset pricing models agrees with the sign of fund flows between 57.8% to 59.6% of the time, and the CAPM dominates other models by a small margin (60.4%). Morningstar ratings, in contrast, predict the direction of flows much better (up to 68% of the time).

To further sharpen the BvB test, we also analyze the spread between flows to top and bottom funds ranked according to various asset pricing models or Morningstar ratings. In all tests, ratings decisively outperform all asset pricing models considered. At the aggregate level, in every single year, funds rated highest by Morningstar received more money than the funds ranked highest according to any asset pricing model. Moreover, when using either raw dollar flows or flows as a fraction of total net assets, the CAPM no longer consistently outperforms other models in explaining flows, including raw return (the 'no-model' benchmark).

Next, we look in depth into BHO's methodology and results. BHO decompose fund returns into components associated with a host of commonly-used risk factors and an alpha. They find that while fund flows respond to all return components, flows only respond weakly to returns originating from exposure to the market factor. BHO conclude that investors care about market risk and therefore discount returns that originate from exposure to market risk.

Our analysis indicates that BHO's findings should be interpreted in a different way. Specifically, BHO's conclusion is based on a panel regression with time fixed effects. While this is the standard method in most of the mutual funds literature, in this particular case, it overweight periods with extreme market returns because the dispersion in the independent variable of interest (i.e., the market-related component of fund returns) is higher in those periods. Also, during the same periods, fund flows are significantly less responsive to fund performance, an empirical fact first documented by Franzoni and Schmalz (2017). Put together, a panel regression with time fixed effects would convey the impression that flows respond less to the market-related component of fund returns even if investors do not use the CAPM.

We use a simulation analysis to provide further support that a panel regression might mismeasure investors' response. In the simulation, investors *do not* use the CAPM and chase all components of past returns equally. Consistent with Franzoni and Schmalz (2017), we also assume that investors react less strongly to returns in periods of extreme market returns. We feed the simulated data into a panel regression, similar to the BHO specification. Even through investors in our simulated economy do not care about systematic risk, the panel regression mirrors BHO's result: flows appear to respond more weakly to market-related returns.

To address this apparent puzzle in BHO, we examine the distribution of the coefficients from period-by-period cross-sectional regressions of fund flows on the different components of fund returns. We find that, in fact, there is no evidence that investors discount fund returns related to market risk exposure or to the other risk factors. For example, if we assign equal weights to all time periods using a Fama-MacBeth regression (Fama and MacBeth (1973)), then we find that mutual fund flows respond equally strongly to all components of past returns.

We also propose our own test of whether investors care about the CAPM. If investors discount market-related returns, then, controlling for raw fund returns, high-beta funds should receive lower (higher) flows when the market return has been positive (negative). However, we find no evidence supporting this proposition.

Given that investors rely heavily on Morningstar ratings, one may wonder whether investors indeed *care* about risk, but outsource risk adjustment to Morningstar. For instance, it is possible that investors rely on Morningstar ratings as a simple way to adjust for total fund return volatility, or to adjust for exposure to certain risk factors (e.g., size and value) because rating calculation takes into account Morningstar style benchmarks.

We show that this is unlikely to be the case. First, we consider the fact that flows appear to penalize funds with higher volatility.[4] We show that flows are negatively correlated *only* with the part of variation in volatility that is related to having a different Morningstar rating. This represents only 3% of the dispersion in return volatility; investors do not adjust for the remaining 97%. Therefore, if investors indeed care about volatility, they should realize that relying on Morningstar ratings is a very ineffective way to account for volatility.

Second, we consider whether investors rely on Morningstar as a way to adjust for size and value exposure. We note that Morningstar ratings only started ranking funds within size and value style categories in June 2002, while all US equity funds were ranked together before that date. [5] If investors use Morningstar for style adjustment, then flows should

[4]For example, see Clifford, Fulkerson, Jordan, and Waldman (2013).

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

have responded less to ratings before the implementation, but we find that investors relied heavily on the rating both before and after this implementation. We also carry out an event study around this methodological change. We find that, before June 2002, investors did not account for style, but instead simply followed the ratings that were based on the ranking across all funds.

In summary, we find no evidence that investors use the CAPM, or any other of the commonly-used factor models, to allocate capital to mutual funds. Rather, they naïvely rely on external rankings as a way to chase past winners.

This paper fits into the literature that examines the relationship between investment flows and mutual fund performance. Early work includes Ippolito (1992), Chevalier and Ellison (1997), Sirri and Tufano (1998), Lynch and Musto (2003), Frazzini and Lamont (2008), Pástor and Stambaugh (2012), Pástor, Stambaugh, and Taylor (2015), Franzoni and Schmalz (2017), Del Guercio and Reuter (2014), and Song (2018), among many others.[6] We contribute to this literature by demonstrating that mutual fund investors behave in a less sophisticated way than asset pricing models would predict.

Two other papers put forward explanations for the results of BvB and BHO. Chakraborty, Kumar, Muhlhofer, and Sastry (2018) argue that investors adjust for market returns but not for other factors because market returns are readily available to investors. To support their claim, they show that in the subsample of sector funds, where both market returns and sector-specific historical returns are presented to investors, flows treat sector-specific returns as a source of risk. Jegadeesh and Mangipudi (2017) contest the validity of the tests proposed by BvB. They assert that estimated alphas of simple factor models are less noisy than estimated alphas of complex models, and therefore are more likely to win a horse race test. For the same reason, they argue that the tests by BHO are contaminated by measurement error and therefore are tilted towards favoring a simple asset pricing model such as the CAPM.

The rest of the paper is organized as follows. Section 2 introduces the Morningstar ratings system. Section 3 describes the dataset and the linear factor models used in this paper. Section 4 shows that mutual fund ratings explain fund flows much better than the CAPM and other commonly-used asset pricing models. Section 5 explores the econometric framework of Barber, Huang, and Odean (2016) and finds no evidence that investors discount market-related returns more than other components of fund returns. Section 6 shows that investors discount volatility only through the Morningstar ratings channel. Section 7 provides concluding remarks. Robustness checks are in the appendices.

[6]See Barber, Huang, and Odean (2016) for a more comprehensive review.

2 Overview of Morningstar Ratings

Mutual funds became increasingly popular in 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 US 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.

In the United States, Morningstar is the undisputed leader of this industry (Del Guercio and Tkac (2008)). Its most well-known product, the five-star rating system, was introduced in 1985 and is widely employed by financial professionals and advisors. Ratings are also used by asset management companies in advertising (Blake and Morey (2000), Morey (2003)). Morningstar ratings have been shown to have a strong independent influence on investors flows (Del Guercio and Tkac (2008), Reuter and Zitzewitz (2015)).[7]

Morningstar explains its rating method in a publicly available manual.[8] Ratings are assigned using a relative ranking system and updated every month. Mutual funds are ranked against funds in their peer group using past risk-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 US 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).[9] Ranking within styles was introduced in June 2002, while all US equity fund were ranked against each other (i.e., without regard for their investment style) before then (see Section 6). The top 10% of funds within each style category are assigned five stars. The following 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-Adjusted Return ($MRAR$):

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

where ER_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

[7]According to Morningstar's company statistics dated December 2017 (kindly provided by Morningstar), its ratings are being subscribed by 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.

[8]The Morningstar manual is available at `https://corporate.morningstar.com/US/documents/MethodologyDocuments/FactSheets/MorningstarRatingForFunds_FactSheet.pdf`.

[9]An additional category, called Leveraged Net Long, has been introduced in the US Equities group as of September 30, 2007. We do not include these funds in our sample.

funds with higher return volatility. No other adjustment is carried out, e.g., 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.[10] 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).[11]

Depending on the age of the fund, separate MRAR measures are calculated using the past three, five, and ten years of monthly excess returns, respectively. 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 to calculate 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 history shorter than three years are not rated.[12] These horizon-specific ratings are then, subject to availability, 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).[13]

[10]Morningstar motivates the MRAR formula using expected utility theory. Specifically, consider an investor with 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.

[11]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 Morningstar's 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 additiona evidence that Morningstar's risk-adjustment leads to rankings that are highly correlated with Sharpe ratio rankings.

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

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

3 Data and Methods

In this section, we describe the mutual fund dataset and the linear factor models used in the study, both of which are standard in the academic literature. To make our results more directly comparable to prior literature, we use the same sample of funds in BHO which spans January 1991 to December 2011.[14] To limit discrepancies are driven by variable construction or other methodological choices, we take the fund flow 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.

3.1 Data

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 ETFs. While funds are often marketed to different clients through different share classes, they invest in the same portfolio and the only difference is typically the fee structure. Therefore all share classes are aggregated together at the fund level.

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

$$\mathrm{F}_{p,t} = \frac{\mathrm{TNA}_{p,t}}{\mathrm{TNA}_{p,t-1}} - (1 + R_{p,t}). \tag{2}$$

Here, $\mathrm{TNA}_{p,t}$ is fund p's total net assets at the end of month t, and $R_{p,t}$ is its total 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 using fund CUSIPs to obtain ratings and fund style. The resulting sample includes 3,432 funds in total.

Table 1 provides descriptive statistics for the final sample consisting of over 250,000 fund-month observations. During our sample period, the average fund has a modestly negative monthly flow of -0.53%, manages $\$1,443.50$ million, and has an average age of 16.87 years. Funds with higher Morningstar ratings tend to be larger and receive higher investor

[14]We thank the BHO authors for generously sharing their data. The dataset of BvB ranges from January 1977 to March 2011. We restrict the sample to mutual funds that start on 1991 because the CRSP database contains monthly total net assets beginning in 1991.

flows. Consistent with the algorithm that Morningstar uses to assign ratings (Section 2), 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.

Table 1. Descriptive statistics for the mutual fund sample.

	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	257,053
Fund size ($million)	500.70	751.88	1293.52	2136.05	3460.10	1443.50
Fund age (years)	16.22	16.67	16.95	17.37	16.50	16.87
Fund flow	−1.54%	−1.23%	−0.69%	0.17%	1.14%	−0.53%
Weighted past return	−0.08%	0.18%	0.36%	0.55%	0.78%	0.37%
Return volatility (1 year)	5.51%	5.05%	4.85%	4.81%	4.89%	4.93%
Return volatility (5 years)	6.28%	5.55%	5.22%	4.94%	4.93%	5.27%
Market beta	0.99	0.96	0.94	0.91	0.90	0.93
Size beta	0.19	0.13	0.12	0.12	0.13	0.13
Value beta	−0.031	0.013	0.038	0.063	0.078	0.038
Momentum beta	−0.011	0.011	0.016	0.023	0.043	0.017
Fraction of positive flows	15.9%	19.4%	29.7%	49.3%	67.0%	33.9%

3.2 Estimating alpha of asset pricing models

The tests in our paper, as well as BvB and BHO, are designed to compare the ability of different signals – asset pricing models and Morningstar ratings – to explain fund flows. For each asset pricing model, we follow BHO to estimate its alpha from past fund returns.

As an example, consider the BHO seven-factor model which augments the FFC four factors with the three industry factors of Pástor and Stambaugh (2002). Following BHO, 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 $t-1$:

$$\begin{aligned} R_{p,\tau} - RF_\tau &= a^{7\text{F}}_{p,t} + b_{p,t}(\text{MKT}_\tau - RF_\tau) + s_{p,t}\text{SMB}_\tau + h_{p,t}\text{HML}_\tau \\ &\quad + u_{p,t}\text{UMD}_\tau + \sum_{k=1}^{3} \gamma^k_{p,t}\text{INDk}_\tau + \epsilon_{p,\tau}, \qquad \tau = t-60, \ldots, t-1. \qquad (3) \end{aligned}$$

Here, $R_{p,\tau}$ is the fund return net of fees in month τ and RF_τ is the one-month Treasury bill rate. MKT, SMB, HML, and UMD are the market, size, value, and momentum factors in Carhart (1997).[15] MKT is the return on the value-weighted market portfolio, SMB, HML, and UMD are the returns on the three factor portfolios in Fama and French (1993) and Carhart (1997). IND1, IND2, and IND3 are three industry factors defined in Pástor and Stambaugh (2002).[16] The regression intercept $a_{p,t}^{7\mathrm{F}}$ estimates the seven factor-adjusted average return, while regression coefficients $b_{p,t}, s_{p,t}, h_{p,t}, u_{p,t}$, and $\{\gamma_{p,t}^k\}_{k=1}^3$ capture fund exposures to the seven factors, respectively.

Following BHO, we then estimate the realized return not explained by factor exposures:

$$\begin{aligned} \hat{\alpha}_{p,t}^{7\mathrm{F}} &= R_{p,t} - RF_t - \Big[\hat{b}_{p,t}(\mathrm{MKT}_t - RF_t) + \hat{s}_{p,t}\mathrm{SMB}_t + \hat{h}_{p,t}\mathrm{HML}_t \\ &+ \hat{u}_{p,t}\mathrm{UMD_t} + \sum_{k=1}^{3}\hat{\gamma}_{p,t}^k \mathrm{INDk}_t\Big], \end{aligned} \tag{4}$$

where $\hat{b}_{p,t}, \hat{s}_{p,t}, \hat{h}_{p,t}, \hat{u}_{p,t}$, and $\hat{\gamma}_{p,t}^k$ are the estimated regression coefficients in Equation (3).

Investors often respond to fund performance slowly (Coval and Stafford (2007)). Therefore, we follow BHO to estimate flow responses to exponential-weighted returns in the past 18 months. For instance, we calculate the seven-factor alpha using

$$\mathrm{ALPHA}_{p,t}^{7\mathrm{F}} = \frac{\sum_{s=1}^{18} e^{-\lambda(s-1)}\hat{\alpha}_{p,t-s}^{7\mathrm{F}}}{\sum_{s=1}^{18} e^{-\lambda(s-1)}}, \tag{5}$$

where $\hat{\alpha}_{p,t}^{7\mathrm{F}}$ is from Equation (4) and the decay parameter $\lambda = 0.20551497$ comes from 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

$$\mathrm{ALPHA}_{p,t}^{\mathrm{CAPM}} = \frac{\sum_{s=1}^{18} e^{-\lambda(s-1)}\hat{\alpha}_{p,t-s}^{\mathrm{CAPM}}}{\sum_{s=1}^{18} e^{-\lambda(s-1)}}, \tag{6}$$

$$\text{where } \hat{\alpha}_{p,t}^{\mathrm{CAPM}} = R_{p,t} - RF_t - \hat{\beta}_{p,t}(\mathrm{MKT}_t - RF_t) \tag{7}$$

and $\hat{\beta}_{p,t}$ is estimated using univariate regressions of fund returns on market returns in the 60 months prior to t. Similarly, we also calculate the exponential-weighted alpha of the

[15] 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`).

[16] To construct the industry factors, Fama-French industry returns are first regressed on the FFC factors. Then, the three industry factors are constructed as the first three principal components of the residuals.

Fama-French three-factor model ($\text{ALPHA}_{p,t}^{\text{FF}}$), the FFC four-factor model ($\text{ALPHA}_{p,t}^{\text{FFC}}$), respectively.

4 Morningstar Ratings Trump CAPM

In this section, we show that Morningstar ratings explain mutual fund flows significantly better than the CAPM and other commonly-used asset pricing models. To this end, we first rely on the diagnostic test proposed by BvB. We then perform additional tests with more power and also quantify the economic magnitude of Morningstar outperformance.

4.1 BvB's test

BvB propose that mutual fund flows can be used to infer which asset pricing models investors use. Their core idea is that mutual fund investors compete with each other to allocate capital into positive net present value (NPV) opportunities. When adjusting for risk using the correct asset pricing model, funds with positive alphas are exactly those with positive NPV and thus should receive positive fund flows. Hence, they argue, by investigating how well the signs of alphas match the directions of flows, it is possible to deduce which asset pricing model investors are using. Based on this test, BvB find that CAPM alphas match flows the best and therefore conclude that the CAPM is the closest to the "true" asset pricing model that investors use.

We now explain the BvB methodology. For each fund p in each month t, let $F_{p,t}$ denote the fund flow and let $\text{ALPHA}_{\text{p,t}}^{\mathcal{M}}$ denote the exponential-weighted alpha estimated using the asset pricing model $\mathcal{M}$. Notice that $\text{ALPHA}_{\text{p,t}}^{\mathcal{M}}$ is calculated using historical returns prior to t, as one can see from equations (5) and (6).[17] Following the method of BvB, for each asset pricing model $\mathcal{M}$, we run the following regression:

$$\text{sign}(F_{p,t}) = \beta_0^{\mathcal{M}} + \beta_1^{\mathcal{M}}\text{sign}(\text{ALPHA}_{\text{p,t}}^{\mathcal{M}}) + \epsilon_{\text{p,t}}, \tag{8}$$

where $\text{sign}(F_{p,t})$ and $\text{sign}(\text{ALPHA}_{\text{p,t}}^{\mathcal{M}})$ take on values in $\{-1, 1\}$. Lemma (2) of BvB shows that a linear transformation of the regression slope, intuitively, is directly related to the frequency in which the alpha and flow signs match each other:

$$\frac{\beta_1^{\mathcal{M}} + 1}{2} = \frac{\Pr(\text{sign}(F_{p,t}) = 1|\text{sign}(\text{ALPHA}_{p,t}^{\mathcal{M}}) = 1) + \Pr(\text{sign}(F_{p,t}) = -1|\text{sign}(\text{ALPHA}_{\text{p,t}}^{\mathcal{M}}) = -1)}{2}. \tag{9}$$

[17]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.

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

In our analysis, we find that the simple heuristic of reallocating capital based on Morningstar ratings explain fund flow signs much better than the CAPM. To set the stage, the last row of Table 1 shows that Morningstar ratings have significant explanatory power on flows. Specifically, only 15.9% of funds with a one-star rating have positive flows in the next month. The fraction of funds with positive flows increases monotonically with ratings, reaching 67.0% for the highest rated, five-star funds.

We now consider the following simple heuristic model: investors increase allocation to funds with ratings $\geq i$ and decrease allocation to those with ratings $< i$. We consider three possible thresholds $i = 3, 4$, and 5. Funds with ratings ≥ 3, ≥ 4, and $= 5$ comprise, respectively, 68.9%, 31.8%, and 7.4% of fund-month observations. We estimate regression (8) for the asset pricing models and our rating-based heuristic models. Following BvB, we double cluster standard errors by fund and by time. The results are shown in the first two columns of Table 2.

Consistent with BvB's findings, we also find that the CAPM performs better than the market-adjusted model, the FF three-factor model, and the FFC four-factor model.[18] We also find that the excess return "model" (fund return in excess of the risk-free rate) performs the worst. However, the rating-based heuristics significantly outperform 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 five-star funds, gets the sign of the flows right 67.95% of the time, while the CAPM gets the flow signs right roughly 60.36% of the time. The difference is approximately 7.6%, which, for comparison, is much larger than the 3.46% difference between the CAPM (60.36%) and the worst performing model (excess returns, 56.90%).

Is this outperformance of rating-based heuristics statistically significant? We follow BvB to conduct pairwise model horse races. For any two models $\mathcal{M}1$ and $\mathcal{M}2$, we run regression

$$\text{sign}(F_{p,t}) = \gamma_0 + \gamma_1 \left(\frac{\text{sign}(\text{ALPHA}^{\mathcal{M}1}_{\text{p,t}})}{\widehat{\text{var}}(\text{ALPHA}^{\mathcal{M}1}_{\text{p,t}})} - \frac{\text{sign}(\text{ALPHA}^{\mathcal{M}2}_{\text{p,t}})}{\widehat{\text{var}}(\text{ALPHA}^{\mathcal{M}2}_{\text{p,t}})} \right) + \xi_{p,t} \qquad (10)$$

[18]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 2. Horse race of different models. The first two columns are estimates of Equation (8) for each model considered. For ease of interpretation, the table reports $(\beta_1^{\mathcal{M}} + 1)/2$ in percent, and models are ordered 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 (10). 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 time.

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	67.95	29.48	5.60	9.57	9.93	10.80	12.70	13.13	11.20
Rating ≥ 4	64.41	36.32	-	10.07	8.27	9.21	12.26	12.82	8.53
Rating ≥ 3	61.01	32.15	-	-	1.43	2.81	5.33	6.22	5.15
CAPM	60.36	25.62	-	-	-	5.59	6.53	7.22	4.67
Market-adjusted	59.64	22.64	-	-	-	-	3.07	4.06	3.68
FF 3-factor	58.76	26.44	-	-	-	-	-	3.37	2.61
FFC 4-factor	58.36	25.83	-	-	-	-	-	-	2.08
Excess return	56.90	11.78	-	-	-	-	-	-	-

where $\widehat{\mathrm{var}}(\mathrm{ALPHA}^{\mathcal{M}1}_{\mathrm{p,t}})$ and $\widehat{\mathrm{var}}(\mathrm{ALPHA}^{\mathcal{M}2}_{\mathrm{p,t}}$ are sample variance of 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 by time. The results are reported in the remaining columns in Table 2. The first two rating-based models both outperform the CAPM with strong statistical significance, with t-statistics of 9.93 and 8.27, respectively. In Appendix A, we show that this finding is robust to using different past return windows to calculate return-based model alphas.

Based on BvB's diagnostic, the test results suggest that Morningstar ratings explain investors' capital reallocation better than the CAPM and all other asset pricing models considered.

4.2 Top- and bottom-ranked funds

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

We examine the difference of fund flows between top- and bottom-ranked funds defined using different performance measures. In each month, we sort funds using each performance measure and use the number of 5-star and 1-star funds to classify top- and bottom-ranked funds. This way, the number of funds in each group is the same. For instance, if there are 150 5-star funds in a month, then the 150 funds with the highest CAPM alpha are defined as top-ranked by CAPM. On average, 7.4% and 6.9% of 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 flows as a fraction of TNA, and the average dollar flows. The results are reported in Panel A of Table 3.[19]

When using this more powerful test, the outperformance of Morningstar ratings is more striking. When using Morningstar, 67.0% of top-ranked funds receive positive flows while only 15.9% of bottom-ranked funds receive positive flows, generating a difference of 51.1%. This is significantly higher than all other measures which generate differences in the 15.6% to 22.8% range. Morningstar also outperforms by a sizable margin when using the other two flow measures based on flow magnitudes rather than signs, indicating that the outperformance is economically significant.

To visualize the economic magnitude of outperformance, Figure 1 plots the annual aggre-

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

Table 3. Flows to top-ranked and bottom-ranked funds

Panel A

Top-ranked: five-star funds and the best 7.4% of funds for each model

Bottom-ranked: one-star funds and the worst 6.9% of funds for each model

	Fraction positive flow			Fund flow (%)			Fund flow ($ Mn)		
	Top	**Bottom**	**Diff**	**Top**	**Bottom**	**Diff**	**Top**	**Bottom**	**Diff**
Morningstar	67.0%	15.9%	51.1%	1.15%	−1.53%	2.68%	37.3	−8.0	45.4
Market-adjusted	48.8%	25.9%	22.8%	0.29%	−1.19%	1.48%	11.6	−9.5	21.1
CAPM	43.9%	23.2%	20.6%	0.04%	−1.38%	1.41%	8.2	−10.6	18.7
FF 3-factor	41.0%	23.8%	17.2%	−0.11%	−1.31%	1.20%	5.3	−9.5	14.8
FFC 4-factor	40.3%	24.6%	15.6%	−0.16%	−1.26%	1.11%	4.1	−8.2	12.2

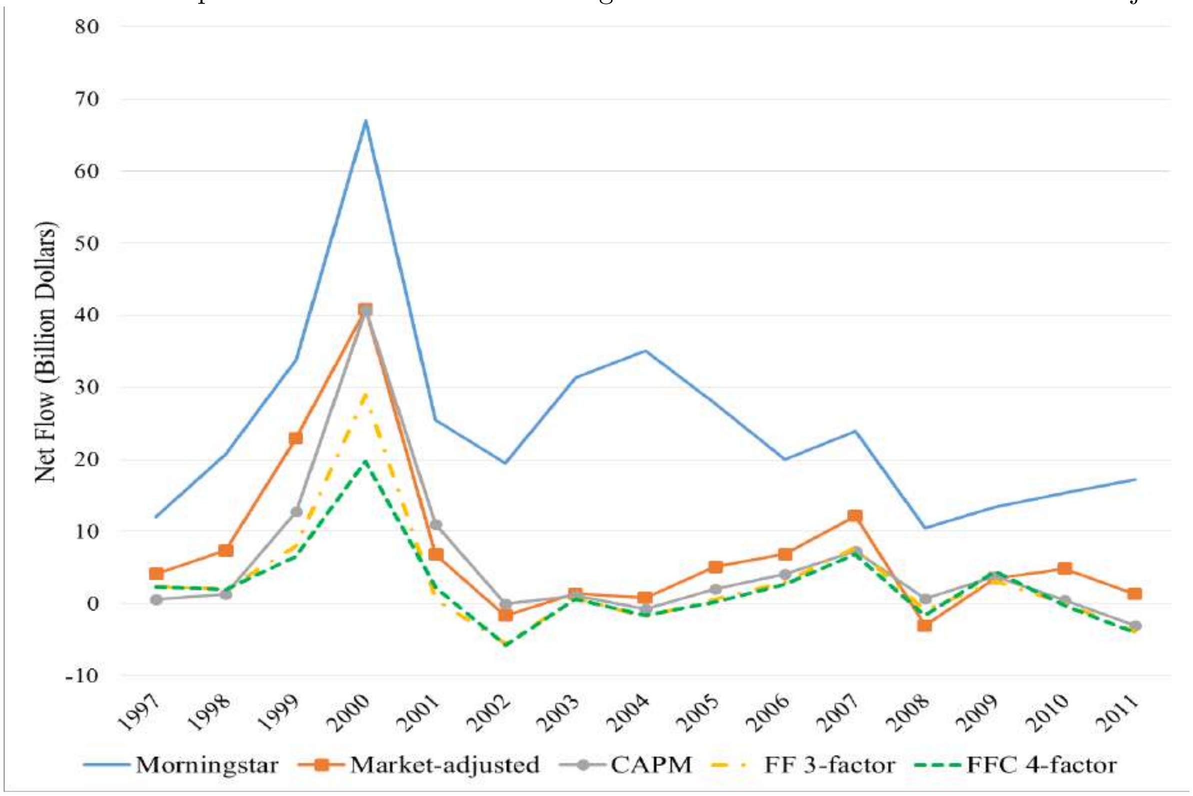
Panel B

Top-ranked: 4- & 5-star funds and the best 31.8% of funds for each model

Bottom-ranked: 1- & 2-star funds and the worst 31.2% of funds for each model

	Fraction positive flow			Fund flow (%)			Fund flow ($ Mn)		
	Top	**bottom**	**Diff**	**Top**	**Bottom**	**Diff**	**Top**	**Bottom**	**Diff**
Morningstar	54.2%	19.3%	34.9%	0.41%	−1.29%	1.71%	12.7	−9.6	22.3
Market-adjusted	46.5%	24.6%	21.8%	0.10%	−1.09%	1.19%	8.5	−11.0	19.6
CAPM	45.9%	23.7%	22.2%	0.07%	−1.14%	1.21%	8.1	−11.7	19.8
FF 3-factor	44.3%	25.2%	19.0%	0.00%	−1.07%	1.08%	6.4	−9.9	16.3
FFC 4-factor	43.9%	25.6%	18.2%	−0.02%	−1.06%	1.04%	5.9	−9.3	15.2

Figure 1. Flows to top-ranked funds. This figure presents annual aggregate new flows to top-ranked funds when ranked according to five different measures of performance and according to the Morningstar rating system. Because funds are ranked within each month, rankings based on raw 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 of 'market-adjusted'.



gate net flows to the top-ranked funds using different performance measures. There are two main takeaways. First, in each year, 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 ($20.3 billion per year on average). Second, the difference between the asset pricing models are much smaller. In particular, the CAPM and market-adjusted return model, which are the two best-performing models in the BvB test (see Table 2), appear to perform similarly. By construction, the only difference between these models is driven by differences in fund market betas, so this result is consistent with investors not adjusting for market beta – the main result in the next section.

In Panel B of Table 3, we also report the results when classifying both 4- and 5-star funds to be top-ranked and 1- and 2-star funds to be bottom-ranked. In this case, 31.8% and 31.2% of observations are top and bottom-ranked, respectively. Morningstar still outperforms, but by a smaller margin, indicating that the outperformance is concentrated in the most extreme rankings.

5 Do Investors Adjust for Market Beta?

Similar to BvB, Barber, Huang, and Odean (2016) (BHO) analyze which asset pricing model appears to better describe how investors allocate money across mutual funds. BHO, however, take a different approach. They decompose fund returns into factor-related returns and an alpha, and estimate how fund flows respond to these different components. Using a pooled regression with time fixed effects (FEs), BHO find that fund flows are much less responsive to a fund's market-related returns than to other components. Since investors appear to discount returns arising from exposure to market risk, BHO conclude that investors presumably use a model akin to the CAPM.

In this section, we suggest a different explanation for BHO's result. The difference in interpretation has to do with the fact that, by construction, panel regressions overweight periods with more dispersion in the independent variable. When estimating flow response to market-related returns, most of the variation in the independent variable is concentrated in periods with extreme market returns when the flow-performance sensitivity is particularly low. Panel regressions thus overweigh periods with low flow-performance sensitivity when estimating response to market-related returns. Once we account for this, there is no clear evidence that investors differentiate market-related returns from returns related to other factors or alphas. In other words, investors do not seem to account for market beta or fund exposures to other factors when allocating capital across mutual funds.

5.1 Replicating BHO's return decomposition results

We briefly explain BHO's methodology for the reader's convenience. For each fund, they use rolling time series regressions to decompose monthly fund excess returns into seven factor-related components (market, size, value, momentum, and the three industry factors of Pástor and Stambaugh (2002)) and a residual, which they refer to as the seven-factor alpha. They account for the slow response of flows to past returns by applying an exponential decay function to each of the return components in the past 18 months. For instance, the relevant market-related return in month t is

$$\mathrm{MKTRET}_{p,t} = \frac{\sum_{s=1}^{18} e^{-\lambda(s-1)} \hat{b}_{p,t-s}(\mathrm{MKT}_{t-s} - RF_{t-s})}{\sum_{s=1}^{18} e^{-\lambda(s-1)}}, \tag{11}$$

where $\hat{b}_{p,t}$ is the fund exposure to the market factor under the seven-factor model in Equation (3), estimated using a time-series regression with the past 60-month returns prior to month t. They also calculate returns related to the fund's size, value, momentum, and three industry tilts, which are labeled SIZRET, VALRET, MOMRET, INDRET1, INDRET2, and

INDRET3, respectively.

To infer investor response to different return components, BHO estimate the following panel regression with time fixed effects:

$$F_{p,t} = b_0 + \mu_t + \gamma X_{p,t} + b_{\text{ALPHA}}\text{ALPHA}^{7\text{F}}_{p,t} + b_{\text{MKTRET}}\text{MKTRET}_{p,t} + b_{\text{SIZRET}}\text{SIZRET}_{p,t}$$
$$+ b_{\text{VALRET}}\text{VALRET}_{p,t} + b_{\text{MOMRET}}\text{MOMRET}_{p,t} + \sum_{k=1}^{3} b_{\text{INDRET}_k}\text{INDRET}k_{p,t} + e_{p,t}, \quad (12)$$

where $F_{p,t}$ is the fractional fund flow in month t, μ_t is the time fixed effects in 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 one year, the log of fund size in month $t-1$, the log of fund age, and lagged fund flows from month $t-19$. The coefficients $b_{\text{ALPHA}}, b_{\text{MKTRET}}, \ldots,$ measure how fund flows respond to different return components. Standard errors are two-way clustered by month and fund.

Using the data provided to us by BHO, we are able to exactly reproduce their key result, which we report in Column (1) of Table 4 (the same result is presented in Table 5 of BHO). In Column (2) we also report the difference between each reported coefficient and the coefficient on the market-related return component. As noted in BHO and reproduced in Column (1) of Table 4, the response coefficient to market-related returns, ($\hat{b}_{\text{MKTRET}} = 0.25$), is significantly lower than the coefficients on all other components of returns. Based on this finding, BHO concluded that investors discount market-related returns more than other components of returns when assessing mutual fund performance, implying that investors appear to be using the CAPM in their capital allocation decisions.

Compared to the methodology of BvB, the econometric specification of BHO has the advantage that it exploits the full variation in fund flows as opposed to simply using signs. However, BHO's test has an important drawback. We argue that their main finding of low response to market-related returns is, at least, partially driven by the time-varying nature of low-performance sensitivity (FPS). In the next section, we show that time-varying FPS causes the estimated average response of fund flows to market-related returns to be downward biased. After adjusting for this effect, we no longer find evidence that investors discount market-related returns more than other return components.

Table 4. Response of fund flows to components of fund returns. This table presents coefficient estimates from panel regressions of percentage fund flow (dependent variable) on the components of a fund's return in Equation (12). The controls include the total expense ratio, a dummy variable for no-load, a fund's return standard deviation over the prior one year, the log of fund size in month $t-1$, the log of fund age, and lagged fund flows from month $t-19$. Columns (1) and (3) are based on pooled regression with time FEs and Fama-MacBeth regression, respectively. Columns (2) and (4) report the difference between the flow-response to MKTRET and the flow-response to other return components. Column (5) shows the change in each of the coefficient estimates by the two different regression methods (Columns (1) and (3)). The t-statistics (double clustered by fund and by month) are in parentheses. *, **, and *** indicate significance at the 10% ,5%, and 1% level, respectively.

	BHO panel regression with time FEs		**Fama-MacBeth regression**		**Change in coefficients**
	(1)	(2)	(3)	(4)	(5)
	Coefficients	**Difference**	**Coefficients**	**Difference**	
ALPHA^{7F}	0.88***	0.63***	1.04***	0.24*	18%
	(32.74)	(10.15)	(39.70)	(1.96)	
MKTRET	0.25***	-	0.80***	-	216%
	(4.52)		(6.65)		
SIZERET	0.76***	0.51***	0.54***	−0.26	−29%
	(14.06)	(6.50)	(3.24)	(−1.27)	
VALRET	0.67***	0.42***	0.93***	0.13	40%
	(10.56)	(4.89)	(5.63)	(0.65)	
MOMRET	1.06***	0.81***	0.65**	−0.15	−38%
	(17.65)	(9.82)	(2.28)	(−0.47)	
INDRET1	0.92***	0.67***	0.76***	−0.04	−17%
	(12.43)	(7.19)	(4.91)	(−0.18)	
INDRET2	0.70***	0.45***	0.98***	0.18	40%
	(7.38)	(4.06)	(3.74)	(0.62)	
INDRET3	0.69***	0.44***	1.14***	0.34	64%
	(7.97)	(4.25)	(3.40)	(0.95)	
Month FE	Yes	-	-	-	-
Controls	Yes	-	Yes	-	-
Observations	257,053	-	257,053	-	-
Adjusted R^2	0.173	-	0.204	-	-

5.2 Assessing the relation between flows and factor-related returns

5.2.1 Investors are less responsive during extreme market movements

To illustrate the relationship between market returns and the sensitivity of fund flows to returns, we reproduce the observation of Franzoni and Schmalz (2017). In particular, we split the entire sample period into ten buckets depending on the past-18-month-weighted excess returns of the market factor. We measure the FPS as the slope from monthly cross-sectional regressions of fund flows on prior 18-month-weighted fund returns, and report the average FPS in each bucket in Figure 2.

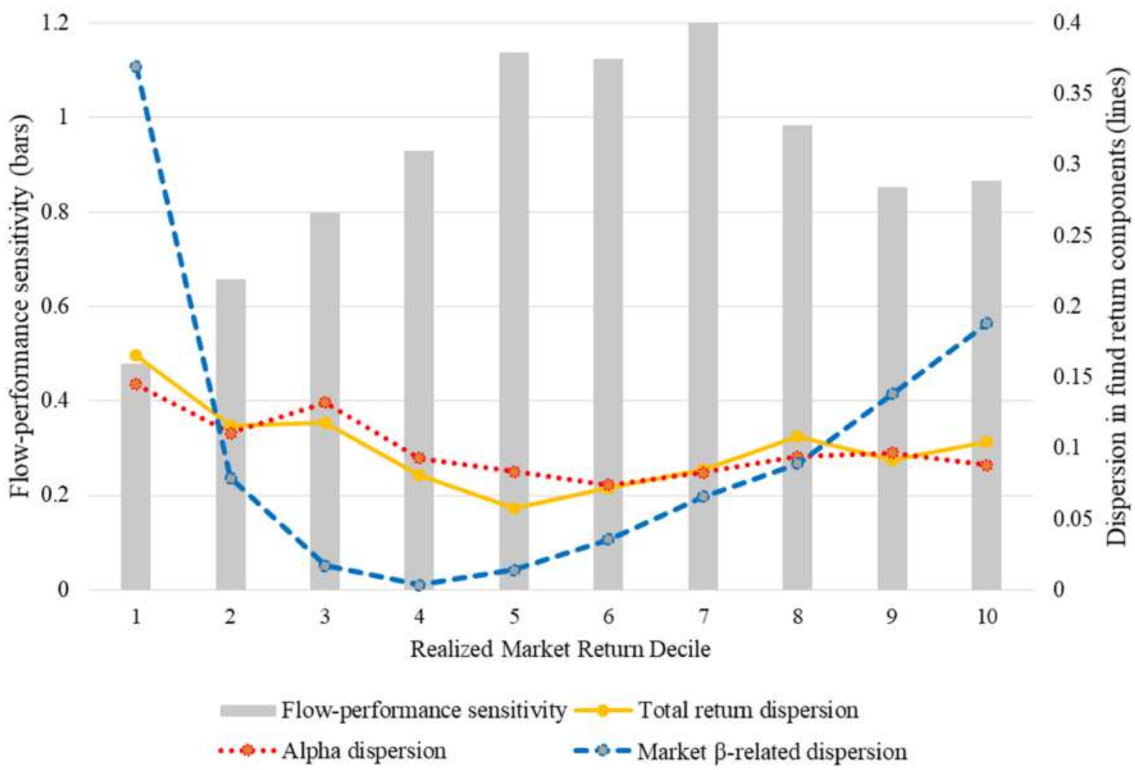
The figure shows that the FPS (left axis) is a hump-shaped function of aggregate market realizations. This is consistent with the finding of Franzoni and Schmalz (2017). The FPS is more than twice as large in moderate states as in the states when the aggregate market has extremely negative returns. While the FPS is a hump-shaped function of past realized market returns, the cross-sectional dispersion in the market-related component of fund returns is an inverse hump-shaped function of it, by construction. In contrast, the cross-sectional dispersion in seven-factor alpha or in other factor-related returns is essentially flat across different market states.[20]

To understand why the empirical patterns depicted in Figure 2 can impact the estimates of the coefficients in the flow-performance relation in Equation 12, we need to consider how cross-sectional regression coefficient estimates are linked to panel regression estimates. Based on the mathematical relationships derived by Pástor, Stambaugh, and Taylor (2017), we can express the coefficient estimate for the response to variable X in a panel regression with time fixed effects as a weighted average of period-by-period cross-sectional regression coefficients, i.e., $\hat{b}_X = \sum_{t=1}^{T} w_t \hat{b}_{X,t}$, where the weight given to each cross-sectional coefficient estimate is directly proportional to the variance of the independent variable X in that cross-section (and to the number of observations in that cross-section).[21] Now, as shown in Figure 2, the FPS is weaker when the dispersion in the market-related return component of fund returns, MKTRET, is higher. Hence, $\hat{b}_{MKTRET,t}$ and $w_{MKTRET,t}$ are negatively correlated, and this causes the panel regression estimate for the response to the market-related return, $\hat{b}_{MKTRET}$, to be smaller than the response coefficients to other fund return components, regardless of whether investor's flows are actually averse to market-related returns. In other words, the

[20]We also find that, after controlling for the market factor, the flow-performance sensitivity does not meaningfully depend on the volatility of other factors. For this reason, the argument made here matters for the response coefficient on MKTRET, but a lot less for the other coefficients.

[21]Specifically, $w_t = (N_t \hat{\sigma}^2_{X_t})/(\sum_{t=1}^{T} N_t \hat{\sigma}^2_{X_t})$, where N_t is the number of observations in period t and $\hat{\sigma}^2_{X_t}$ is the sample variance of the independent variable.

Figure 2. Flow-performance sensitivity in different market states. We split the entire sample period into ten market-state buckets depending on the past-18-month-weighted excess returns of the aggregate market. We then measure the flow-performance sensitivity (FPS) each month as the estimated coefficient from the monthly cross-section regressions of percentage flows on the past-18-month-weighted fund returns. We also calculate the monthly cross-sectional standard deviation of the fund market-related returns, the BHO 7F-alphas, and the total fund returns, respectively. The grey bars (the left axis) present the time-series averages of the FPS for each of the ten market-state buckets. The blue, red, and yellow lines (the right axis) show the time-series averages of the cross-sectional variation in the market-related returns, the BHO 7F-alphas, and the total returns for each market-state buckets, respectively.



pooled regression estimate of b_{MKTRET} is likely downward-biased relative to other coefficient estimates in Equation (12).

5.2.2 Simulation: why the evidence in favor of the CAPM may be spurious

We perform a simulation exercise to verify that the issue discussed above can significantly impact the estimated flow response to market-related returns. We simulate a panel of fund returns based on betas, alphas, and market returns that are drawn from distributions modeled after real data. We assume that investors only respond to the total return of each fund and do not differentiate between different return components. We simulate fund flows to past returns under two scenarios. In the first scenario, the FPS is constant across all periods. In the second scenario, the FPS is time-varying and, as in the data, it is a hump-shaped function of realized market returns. We then run panel regressions of simulated flows on different components of fund returns. Simulation details are explained in Appendix B.

By construction, investors only respond to total fund returns, so we expect to find the same flow response to each return component. This in indeed the case in the first scenario in which the FPS is constant over time. However, in the second scenario in which the FPS is time-varying, the regression coefficient on the market-related return is 58% smaller than the coefficient on the alpha component (Panel A of Table B.I in Appendix B). This confirms that the panel regression coefficient on MKTRET is significantly downward biased relative to the other coefficients.

5.2.3 Evidence in favor of the CAPM fades in Fama-MacBeth procedure

As a remedy to this problem, we propose using the Fama-MacBeth procedure (FMB) which weights the flow-response coefficients across all periods equally (Fama and MacBeth (1973)). In the simulation exercise in Appendix B, we verify that the FMB-estimated flow-response coefficients to all return components are equal, even in the scenario with time-varying FPS.

Therefore, we re-estimate the BHO exercise using FMB. That is, for each month, we run cross-sectional regressions of fund flows on the eight components of fund returns and controls in Equation (12). We then calculate the time-series averages of the estimated cross-sectional coefficients. We report the results in column (3) of Table 4. In column (5), we report the changes of coefficient estimates when switching from panel regression to FMB.

As expected, the most significant change is in the coefficient on market-related returns, becoming about three times as large when using FMB (from 0.25 to 0.80). The changes in the other coefficients are much smaller in magnitude and exhibit no clear pattern, with 3 out of

7 decreasing and the other 4 increasing. While the market-related coefficient is significantly smaller than all other coefficients in the panel regression, in the FMB specification, it is no longer different from the other coefficients at the conventional 5% confidence level (column 4). Moreover, when using FMB, the coefficient on the market-related return has a higher point estimates than the size-, momentum-, and the first industry factor-related returns.

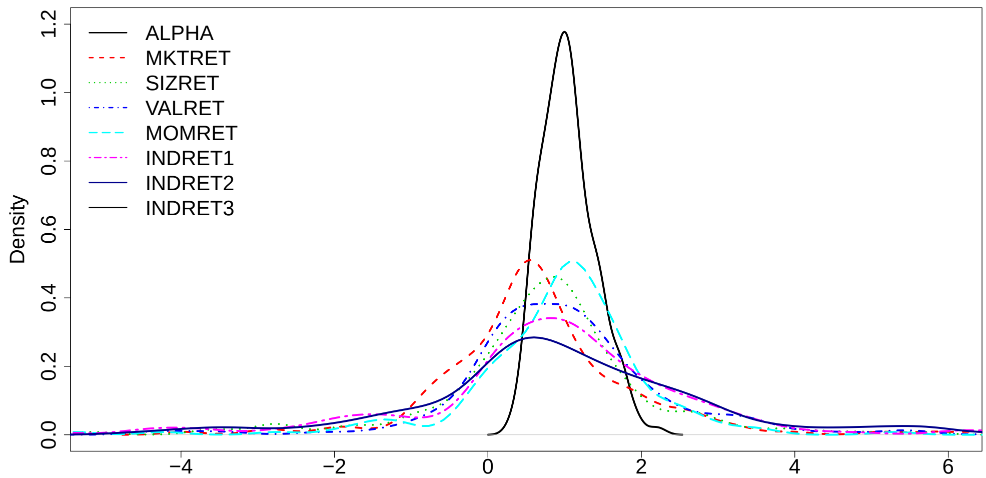
5.2.4 Period-by-period evidence and discussion

Based on the above discussion and on the simulation exercise carried out in the appendix, it is clear that a standard pooled panel regression is not appropriate in this context. As we have seen, the results change substantially when using the FMB procedure. Since FMB regressions and pooled panel regressions are simply two different schemes for weighting period-by-period cross-sectional coefficient estimates (Pástor et al. (2017)), one may wonder whether a different weighting scheme could be used. In practice, because we are trying to estimate the response of fund flows to 8 different components of fund returns, 7 of which have cross-sectional dispersions that vary substantially over time (because they depend on realized factor returns), it does not seem possible to propose a weighting scheme that is simultaneously optimal for all fund return components. Instead, it seems that a simpler and more direct way to gain insight on this issue is to look at the distributions of the coefficient estimates across all cross-sectional regressions.

In Figure 3, we plot the kernel density of period-by-period cross-sectional regression coefficients for different return components. While the distribution of the coefficient on the factor-adjusted return, ALPHA^{7F}, is more concentrated, the coefficients on all factor-related components are all highly dispersed and not clearly different from each other. We have shown in Column (4) of Table 2 that one cannot reject the null that the coefficient on the market-related return is different from the coefficient on the other factor-related returns. The coefficient of market-related returns is only different from that of the alpha measure at the 10% confidence level.

One may wonder why the distribution of the ALPHA^{7F} coefficients from cross-sectional regressions is much more concentrated than that of the coefficients for the other factors. The explanation for why this happens is as follows. In any given month, if we sort funds based on their total past return, funds with high (low) returns always tend to have high (low) alphas, but funds with high (low) returns do not necessarily also have high (low) factor-related returns for all of the 7 factors in all months. Hence, as investors move money from low-return funds into high-return funds, they always happen to be moving money from low-alpha funds into high-alpha funds. The same cannot be said for the other components on fund returns.

Figure 3. Density of the period-by-period cross-sectional coefficient estimates. This figure presents the density of the coefficient estimates from cross-sectional regression of percentage fund flow on the components of a fund's return: a fund's BHO 7F alpha and seven factor-related returns.



In Table A.II of Appendix A, we report results that support the intuition for the explanation given above. The table shows summary statistics for the cross-sectional Spearman's rank correlation between total fund return and its eight components, i.e., ALPHA^{7F}, MKTRET, etc. Only alpha has a positive correlation with the total fund return in all the months in the sample, ranging from a minimum of 0.34 to maximum of 0.94 and averaging 0.71. On the contrary, the factor-related return components are not always highly correlated with the total fund return. The average cross-sectional correlation between the factor-related components and the total fund return is between -0.03 and 0.26 and is negative in at least 10% of the months in the sample for all factor-related components. This happens because the dispersion of each factor-related component of fund returns varies substantially relative to the dispersion of the other 6 factor-related from one period to the next (because these dispersions depend on the magnitude of the associated factor return realizations). Hence, in any given month, some of the 7 factor-related returns are not a significant determinant of the dispersion in total fund returns, leading to a low cross-sectional correlation between those factor-related returns and the total fund returns. In contrast, the cross-sectional dispersion in alphas is always relatively large and stable from one period to the next. This observation also explains why, in Columns (1) and (3) of Table 4, the t-statistic of the ALPHA^{7F} coefficient is significantly larger that of the other coefficients.

BHO also offer several robustness checks with different subsamples.[22] We repeat their exercises but using the FMB regression approach instead. In all of these additional exercises, we again find that one cannot reliably conclude that investors discount market-related returns more than other factor-related returns. The results are presented in Table A.I of Appendix A. Therefore, based on these tests, we argue that there is no clear evidence that investors treat market-related returns differently than they treat other components of fund returns.

In the next section, we propose an additional and more direct test whose results suggest that mutual fund investors do not behave as if they account for market beta when they allocate capital among mutual funds.

5.3 A new test: Do investors respond to market beta?

In the previous section, we showed that the econometric test proposed by BHO might deliver spurious evidence in support of the CAPM because the dispersion in the market-related component of fund returns varies systematically over time with the FPS. Note, however, that market beta itself is less likely to be affected by this problem, because the dispersion in market betas across funds is relatively stable over time.

In this section, we propose a simple test for whether investors use market beta to guide their investments. The logic of the test is if investors would care about market beta, then when the market has positive returns they would discount returns of high beta funds. This relation predicts that at times with positive market returns, the correlation between flows and beta is negative, controlling for observed returns. In contrast, when the market has negative returns investors who care about market beta would understand that funds with high beta have low returns because of their market exposure, and therefore would not penalize them with low flows. In other words, at times of negative market return, the relation between flows and beta should be positive, given the observed returns of funds.

To this end, we first estimate the following regression with time FEs:

$$F_{p,t} = \nu_t + \psi \text{RET}_{p,t} + \phi \hat{\beta}_{p,t} + \xi \text{Rating}_{p,t} + \gamma X_{p,t} + \epsilon_{p,t}, \tag{13}$$

where ν_t is the time fixed effect, $\text{RET}_{p,t}$ is the weighted average of the 18-month returns prior to month t, $\hat{\beta}_{p,t}$ is the estimated market beta in the CAPM from time $t-60$ to $t-1$, $\text{Ratings}_{p,t}$ is the Morningstar rating, and $X_{p,t}$ is a vector of controls as in Equation (12).

[22] BHO also conduct a nonlinear pairwise test of asset pricing models in their Table 4. They find that the CAPM and the market-adjusted model clearly win against all other models with more factors, and that the CAPM slightly beats the market-adjusted model. We redo their exercise using the Fama-MacBeth regression. We find that the CAPM does not outperform the market-adjusted model if one uses this econometric approach. Detailed results for this analysis are presented in Appendix A.

Table 5. Response of fund flows to market beta. This table presents coefficient estimates from panel regressions of percentage fund flow on past returns and market beta in Equation (13). The t-statistics (double clustered by fund and by month) are in parentheses. *, **, and *** indicate significance at the 10% ,5%, and 1% level, respectively.

	Panel regression			**Fama-MacBeth regression**		
	All	**+MKT**	**−MKT**	**All**	**+MKT**	**−MKT**
	(1)	(2)	(3)	(4)	(5)	(6)
Weighted past return	0.68***	0.85***	0.53***	0.81**	0.93***	0.59***
	(23.39)	(24.32)	(16.48)	(31.11)	(29.07)	(20.87)
Market beta	−0.000080	−0.00039	0.00015	0.000085	0	0.00023
	(−0.30)	(−1.27)	(0.46)	(0.71)	(0.060)	(1.37)
Ratings	0.0049***	0.0051***	0.0043***	0.0047***	0.0050***	0.0042***
	(29.03)	(28.00)	(18.67)	(47.22)	(45.48)	(22.70)
Month FE	Yes	Yes	Yes	-	-	-
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs	257,053	167,936	89,117	257,053	167,936	89,117
Adjusted R^2	0.206	0.222	0.190	0.242	0.250	0.220

Inference is unchanged if we control for Morningstar ratings by means of fixed effects instead of a rating variable. Standard errors are double clustered by time and by fund. We also estimate the same equation by means of a Fama-MacBeth regression. Columns (1) and (4) show results for the entire sample whereas Columns (2)-(3) and (5)-(6) present results for subsamples based on the signs of the weighted average of the past 18-month returns of the aggregate market. The results are displayed in Table 5.

The table suggests that, controlling for past fund returns, market beta does not influence fund flows in a significant manner for either the entire sample or any of the subsamples. This result holds with both pooled regressions with time FEs and with the FM procedure. In other words, investors do not seem to adjust for market beta when they allocate flows across mutual funds. In Table A.IV of Appendix A, we show that the results are robust to controlling for month-style or month-style-rating fixed effects. This implies that the CAPM is unlikely to be the model that investors use.

The results presented in Table 5, although based on a simple model specification, provide a meaningful sanity check for the other results we present through the paper.

6 Why Do Investors Rely on Morningstar?

So far, our results show unequivocally that flows follow Morningstar ratings rather than factor-adjusted alphas. This suggests that investors do not care about (or do not understand) risk and just naïvely follow Morningstar. However, it is still possible that investors *do care about* risk, but outsource the adjustment to Morningstar.

We explore this issue in the current section by applying the following logic. First, we show that although Morningstar rating is the main driver of flows, investors also chase recent winners. Second, while investors discriminate among funds with the same rating based on very recent returns, there is little evidence that they make any risk adjustment. The calculation of Morningstar ratings penalizes high volatility funds, and ratings explain 3% of fund volatility. Investors, however, do not make any further adjustment for the remaining 97% of volatility. Third, we present evidence that investors do not rely on Morningstar as an indirect way to adjust for size and value via style benchmarking.

Thus, it appears from the following analysis that mutual fund investors either do not care or do not understand systematic risk and rely on Morningstar ratings simply because they are a simple way to identify and chase past winners.

6.1 Investors also respond to recent fund returns

As a first step in understanding the role of Morningstar in the eyes of investors, we evaluate whether investors rely *solely* on Morningstar. From our earlier analysis, it appears that these ratings are not the sole determinant of fund flows. Past returns have been the most cited and studied determinant of fund flows (Christoffersen, Musto, and Wermers (2014); also see Table 5), and thus may be a good candidate for a signal that investors use beyond Morningstar.

In order to allow for flexible dependence of flows on past returns, we regress fund flows on 50 lags of past monthly returns, while controlling for a vector $X_{p,t}$ of controls as in previous specifications:[23]

$$F_{p,t} = b_0 + b_1 R_{p,t-1} + b_2 R_{p,t-2} + ... + b_{50} R_{p,t-50} + \gamma X_{p,t} + \epsilon_{p,t}. \tag{14}$$

[23] In unreported regressions, we estimate several variations of Equation 14 where we include past returns from windows of different length. If we do not control for star ratings, approximately 50 lags of monthly returns have coefficients that are positive and somewhat statistically significant. However, consistent with BHO's exponential decay specification (see Equation 5), the magnitude (and significance) of the coefficients drops very quickly after the first three lags, and the coefficients past the first 18 lags are about 85% smaller than that on the first month. Once we control for the star ratings, only the first 18 lags are still significant predictors of fund flows. Hence, we choose to include 50 lags in the main specification.

Table 6. Explanatory power of ratings and past fund returns on fund flows. This table reports the adjusted R^2 for several variations of the flow-performance model in Equation 14, and highlights the marginal R^2 of past returns and Morningstar ratings fixed effects. In column (1) the independent variables are the same controls used in the other regression specifications, i.e., a dummy for no-load, the fund's standard deviation over the prior year, the log of fund size in month t-1, the log of fund age, and the lagged fund flow from month t-19. In column (2), we include 50 lagged monthly returns for each fund. Style fixed effects refer to the Morningstar style category classification of each fund.

Independent Variables:	Only Controls	Controls and 50 Past Returns	
	Adjusted R^2	Adjusted R^2	Marginal R^2 of Returns
	(1)	(2)	(3)
No FE	0.033	0.069	0.036
Month FE	0.067	0.107	0.040
Month-Style FE	0.099	0.127	0.028
Month-Rating FE	0.150	0.169	0.019
Month-Style-Rating FE	0.175	0.188	0.013
	Marginal R^2 of Ratings		
	(4)	(5)	
Month-Rating vs Month	0.083	0.062	
Month-Style-Rating vs Month-Style	0.077	0.062	

In Table 6 we report the adjusted R^2 from ten variations of the flow-performance specifications. To set the stage, in Column (1) we control only for the same fund-level characteristics we used as controls in other flow-performance regressions throughout the paper. In Column (2), we include 50 lags of past returns. Outliers may artificially decrease the R^2 of past returns; for this reason, we winsorize the most extreme 1% of fund returns within each month. We then include month fixed effects, style fixed effects, and Morningstar rating fixed effects and their combinations.

Overall, the main message of the table is that while Morningstar ratings are the main explanatory variable for fund flows, recent past returns explain capital allocation beyond Morningstar. This means that investors are not totally submitting their investment judgement to Morningstar, but rather add their own refinement on top of Morningstar ratings.

6.2 Investors *do not* independently penalize funds for high volatility

We wish to understand to what degree Morningstar adjusts for risk and whether investors are content with this adjustment or adjust further. It is a known result in the literature that capital flows to volatile funds are lower (Clifford et al. (2013)). We also know that Morningstar penalizes funds for high volatility (see Section 2). The question is whether these adjustments are independent. If these are independent adjustments, it would mean that investors express their independent preferences to avoid volatile funds. If, however, these are the same adjustment, it would mean that investors are content with the way the Morningstar adjusts for risk. (We know already from the previous analysis that investors make adjustment should they desire.)

We start by estimating the following regression model:

$$F_{p,t} = b_0 + \xi \text{Rating}_{p,t} + \phi \text{Vol}^5_{p,t} + \pi \text{Vol}^1_{p,t} + \gamma Y_{p,t} + \nu_t + \epsilon_{p,t}, \tag{15}$$

where $\text{Ratings}_{p,t}$ is the Morningstar rating, $\text{Vol}^5_{p,t}$ and $\text{Vol}^1_{p,t}$ are the monthly standard deviations of fund returns, estimated over the prior 5 years and prior 1 year, respectively, and $Y_{p,t}$ is a vector of controls that include the total expense ratio, a dummy variable for no-load, the log of fund size, the log of fund age prior to month t, market beta over the prior 5 years, and lagged fund flows from month $t-19$. We also include time fixed effects, and the standard errors are double clustered by time and by fund. The results are reported in Table 7. We also estimate the regression model (15) with the Fama-MacBeth procedure or controlling for time-style fixed effects, and we get similar results in Tables A.V and A.VI of Appendix A.

In the first three specifications, where we do not control for Morningstar ratings, return volatility has a negative and statistically significant coefficient. Interestingly, in Column (3), we find that the five-year volatility is statistically more important than the one-year volatility in predicting negative flows. This result is consistent with what we would expect to observe if investors used Morningstar ratings to direct their flows, because Morningstar uses up to 10 years of past returns to assign the rating.

We conjecture that the negative effect of a fund's volatility on future flows is not due to the fact that investors actually research or calculate fund return volatility and use that information to direct flows, but it is rather due to the fact that Morningstar takes volatility into account when assigning ratings. In Column (4) we confirm that, controlling for a fund's return, volatility is a significant negative predictor of fund ratings. This is, of course, consistent with how the Morningstar Risk-Adjusted Return formula (see Equation (1)), which is used to assign ratings. Based on this evidence, controlling for a fund's rating

Table 7. Response of fund flows to return volatility through Morningstar ratings. This table presents coefficient estimates from panel regression of percentage fund flow on a fund's return volatility over the prior 5 years or 1 year in Equation (14). The t-statistics (double clustered by fund and by month) are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	Flow	Flow	Flow	Ratings	Flow	Flow	Vol^5	Flow
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Vol^5	−0.055***		−0.050***	−7.26***	−0.0080	−0.011		
	(−4.55)		(−3.53)	(−9.65)	(−0.63)	−(0.86)		
$\text{Vol}^5_{\text{predicted}}$								−2.02***
								(−35.47)
$\text{Vol}^5_{\text{residual}}$								0.0037
								(0.31)
Vol^1		−0.049***	−0.012	−5.31***	0.027	0.0166		
		(−2.74)	(−0.54)	(−4.18)	(1.56)	(0.85)		
Ratings					0.0064***		−0.0032***	
					(33.81)		(−11.77)	
Month FE	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Month-rating FE	No	No	No	No	No	Yes	No	No
Controls	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Number of obs	257,053	257,053	257,053	257,053	257,053	257,053	257,053	257,053
Adjusted R^2	0.067	0.066	0.068	0.22	0.14	0.15	0.030	0.14

seems imperative. Once we add the Morningstar ratings in the regression (Column (5)), return volatility loses its ability to predict future flows for both return horizons considered. Moreover, the coefficient estimates are so small that, even if the effects were statistically significant, the economic meaning would be negligible. On the other hand, the coefficient on the ratings is highly significant. A one-notch increase in ratings is on average associated with a 0.64% increase in fund flows in the next month. Alternatively, if we include month-rating fixed effects (Column (6)), the effect of fund return volatility on fund flows becomes again insignificant.

We estimate two additional regressions to help us interpret these results. The fact that, controlling for Morningstar ratings, volatility is no longer a significant predictor of fund flows might stem from two mutually-exclusive reasons. The first is that investors might want to account for fund return volatility when allocating across funds, but delegate the calculation of fund volatility to Morningstar. Alternatively, investors do not actually intend to account for fund return volatility, and the negative correlation between flows and volatility is only due to the fact that Morningstar's formula takes volatility into account.

To evaluate which of the two potential explanations better describes investors' behavior, we decompose return volatility into a component that is correlated with Morningstar ratings, and a component that is orthogonal to the ratings but may reflect investors' preferences beyond the ratings. We focus on five-year volatility because, as seen in Column (3), it is more strongly related to fund flows than the one-year volatility; we find similar results if we use the one-year volatility instead. We execute this test in two stages. In the first stage, we run a regression of volatility on fund ratings and report the results in Column (7). Consistent with Column (5) and with the Morningstar Risk-Adjusted Return formula, the correlation is negative and highly significant. In the second stage, reported in Column (8), we use the predicted value of volatility and the residual to explain fund flows. The residual part, however, does not explain flows.

Hence, it appears that Morningstar is the exclusive channel through which investors regulate flows to volatile funds. Since Morningstar accounts for total volatility, this is an adjustment based on total volatility, not systematic risk necessarily.

An important question is how significant this adjustment is. Table 7 has the answer. Column (5) shows that Morningstar ratings explain only 3% of fund return volatility (see R^2). Since investors do not make any further adjustments beyond Morningstar ratings, then 97% of fund return volatility is not penalized for by either Morningstar or investors. It appears that even though investors can make refinements to Morningstar's ratings at will, they do not do so for volatility or risk. This result support the proposition that mutual fund investors do not care about risk or do not understand risk.

In Table A.VII we split our sample into five subgroups based on Morningstar rating assignments. We estimate Equation (15) for each of the five rating groups. Consistent with the results presented in 7, within each rating group, return volatility is no longer negatively correlated with future flows.

6.3 Investors chase the 'stars', regardless of style benchmarking

One may suspect that investors rely so much on Morningstar ratings, not only because of rating, but also because that ratings are performed within style category (size-value), which may fit investors' risk preferences. Adjusting for investment style might be a simple way for investors to distinguish between lucky and skilled managers and possibly to account for a desired exposure to certain risk factor. Our tests suggest that this is not the case.

Besides ranking funds based on historical performance, Morningstar also classifies diversified US equity mutual funds into style categories based on the so-called Morningstar style box. Each fund is assigned to one of 9 style categories based on its size tilt (Small, Mid-Cap, or Large) and value tilt (Value, Blend, or Growth).[24]

A potentially important fact to consider when evaluating the evidence presented in this study is that a fund's star rating is assigned based on the performance of that fund ranked against the performance of other funds with the same style assignment (see Section 2 for more details). Based on this fact, one may hypothesize that the reason why investors seem to rely so strongly on Morningstar ratings is that the rating system helps them to somehow adjust for risk. There are two main non-mutually exclusive hypotheses to consider.

First, it is possible that investors intend to adjust for factor exposures, e.g., size, value, and momentum. However, it might be difficult and costly for investors to carry out these adjustments on their own and therefore they might want to 'outsource' this task to an external specialized entity, specifically, Morningstar. One argument against this hypothesis is that only a small amount of the dispersion in risk factor loadings across funds is captured by differences in loadings across the style categories.[25] More importantly, the BvB test, which we replicated and extended in our analysis, show that the market-adjusted model and the CAPM do a better job at explaining fund flows than the Fama-French/Carhart models.

[24] This information is usually presented together with a fund's star rating in fund summary and marketing material. In order to assign 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)).

[25] An counter-argument against this line of reasoning would be that some of the variation in loadings that is not captured by differences in loadings across style boxes is likely to be noise due to estimation error. In any case, relying on Morningstar style assignments as a way to adjust for risk exposure would be at best a partial and noisy solution.

In order words, investors' flows do not appear to penalize funds for exposure to size, value, and momentum.

The second hypothesis is related to the idea that some investors consider concepts such as small and large, or value and growth, as sub-asset classes to which they might want to gain exposure. In this context, a fund's style tilt might be considered a benchmark against which to evaluate the manager's skill. If so, it is possible that investors' flows are highly correlated with Morningstar ratings because investors want to invest with the best-performing fund managers within each style.

In order to gauge the merits of the hypotheses above, we exploit a change in the way Morningstar assigns ratings to funds. As mentioned, star ratings are assigned to funds based on their performance relative to other funds in their assigned Morningstar style category. However, this form of style benchmarking has been implemented only since June 2002.[26] Before that date, US equity funds were ranked against all other funds (despite the fact that Morningstar style categories existed before that date). Therefore, before June 2002, mediocre fund managers whose benchmark's style happened to outperform other styles were mechanically assigned high star ratings.

In Panel A of Figure 4, for each month in the sample, we plot the fraction of funds whose star rating changes. The ratings are updated every month. The figure shows that about 11% of funds experience a change in rating every month.[27] In June 2002, the date on which the methodological change was implemented, 54% of funds experienced a change in rating. Adjusting for the average fraction of monthly rating changes (11%), this means that in June 2002 about 43% of ratings changed because of the change in methodology. Hence, in May 2002, about 43% of funds had ratings that were higher or lower just because their style benchmark happened to have had relatively high or low past returns, respectively.

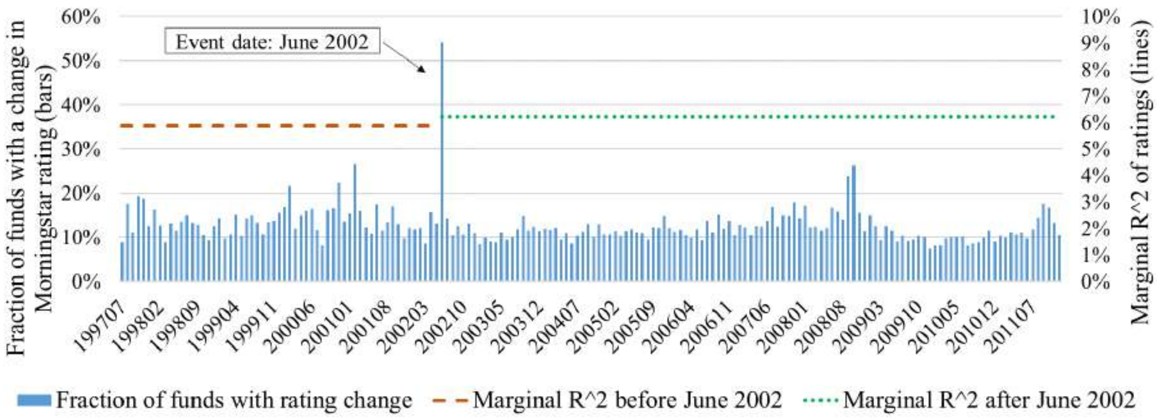
We design two tests based on the change in rating assignment methodology. First, we estimate the tests described in Section 6.1 in two sub-periods, i.e., the periods before and after the methodological change. In particular, we want to compare the marginal R^2 of the

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

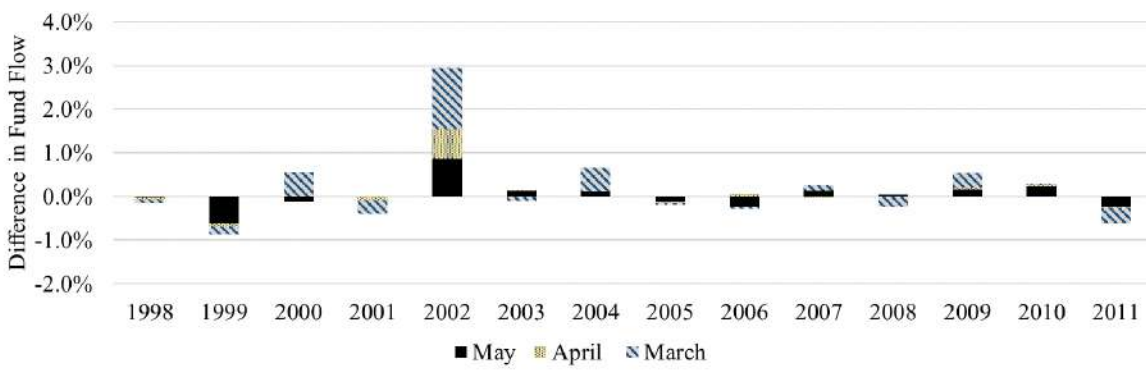
[27] This is not surprising because, given that funds are assigned to 5 groups and therefore there are 4 thresholds, at any given time a large number of funds is just above or just below one of the thresholds, and therefore a small difference in performance in the last month can make the rating change at the beginning of the current month. Moreover, there are a number of other methodological details that 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 of them to cross one of the thresholds even if their relative performance did not change. In unreported analysis, we find that about 40% of the changes in rating are reversed within two months (except for the June 2002 changes, of which only 15% reverted within 2 months).

Figure 4. Event study: change in Morningstar methodology. In June 2002, Morningstar started to rank funds within style categories as opposed to across all US equity fund. In Panel A, we show the fraction of funds with a change in rating, as well as the marginal R^2 of ratings in flow-performance regressions before and after the event. In Panel B and C we present the results of an event study based on the change in the methodology. Years other than 2002 serve as placebo tests. Please refer to the text for datails.

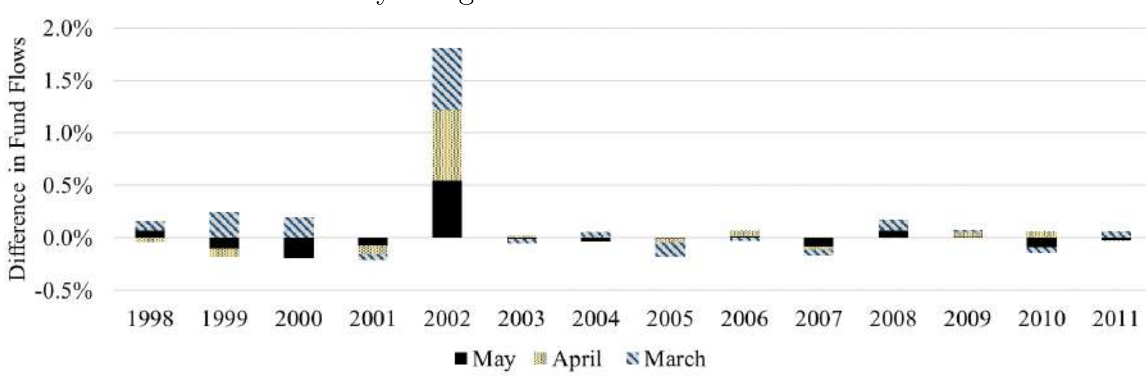
Panel A: Rating changes by month



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



ratings in the two periods. In the full sample, this figure is 6.2% (Column (5) of Table 6). The marginal R^2 for the two subperiods is plotted in Panel A of Figure 4 (right axis). We find that the figure is almost identical before and after June 2002, namely, it is 5.9% and 6.2%, respectively. This indicates that, before June 2002, ratings explained flows virtually as much as after June 2002, despite the fact that pre-June 2002 ratings did not adjust for style.

The second test we conduct is an event study around the methodological change. Before June 2002, several funds had ratings that were high or low because their style benchmark happened to have had high or low returns, respectively. By contrast, starting in June 2002, ratings reflected a fund's performance relative to its benchmark. Based on these facts, in this test, we focus on the three months immediately before the methodological change (i.e., May, April, and March), and use the June 2002 fund ratings as a quasi-counterfactual rating assignment that accounts for style returns. In other words, for each funds in these 3 months, we have 2 sets of star ratings: the actual rating, which does not adjust for style return, and the June rating, which does.

We start by calculating the average percentage flow in each month for funds sorted into 5 groups based on the actual current ratings and into 5 other groups based on the counterfactual June rating. Then, we calculate the following measure:

$$(\overline{\text{flow}}_{\text{current high}} - \overline{\text{flow}}_{\text{current low}}) - (\overline{\text{flow}}_{\text{June high}} - \overline{\text{flow}}_{\text{June low}}), \tag{16}$$

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 in flows between high-rated and low-rated funds based on the June rating. We calculate this measure for flows to funds in May, April, and March of 2002. We also compute the same measure in all other years in the sample as placebo tests.[28] If the measure is negative, it means that investors are adjusting a fund's past performance for its style return. If the measure is positive, it means that funds with higher returns are being rewarded with higher flows despite the fact that part of those higher returns are attributable to high style benchmark returns. We present the results in the last two panels of Figure 4. 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

[28]One might be concerned that this test may be subject to look-ahead bias. In May, April, and March, we sort funds based on their (future) June rating, which partially depends on future information about fund performance. If funds flows predict the cross-section of future fund performance (at 1, 2 and 3 month horizon), then the measure we compute using equation 16 might be biased. 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 the figure 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.

and 1- and 2-star funds are considered low-rated.

The results of the event study show very clearly that 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 suggests that investors followed the actual ratings, and failed to distinguish between ratings that were high because of relative outperformance and ratings that were high because of high average style returns. This evidence suggests that, on average, investors understand the rating system in a naive way. Based on this interpretation of the results, Morningstar's choice to change its methodology to account for style tilts appears justified.

7 Conclusion

The key to understanding investor behavior and market prices is to understand how investors perceive risk. Two recent studies, Berk and van Binsbergen (2016) and Barber et al. (2016), took on this task by studying the drivers of mutual fund flows. The idea is that by allocating funds across active mutual funds, investors reveal their preferences and dislikes. Both studies find that investors appear to behave as if they use the CAPM.

In this paper, we contrast the results of these studies with another line of research from the mutual fund literature that finds that mutual fund flows respond strongly to external rankings (e.g., Del Guercio and Tkac (2008), Kaniel and Parham (2017)). Our results show that mutual fund investors primarily follow external (Morningstar) ratings and then recent past returns. We find no clear evidence they pay attention to whether past returns were generated by the systematic component of any of the commonly-used asset pricing models. We use the test proposed by Berk and van Binsbergen (2016) to show that Morningstar ratings dominate alphas from any other commonly-used asset pricing models. We also show that the tests run by Barber et al. (2016) are not robust to specification and thus are not conclusive. Finally, it is not plausible that Morningstar ratings serve as a proxy for alpha (of the CAPM or of another asset pricing model) since these ratings do not account for systematic exposure to any risk factor.

Where do our results leave the study of investor behavior and asset pricing? It is clear that mutual fund investors do not use any of the commonly-used asset pricing models for their investment decisions. Mutual fund flows indicate that investors pursue easy-to-follow signals (Morningstar ratings and recent returns), which are ultimately not informative about systematic risk. Using the same logic that guided Berk and van Binsbergen (2016), we can conclude that mutual fund investors do not use any of the asset pricing models that are commonly used in academia.

References

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

Agarwal, Vikas, Lei Jiang, and Quan Wen, 2018b, Why do mutual funds hold lottery stocks?, Working paper. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3164692`.

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.

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.

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, 2018, The revealed preference of sophisticated investors, *European Financial Management* 23, 839–872.

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. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2869426`.

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 EK, David K Musto, and Russ Wermers, 2014, Investor flows to asset managers: Causes and consequences, *Annu. Rev. Financ. Econ.* 6, 289–310.

Chuprining, Oleg, and Thomas Ruf, 2018, Rent seeking and mutual fund inceptions, Working paper. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3189986.

Clifford, Christopher P., Jon A. Fulkerson, Bradford D. Jordan, and Steve Waldman, 2013, Risk and fund flows, Working paper. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1752362.

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.

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

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

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

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. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3160271.

Hartzmark, Samuel M., and Abigail Sussman, 2018, Do investors value sustainability? A natural experiment examining ranking and fund flows, Working paper, University of Chicago.

Ippolito, Richard A., 1992, Consumer reaction to measures of poor quality: Evidence from mutual fund industry, *Journal of Law and Economics* 35, 45–70.

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. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3090541`.

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.

Lynch, Anthony W., and David K. Musto, 2003, How investors interpret past fund returns, *Journal of Finance* 58, 2033–2058.

Morey, Mattew R., 2003, The kiss of death: A 5-Star morningstar mutual fund rating?, Working paper. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=455240`.

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

Pástor, Ľuboš, and Robert F. Stambaugh, 2012, On the size of the active management industry, *Journal of Political Economy* 120, 740–781.

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

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

Reuter, Jonathan, and Eric Zitzewitz, 2015, How much does size erode mutual fund performance? a regression discontinuity approach, Working paper. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1661447`.

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.

Song, Yang, 2018, The mismatch between mutual fund scale and skill, Working paper, University of Washington, Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3065700.

Xiao, Liu, Wei Zhang, and Ajay Kalra, 2018, The costly zero bias in target retirement fund choice, Working paper. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3214811.

Appendix A Additional Results

A.1 Robustness of Table 2 to different past return windows

When conducting the BvB model horse races, we follow BHO to use the exponential weighted average of past 18 month alphas for the return-based models. Here, we redo the exercise using equal weighted-alphas from past 1, 6, 12, 24, 36, 60, or 120 months. Our main finding is robust to using different past windows: Morningstar ratings explain fund flows much better than all other models.

Using past 1 month									
Model	Estimate of $(\beta_1^{\mathcal{M}})/2$	Univariate t-stat	Rating ≥ 4	Rating ≥ 3	Excess return	Rating CAPM	Market-adjusted	FFC 4-factor	FF 3-factor
Rating ≥ 5	67.92	29.15	5.44	9.44	18.53	21.11	22.04	22.70	22.65
Rating ≥ 4	64.44	36.18	-	10.11	18.19	23.40	24.69	25.89	25.80
Rating ≥ 3	61.02	32.13	-	-	13.91	18.08	19.15	20.54	20.49
Ex return	53.42	9.23	-	-	-	0.02	1.13	1.64	1.66
CAPM	53.41	11.01	-	-	-	-	3.03	2.74	3.28
Market-adj	52.93	9.51	-	-	-	-	-	0.74	0.85
FFC 4-factor	52.75	10.87	-	-	-	-	-	-	0.07
FF 3-factor	52.74	10.55	-	-	-	-	-	-	-

Using past 6 month									
Model	Estimate of $(\beta_1^{\mathcal{M}})/2$	Univariate t-stat	Rating ≥ 4	Rating ≥ 3	CAPM	Market-adjusted	FF 3-factor	FFC 4-factor	Excess return
Rating ≥ 5	67.92	29.15	5.44	9.44	15.30	16.82	17.57	17.71	15.39
Rating ≥ 4	64.44	36.18	-	10.11	15.82	17.54	18.76	18.81	13.54
Rating ≥ 3	61.02	32.13	-	-	9.52	11.27	12.46	12.83	9.62
CAPM	57.01	20.90	-	-	-	4.64	5.17	5.57	4.05
Market-adj	56.19	18.60	-	-	-	-	1.28	2.11	2.19
FF 3-factor	55.90	20.54	-	-	-	-	-	2.21	1.63
FFC 4-factor	55.69	20.18	-	-	-	-	-	-	1.23
Ex return	55.11	12.04	-	-	-	-	-	-	-

Using past 12 month

Model	Estimate of $(\beta_1^{\mathcal{M}})/2$	Univariate t-stat	Rating ≥ 4	Rating ≥ 3	CAPM	Market-adjusted	FF 3-factor	FFC 4-factor	Excess return
Rating ≥ 5	67.92	29.15	5.44	9.44	13.38	14.77	15.48	15.67	15.52
Rating ≥ 4	64.44	36.18	-	10.11	13.44	15.02	16.19	16.19	13.27
Rating ≥ 3	61.02	32.13	-	-	6.48	8.46	9.14	9.49	9.79
CAPM	58.25	22.91	-	-	-	4.93	4.70	5.02	6.58
Market-adj	57.28	19.60	-	-	-	-	0.40	1.07	4.55
FF 3-factor	57.17	21.89	-	-	-	-	-	1.50	4.58
FFC 4-factor	56.99	21.68	-	-	-	-	-	-	4.48
Ex return	54.39	8.32	-	-	-	-	-	-	-

Using past 24 month

Model	Estimate of $(\beta_1^{\mathcal{M}})/2$	Univariate t-stat	Rating ≥ 4	Rating ≥ 3	CAPM	Market-adjusted	FF 3-factor	FFC 4-factor	Excess return
Rating ≥ 5	67.92	29.15	5.44	9.44	13.94	14.30	15.14	15.31	16.91
Rating ≥ 4	64.44	36.18	-	10.11	14.84	15.18	16.40	16.79	14.82
Rating ≥ 3	61.02	32.13	-	-	7.17	7.86	8.55	8.84	10.60
CAPM	58.17	22.38	-	-	-	1.91	2.65	2.73	6.85
Market-adj	57.71	20.21	-	-	-	-	0.70	0.75	5.68
FF 3-factor	57.50	20.46	-	-	-	-	-	0.06	5.46
FFC 4-factor	57.49	21.55	-	-	-	-	-	-	5.62
Ex return	53.85	7.35	-	-	-	-	-	-	-

Using past 36 month

Model	Estimate of $(\beta_1^{\mathcal{M}})/2$	Univariate t-stat	Rating ≥ 4	Rating ≥ 3	Market-adjusted	CAPM	FFC 4-factor	FF 3-factor	Excess return
Rating ≥ 5	67.92	29.15	5.44	9.44	13.31	13.93	14.55	14.43	16.93
Rating ≥ 4	64.44	36.18	-	10.11	14.25	15.01	16.06	15.82	15.64
Rating ≥ 3	61.02	32.13	-	-	6.94	7.65	8.18	8.15	10.74
Market-adj	58.05	19.53	-	-	-	0.85	1.27	1.39	6.10
CAPM	57.86	20.21	-	-	-	-	0.76	0.97	6.22
FFC 4-factor	57.65	20.05	-	-	-	-	-	0.46	5.73
FF 3-factor	57.59	19.63	-	-	-	-	-	-	5.39
Ex return	54.11	8.06	-	-	-	-	-	-	-

Using past 60 month									
Model	Estimate of $(\beta_1^{\mathcal{M}})/2$	Univariate t-stat	Rating ≥ 4	Rating ≥ 3	Market-adjusted	CAPM	FFC 4-factor	FF 3-factor	Excess return
Rating ≥ 5	67.92	29.15	5.44	9.44	11.74	12.58	14.16	14.39	12.67
Rating ≥ 4	64.44	36.18	-	10.11	13.51	13.85	16.99	16.79	12.10
Rating ≥ 3	61.02	32.13	-	-	6.57	7.30	9.19	9.43	7.38
Market-adj	58.19	19.28	-	-	-	2.31	3.67	4.10	3.94
CAPM	57.68	17.67	-	-	-	-	1.83	2.69	3.14
FFC 4-factor	57.16	17.74	-	-	-	-	-	1.66	1.88
FF 3-factor	56.83	16.14	-	-	-	-	-	-	1.23
Ex return	56.11	12.12	-	-	-	-	-	-	-

Using past 120 month									
Model	Estimate of $(\beta_1^{\mathcal{M}})/2$	Univariate t-stat	Rating ≥ 4	Rating ≥ 3	CAPM	Market-adjusted	Excess return	FFC 4-factor	FF 3-factor
Rating ≥ 5	67.92	29.15	5.44	9.44	9.64	11.08	8.04	12.38	12.83
Rating ≥ 4	64.44	36.18	-	10.11	12.21	14.89	6.87	16.67	17.06
Rating ≥ 3	61.02	32.13	-	-	6.34	8.51	4.34	9.21	10.04
CAPM	55.63	9.09	-	-	-	3.70	1.11	2.47	3.32
Market-adj	54.62	7.60	-	-	-	-	0.04	0.58	1.80
Ex return	54.58	7.14	-	-	-	-	-	0.29	1.14
FFC 4-factor	54.33	7.32	-	-	-	-	-	-	2.09
FF 3-factor	53.62	5.88	-	-	-	-	-	-	-

A.2 Robustness checks of Section 5

We verify that our findings in Section 5 are robust to using different fixed effects specifications and sub-samples. In Table A.I, we first reproduce Columns (2) to (9) in BHO's Table 5 using panel regressions in Panel A, and then run the same regressions using the Fama-MacBeth procedure in Panel B. The change of coefficients are then reported in Panel C. Specifications in Columns (1) and (2) use all funds but include different fixed effects from the main specification in Table 4. Columns (3) and (4) split the sample by median fund sizes; Columns (5) and (6) split the sample by median fund age, and Columns (7) and (8) split the sample by median fund return.

These tests confirm our main finding that the sensitivity of flows to the market-related return component is much higher when using the Fama-MacBeth procedure. The coefficients on the market-related component increase by over 200% for all specifications except for the small fund sample (Column (3)). The coefficient changes for other return components are smaller and similar in magnitude to our findings in Table 4.

Table A.I. Robustness checks of the findings in Table 4. Panel A reproduces the BHO panel regressions in Columns (2) to (9) in BHO's Table 5; Panel B estimates the same regressions using Fama-MacBeth procedure, and Panel C presents the change in coefficients.

Panel A: BHO Panel regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All funds	All funds	Small funds	Big funds	Young funds	Old funds	Below-median ret	Above-median ret
$ALPHA^{7F}$	0.79***	0.74***	0.84***	0.90***	0.92***	0.87***	0.70***	0.89***
	(26.30)	(27.33)	(25.55)	(29.97)	(25.50)	(31.00)	(19.47)	(24.75)
MKTRET	0.21***	0.19***	0.24***	0.26***	0.25***	0.25***	0.16***	0.24***
	(4.14)	(4.73)	(4.21)	(4.64)	(4.42)	(4.60)	(2.81)	(4.36)
SIZERET	0.69***	0.64***	0.53***	0.89***	0.73***	0.78***	0.71***	0.59***
	(12.92)	(13.31)	(8.67)	(14.27)	(11.69)	(12.50)	(11.33)	(7.88)
VALRET	0.59***	0.57***	0.70***	0.65***	0.70***	0.65***	0.52***	0.68***
	(10.35)	(10.72)	(10.58)	(9.80)	(9.96)	(10.20)	(7.16)	(9.63)
MOMRET	0.94***	0.85***	0.93***	1.11***	1.20***	1.00***	0.91***	1.00***
	(15.16)	(17.02)	(13.93)	(15.58)	(15.03)	(16.60)	(11.92)	(13.32)
INDRET1	0.82***	0.84***	0.91***	0.92***	0.94***	0.92***	0.71***	0.95***
	(11.23)	(11.17)	(11.01)	(10.89)	(9.28)	(11.92)	(7.58)	(9.64)
INDRET2	0.59***	0.63***	0.68***	0.71***	0.69***	0.70***	0.54***	0.74***
	(7.06)	(7.08)	(5.92)	(6.61)	(5.57)	(6.81)	(5.23)	(5.74)
INDRET3	0.64***	0.48***	0.74***	0.66***	0.73***	0.68***	0.43***	0.78***
	(7.83)	(5.91)	(6.60)	(7.30)	(6.60)	(7.13)	(4.14)	(7.61)
Month FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Month-style FE	Yes	No	No	No	No	No	No	No
Month-style-rat FEs	No	Yes	No	No	No	No	No	No
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	257,053	248,463	257,053		257,053		257,053	
Adj. R-squared	0.190	0.216	0.175		0.173		0.175	

Panel B: Fama-MacBeth regressions								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All funds	All funds	Small funds	Big funds	Young funds	Old funds	Below-median ret	Above-median ret
ALPHA^{7F}	1.07***	0.89***	0.92***	1.10***	1.14***	1.02***	0.96***	1.19***
	(38.58)	(33.13)	(31.74)	(39.31)	(36.02)	(37.40)	(27.72)	(32.64)
MKTRET	0.79***	0.75***	0.41***	1.00***	0.92***	0.81***	0.59***	1.08***
	(6.57)	(6.82)	(2.76)	(7.67)	(5.12)	(6.82)	(3.87)	(6.41)
SIZERET	0.70***	0.56***	0.18	0.71***	−0.09	0.79***	0.41***	0.43**
	(3.98)	(3.80)	(1.04)	(4.00)	-(0.34)	(4.19)	(2.74)	(1.99)
VALRET	1.01***	0.68***	0.76***	1.04***	0.63**	1.00***	0.60***	1.15***
	(5.43)	(4.94)	(3.68)	(6.38)	(2.45)	(5.92)	(4.36)	(4.82)
MOMRET	0.75***	0.66**	0.19	0.88***	0.46	0.77***	0.83**	0.51
	(2.78)	(2.52)	(0.58)	(3.13)	(1.03)	(3.26)	(2.48)	(1.54)
INDRET1	0.75***	0.64***	0.79***	0.76***	0.88***	0.91***	0.66***	1.12***
	(4.57)	(4.09)	(5.20)	(3.79)	(4.16)	(4.84)	(3.89)	(5.16)
INDRET2	1.11***	0.90***	0.69**	1.20***	0.57*	1.17***	0.67*	1.42***
	(4.55)	(4.35)	(2.52)	(3.80)	(1.95)	(4.20)	(1.85)	(5.11)
INDRET3	1.15***	0.83***	0.87***	1.22***	0.99***	1.11***	0.81***	1.56***
	(3.66)	(2.67)	(2.63)	(3.23)	(2.69)	(3.69)	(2.98)	(4.00)
Month FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Month-style FE	Yes	No	No	No	No	No	No	No
Month-style-rat FE	No	Yes	No	No	No	No	No	No
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	257,053	248,463	257,053		257,053		257,053	
Adj. R-squared	0.179	0.125	0.196	0.234	0.231	0.217	0.152	0.166

Panel C: Change in Coefficients								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All funds	All funds	Small funds	Big funds	Young funds	Old funds	Below-median ret	Above-median ret
ALPHA^{7F}	+35%	+21%	+9%	+22%	+24%	+18%	+37%	+33%
MKTRET	+283%	+286%	+75%	+291%	+266%	+219%	+263%	+344%
SIZERET	+2%	-13%	-65%	-20%	-112%	+1%	-43%	-28%
VALRET	+71%	+20%	+9%	+60%	-9%	+53%	+14%	+69%
MOMRET	-20%	-23%	-79%	-20%	-62%	-23%	-8%	-49%
INDRET1	-8%	-24%	-13%	-17%	-6%	-1%	-6%	+19%
INDRET2	+86%	+42%	+1%	+67%	-18%	+67%	+25%	+92%
INDRET3	+79%	+73%	+17%	+84%	+36%	+63%	+89%	+101%
Month FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Month-style FE	Yes	No	No	No	No	No	No	No
Month-style-rat FE	No	Yes	No	No	No	No	No	No
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	257,053	248,463	257,053		257,053		257,053	

Table A.II. Correlation between total fund return and its components. In this table, we report the correlation between total fund return and the 8 components into which it is decomposed using in Equation (12). Every month, we calculate the cross-sectional Spearman's rank correlation between the total fund return and the 8 return components. We report the mean, minimum, 10^{th} percentile, median, 90^{th} percentile, and maximum values of the correlation measure across the 175 months in the sample.

	Spearman's rank correlation with total fund return						
	Obs (months)	Mean	Min	P10	Median	P90	Max
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ALPHA^{7F}	175	0.71	0.34	0.51	0.72	0.85	0.94
MKTRET	175	0.04	-0.50	-0.29	0.05	0.34	0.55
SIZERET	175	0.22	-0.54	-0.06	0.22	0.58	0.81
VALRET	175	0.26	-0.47	-0.16	0.26	0.69	0.81
MOMRET	175	0.09	-0.59	-0.32	0.14	0.41	0.59
INDRET1	175	0.15	-0.51	-0.19	0.17	0.46	0.63
INDRET2	175	-0.03	-0.46	-0.35	-0.03	0.27	0.53
INDRET3	175	-0.01	-0.49	-0.26	0.02	0.20	0.53

Table A.III. Results of horse race between CAPM and Market-adjusted return (MAR).

	BHO panel regression with time FEs	Fama-MacBeth regression
	CAPM – MAR	**CAPM – MAR**
Sum of coefficient differences	7.41***	0.62
t-stat	(3.46)	(0.032)
% of coefficient difference > 0	77.78%	46.7%
Binomial p-value	< 1%	> 10%

A.3 Horse-race results of BHO

BHO also conduct a nonlinear pairwise test of asset pricing models in their Table 4. We again summarize their methodology for the reader's convenience. To compare two asset pricing models, in each period, funds are sorted into deciles using both models. Then BHO runs a panel regression with time fixed-effects on fund flows:

$$F_{p,t} = a + \sum_i \sum_j b_{i,j} D_{i,j,p,t} + cX_{p,t} + \mu_t + \epsilon_{p,t}, \tag{17}$$

where $D_{i,j,p,t}$ is a dummy variable indicating that fund p is ranked ith decile by model 1 and jth decile in model 2 (10th decile means the highest alpha), $X_{p,t}$ are a vector of controls, and μ_t are time fixed effects. The authors then compute test statistic $\hat{\theta} = \sum_{i<j} \hat{b}_{i,j} - \sum_{i>j} \hat{b}_{i,j}$ using the cases where the two models rank funds differently. If $\hat{\theta}$ is statistically larger than zero, then this indicates that flows are more responsible to the ranking by model 2 than model 1, and vice versa.

In Table A.III, we reproduce the horse race between CAPM and the market-adjusted model. While BHO find that CAPM beats the market-adjusted model, we find that the outperformance of CAPM disappears once we use the Fama-MacBeth regression. This is consistent with our analysis in Section 5 that investors do not behave as if they adjust for market beta.

A.4 Robustness checks of Tables 5 and 7

We first show that the result in Table 5 of Section 5.3 that investors do not discount for market beta is robust to controlling for month-style fixed effects or controlling for month-style-rating fixed effects. That is, we estimate Equation (13) with month-style FEs or month-

Table A.IV. Response of fund flows to market beta: controlling for month-style FEs and month-style-rating FEs. This table presents coefficient estimates from panel regressions of percentage fund flow on past returns and market beta in Equation (13). The t-statistics (double clustered by fund and by month) are in parentheses. *, **, and *** indicate significance at the 10% ,5%, and 1% level, respectively.

	All	**+MKT**	**−MKT**	**All**	**+MKT**	**−MKT**
	(1)	(2)	(3)	(4)	(5)	(6)
Weighted past return	0.71***	0.88***	0.55***	0.74***	0.91***	0.56***
	(23.99)	(23.11)	(17.61)	(24.37)	(23.33)	(17.78)
Market beta	0.000013	−0.00014	0.000051	−0.000047	−0.00025	0
	(0.048)	(−0.45)	(0.16)	(−0.18)	(−0.84)	(−0.028)
Ratings	0.0049***	0.0052***	0.0042***			
	(28.42)	(28.00)	(18.45)			
Month-style FE	Yes	Yes	Yes	No	No	No
Month-style-rat FE	No	No	No	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs	257,053	167,936	89,117	257,053	167,936	89,117
Adjusted R^2	0.22	0.23	0.20	0.23	0.24	0.21

style-rating FEs. The results are presented in Table A.IV. As one can see, market beta is not a significant determinant of fund flows under these alternative specifications.

We also estimate Equation (14) by the Fama-MacBeth procedure or controlling for month-style fixed effects, and we report the results in Tables A.V and A.VI, respectively. We confirm the results in Table 7 that the negative correlation between fund flows and return volatility is a byproduct of the fact that Morningstar accounts for a fund's return volatility when assigning ratings.

Table A.V. Response of fund flows to return volatility through Morningstar ratings: Fama-MacBeth procedure This table presents coefficient estimates from the Fama-MacBeth regression of percentage fund flow on a fund's return volatility over the prior 5 years or 1 year in Equation (14). The t-statistics (double clustered by fund and by month) are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	Flow	Flow	Flow	Ratings	Flow	Flow
	(1)	(2)	(3)	(4)	(5)	(6)
Vol^5	−0.046***		−0.030***	−8.20***	0.020**	
	(−4.47)		(−2.97)	(−16.89)	(2.06)	
$Vol^5_{predicted}$						−2.02***
						(−49.72)
$Vol^5_{residual}$						0.036***
						(3.41)
Vol^1		−0.054***	−0.024	−5.74***	0.022	
		(−3.36)	(−1.29)	(−4.83)	(1.44)	
Ratings					0.0065***	
					(43.28)	
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs	257,053	257,053	257,053	257,053	257,053	257,053
Adjusted R^2	0.091	0.094	0.098	0.21	0.17	0.17

Table A.VI. Response of fund flows to return volatility through Morningstar ratings: controlling for time-style FEs. This table presents coefficient estimates from panel regression of percentage fund flow on a fund's return volatility over the prior 5 years or 1 year in Equation (14). The t-statistics (double clustered by fund and by month) are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	Flow	Flow	Flow	Ratings	Flow	Flow
	(1)	(2)	(3)	(4)	(5)	(6)
Vol^5	−0.057***		−0.052***	−8.24***	−0.0022	
	(−4.90)		(−3.84)	(−9.34)	(−0.19)	
$\text{Vol}^5_{\text{predicted}}$						−1.96***
						(−36.86)
$\text{Vol}^5_{\text{residual}}$						0.0094
						(0.85)
Vol^1		−0.052***	−0.014	−6.33***	0.031	
		(−2.52)	(−0.60)	(−4.26)	(1.53)	
Ratings					0.0063***	
					(35.54)	
Month-style FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs	257,053	257,053	257,053	257,053	257,053	257,053
Adjusted R^2	0.098	0.097	0.098	0.22	0.16	0.16

Table A.VII. Response of fund flows to return volatility within Morningstar ratings groups. This table presents coefficient estimates from panel regression of percentage fund flow on a fund's return volatility over the prior 5 years or 1 year for each of the five Morningstar ratings groups. The t-statistics (double clustered by fund and by month) are in parentheses. *, **, and *** indicate significance at the 10% ,5%, and 1% level, respectively.

	Rating 1	Rating 2	Rating 3	Rating 4	Rating 5
	(1)	(2)	(3)	(4)	(5)
Vol^5	−0.026	−0.016	0.015	−0.025	−0.043
	(−1.35)	(−1.01)	(0.88)	(−0.99)	(−0.99)
Vol^1	−0.0098	0.058***	0.027	0.0073	−0.030
	(−0.35)	(2.61)	(1.26)	(0.26)	(−0.66)
Month FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of obs	17,024	60,416	92,131	60,613	18,279
Adjusted R^2	0.071	0.050	0.049	0.054	0.086

Appendix B Simulation Exercise

B.1 Simulation for bias in flow-performance regressions when the sensitivity of flows to differences in fund returns varies over time

In Section 5 we argued that coefficient estimates in the baseline BHO regression,

$$\begin{aligned} F_{p,t} &= b_0 + \mu_t + \gamma X_{p,t} + b_{\text{ALPHA}}\text{ALPHA}^{7\text{F}}_{p,t} + b_{\text{MKT}}\text{MKTRET}_{p,t} + b_{\text{SMB}}\text{SIZRET}_{p,t} \\ &+ b_{\text{HML}}\text{VALRET}_{p,t} + b_{\text{MOM}}\text{MOMRET}_{p,t} + \sum_{k=1}^{3} b_{\text{IND}_k}\text{INDRET}k_{p,t} + e_{p,t}, \end{aligned} \quad (18)$$

are likely to be biased when estimated using a panel regression with time fixed effects. In particular, we argued that the coefficient on the market-related portion of a fund's return is likely downward biased, leading to spurious evidence in favor of the CAPM. We showed that, when estimating the model using the Fama-MacBeth procedure, the evidence in favor of the CAPM largely disappeared.

In this Appendix, we present a simple simulation exercise to illustrate the intuition behind the reason why the panel regression estimates are misleading. Specifically, we simulate an economy where fund flows respond only to cross-sectional differences in raw fund returns, thus ignoring whether a fund's return is attributable to alpha or to factor loadings. In this setting, we show that if - as observed in the data - the flow-performance sensitivity (FPS) is weaker after extreme market returns, then panel regression estimates would incorrectly suggest that flows are averse to returns attributable to loadings on the market factor.

The simulated economy is generated as follows. There are 500 funds and 175 time periods. Consistent with the standard return attribution method, the return of each fund in each period is given by its factor loadings times the realized factors returns (MKTRET), plus the fund's alpha. For simplicity and without loss of generality, the only factor we consider in this simulation is the excess market return, and betas and alphas are uncorrelated across funds. Each fund's alpha and beta are drawn from normal distributions with mean of 0 and 1 and with standard deviation of 0.02 and 0.2, respectively. The excess market return, modeled after the market factor in the Fama and French (1993) model, is drawn each period from a normal distribution with mean of 0.006 and standard deviation of 0.045. Each period t, fractional flows to a given fund equal the prior period's fund return times the flow-performance coefficient γ_t plus an error term drawn from a normal distribution with mean of 0 and standard deviation of 0.015. We consider two possible scenarios. In scenario

A, the FPS is constant, i.e., $\gamma_t = \gamma = 0.7$. In scenario B, the FPS is weaker after extreme market return realizations, i.e., it is a hump-shaped function of the past market return. As mentiorned previously, this assumptions is consistent with empirical evidence presented by Franzoni and Schmalz (2017). In scenario B, the FPS in each period t is determined as

$$\gamma_t = 0.87 + 0.40 * I[mkt_{t-1} > 0] - 13.23 * mkt_{t-1} * I[mkt_{t-1} > 0] \\ +8.65 * mkt_{t-1} * I[mkt_{t-1} <= 0], \tag{19}$$

where $mkt_{t-1}$is the lagged excess market return. The coefficients in this model are estimated using the data used to generate Figure 2. Specifically, we regress each cross-sectional FPS coefficient (i.e., the coefficient from a cross-sectional regression of fund flows on fund returns) on the lagged realized market return, allowing for a different slope and intercept for positive and negative values of the market return. Consistent with Figure 2 and with Franzoni and Schmalz (2017), Equation 19 implies that the FPS is a hump-shaped function of realized market returns.

In each of the two scenarios, we estimate two versions of the flow-performance regression using the data generated by the simulation, i.e., a regression of fund flows on total fund returns; and a regression of fund flows on the two components of the fund return, alpha and MKTRET. We estimate these equations using both panel regressions with time fixed effects and Fama-MacBeth regressions. In Panel A and Panel B of Table B.I we report, respectively, the mean and median values across 1,000 simulations.

The first three columns of the table report the results for the scenario in which the FPS is constant over time. The true value of the flow-performance coefficient, in this case, is set to 0.7. The simulation results clearly show that, in this scenario, both the panel regressions and the Fama-MacBeth procedure are able to estimate the true parameter with high precision. The results are dramatically different in the scenario in which the FPS varies over time and is a hump-shaped function of realized market returns, which is the scenario closest to reality. In this case, the true flow-performance sensitivity coefficient is close to 0.68. In particular, this coefficient applies to both components of the total fund return, i.e., alpha and MKTRET, because fund flows are generated as a function of total fund returns, irrespective of the source of the return. Yet, column (6) shows that the mean and median coefficient on the market-related component of fund returns estimated by the panel regression is significantly lower than the true coefficient. In the same regression, however, the estimated coefficient on the alpha component is not significantly different from its true value. The low estimate of the coefficient on MKTRET might incorrectly be interpreted as evidence that investors are using the CAPM to evaluate fund returns, thus 'discounting' the portion of a fund's return that is attributable to a high market factor loading. In this simulation exercise, however, this is impossible because, by construction, flows only respond to (total) raw fund returns. The reason why the coefficient on MKTRET is downward-biased is that, by construction, the cross-sectional dispersion in MKTRET is significantly higher in period when the FPS is weaker, and these periods tend to be overweighted in the panel regression. On the other hand, the coefficients estimated using the Fama-MacBeth procedure do not appear to be significantly biased, because they weight each time period equally.

In general, the exercise presented in this Appendix indicates that caution should be used when interpreting the results from panel regressions of fund flows on different components of fund returns. More specifically, these results suggest that, given the hump-shaped relation between the FPS and realized returns, panel regressions are likely to deliver spurious evidence in favor of the CAPM. Hence, as we argued in Section 5, a better way to study the relationship between fund flows and fund return components is to weight each period equally (i.e., by using the Fama-MacBeth procedure) or to examine the distribution of coefficients from period-by-period cross-sectional regressions, as we do in Figure 3.

Table B.I. Results for the simulated economy described in Appendix B In this table we report the coefficients for regressions of fund flows on lagged fund returns in a simulated economy under two different scenarios. In the first scenario, the flow-performance sensitivity (FPS) is constant over time. In the second scenario, the FPS varies over time and is a hump-shaped functions of past realized market returns. We estimate two regressions. In both cases, the dependent variable is the simulated fund flow. In regression model 1, the independent variable is the total fund return. In regression model 2, the explanatory variables are the two components of the total fund returns, i.e., the fund's alpha and the fund's beta times the realized market return (MKTRET). In Panel A (B), we report mean (median) coefficient values across 1,000 simulations.

Panel A: Mean values across 1,000 simulations

	Constant FPS			Variable FPS		
	Regression 1	Regression 2		Regression 1	Regression 2	
	Tot Return	Alpha	MKTRET	Tot Return	Alpha	MKTRET
	(1)	(2)	(3)	(4)	(5)	(6)
True value	0.7000	0.7000	0.7000	0.6816	0.6816	0.6816
Panel regression	0.6997	0.6999	0.6988	0.6139	0.6818	0.2872
Fama-MacBeth	0.6999	0.6999	0.6979	0.6816	0.6817	0.6806

Panel B: Median values across 1,000 simulations

	Constant FPS			Variable FPS		
	Regression 1	Regression 2		Regression 1	Regression 2	
	Tot Return	Alpha	MKTRET	Tot Return	Alpha	MKTRET
	(1)	(2)	(3)	(4)	(5)	(6)
True value	0.7000	0.7000	0.7000	0.6814	0.6814	0.6814
Panel regression	0.6997	0.6995	0.6983	0.6146	0.6811	0.2910
Fama-MacBeth	0.7001	0.6996	0.7045	0.6812	0.6808	0.6837